DISTRIBUTED ANOMALY DETECTION SCHEME BASED ON LIGHTWEIGHT DATA AGGREGATION IN WIRELESS SENSOR NETWORK

NURFAZRINA BINTI MOHD ZAMRY

UNIVERSITI TEKNOLOGI MALAYSIA

DISTRIBUTED ANOMALY DETECTION SCHEME BASED ON LIGHTWEIGHT DATA AGGREGATION IN WIRELESS SENSOR NETWORK

NURFAZRINA BINTI MOHD ZAMRY

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy in (Computer Science)

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

APRIL 2022

ACKNOWLEDGEMENT

In the name of Allah, the most gracious, the most merciful. Praise to Allah.

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Assoc. Professor Dr. Anazida Zainal, for encouragement, guidance, advices, motivation and friendship. I am also very thankful to my external supervisor Dr. Murad A. Rassam for his guidance and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Ministry of Higher Education (MOHE) and Universiti Teknologi Malaysia (UTM) for funding my Ph.D. study. I must also acknowledge KULIM (M) for their invaluable assistance given throughout my study.

My fellow postgraduate friends should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family members.

Last but certainly not least, to my family members especially my late ‘Abah’, ‘Atok’ and ‘Pakteh’ for their prayer and moral support. Special thanks to best friends for their continuous support and the inspiration throughout the journey. Thank you for being there with me during the joy and sorrow. Thank you especially to my Ph.D friends for sharing times and knowledge to uplift my motivation to finish my Ph.D journey. Without them, this journey might be lonely and tougher.

ABSTRACT

Wireless Sensor Networks (WSNs) have been used in many domains for instance in business applications, industrial applications, and military applications to monitor a phenomenon, track an object, or control a process. As the sensor nodes communicate continuously from the target phenomenon to the base station, hundreds of thousands of multivariate data are collected from sensor nodes will be analysed at the endpoint called base station or sink node for decision making. Unfortunately, data is not usually accurate and reliable which will affect the decision making at the base station. There are many reasons that cause inaccurate and unreliable data such as malicious attack, harsh environment as well as sensor node failure. In the worst-case scenario, the node failure will also lead to the dysfunction of the entire network. Thus, anomaly detection is used to ensure that the data acquired at the endpoint is accurate. On the other hand, as sensor nodes possess resources constraint in terms of energy, processing, and storage, therefore, anomaly detection techniques must be designed in a lightweight manner. Meanwhile, existing anomaly detection techniques pose weaknesses as these have high computational and communication cost, ignore multivariate data and features' correlation, and some are parameter-dependent. The purpose of this research is to design and develop an efficient and effective anomaly detection scheme for WSN by minimizing the resource constraint in WSNs. This purpose can be achieved by first, applying the feature selection method to select significant features for minimizing the resource utilization. Second, designing an efficient network structure of WSNs architecture based on a data aggregation scheme for reducing data transmission in the network. Third, designing lightweight anomaly detection scheme (CESVM-DR) using One-class Support Vector Machine (OCSVM) anomaly detection scheme and incorporating dimension reduction technique based on Candid Covariance-Free Incremental Principal Component Analysis (CCIPCA) to minimise the computational complexity of covariance matrix in CESVM. Lastly, enhancing the efficiency and effectiveness of the anomaly detection scheme by designing the distributed anomaly detection scheme (DCESVM-DR). The effectiveness and efficiency of the proposed anomaly detection schemes were tested using real-world datasets as well as soil data collected from the palm oil plantation. The results show the proposed CESVM-DR anomaly detection scheme with an average of 92%–100% detection accuracy using the real datasets while minimizing the computational complexity and energy overhead. Meanwhile, exploiting the correlation between sensor nodes to detect the anomalies on the DCESVM-DR has enhanced the effectiveness results as well as minimised the memory complexity and energy

ABSTRAK

Rangkaian Penderia Tanpa Wayar (WSN) telah digunakan dalam banyak domain contohnya dalam aplikasi perniagaan, industri dan ketenteraan untuk memantau fenomena, menjejak objek atau mengawal proses. Apabila nod pengesan berkomunikasi secara berterusan daripada fenomena sasaran ke stesen pangkalan, ratusan ribu data berbilang variasi dikumpulkan daripada nod pengesan akan dianalisis pada titik akhir yang dipanggil stesen pangkalan atau nod sink untuk membuat keputusan. Malangnya, data biasanya tidak tepat dan boleh dipercayai yang akan menjejaskan pembuatan keputusan di stesen pangkalan. Terdapat banyak sebab yang ditimbulkan oleh data yang tidak tepat dan tidak boleh dipercayai seperti serangan berniat jahat, persekitaran yang kasar serta kegagalan nod sensor. Dalam senario terburuk, kegagalan nod, sebaliknya, juga akan menyebabkan kepada keseluruhan rangkaian tidak berfungsi. Oleh itu, pengesanan anomali digunakan untuk memastikan bahawa data yang diperoleh pada titik akhir adalah tepat. Sebaliknya, memandangkan nod sensor mempunyai kekangan sumber dari segi tenaga, pemprosesan dan penyimpanan, oleh itu, teknik pengesanan anomali mesti direka dengan cara yang ringan. Sementara itu, teknik pengesanan anomali sedia ada terhad dari segi kos pengiraan dan komunikasi yang tinggi, mengabaikan data berbilang variasi dan korelasi ciri manakala sesetengahnya bergantung kepada parameter. Tujuan penyelidikan ini adalah untuk merekabentuk dan membangunkan skim pengesanan anomali yang cekap dan berkesan untuk WSN dengan meminimumkan kekangan sumber dalam WSN. Tujuan ini boleh dicapai dengan pertama sekali, menggunakan kaedah pemilihan ciri untuk memilih ciri penting untuk meminimumkan penggunaan sumber. Kedua, merekabentuk struktur rangkaian yang cekap seni bina WSN berdasarkan skema pengagregatan data untuk mengurangkan penghantaran data dalam rangkaian. Ketiga, merekabentuk skim pengesanan anomali ringan (CESVM-DR) menggunakan skema pengesanan anomali Mesin Vektor Sokongan Satu (OCSVM) dan menggabungkan teknik pengurangan dimensi berdasarkan Analisis Komponen Utama Tanpa Kovarian Tambahan (CCIPCA) untuk meminimumkan kerumitan pengiraan matriks kovarians dalam CESVM. Akhir sekali meningkatkan kecekapan dan keberkesanan skim pengesanan anomali dengan merekabentuk skim pengesanan anomali teragih (DCESVM-DR). Keberkesanan dan kecekapan skim pengesanan anomali yang dicadangkan diuji menggunakan set data dunia sebenar serta data tanah yang dikumpul dari lading sawit. Keputusan menunjukkan skim pengesanan anomali CESVM-DR yang dicadangkan dengan purata ketepatan pengesanan 92-100% menggunakan set data sebenar sambil meminimumkan kerumitan pengiraan dan overhed tenaga. Sementara itu, mengeksploitasi korelasi antara nod sensor untuk mengesan anomali pada DCESVM-DR telah meningkatkan hasil keberkesanan serta meminimumkan kerumitan memori dan tenaga.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	-	Detection Accuracy
ADC	-	Analog Digital Convertor
ADSs	-	Anomaly Detection Systems
ANN	-	Artificial Neural Network
APCADR	-	Adaptive And Efficient Dimension Reduction Model
BS	-	Base Station
CCIPCA	-	Candid Covariance-Free Incremental Principle Component Analysis
CESVM	-	Centered ellipsoidal SVM
CESVM-DR	-	Centered ellipsoidal SVM - Dimension Reduction
CH	-	Cluster Heads
CMD	-	Compact Matrix Decomposition
DADMA	-	Data Aggregation and Dilution by Modulus Addressing
DB-AD	-	A distance-based anomaly detection
DBNs	-	Deep Belief Networks
DDCD	-	Data Density Correlation Degree
DR	-	Detection Rate
DSC	-	Distributed Source Coding
DT	-	Decision Tree
DWT	-	Discrete Wavelet Transform
EAOD	-	Ellipsoidal SVM-Based Adaptive Detection
ECBDA	-	Energy-efficient Cluster-Based Data Aggregation scheme
EECA	-	Energy-Efficient Clustering Algorithm
EEHA	-	Energy-efficient and high-accuracy
EM	-	Expectation-Maximization
EOOD	-	Ellipsoidal SVM-Based Online Detection
EPFL	-	Ecole Polytechnique Fédérale de Lausanne
FNR	-	False Negative Rate
FPR	-	False Positive Rate
FS	-	Feature selection

GA	-	Genetic Algorithm
GAF	-	Geographical Adaptive Fidelity
GanoDROP	-	Ganoderma and Disease Research for Oil Palm Unit
GeRaF	-	Geographic Random Forwarding
GNM	-	Global Normal Reference Model
GSB	-	Grand St. Bernard
IBRL	-	Intel Berkeley Research Lab
IDS	-	Intrusion detection systems
IoT	-	Internet of Things
IPS	-	Intrusion Prevention System
k-NN	-	kth-Nearest Neighbour
KPCA	-	Kernel-Based PCA
LEACH	-	Low-energy adaptive clustering hierarchy
LNM	-	Local Normal Reference Model
LUCE	-	Lausanne Urban Canopy Experiment
MD	-	Mahalanobis Distance
MSE	-	Mean Square Error
MTS	-	Mahalanobis Taguchi System
NAMOS	-	Networked Aquatic Microbial Observing System
NN	-	Neural Networks
OCPCC	-	One-Class Principal Component Classifier
OCSVM	-	One-Class Support Vector Machine
PCA	-	Principal Component Analysis
PCCAD	-	Principal Component Classifier based Anomaly Detection
PDG	-	Patrouille des Glaciers
PSO	-	Particle Swarm Optimization
QS-SVM	-	Quarter-sphere SVM
RBF	-	Radial basis function
RFS	-	Ranking-based Feature Selection
RMSE	-	Root Mean Square Error
RO	-	Retained Originality
RSES	-	Rough Set Exploration System
RST	-	Rough Set Theory

SDA	-	Secure Data Aggregation
SN	-	Sensor Nodes
SOM	-	Self-Organizing Maps
SVM	-	Support Vector Machine
TAG	-	Tiny AGgregation
TiNA	-	Temporal coherency-aware in-Network Aggregation
UNPCA	-	Unsupervised Principle Component Analysis
UTM	-	Universiti Teknologi Malaysia
WEKA	-	Waikato Environment for Knowledge Analysis
WSMART	-	Witness based Slice and Mix (WSMART) scheme
WSNs	-	Wireless Sensor Networks
XML	-	Extensible Mark-up Language

LIST OF SYMBOLS

m	-	Number Of Data Measurements/ A Low-Rank Approximation Of Kernel Gram
N, n	-	Number Of Variables
μ	-	Mean
σ	-	Standard Deviation & Width Of The Kernel Function
f	-	Normal Random Distribution Function Of Artificial Anomalies
d, D	-	The Dimension Of Data Vector With Reducing The Dimension
q	-	Number Of Neighbours/ Nodes In The Cluster
p	-	The Dimension Of The Data Vector/ Optimal Election Probability Of A Node To Become Cluster Head
P	-	Linear Optimization In EOOD Scheme
e	-	Encoding The Data Observation Using DWT
l	-	Anomaly Detection Computation In SOM
k	-	Communication Of Wavelet Coefficient To The Central Node
F	-	Algorithm's Fitness Function
$\mathcal{S}$	-	The Set Of Sets Corresponding To The Discernibility Function
ϑ	-	A Weighted Multiset
w	-	Maps Each Element In Θ To A Non-Zero Value
ρ	-	A Weighting Between Subset Cost And Hitting Fraction
r	-	Relevant In The Case Of Approximate Solutions & Degree Of Polynomial Kernel
c	-	The Cost Of An Attribute Subset
a	-	The Attribute That Maximizes ∑W(S),
E_0	-	Initial Energy
E_{elec}	-	Etx & Erx
E_{fs}	-	Free Space Energy

E_{amp}	-	Multipath Fading Energy
E_{DA}	-	Data Aggregation Energy
d_0	-	Threshold Distance
δ	-	Number Of Consecutive Rounds
$E_{residual}$	-	Nodes' Current Energy
G	-	Set Of Nodes That Have Not Been Cluster-Heads In The Last
λ	-	Threshold Calculation A Relay Node Will Be Selected Among The Cluster Heads
μ_{CH}	-	The Average Distance Of All Cluster Heads
$T(n)$	-	Threshold For Cluster Head Selection
ϕ	-	Linear Mapping Function
R	-	Minimum Effective Radius/ Radius
Kc	-	Centered Kernel Matrix
Σ^{-1}	-	Inverse Of The Covariance Matrix
Λ,V	-	Diagonal Matrix With Positive Eigenvalues
P,D	-	The Eigenvector Matrix Corresponding To The Positive Eigenvalues
α_i	-	Lagrangian Multiplier
$f\ (x)$	-	Decision Function
$Md(x)$	-	Distance Measure
x_{new}	-	New Standardize Data
k_{Linear}	-	Linear Kernel Function
$k_{(RBF)}$	-	RBF /Gaussian Kernel Function
k_{Poly}	-	Polynomial Kernel Function
v	-	Regularization Parameter
RO	-	Retained Originality
$Error, e$	-	Approximation Error
ξ	-	Root Mean Square Error
$S(t)$	-	The Original Data For Each CH Nodes
K	-	Represents The Number Of Nodes Involves During The Communication
2	-	The Two-Way Communication

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

With the advancement of digital technology from the past few decades, every digital equipment and appliance are expected to be embedded with tiny yet powerful devices called sensor node. Furthermore, the wireless communication between physical item and sensor to exchange information for smart living in the future has been coined as the Internet of Things (IoT). In the world of modern wireless telecommunications, IoT is a revolutionary paradigm that is rapidly growing. (Atzori et al., 2010). When these sensor nodes communicate together to collect large amount of data from targeted area via the wireless channel, they are called Wireless Sensor Networks (WSNs). Businesses, industries, and the military have utilised WSNs to track an object or monitor a phenomena. Besides, many types of research areas have emerged from the WSNs domain such as from routing protocol, security, and privacy to data mining and many other. Nevertheless, currently researches are concerned in improving the performance of the WSNs technologies (Ayadi, Ghorbel, Obeid, *et al.*, 2017).

In general, the sensor nodes are equipped with sensing, processing, radio, and power unit, yet they have limited resource in term of energy, computation, and storage (Gao et al., 2020). Frequently, a large number of the sensor nodes are deployed widely in the target environment and continuously communicate the phenomenon measurement like ambient temperature, relative humidity, soil moisture, and wind speed to the base station. Therefore, in most situations, sensor data need to collect accurate and reliable measurement for data analysis and decision-making especially in a critical domain such as in meteorology station, the military application as well as security monitoring.

Unfortunately, the raw data collected from WSNs communication usually are not reliable and inaccurate due to imperfect nature of WSNs (Ayadi, Ghorbel, Obeid, *et al.*, 2017). According to Zhang et al., (2008) the reasons that lead to the unreliable and inaccurate data is due to sensor nodes are deployed in harsh and unattended environment, and these sensors are also vulnerable to malicious attacks. In addition, resource constraint imposed by sensor nodes makes the device fails to operate properly thus reduces data accuracy. Therefore, data collected from these sensor nodes are often generates missing data, duplicated or error records. In order to ensure the collected data is reliable and accurate for data analysis and decision-making, one of the solutions is to detect the erroneous data, malicious attack or the changes in the environment namely anomaly or outlier detection (These terms will be used interchangeable throughout this thesis). Anomaly detection is one of the potential approaches that can be considered as a solution. Anomaly detection is defined by Chandola *et al.*, (2009) as the process of identifying data patterns that vary from anticipated behaviour. When it comes to WSNs, anomaly detection has been widely employed across a wide range of industries such as the military and environmental sectors. (Akyildiz et. al., 2002). This is due to the characteristic of low-cost, small in size and multi-functional sensor nodes; it helps to achieve the need of fast and cheap data collection.

Another wide Implementation of WSNs is for agricultural in the respective works i.e., monitor the irrigation Morais et. al., 2005); monitor micro-climate in the crop field (Baggio, 2005); to detect pathological symptoms presence of in oil palm (Shafri and Anuar, 2008 and Mazliham et. al., 2007); control the irrigation in (Maurya and Jain, 2016); to investigate the effects of the environmental conditions(Ferentinos et al., 2017); to detect any abnormal situation in the meteorological data (Salim et al., 2020). For instance, implementing early detection on palm oil disease namely Ganoderma Boninense (G.Boninence) can reduce billions of Ringgit loss (Zain et. al., 2013; Cooper et. al., 2011; Hushiarian et. al., 2013). Therefore, one of the potential research utilizing the concept of WSNs is to collect potential data from palm oil plantation and implementing anomaly detection method to detect the palm oil disease Hushiarian et. al., (2013); Abdullah et. al., (2012); Markom et. al. (2009). Moreover, due to the geographical structure of the plantation field, deploying this sensor node again helps to eliminate human intervention to take the samples from the field.

Although WSN has been widely used to detect anomalies in various fields, it has many challenges that affect the efficiency and effectiveness of a detection. Again, these issues may arise from the resource constraint of the sensor nodes, including low energy, processing capability, memory or storage limitation. These limitations need to be considered during designing the desired solution using anomaly detection in WSNs domain. Possible approaches to tackle the energy consumption for WSN is by implementing data aggregation techniques and dimension reduction (Chitradevi *et al.*, 2010; Ullah and Youn, 2020; Ullah *et al.*, 2021). As defined by Rajagopalan and Varshney (2006), data aggregation is the process of eliminating duplicate transmissions and delivering fused information to the base station by combining data from numerous sensors. Data aggregation helps to reduce the transmission of data within the network to consume fewer data communication and prolong the network lifetime. This approach has been used (Chitradevi *et al.*, 2010; Bharuka and Jinwala, 2014; Nisha *et al.*, 2014; Otoum *et al.*, 2018) in recent anomaly detection researches in developing more effective and efficient detection in term of energy usage.

1.2 Background of the Problem

As mentioned, the unique characteristic of sensor nodes needs few considerations on designing an adequate anomaly detection solution for in WSNs domain. Therefore, applying traditional techniques like cryptography, authentication or complicated detection technique for data analysis to WSNs domain is unsuitable as these sensor nodes are highly constrained in resources. Sensor networks have a number of significant issues, including energy constraints, limited computational power, restricted memory, and data security, for which academics have proposed many solutions (Ur-Rehman et al., 2014). An adequate solution design in anomaly detection measured in terms of their detection effectiveness as well as efficiency in leveraging the network's limited resources (Rassam et. al., 2013a). Effective anomaly detection associated with detection accuracy, detection rate, and false alarms whereas efficient anomaly detection associated with energy consumption and memory utilization. These characteristics often used as performance measures in several anomaly detection types of research (Zhang et. al., 2013; Rassam et. al., 2014; Ghorbel et. al. 2017; Ayadi et al., 2020). In other words, the lightweight detection is an adequate solution to design anomaly detection in WSNs by utilizing less energy consumption and low processing and storage usage. Therefore, lightweight detection is a desirable characteristic of the anomaly detection scheme. Several solution approaches have been described in the previous section including the use of WSNs to collect required data from the fields for various detection. Figure 1.1 shows the scenario that leads to problem addressed in this research.

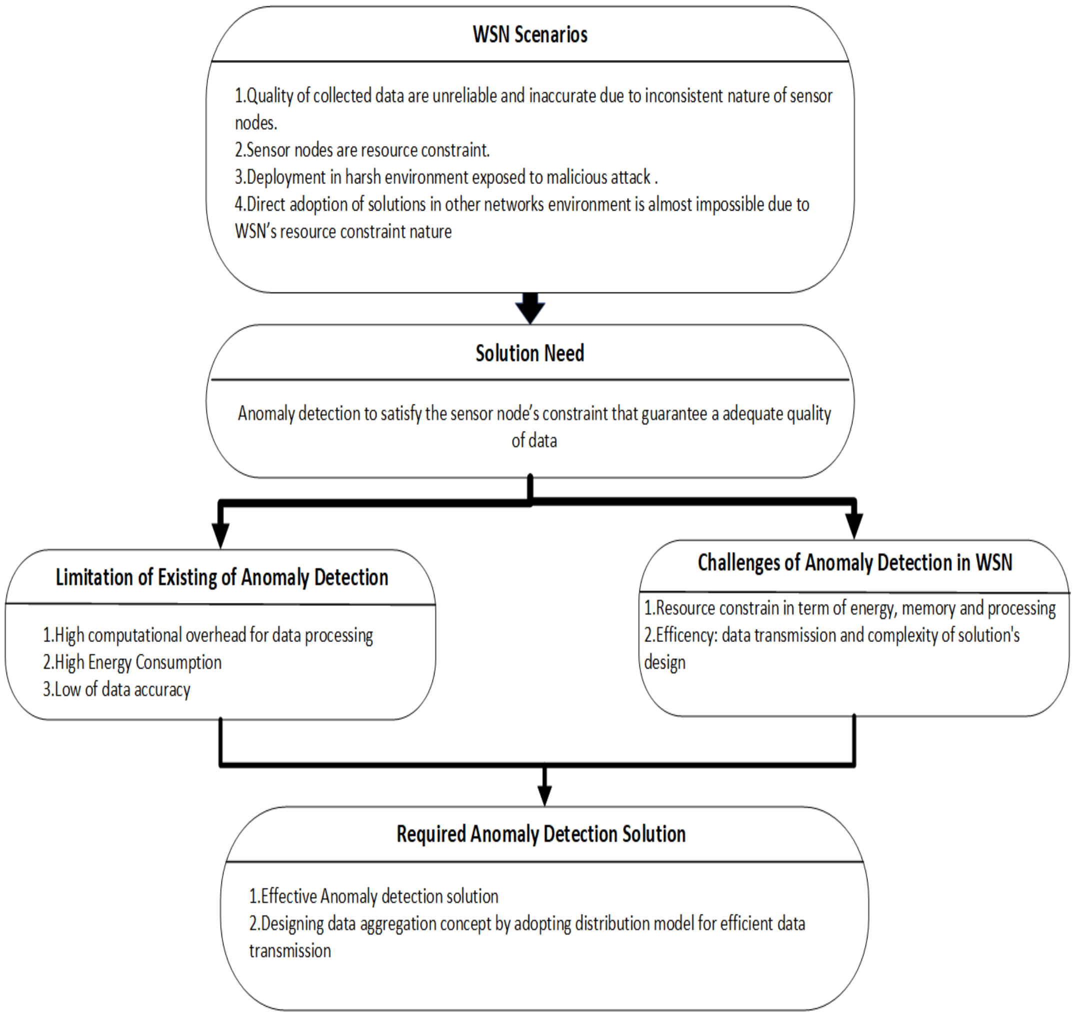

Figure 1.1 Scenario Leading to the Problem

Figure 1.1 presents the scenario in WSNs, the solution needs, challenges, and limitation as well as the desired solution needs to design anomaly detection scheme in WNS. In WSNs domain, the desired solution in designing anomaly detection is to have efficient detection while preserving resources and effective detection while preserving accuracy. According to Rassam et. al. (2013a), there are five requirements need to be satisfied in designing and developing an efficient and effective anomaly detection model. These five main requirements for detection models are termed as RODAC and it includes; reduction of data, online detection, distributed detection, adaptive detection, and correlation exploitation. Online detection is the mean to ensure real-time or near real-time detection. Meanwhile, distributed detection is to ensure optimum resources utilization especially for data transmission over the network. Meanwhile, reduction of data, preserve energy during data transmission. Correlation exploitation in sensor data in neighbourhoods can limit the data transmission. One of the factors that leads to quick sensor depletion is caused by data transmission (Rassam et. al. 2014). The transmission of one data bit requires as much energy as processing a thousand bits in a sensor. Hence, designing the energy-efficient solution is crucial in WSN in application such as anomaly detection. The dynamic changes of data sensed by the sensor need to take into account as it will affect the detection effectiveness of the anomaly detection model. Hence, adaptive detection is a recommended feature to enhance the detection effectiveness in anomaly detection, especially in a dynamic environment. Therefore, one needs to consider this requirement when designing an effective and efficient anomaly detection model. The complexity of the anomaly detection model or schemes also discussed in Ayadi *et al.*, (2017) and Rassam *et al.*, (2018).

Recently, Safaei *et al.* (2020) and Chander and Kumaravelan (2021) have discussed the techniques and challenges related to anomaly detection in WSNs. Both studies indicate that in designing and developing and effective and efficient anomaly detection, these factors must be considered: resource constraints, robust communication with minimum communication, and computational complexity, while taking into account the high-dimensional data. Recent lightweight anomaly detection schemes show tolerable performance evaluation in terms of accuracy or detection rate. However, the false alarms are higher in some schemes. Current techniques present between 81% and 98% detection rate with false alarm rates between 2% and 40% (Rassam *et al.*, 2012; Cheng and Zhu, 2015; Ayadi *et al.*, 2017; Ghorbel *et al.*, 2017; Safaei *et al.*, 2020). Meanwhile, Chen and Li (2019) reported AUC values between 95% and 98% and false alarm rates between 5% and 40% using different kernel widths.

Due to the restricted energy and harsh deployment conditions in WSNs, raw sensor observations frequently have low accuracy. On the other hand, Anomalies can be triggered by a malfunctioning sensor node, as well as network security attacks or unusual phenomena in the monitoring scope (Feng et al., 2017). Moreover, WSNs are open to both internal attacks and external attack stated by Yu et al. (2012). Internal attacks include eavesdropping, injecting fractional data, and fabricating non-existent records in order to disrupt the normal operation of the entire network, whereas external attacks involve the invader breaking through some traditional safeguards in order to capture sensor nodes and learn vital information from them. Due to the inconsistent nature of sensor nodes, the quality of the data collected by sensor nodes is unreliable and inaccurate. Therefore, applying any security solution or technique to the data or to the sensor node can ensure the high data quality. However, according to Akyildiz et al. (2002) and Xie et al. (2011), though WSNs is derived from the ad-hoc network, adopting any ad-hoc network detection schemes into WSN is not feasible due to the resource constraint of sensor nodes. Designing adequate solution specifically for WSNs need to consider its' limitations and challenges.

Meanwhile, the large amount of data collected from the sensor nodes composes of irrelevant and redundant features (Randhawa and Jain, 2017; Xue et al., 2018). These numerous amounts of data lead to greater resource consumption and also affects the detection effectiveness. This massive quantity of data, which comprises irrelevant and redundant characteristics, causes delayed training and testing, increased resource usage, and a low detection rate. (Li et al., 2009; Manbari et al., 2019). In most situations, normal data is more commonly presented in the WSNs environment as compared to abnormal data. Furthermore, the abnormal or anomalous data usually can be recognized by the unique pattern yet camouflaged by the normal data. Additionally, most of the features in the datasets are redundant or irrelevant, which can lead to higher training time and low detection performance (Xue et al., 2018). Therefore, feature selection is often performed on the collected data to select and filter the unimportant and irrelevant features (Kumar and Sonajharia, 2014; Xue et al., 2018). Feature selection is a pre-processing task that usually performed before data is fed to the anomaly detection scheme. Therefore, removing the irrelevant and redundant features can improve the classifier performance as well as the detection effectiveness. Feature selection is crucial as there are too many features in the data or features that are noisy, irrelevant and redundant, as this will affect detection accuracy and speed.

Moreover, multivariate data is also needed to be considered when designing an anomaly detection scheme as multivariate data are always sensed in the target phenomenon (Aldweesh *et al.*, 2020; Safaei, Asadi, *et al.*, 2020; M. A. Rassam *et al.*, 2013b). Unfortunately, the multivariate data detection is energy consuming thus reducing the data dimension may help on minimizing energy utilization. For effective analytical results, data reduction is done in order to obtain quality of knowledge without affecting the integrity of the original data. (Randhawa and Jain, 2017). Moreover, dimensionality reduction can help to minimize the space required to store the data when the number of dimensions increases (Poornima and Paramasivan, 2020). There are many types of research have proposed dimension reduction scheme when designing anomaly detection model, for instance, Ullah *et al.*, (2021), Ghorbel *et al.*, (2017), Erfani *et al.*, (2016), Rassam *et al.*, (2012), Takianngam and Usaha, (2011) and Siripanadorn *et al.*, (2010).

Apart from that, data transmission process is more energy consuming compared to computation process (Randhawa and Jain, 2017). This means large amount of energy is required for communication process (Yue et al., 2018). Moreover, centralized data communications by directly sending whole data to the sink or base station is also an inefficient solution that can also rapidly drain the sensor energy. Therefore, data aggregation is a widely adopted method to effectively reduce the data transmission volume and improve the lifetime of WSNs (Wan et al., 2019; Li et al., 2018). Energy consumption can be reduced by aggregating the collected using aggregate function before forwarding to the base station compared to centralize solution (Gomathi and Krishnan, 2020; Liu et al., 2020). Thus, this data aggregation solution is considered as one of the key solutions for energy utilization by reducing the number of communication while transferring the whole data to base station. This aggregation concept has been adopted in distributed anomaly detection solution (Ullah et al., 2021; Otoum et al., 2018; Nisha et al., 2014; Bharuka and Jinwala, 2014; Chitradevi et al., 2010;) for energy efficient. Usually in distributed environment a special node called cluster head is used to aggregate the data sent by sensor node within the specific cluster before sending the aggregated data to base station (Ullah et al., 2021).

Accurate data collection on the network is crucial in some scenario, such as health applications (Safaei, Asadi, *et al.*, 2020). As previously discussed, a various factor might lead to inaccurate WSN sensory data. Measurements that deviate significantly from the normal pattern of sensed data are classified as outliers in the field of WSNs (Safaei, *et al.*, 2020). Outliers are also defined as anomalies or divergences that exhibit unexpected behaviour when compared to the majority of sensory data. Therefore, the necessity to identify/detect outliers in deployed sensor nodes in WSNs is crucial. This detecting process is known as sensor outlier/anomaly detection.

Meanwhile, neighbourhood correlation has advantages of detecting anomalous data. The distinction between anomalous data and significant events can be made by observing that anomalous are likely to be spatially unrelated, but significant event measurements are likely to be spatially associated (Yang Zhang *et al.*, 2010; Rassam *et al.*, 2018; Kumar and Chaurasiya, 2019). Moreover, the correlation between the nodes can reduce the false alarm rate due to the information exchanged between the nodes to distinguish between the event and anomalous data. This requirement is closely related to distributed detection and correlation exploitation as they can collaborate to enhance the detection effectiveness and efficiency. Sun et. al., (2013) have suggested for system monitoring modules (SMM) should be integrated with intrusion detection modules (IDM) in the context of WSNs. SMM is used to monitoring the important event in the sensor environments by exploiting node correlations (Banu and Balasubadra, 2018; Rajasegarar *et al.*, 2014; Kannadhasan *et al.*, 2014; Arthi A, 2014; Francis and Babu, 2014).

The existing distributed anomaly detection techniques such as Ullah et al. (2021), Rassam et al. (2018), Ghorbel and Abid (2015), and Zhang et al. (2013) are developed with the intention to attain effective and efficient detection with few consideration factors. Firstly, data redundancy from sensor nodes to the intermediate node can affect not only the excessive transmission but also the computational overhead (Kumar and Chaurasiya, 2019). Thus, by transmitting some useful data can save energy. Secondly, when adapting the distributed solution, adequate size of the summary often called the normal reference model to transmit to central location need to be as small as possible during the data transmission in order to cope with resource constraint. The normal reference model is the normal state of the specific dataset used in anomaly detection to detect anomalies obtained by performing the data training process (Maya et al., 2019). Lastly, efficient transmission structure to ensure data can transmit to the base station even when some nodes are malfunctioning or faulty therefore accurate decision making can be achieved. Guo *et al.* (2014) and Widhalm *et al.* (2021) categorised faulty node into function or data fault and hard or soft faults respectively. The second category frequently refers to anomalies or outliers that may be detected by utilising anomaly detection. Meanwhile, the first category is erroneous data, which may be addressed using data aggregation techniques (Guo *et al.*, 2014; Shial *et al.*, 2020).

In addition, designing efficient anomaly detection needs to consider the complexity of the technique. One class classifier like One-Class Support Vector Machine (OCSVM) (Zhang *et al.*, 2013; Rajasegarar *et al.*, 2008b;) and One-Class Principal Component Classifier (OCPCC) (Rassam et al., 2014) based anomaly detection are favourable in the case of anomaly detection in WSNs (Rassam et al., 2014). One class setting assumes that the data have only one label which is normal data label. When the data is not fitted with the training normal data, it is considered as anomalous. As only one class is utilized during the anomaly detection procedure, thus giving the advantages in term of processing, storage as well as the training time. Furthermore, OCSVM and OCPCC's one-class learning methods and unsupervised approaches are ideal for datasets with no ground truth since these techniques do not require pre-labelled data, which is difficult or expensive to provide. (Chander and Kumaravelan, 2021; Rassam et al., 2014; Shahid et al., 2013).

A recent review by Chander and Kumaravelan (2021) highlighted the limitations, issues, and requirements of the existing anomaly detection approaches in WSNs. In general, anomaly detection limitation includes ignoring multivariate data and attributes' correlation, making use of user-specified pre-set threshold values, not being able to distinguish between events and errors, challenges of streaming data, the appropriate transmission of the reference model, and space requirement to store batches of data for a period of time. Meanwhile, some specific anomaly detection techniques suffer from parameter selection or tuning and high computational cost. To address these limitations, Chander and Kumaravelan (2021) highlighted the key design elements needed, which include a combination of unsupervised learning scheme, distributed approach, multivariate data, online mode, spatio-temporal correlation, adaptability, automated communication, differentiated event and outlier, ability to detect multiple anomalous types, intelligent strategy, and low computational complexity with high detection rate.

1.3 Problem Statement

Detection effectiveness and efficiency need to be considered when designing and developing anomaly detection scheme. Anomaly detection suffers from computational complexity when a large number of unrelated and unimportant multivariate data features are processed. As a result, detection accuracy can decrease while energy depletes when transmitting big amount of data. Data accuracy can also be affected when a malfunction or faulty node transmits the faulty or malicious data to sink or base station. Thus, this can lead to low detection accuracy. Therefore, feature selection to filter the unrelated features as well as dimension reduction are required for efficient and effective detection

Meanwhile, detection efficiency can be achieved by minimizing energy consumption. In WSNs, processes of sensing, processing and transmitting data may consume a lot of energy which leads to high communication overhead. Besides, excessive energy consumption may happen when sensor nodes are located far from the base station. Therefore, data aggregation technique can be incorporated in the proposed distributed anomaly detection scheme while utilizing spatial neighbourhood correlation to reduce the excessive energy consumption. Moreover, the spatial correlation between the nodes can enhance the detection effectiveness when anomalous data can be distinguished from common events in the network. The research hypothesis of this research is:

The effectiveness and efficiency of the solution approach for detecting anomalous data in WSNs can be achieved by utilizing feature selection, dimension reduction and distributed detection to ensure efficient use of energy.

1.4 Research Question

The following questions are addressed in this research:

1. How to eliminate the irrelevant and unimportant data to produce highly accurate detection while reducing energy consumption?

2. What is the suitable technique to reduce transmission and data processing in WSN for energy consumption while prolonging the lifetime of WSN?

3. What is the impact of transmitting long distance and big amount data on high energy consumption?

4. How to design a local lightweight anomaly detection scheme for multivariate data in order to further reduce the energy consumption?

5. How to further enhance the effectiveness and efficiency of the anomaly detection scheme by utilizing distributed and spatial-correlation approaches?

1.5 Research Purpose

The purpose of this research is to design and develop a lightweight anomaly detection scheme for WSNs using data aggregation approach in distributed manner with high accuracy and low false alarm, low communication overhead and low computational complexity.

1.6 Research Objective

The main objectives of this research are:

1. To design an effective WSN data transmission scheme by selecting significant features using Rough Set Theory and Ranking method.

2. To design and implement a hybrid Low-Energy Adaptive Clustering Hierarchy (LEACH) data aggregation techniques to minimize energy consumption

3. To design and develop lightweight local anomaly detection scheme by reducing data dimension using one class support vector machine technique utilizing the selected features obtained in objective (1).

4. To enhance the efficiency of the anomaly detection scheme in objective (3) by developing a distributed anomaly detection scheme to distribute the energy consumption and detection process while utilising spatial correlation of the sensor nodes to improve detection effectiveness.

1.7 Scope and Assumption

This study is limited to the following:

1. Data used in the study obtained from the following resources:

   a) Palm oil soil data is collected from Malaysia plantation. These data will be used to illustrate the effectiveness of the proposed RST-Ranking feature selection technique

   b) Intel Berkeley Research Lab (IBRL), Sensorscope: Environmental Data for Wireless Sensor Networks includes Grand St. Bernard (GSB), Lausanne Urban Canopy Experiment (LUCE) and c) Patrouille des Glaciers (PDG), and Networked Aquatic Microbial Observing System

(NAMOS) will be used to validate the proposed scheme. These datasets are largely used by WSNs researcher especially in the anomaly detection area (Rassam et. al., 2013b; Zhang et. at., 2013; Rajasegarar et. al., 2007).

2. Data aggregation is presented by combining LEACH-C and LEACH-R clustering-based structure and spatial-correlation between the nodes and each cluster as aggregator nodes to collect the data from the cluster.

3. Network metrics such as energy consumption, node failure rate and data transmission rate, are used to measure the efficiency of the proposed data aggregation scheme as used by other researchers in the domain (Heinzelman *et al.*, 2002; Wang and Zhu, 2012; Ullah *et al.*, 2021).

4. The performance measurements of this proposed anomaly detection scheme are based on effectiveness and efficiency. Effectiveness will be measured based on detection accuracy, detection rate, and false alarm while the efficiency will be measured by energy consumption and memory utilization.

1.8 Significance of the Research

The significance of the research is outlined as follows:

1. The unreliable and inaccurate data includes of the harsh and unattended environment, as well as sensor nodes are vulnerable to malicious attacks affects the decisions making at the base station in WSNs thus anomaly detection scheme is suggested for accurate data analysis.

2. Large volume of data sensed by the sensor nodes composes of irrelevant and redundant features as well as the multivariate data features leads to greater resource consumption and also affects the detection effectiveness.

3. High energy consumption during the data transmission in the network leads to energy depletion thus decreases the network lifetime. Therefore, enhancing

data aggregation scheme helps to prolong the network lifetime and reduce the communication overhead.

4. Designing efficient and effective anomaly detection based on feature selection, dimension reduction, and data aggregation technique helps in reducing the energy consumption, therefore, prolong the network lifetime.

5. Taking the advantages of the spatial correlation between the nodes to enhance the detection efficiency and effectiveness, thus minimizing the data transmission as well as the anomalous data and important event can be discriminated.

6. Application of this research findings can benefit few application domains. One example is palm oil industry in term of fast detection of palm oil tree diseases such as Ganoderma Boninence fungus in the soil. Detecting the presence of Ganoderma Boninence at the early stage can lead to early treatment of the infected trees and reduce the revenue loss.

1.9 Definition of Terms

The definition of the terms used in this research as follows:

1. Detection Effectiveness;

   The detection accuracy and false alarm rate reflect the effectiveness of a detection method. The detection accuracy rate is calculated by dividing the number of correctly identified anomalies by the total number of anomalies. A false alarm is made up of two parts: a false positive and a false negative. A false positive occurs when a legitimate event is mistakenly classified as an abnormality, whereas a false negative occurs when a true anomaly is missed. The false alarm rate is calculated by dividing the number of false alarms by the number of reported anomalies (Xie *et. al.*, 2011).

2. Detection Efficiency;

   Detection efficiency describes the minimum usage of the resource including energy consumption and memory capacity of sensor nodes in the network. Communication overhead, computational complexity, and memory utilization are measurements used to evaluate detection efficiency.

3. Lightweight detection;

   Lightweight detection describes the reduction of computation complexity during the process of anomaly detection that leads to reduction of energy consumption. By reducing thus prolonging the network lifetime.

4. Robustness;

   Robustness describes the success of data transmission in the network structure to the final destination in case of link/device failures. The robust structure can increase the detection accuracy.

1.10 Thesis Organization

This thesis is organized into seven chapters. Chapter 1 is the introduction of the research and it provides the problem background, objective, the importance of the research and its aim. Chapter 2 is the literature review which discusses the previous work of related to feature selection, data aggregation, WSNs, anomaly detection and Ganoderma disease as the case study for anomaly detection scheme. The research methodology that outlines the research framework into three phases is presented in Chapter 3. Chapter 4 discusses the proposed energy efficient scheme achieved by using feature selection and data aggregation approaches. The lightweight anomaly detection based on one-class support vector machine (OCSVM) and Candid Covariance-Free Incremental Principal Component Analysis (CCIPCA) and the distributed anomaly detection scheme are presented in Chapter 5 and Chapter 6 respectively. Lastly, the Chapter 7 concludes this thesis.

REFERENCES

Abbas, Z. and Burney, A. (2016) 'A Survey of Software Packages Used for Rough Set Analysis', *Journal of Computer and Communications*, 4(July), pp. 10–18.

Abdallah, W., Sarhan, S. and Elmougy, S. (2016) 'REDUCING ATTRIBUTES of Reducing Attributes Of Facebook Users Using Rough Set Theory', *International Journal of Intelligent Computing and Information Sciences*, 16(4), pp. 29–40.

Abdullah, a. H., Adom, a. H., Shakaff, a. Y. M., Ahmad, M. N., Zakaria, a., Saad, F. S. a., Isa, C. M. N. ., Masnan, M. J. and Kamarudin, L. M. (2012) 'Hand-Held Electronic Nose Sensor Selection System for Basal Stamp Rot (BSR) Disease Detection', *2012 Third International Conference on Intelligent Systems Modelling and Simulation*. Ieee, pp. 737–742.

Abid, A., Khediri, S. El and Kachouri, A. (2021) 'Improved approaches for density-based outlier detection in wireless sensor networks', *Computing*. Springer Vienna.

Agrawal, P., Abutarboush, H. F., Ganesh, T. and Mohamed, A. W. (2021) 'Metaheuristic Algorithms on feature selection: A survey of one decade of research (2009-2019)', *IEEE Access*, 9, pp. 1–1.

Ahmed, M., Naser Mahmood, A. and Hu, J. (2016) 'A survey of network anomaly detection techniques', *Journal of Network and Computer Applications*. Elsevier, 60, pp. 19–31.

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. (2002) 'Wireless sensor networks : a survey', *Computer Networks*, 38, pp. 393–422.

Al-Shalabi, M., Anbar, M., Wan, T. C. and Khasawneh, A. (2018) 'Variants of the low-energy adaptive clustering hierarchy protocol: Survey, issues and challenges', *Electronics (Switzerland)*, 7(8), pp. 1–28.

Al-Baz, A. and El-Sayed, A. (2017) 'A new algorithm for cluster head selection in LEACH protocol for wireless sensor networks', *International Journal of Communication System*, 31(1), pp. 1–33.

Alam, S., Sonbhadra, S. K., Agarwal, S. and Nagabhushan, P. (2020) 'One-class support vector classifiers: A survey', *Knowledge-Based Systems*. Elsevier

B.V., 196, p. 105754.

Aldweesh, A., Derhab, A. and Emam, A. Z. (2020) ‘Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues’, *Knowledge-Based Systems*. Elsevier B.V., 189, p. 105124.

Ali, A., Ming, Y., Chakraborty, S. and Iram, S. (2017) ‘A comprehensive survey on real-time applications of WSN’, *Future Internet*, 9(4).

Alikhashashneh, E. A. and Al-Radaideh, Q. A. (2013) ‘Evaluation of discernibility matrix based reduct computation techniques’, in *5th International Conference on Computer Science and Information Technology, CSIT 2013 - Proceedings*, pp. 76–81.

Almeida, F. R., Brayner, A., Rodrigues, J. P. C. and Bessa Maia, J. E. (2017) ‘Improving multidimensionalwireless sensor network lifetime using pearson correlation and fractal clustering’, *Sensors (Switzerland)*, 17(6), pp. 1–24.

Alshammari, H., Ghorbel, O., Aseeri, M. and Abid, M. (2018) ‘Non-Negative Matrix Factorization (NMF) for outlier detection in Wireless Sensor Networks’, *2018 14th International Wireless Communications and Mobile Computing Conference, IWCMC 2018*. IEEE, pp. 506–511.

Amidi, A., Hamm, N. A. S. and Meratnia, N. (2013) ‘Wireless Sensor Networks and Fusion of Contextual Information for Weather Outlier Detection’, in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 37–41.

Anastasi, G., Conti, M., Di Francesco, M. and Passarella, A. (2009) ‘Energy conservation in wireless sensor networks: A survey’, *Ad Hoc Networks*. Elsevier B.V., 7(3), pp. 537–568.

Anisi, M. H. (2011) ‘Energy-Efficient Data Collection in Wireless Sensor Networks’, *Wireless Sensor Network*, 03(10), pp. 329–333.

Ardakani, S. P. (2018) ‘Data Aggregation Routing Protocols in Wireless Sensor Networks : A Taxonomy’, *International Journal of Computer Networks & Communications*, 9(March 2017), pp. 90–107.

Arjunan, S. and Pothula, S. (2019) ‘A survey on unequal clustering protocols in Wireless Sensor Networks’, *Journal of King Saud University - Computer and Information Sciences*. King Saud University, 31(3), pp. 304–317.

Arthi.A (2014) ‘Analysis of Secure In-Network Aggregation for Anomaly Detection in Wireless Sensor Networks’, *International Journal of Innovative Research in*

Computer and Communication Engineering, 2(1).

Arumugam, G. S. and Ponnuchamy, T. (2015) 'EE-LEACH: development of energy-efficient LEACH Protocol for data gathering in WSN', *EURASIP Journal on Wireless Communications and Networking*, 2015(1), p. 76.

Asy Syura, M. and Tsan, F. Y. (2012) 'The Impact of Organic Fertilizer Application to Oil Palm Production in FELDA Maokil 7: A Case Study', in *International Plantation Industry Conference & Exibition*, pp. 1–8.

Atzori, L., Iera, A. and Morabito, G. (2010) 'The Internet of Things : A survey', *Computer Networks*. Elsevier B.V., 54(15), pp. 2787–2805.

Awan, A. M. and Sap, M. N. (2006) 'A Framework for Predicting Oil-Palm Yield from Climate Data'.

Ayadi, A., Ghorbel, O., BenSalah, M. S. and Abid, M. (2020) 'Spatio-temporal correlations for damages identification and localization in water pipeline systems based on WSNs', *Computer Networks*. Elsevier B.V., 171(December 2019), p. 107134.

Ayadi, A., Ghorbel, O., M. S. Bensaleh, Obeid, A. and Abid, M. (2017) 'Performance of outlier detection techniques based classification in Wireless Sensor Networks', in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, pp. 687–692.

Ayadi, A., Ghorbel, O., Obeid, A. M. and Abid, M. (2017) 'Outlier detection approaches for wireless sensor networks: A survey', *Computer Networks*. Elsevier B.V., 129, pp. 319–333.

Azim, A. and Islam, M. M. (2009) 'Hybrid LEACH: A relay node based low energy adaptive clustering hierarchy for wireless sensor networks', in *Communications (MICC), 2009 IEEE 9th Malaysia International Conference on*, pp. 911–916.

Baggio, A. (2005) 'Wireless sensor networks in precision agriculture.', in *ACM Workshop on Real-World Wireless Sensor Networks (REALWSN 2005)*.

Banu, S. S. and Balasubadra, K. (2018) 'Revealing Malicious Events Against Inside Attacker in Wireless Sensor Networks', 14, pp. 134–143.

Barakkath Nisha, U., Subair, A. and Yasir Abdullah, R. (2020) 'An Efficient Algorithm for Anomaly Detection in Wireless Sensor Networks', *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, (Icosec), pp. 925–932.

Batra, P. K. and Kant, K. (2016) 'LEACH-MAC : a new cluster head selection algorithm for Wireless Sensor Networks', *Wireless Networks*. Springer US, pp. 49–60.

Bazan, J. G. and Szczuka, M. (2005) 'The Rough Set Exploration System', pp. 37–56.

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B. and Rätsch, G. (2008) 'Support vector machines and kernels for computational biology', *PLoS Computational Biology*, 4(10).

Bengio, Y. (2009) *Learning Deep Architectures for AI*.

Bensaleh, M. S., Saida, R., Kacem, Y. H. and Abid, M. (2020) 'Wireless Sensor Network Design Methodologies: A Survey', *Journal of Sensors*, 2020.

Bharuka, K. and Jinwala, D. C. (2014) 'A Secure Data Aggregation protocol for Outlier Detection in Wireless Sensor Networks Using Aggregate Message Authentication Code', in *Industrial and Information Systems (ICIIS), 2014 9th International Conference on*. IEEE, pp. 1–6.

Bhavsar, P., Jhunjhunwala, P. and D'Mello, L. (2020) 'Attribute Reduction for Medical Data Analysis Using Rough Set Theory', in *Advanced Computing Technologies and Applications, Algorithms for Intelligent Systems*, pp. 325–335.

Bosman, H. H. W. J., Iacca, G., Tejada, A., Wörtche, H. J. and Liotta, A. (2015) 'Ensembles of incremental learners to detect anomalies in ad hoc sensor networks', *Ad Hoc Networks*, 35, pp. 14–36.

Branch, J. W., Giannella, C., Szymanski, B., Wolff, R. and Kargupta, H. (2012) *In-network outlier detection in wireless sensor networks, Knowledge and Information Systems*.

Buratti, C., Conti, A., Dardari, D. and Verdone, R. (2009) 'An overview on wireless sensor networks technology and evolution.', *Sensors (Basel, Switzerland)*, 9(9), pp. 6869–96.

Burney, A. and Abbas, Z. (2015) 'Applications of Rough Sets in Health Sciences and Disease Diagnosis', *Recent Researches in Applied Computer Science*, (May 2015), pp. 153–161.

Cai, S., Gallina, B., Nyström, D. and Seceleanu, C. (2018) 'Data aggregation processes : a survey , a taxonomy , and design guidelines', *Computing*. Springer Vienna.

Carvalho, C., Gomes, D. G., Agoulmine, N. and de Souza, J. N. (2011) 'Improving prediction accuracy for WSN data reduction by applying multivariate spatio-temporal correlation', *Sensors,* 11(11), pp. 10010–10037.

Castillo, E., Peteiro-Barral, D., Berdiñas, B. G. and Fontenla-Romero, O. (2015) 'Distributed One-Class Support Vector Machine', *International Journal of Neural Systems*, 25(07), p. 1550029.

Chan, L., Gomez Chavez, K., Rudolph, H. and Hourani, A. (2020) 'Hierarchical routing protocols for wireless sensor network: a compressive survey', *Wireless Networks*. Springer US, 0123456789.

Chander, B. and Kumaravelan, G. (2021) 'Outlier Detection Strategies for WSNs : A Survey', *Journal of King Saud University - Computer and Information Sciences*. King Saud University.

Chandola, V. (2009) 'Anomaly Detection : A Survey', *ACM Computing Surveys,* 41(3), pp. 1–72.

Chandola, V., Banerjee, A. and Kumar, V. (2009) 'Anomaly detection: A Survey', *ACM Computing Surveys*, 41(3), pp. 1–58.

Chandola, V., Banerjee, A. and Kumar, V. (2012) 'Anomaly Detection for Discrete Sequences : A Survey', *IEEE Transactions on Knowledge and Data Engineering*, 24(5), pp. 823–839.

Chandrashekar, G. and Sahin, F. (2014) 'A survey on feature selection methods', *Computers and Electrical Engineering*. Elsevier Ltd, 40(1), pp. 16–28.

Chen, F., Wen, F. and Jia, H. (2010) 'Algorithm of data compression based on multiple principal component analysis over the WSN', *2010 6th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2010*, pp. 2–5.

Chen, Y. and Li, S. (2019) 'A Lightweight Anomaly Detection Method Based on SVDD for Wireless Sensor Networks', *Wireless Personal Communications*. Springer US, 105(4), pp. 1235–1256.

Cheng, P. and Zhu, M. (2015) 'Lightweight Anomaly Detection for Wireless Sensor Networks', *International Journal of Distributed Sensor Networks*, 2015(1).

Cheng, Y., Yu, L., Xu, Y., Lu, H., Cracknell, A. P., Kanniah, K. and Gong, P. (2018) 'Mapping oil palm extent in Malaysia using ALOS-2 PALSAR-2 data', *International Journal of Remote Sensing*, 39(2), pp. 432–452.

Chitradevi, N., Palanisamy, V., Baskaran, K. and Nisha, U. B. (2010) 'Outlier aware data aggregation in distributed wireless sensor network using robust principal component analysis', *2010 Second International conference on Computing, Communication and Networking Technologies*. Ieee, pp. 1–9.

Cho, S., Han, L., Joo, B. and Han, S. (2014) 'P-LEACH: An Efficient Cluster-Based Technique to Track Mobile Sinks in Wireless Sensor Networks', *International Journal of Distributed Sensor Networks,* 2014.

Chouikhi, S., El Korbi, I., Ghamri-Doudane, Y. and Azouz Saidane, L. (2015) 'A survey on fault tolerance in small and large scale wireless sensor networks', *Computer Communications*. Elsevier Ltd., 69, pp. 22–37.

Cooper, R. M., Flood, J. and Rees, R. W. (2011) 'Ganoderma boninense in Oil Palm Plantations : Current Thinking on Epidemiology , Resistance and Pathology *', in *Sustainable Agriculture - An Insight into Ganoderma,* pp. 515–526.

Daanoune, I., Abdennaceur, B. and Ballouk, A. (2021) 'A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks', *Ad Hoc Networks*. Elsevier B.V., 114(July 2020), p. 102409.

Díaz, S. E., Pérez, J. C., Mateos, A. C., Marinescu, M.-C. and Guerra, B. B. (2011) 'A novel methodology for the monitoring of the agricultural production process based on wireless sensor networks', *Computers and Electronics in Agriculture*. Elsevier B.V., 76(2), pp. 252–265.

Dobbin, K. K. and Simon, R. M. (2011) 'Optimally splitting cases for training and testing high dimensional classifiers', *BMC Medical Genomics,* 4.

Eddine, D., Bouabdallah, A. and Lakhlef, H. (2018) 'Internet of things security : A top-down survey', *Computer Networks*. Elsevier B.V., 141, pp. 199–221.

Ejaz, W. and Anpalagan, A. (2019) 'Dimension Reduction for Big Data', in *SpringerBriefs in Electrical and Computer Engineering,* pp. 31–37.

Erfani, S. M., Baktashmotlagh, M., Rajasegarar, S., Karunasekera, S. and Leckie, C. (2015) 'R1SVM: A randomised nonlinear approach to large-scale anomaly detection', in *Proceedings of the National Conference on Artificial Intelligence*, pp. 432–438.

Erfani, S. M., Rajasegarar, S., Karunasekera, S. and Leckie, C. (2016) 'High-Dimensional and Large-Scale Anomaly Detection using a Linear One-Class SVM with Deep Learning', *Pattern Recognition*. Elsevier, 58, pp. 121–134.

Erhan, L., Ndubuaku, M., Di Mauro, M., Song, W., Chen, M., Fortino, G., Bagdasar, O. and Liotta, A. (2021) 'Smart anomaly detection in sensor systems: A multi-perspective review', *Information Fusion*. Elsevier B.V., 67(September 2020), pp. 64–79.

Fan, K.-W., Liu, S. and Sinha, P. (2007) 'Structure-Free Data Aggregation in Sensor Networks', *IEEE Transactions on Mobile Computing*, 6(8), pp. 929–942.

Fan, Y. and Chen, A. L. P. (2010) 'Efficient and Robust Schemes for Sensor Data Aggregation Based on Linear Counting', 21(11), pp. 1675–1691.

Fan, Y. and Chen, A. L. P. (2012) 'Energy Efficient Schemes for Accuracy-Guaranteed Sensor Data Aggregation Using Scalable Counting', *Transactions on Knowledge and Data Engineering*, 24(8), pp. 1463–1477.

Fasolo, E., Rossi, M., Widmer, J. and Zorz, M. (2007) 'In-network aggregation techniques for wireless sensor networks: a survey', *Wireless Communications,IEEE*, 14(2), pp. 70–87.

Fawzy, A., Mokhtar, H. M. O. O. and Hegazy, O. (2013) 'Outliers detection and classification in wireless sensor networks', *Egyptian Informatics Journal*. Ministry of Higher Education and Scientific Research, 14(2), pp. 157–164.

Feng, Z., Fu, J., Du, D., Li, F. and Sun, S. (2017) 'A new approach of anomaly detection in wireless sensor networks using support vector data description', *International Journal of Distributed Sensor Networks*, 13(1).

Ferentinos, K. P., Katsoulas, N., Tzounis, A., Bartzanas, T. and Kittas, C. (2017) 'Wireless sensor networks for greenhouse climate and plant condition assessment', *Biosystems Engineering*. Elsevier Ltd, 153, pp. 70–81.

Fowotade, S. A., Yusof, N. A., Abdullah, J., Sulaiman, Y. and Abd Rahman, S. F. (2019) 'Enhanced electrochemical sensing of secondary metabolites in oil palms for early detection of Ganoderma boninense based on novel nanoparticle-chitosan functionalized multi-walled carbon nanotube platform', *Sensing and Bio-Sensing Research*. Elsevier, 23(November 2018), p. 100274.

Francis, A. and Babu, H. K. (2014) 'Extended Faulty Nodes and Event Identification for WSNs Using IDM and SMM', 3(3), pp. 5692–5695.

Gao, C., Chen, Y., Wang, Z., Xia, H. and Lv, N. (2020) 'Anomaly detection frameworks for outlier and pattern anomaly of time series in wireless sensor networks', *Proceedings - 2020 International Conference on Networking and Network Applications, NaNA 2020*, pp. 229–232.

Garcia-Sanchez, A.-J., Garcia-Sanchez, F. and Garcia-Haro, J. (2011) 'Wireless sensor network deployment for integrating video-surveillance and data-monitoring in precision agriculture over distributed crops', *Computers and Electronics in*

Agriculture. Elsevier B.V., 75(2), pp. 288–303.

García, S., Ramírez-gallego, S., Luengo, J., Benítez, J. M. and Herrera, F. (2016) 'Big data preprocessing : methods and prospects', *Big Data Analytics*, 1(9), pp. 1–22.

Ghorbel, O. and Abid, M. (2015) 'A novel outlier detection model based on One Class Principal Component Classifier in Wireless Sensor Networks', pp. 70–76.

Ghorbel, O., Ayadi, A. and Loukil, K. (2017) 'Classification Data using Outlier Detection Method in Wireless Sensor Networks', *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, pp. 699–704.

Ghorbel, O., Ayedi, W., Snoussi, H. and Abid, M. (2015) 'Fast and efficient outlier detection method in wireless sensor networks', *IEEE Sensors Journal*, 15(6), pp. 3403–3411.

Gogoi, P., Bhattacharyya, D. K., Borah, B. and Kalita, J. K. (2011) 'A Survey of Outlier Detection Methods in Network Anomaly Identification', *The Computer Journal*, 54(4), pp. 570–588.

Gomathi, S. and Gopala Krishnan, C. (2020) 'Malicious Node Detection in Wireless Sensor Networks Using an Efficient Secure Data Aggregation Protocol', *Wireless Personal Communications*. Springer US, 113(4), pp. 1775–1790.

Gómez-Verdejo, V., Arenas-García, J., Lázaro-Gredilla, M. and Navia-Vázquez, Á. (2011) 'Adaptive one-class support vector machine', *IEEE Transactions on Signal Processing*, 59(6), pp. 2975–2981.

Guo, S., Zhang, H., Zhong, Z., Chen, J., Cao, Q. and He, T. (2014) 'Detecting Faulty Nodes with Data Errors for Wireless Sensor Networks', *ACM Transactions on Sensor Networks*.

Hancer, E., Xue, B. and Zhang, M. (2020) 'A survey on feature selection approaches for clustering', *Artificial Intelligence Review*. Springer Netherlands, 53(6), pp. 4519–4545.

Hassanien, A. E., Abdelhafez, M. E. and Own, H. S. (2008) 'Rough Sets Data Analysis in Knowledge Discovery : A Case of Kuwaiti Diabetic Children Patients', *Advances in Fuzzy Systems*, 2008.

He, Q., Zhang, Q., Wang, H. and Zhang, C. (2020) 'Local Similarity-Based Fuzzy Multiple Kernel One-Class Support Vector Machine', *Complexity*, 2020(1).

Heinzelman, W. B., Chandrakasan, A. P., Member, S. and Balakrishnan, H. (2002)

‘An Application-Specific Protocol Architecture for Wireless Microsensor Networks’, 1(4), pp. 660–670.

Heinzelman, W. R., Chandrakasan, A. and Balakrishnan, H. (2000) ‘Energy-Efficient Communication Protocol for Wireless Microsensor Networks’, 00(c), pp. 1–10.

Hofmann, T., Schölkopf, B. and Smola, A. J. (2008) ‘Kernel methods in machine learning’, *Annals of Statistics,* 36(3), pp. 1171–1220.

Hsu, H., Hsieh, C. and Lu, M. (2011) ‘Hybrid feature selection by combining filters and wrappers’, *Expert Systems With Applications*. Elsevier Ltd, 38(7), pp. 8144–8150.

Huang, W., Ling, Y. and Zhou, W. (2018) ‘An Improved LEACH Routing Algorithm for Wireless Sensor Network’, *International Journal of Wireless Information Networks*. Springer New York LLC, 25(3), pp. 323–331.

Hushiarian, R., Yusof, N. A. and Dutse, S. W. (2013) ‘Detection and control of Ganoderma boninense: strategies and perspectives.’, *SpringerPlus,* 2(1), p. 555.

Idris, A. S., Kushairi, A., Ismail, S. and Ariffin, D. (2004) ‘Selection For Partial Resistance In Oil Palm Progenies To Ganoderma Basal Stem Rot’, *Journal of Oil Palm Research*, 16(2), pp. 12–18.

Idris, A. S., Mazliham, M. S., Madihah, A. . and Kushairi, A. (2010) ‘Field Recognition and Diagnosis Tools of Ganoderma Disease in Oil Palm’, in *Second International Seminar Oil Palm Disease Advance in Ganoderma Research and Management,* pp. 1–13.

Idris, A. S., Yamaoka, M., Hayakawa, S., Basri, M. W., Noorhasimah, I. and Ariffin, D. (2003) ‘PCR Technique For Detection Of Ganoderma’, *MPOB Information Series,* pp. 1–4.

Inbarani, H. H., Azar, A. T. and Jothi, G. (2014) ‘Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis’, *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd, 113(1), pp. 175–185.

Jerbi, W., Guermazi, A. and Trabelsi, H. (2016) ‘O-LEACH of Routing Protocol for Wireless Sensor Networks’, in *13th International Conference Computer Graphics, Imaging and Visualization*. IEEE, pp. 399–404.

Jia, X., Rao, Y., Shang, L. and Li, T. (2020) ‘Similarity-based attribute reduction in rough set theory: a clustering perspective’, *International Journal of Machine*

Learning and Cybernetics. Springer Berlin Heidelberg, 11(5), pp. 1047–1060.

Jimenez, M., Jimenez, A., Lozada, P., Jimenez, S. and Jimenez, C. (2013) ‘Using a Wireless Sensors Network in the Sustainable Management of African Palm Oil Solid Waste’, *2013 10th International Conference on Information Technology: New Generations*. Ieee, pp. 133–137.

Johnson, D. S. (1974) ‘Approximation algorithms for combinatorial problems’, *Journal of Computer and System Sciences*, 9(3), pp. 256–278.

Jovi, A., Brki, K. and Bogunovi, N. (2015) ‘A review of feature selection methods with applications’, in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. MIPRO, pp. 1200–1205.

Kandris, D., Nakas, C., Vomvas, D. and Koulouras, G. (2020) ‘Applications of wireless sensor networks: An up-to-date survey’, *Applied System Innovation*, 3(1), pp. 1–24.

Kannadhasan, S., Sangeetha, J. and Niranjana, K. (2014) ‘A Novel Approach Based Anomaly Detection for Secure Data Aggregation Using HEED Algorithm’, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2(3), pp. 749–751.

Kasno, A. and Subardja, D. (2006) ‘Soil Fertility and Nutrient Management on Spodosol for Oil Palm’, (98), pp. 1–8.

Kaur, M. and Munjal, A. (2020) ‘Data aggregation algorithms for wireless sensor network: A review’, *Ad Hoc Networks*. Elsevier B.V., 100.

Khan, O., Khan, F. G., Nazir, B. and Wazir, U. (2016) ‘Energy Efficient Routing Protocols in Wireless Sensor Networks : A survey’, *International Journal of Computer Science and Information Security*, 14(6), pp. 398–406.

Kim, G., Lee, S. and Kim, S. (2014) ‘A novel hybrid intrusion detection method integrating anomaly detection with misuse detection’, *Expert Systems with Applications*. Elsevier Ltd, 41(4), pp. 1690–1700.

Kumar, S. and Chaurasiya, V. K. (2019) ‘A Strategy for Elimination of Data Redundancy in Internet of Things (IoT) Based Wireless Sensor Network (WSN)’, *IEEE Systems Journal*, 13(2), pp. 1650–1657.

Kumar, V., Sharma, V. and Sachdeva, M. (2016) ‘A survey on LEACH and other ’ s routing protocols in wireless sensor network’, *Optik - International Journal for Light and Electron Optics*. Elsevier GmbH., 127(16), pp. 6590–6600.

Kumar, V. and Sonajharia, M. (2014) ‘Feature Selection: A literature Review’, *The Smart Computing Review*, 4(3), pp. 211–229.

Lalwani, P., Das, S., Banka, H. and Kumar, C. (2016) ‘CRHS : clustering and routing in wireless sensor networks using harmony search algorithm’, *Neural Computing and Applications*. Springer London.

Laskov, P., Schäfer, C., Kotenko, I. and Müller, K.-R. (2004) ‘Intrusion detection in unlabeled data with quarter-sphere Support Vector Machines’, in *Detection of Intrusion and Malware & Vulnerability Assessment*.

Lee, J. Y., Jung, K. D., Moon, S. J. and Jeong, H. Y. (2017) ‘Improvement on LEACH protocol of a wide-area wireless sensor network’, *Multimedia Tools and Applications*. Multimedia Tools and Applications, 76(19), pp. 19843–19860.

Lelong, C. C. D., Roger, J.-M., Brégand, S., Dubertret, F., Lanore, M., Sitorus, N. a, Raharjo, D. a and Caliman, J.-P. (2010) ‘Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data.’, *Sensors (Basel, Switzerland)*, 10(1), pp. 734–47.

Li, C., Tan, G., Wu, J., Zhang, Z. and Xu, L. (2011) ‘Analyzing Cluster-head Selection Mechanisms and Improving the LEACH’, (61001068), pp. 747–750.

Li, G., Peng, S., Wang, C., Niu, J. and Yuan, Y. (2019) ‘An Energy-Efficient Data Collection Scheme Using Denoising Autoencoder in Wireless Sensor Networks’, *Tsinghua Science and Technology*, 24(1), pp. 86–96.

Li, H., Lin, K. and Li, K. (2011) ‘Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks’, *Computer Communications*. Elsevier B.V., 34(4), pp. 591–597.

Li, X., Liu, W., Xie, M., Liu, A., Zhao, Ming, Xiong, N. N., Zhao, Miao and Dai, W. (2018) ‘Differentiated data aggregation routing scheme for energy conserving and delay sensitive wireless sensor networks’, *Sensors (Switzerland)*, 18(7).

Li, Y., Li, T. and Liu, H. (2017) ‘Recent advances in feature selection and its applications’, *Knowledge and Information Systems*. Springer London, 53(3), pp. 551–577.

Li, Y., Wang, J., Tian, Z., Lu, T. and Young, C. (2009) ‘Building lightweight intrusion detection system using wrapper-based feature selection mechanisms’, *Computers & Security*. Elsevier Ltd, 28(6), pp. 466–475.

Liao, M.-S., Chuang, C.-L., Lin, T.-S., Chen, C.-P., Zheng, X.-Y., Chen, P.-T., Liao, K.-C. and Jiang, J.-A. (2012) ‘Development of an autonomous early warning system for Bactrocera dorsalis (Hendel) outbreaks in remote fruit orchards', *Computers and Electronics in Agriculture*. Elsevier B.V., 88, pp. 1–12.

Lin, C., Wu, G., Xia, F., Li, M., Yao, L. and Pei, Z. (2012) 'Energy efficient ant colony algorithms for data aggregation in wireless sensor networks', *Journal of Computer and System Sciences*. Elsevier Inc., 78(6), pp. 1686–1702.

Liu, B. (2018) *Wireless Sensor Network Applications in Precision Agriculture, Journal of Agricultural Systems, Technology, and Management*.

Liu, W., Hua, G. and Smith, J. R. (2014) 'Unsupervised one-class learning for automatic outlier removal', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Liu, X., Yu, J., Li, F., Lv, W., Wang, Y. and Cheng, X. (2020) 'Data Aggregation in Wireless Sensor Networks: From the Perspective of Security', *IEEE Internet of Things Journal*. IEEE, 7(7), pp. 6495–6513.

Lorena, L. H. N., Carvalho, A. C. P. L. F. and Lorena, A. C. (2015) 'Filter Feature Selection for One-Class Classification', *Journal of Intelligent and Robotic Systems: Theory and Applications*.

Maivizhi, R. and Yogesh, P. (2020) 'Secure in-network aggregation in wireless sensor networks', *International Journal of Intelligent Information Technologies*, 16(1), pp. 49–74.

Maldonado, S. and Weber, R. (2009) 'A wrapper method for feature selection using Support Vector Machines', *Information Sciences*. Elsevier Inc., 179(13), pp. 2208–2217.

Manbari, Z., AkhlaghianTab, F. and Salavati, C. (2019) 'Hybrid fast unsupervised feature selection for high-dimensional data', *Expert Systems with Applications*. Elsevier Ltd, 124, pp. 97–118.

Mao, Y. and Yang, Y. (2019) 'A Wrapper Feature Subset Selection Method Based on Randomized Search and Multilayer Structure', *BioMed Research International*, 2019.

Maraiya, K., Kant, K. and Gupta, N. (2011) 'Efficient Cluster Head Selection Scheme for Data Aggregation in Wireless Sensor Network', *International Journal of Computer Applications*, 23(9), pp. 10–18.

Marappan, P. and Rodrigues, P. (2016) 'An energy efficient routing protocol for correlated data using CL-LEACH in WSN', *Wireless Networks*. Springer US, 22(4), pp. 1415–1423.

Markom, M. a., Shakaff, a. Y. M., Adom, a. H., Ahmad, M. N., Hidayat, W., Abdullah, a. H. and Fikri, N. A. (2009) 'Intelligent electronic nose system for basal stem rot disease detection', *Computers and Electronics in Agriculture*, 66(2), pp. 140–146.

Masdari, M., Bazarchi, S. M. and Bidaki, M. (2013) 'Analysis of Secure LEACH-Based Clustering Protocols in Wireless Sensor Networks', *Journal of Network and Computer Applications*. Elsevier, 36(4), pp. 1243–1260.

Mashere, M. P., Barve, S. S. and Ganjewar, P. D. (2015) 'Data Reduction in Wireless Sensor Network : A Survey', *International Journal of Computer Science and Technology*, 8491, pp. 86–88.

Matos, T. B., Brayner, A. and Maia, J. E. B. (2010) 'Towards in-network data prediction in wireless sensor networks', p. 592.

Maurya, P. and Kaur, A. (2016) 'A Survey on Descendants of LEACH Protocol', *I.J. Information Engineering and Electronic Business*, 2(March), pp. 46–58.

Maurya, S. and Jain, V. K. (2016) 'Fuzzy based energy efficient sensor network protocol for precision agriculture', *Computers and Electronics in Agriculture*. Elsevier B.V., 130, pp. 20–37.

Maya, S., Ueno, K. and Nishikawa, T. (2019) 'dLSTM: a new approach for anomaly detection using deep learning with delayed prediction', *International Journal of Data Science and Analytics*. Springer International Publishing, 8(2), pp. 137–164.

Mohd Su'ud, M., Loonis, P. and Idris, A. S. (2007) 'Towards Automatic Recognition and Grading of Ganoderma Infection Pattern Using Fuzzy Systems', pp. 30–35.

Morais, R., Valente, A. and Serôdio, C. (2005) 'A Wireless Sensor Network for Smart Irrigation and Environmental Monitoring : A Position Article', (July), pp. 845–850.

Moshtaghi, M., Bezdek, J. C., Fellow, L. and Leckie, C. (2015) 'Evolving Fuzzy Rules for Anomaly Detection in Data Streams', 23(3), pp. 688–700.

Moshtaghi, M., Bezdek, J. C., Havens, T. C., Leckie, C., Karunasekera, S., Rajasegarar, S. and Palaniswami, M. (2012) 'Streaming analysis in wireless sensor networks'.

Moshtaghi, M., Havens, T. C., Bezdek, J. C., Park, L., Leckie, C., Rajasegarar, S., Keller, J. M. and Palaniswami, M. (2011) 'Clustering ellipses for anomaly detection', *Pattern Recognition*. Elsevier, 44(1), pp. 55–69.

Moshtaghi, M., Leckie, C., Karunasekera, S., Bezdek, J. C., Rajasegarar, S. and Palaniswami, M. (2011) 'Incremental Elliptical Boundary Estimation for Anomaly Detection in Wireless Sensor Networks', *2011 IEEE 11th International Conference on Data Mining*. Ieee, (1), pp. 467–476.

Moshtaghi, M., Leckie, C., Karunasekera, S. and Rajasegarar, S. (2014) 'An adaptive elliptical anomaly detection model for wireless sensor networks', *Computer Networks*. Elsevier B.V., 64, pp. 195–207.

Muhamad, H., Ramli, M. I., Zakaria, Z. and Sahid, I. (2013) 'The Fate Of Diuron In Soil In A Malaysian Oil Palm Plantation', *Journal of Oil Palm Research*, 25(April), pp. 149–158.

Nancy, P., Muthurajkumar, S., Ganapathy, S., Santhosh Kumar, S. V. N., Selvi, M. and Arputharaj, K. (2020) 'Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks', *IET Communications*, 14(5), pp. 888–895.

Nisha, U. B., Umamaheswari, N., Venkatesh, R. and Abdullah, R. Y. (2014) 'Statistical Based Outlier Detection in Data Aggregation for Wireless Sensor Networks', *Journal of Theoretical and Applied Information Technology*, 59(3), pp. 770–780.

Øhrn, A. and Rowland, T. (2000) 'Rough sets: A knowledge discovery technique for multifactorial medical outcomes', *American Journal of Physical Medicine and Rehabilitation*, 79(1), pp. 100–108.

Ojha, T., Misra, S. and Raghuwanshi, N. S. (2015) 'Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges', *Computers and Electronics in Agriculture*. Elsevier B.V., 118, pp. 66–84.

Olson, D. L. and Delen, D. (2008) 'Rough Sets', in *Advanced Data Mining Techniques*. Springer, pp. 87–109.

Osanaiye, O., Ogundile, O., Aina, F. and Periola, A. (2019) 'Feature selection for intrusion detection system in a cluster-based heterogeneous wireless sensor network', *Facta universitatis - series: Electronics and Energetics*, 32(2), pp. 315–330.

Otoum, S., Kantarci, B. and Mouftah, H. (2018) 'Adaptively supervised and intrusion-aware data aggregation for wireless sensor clusters in critical infrastructures', *IEEE International Conference on Communications*. IEEE, 2018-May, pp. 1–

6.

Pandey, V., Kaur, A. and Chand, N. (2010) ‘A review on data aggregation techniques in wireless sensor network’, 1(2), pp. 1–8.

Pawlak, Z. (1998) ‘Rough set theory and its applications’, *Journal of Telecommunications and Information Technology*, 29(7), pp. 7–10.

Pięta, P., Szmuc, T. and Kluza, K. (2019) ‘Comparative Overview of Rough Set Toolkit Systems for Data Analysis’, *MATEC Web of Conferences*, 252, p. 03019.

Poornima, I. G. A. and Paramasivan, B. (2020) ‘Anomaly detection in wireless sensor network using machine learning algorithm’, *Computer Communications*. Elsevier B.V., 151(December 2019), pp. 331–337.

Punmiya, R., Zyabkina, O., Choe, S. and Meyer, J. (2019) ‘Anomaly detection in power quality measurements using proximity-based unsupervised machine learning techniques’, *2019 Electric Power Quality and Supply Reliability Conference and 2019 Symposium on Electrical Engineering and Mechatronics, PQ and SEEM 2019*. IEEE.

Rajagopalan, R. and Varshney, P. K. (2006) ‘Data aggregation techniques in sensor networks : A survey’, *IEEE Communications Surveys and Tutorials*, 8(4).

Rajasegarar, S., Bezdek, J. C., Leckie, C. and Palaniswami, M. (2009) ‘Elliptical anomalies in wireless sensor networks’, *ACM Transactions on Sensor Networks*, 6(1), pp. 1–28.

Rajasegarar, S., Gluhak, A., Ali, M., Nati, M., Moshtaghi, M., Leckie, C. and Palaniswami, M. (2014) ‘Ellipsoidal neighbourhood outlier factor for distributed anomaly detection in resource constrained networks’, *Pattern Recognition*. Elsevier, 47(9), pp. 2867–2879.

Rajasegarar, S., Leckie, C. and Bezdek, J. C. (2010) ‘Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks’, *IEEE Transactions on Information Forensics and Security*, 5(3), pp. 518–533.

Rajasegarar, S., Leckie, C. and Palaniswami, M. (2008a) ‘Anomaly Detection In Wireless Sensor Networks’, *IEEE Wireless Communications*, 15(4), pp. 34–40.

Rajasegarar, S., Leckie, C. and Palaniswami, M. (2008b) ‘CESVM: Centered hyperellipsoidal support vector machine based anomaly detection’, in *IEEE*

International Conference on Communications, pp. 1610–1614.

Rajasegarar, S., Leckie, C. and Palaniswami, M. (2014) 'Hyperspherical cluster based distributed anomaly detection in wireless sensor networks', *Journal of Parallel and Distributed Computing*, 74(1), pp. 1833–1847.

Rajasegarar, S., Leckie, C., Palaniswami, M. and Bezdek, J. (2006) 'Distributed Anomaly Detection in Wireless Sensor Networks', *2006 10th IEEE Singapore International Conference on Communication Systems*, pp. 1–5.

Rajasegarar, S., Leckie, C., Palaniswami, M. and Bezdek, J. C. (2007) 'Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks', 1, pp. 3864–3869.

Randhawa, S. and Jain, S. (2017) *Data Aggregation in Wireless Sensor Networks: Previous Research, Current Status and Future Directions, Wireless Personal Communications*. Springer US.

Rani, G. and Chakraverty, S. (2018) *A rough set based approach for web user profiling, Communications in Computer and Information Science*. Springer Singapore.

Ranjani, S. S., Krishnan, S. R. and Thangaraj, C. (2012) 'Energy-efficient Cluster Based Data Aggregation for Wireless Sensor Networks', *2012 International Conference on Recent Advances in Computing and Software Systems*. Ieee, (2), pp. 174–179.

Rashid, B. and Husain, M. (2016) 'Applications of wireless sensor networks for urban areas : A survey', *Journal of Network and Computer Applications*. Elsevier, 60, pp. 192–219.

Rassam, M. A., Maarof, M. A. and Zainal, A. (2012) 'A Survey of Intrusion Detection Schemes in Wireless Sensor Networks', *American Journal of Applied Sciences 9*, 9(10), pp. 1636–1652.

Rassam, M. A., Maarof, M. A. and Zainal, A. (2014) 'Adaptive and online data anomaly detection for wireless sensor systems', *Knowledge-Based Systems*, 60, pp. 44–57.

Rassam, M. A., Maarof, M. A. and Zainal, A. (2018) 'A distributed anomaly detection model for wireless sensor networks based on the one-class principal component classifier', *International Journal of Sensor Networks*, 27(3), pp. 200–214.

Rassam, M. A. and Zainal, A. (2015) 'Principal Component Analysis-Based Data

Reduction Model For Wireless Sensor Networks', *International Journal of Ad Hoc and Ubiquitous Computing*, 18(1–2), pp. 85–101.

Rassam, M. A., Zainal, A. and Maarof, M. A. (2013) 'An adaptive and efficient dimension reduction model for multivariate wireless sensor networks applications', *Applied Soft Computing Journal*. Elsevier B.V., 13(4), pp. 10087–10122.

Rassam, M. a, Zainal, A. and Maarof, M. A. (2013) 'Advancements of data anomaly detection research in wireless sensor networks: a survey and open issues.', *Sensors (Basel, Switzerland)*, 13, pp. 10087–10122.

Rault, T. and Abdelmadjid Bouabdallah, Y. C. (2014) 'Energy Efficiency in Wireless Sensor Networks : a top-down survey'.

Raza, M. S. and Qamar, U. (2017) *Understanding and using rough set based feature selection: Concepts, techniques and applications*, *Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications*.

Razaque, A., Abdulgader, M., Joshi, C., Amsaad, F. and Chauhan, M. (2016) 'P-LEACH: Energy Efficient Routing Protocol for Wireless Sensor Networks', in *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*. IEEE.

Razzaque, M. A. (2014) 'Compression in wireless sensor networks : a survey and comparative evaluation', *ACM Transactions on Sensor Networks*, 10(1).

Rodriguez-galiano, V. F., Luque-espinar, J. A., Chica-olmo, M. and Mendes, M. P. (2018) 'Feature selection approaches for predictive modelling of groundwater nitrate pollution : An evaluation of fi lters , embedded and wrapper methods', *Science of the Total Environment*. Elsevier B.V., 624, pp. 661–672.

Rooshenas, A., Rabiee, H. R., Movaghar, A. and Naderi, M. Y. (2010) 'Reducing the data transmission in Wireless Sensor Networks using the Principal Component Analysis', *2010 Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. Ieee, pp. 133–138.

Roweis, S. T. and Saul, L. K. (2000) 'Nonlinear Dimensionality Reduction by Locally Linear Embedding', 290(December), pp. 2323–2327.

Roy, S., Conti, M., Setia, S. and Jajodia, S. (2014) 'Secure Data Aggregation in Wireless Sensor Networks: Filtering out the Attacker's Impact', *IEEE Transactions on Information Forensics and Security*, 9(4), pp. 681–694.

Sabor, N., Sasaki, S., Abo-Zahhad, M. and Ahmed, S. M. (2017) 'A comprehensive survey on hierarchical-based routing protocols for mobile wireless sensor networks: Review, taxonomy, and future directions', *Wireless Communications and Mobile Computing*, 2017.

Safaei, Mahmood, Asadi, S., Driss, M., Boulila, W., Alsaeedi, A., Chizari, H., Abdullah, R. and Safaei, Mitra (2020) 'A systematic literature review on outlier detection inwireless sensor networks', *Symmetry*, 12(3), pp. 1–41.

Safaei, Mahmood, Driss, M., Boulila, W., Sundararajan, E. A. and Safaei, Mitra (2021) 'Global Outliers Detection in Wireless Sensor Networks: A Novel Approach Integrating Time-Series Analysis, Entropy, and Random Forest-based Classificatio', *Journal of Software: Practice and Experiance*, pp. 1–37.

Safaei, Mahmood, Ismail, A. S., Chizari, H., Driss, M., Boulila, W., Asadi, S. and Safaei, Mitra (2020) 'Standalone noise and anomaly detection in wireless sensor networks: A novel time-series and adaptive Bayesian-network-based approach', *Software - Practice and Experience*, 50(4), pp. 428–446.

Said Elsayed, M., Le-Khac, N. A., Dev, S. and Jurcut, A. D. (2020) 'Network Anomaly Detection Using LSTM Based Autoencoder', *Q2SWinet 2020 - Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pp. 37–45.

Salem, A. O. A. and Shudifat, N. (2019) 'Enhanced LEACH protocol for increasing a lifetime of WSNs', *Personal and Ubiquitous Computing*. Personal and Ubiquitous Computing, 23(5–6), pp. 901–907.

Salim, A., Osamy, W. and Khedr, A. M. (2014) 'IBLEACH: Intra-balanced LEACH protocol for wireless sensor networks', *Wireless Networks*, 20(6), pp. 1515–1525.

Salim, C., Mitton, N., Salim, C., Mitton, N., Learning, M., Data, B., Salim, C. and Mitton, N. (2020) 'Machine Learning Based Data Reduction in WSN for Smart Agriculture To cite this version : HAL Id : hal-02463167 Machine Learning Based Data Reduction in WSN for Smart Agriculture'.

Salmabadi, H., Sarram, M. A. and Adibnia, F. (2015) 'An improvement on LEACH protocol (EZ-LEACH)', in *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pp. 3–7.

Samadi Bonab, M., Ghaffari, A., Soleimanian Gharehchopogh, F. and Alemi, P. (2020) 'A wrapper-based feature selection for improving performance of intrusion detection systems', *International Journal of Communication Systems*, 33(12), pp. 1–26.

Scholkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C. (2001) 'Estimating the support of a high dimensional distribution', *Journal of Neural Computation*, 13, pp. 1443–1471.

Selvin, S. V. and Kumar, S. M. (2012) 'Tree Based Energy Efficient and High Accuracy Data Aggregation for Wireless Sensor Networks', *Procedia Engineering*, 38, pp. 3833–3839.

Seo, S., Kang, J. and Ryu, K. H. (2005) 'Multivariate stream data reduction in sensor network applications', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* 3823 LNCS, pp. 198–207.

Sha, C., Wang, R., Huang, H. and Sun, L. (2010) 'Energy efficient clustering algorithm for data aggregation in wireless sensor networks', *The Journal of China Universities of Posts and Telecommunications*. The Journal of China Universities of Posts and Telecommunications, 17(December), pp. 104–122.

Shafri, H. Z. M. and Anuar, M. I. (2008) 'Hyperspectral Signal Analysis for Detecting Disease Infection in Oil Palms', *2008 International Conference on Computer and Electrical Engineering*. Ieee, pp. 312–316.

Shafri, H. Z. M. and Hamdan, N. (2009) 'Hyperspectral Imagery for Mapping Disease Infection in Oil Palm Plantation Using Vegetation Indices and Red Edge Techniques H', *American Journal of Applied Sciences 6*, 6(6), pp. 1031–1035.

Shahid, N., Naqvi, I. H. and Qaisar, S. Bin (2012) 'Quarter-Sphere SVM : Attribute and Spatio-Temporal Correlations based Outlier & Event Detection in Wireless Sensor Networks', pp. 2048–2053.

Shahid, N., Naqvi, I. H. and Qaisar, S. Bin (2013) 'One-class support vector machines: analysis of outlier detection for wireless sensor networks in harsh environments', *Artificial Intelligence Review*, 43(4), pp. 515–563.

Shahid, N., Naqvi, I. H. and Qaisar, S. Bin (2015) 'One-class support vector machines: analysis of outlier detection for wireless sensor networks in harsh environments', *Artificial Intelligence Review*, 43(4), pp. 515–563.

Shakaff, A. Y. M., Ahmad, M. N., Markom, M. A., Adom, A. H., Hidayat, W., Abdullah, A. H. and Fikri, N. A. (2009) 'E-nose for Basal Stem Rot Detection', *UniMAP*, pp. 4–6.

Shariff, A. R. M., Bejo, S. K., Ahmad, K., Hashim, I. C. and Muharam, F. (2018)

‘Classification for non infected and infected ganoderma boninense of oil palm trees using ALOS PALSAR-2 backscattering coefficient’, *IOP Conference Series: Earth and Environmental Science*, 169, p. 012066.

Sharma, A. B., Golubchik, L. and Govindan, R. (2010) ‘Sensor Faults : Detection Methods and Prevalence in Real-World Datasets’, *ACM Transactions on Sensor Networks*, 6(3), pp. 1–39.

Sheikhpour, R., Sarram, M. A., Gharaghani, S. and Chahooki, M. A. Z. (2017) ‘A Survey on semi-supervised feature selection methods’, *Pattern Recognition*. Elsevier, 64(February 2016), pp. 141–158.

Shi, F. and Li, Q. (2018) ‘A Survey of Data Semantization in Internet of Things’, pp. 1–20.

Shial, R. K., Sankar Gouda, B., Pattanaik, S. R. and Sethi, N. (2020) ‘A Centralized Faulty Node Detection Algorithm Based on Statistical Analysis in WSN’, *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*.

Shokouhi, A., Marzieh, R., Farahnaz, B., Asghar, A., Hosseinabadi, R. and Sangaiah, A. K. (2018) *Survey on clustering in heterogeneous and homogeneous wireless sensor networks, The Journal of Supercomputing*. Springer US.

Al Shorman, A., Faris, H. and Aljarah, I. (2020) ‘Unsupervised intelligent system based on one class support vector machine and Grey Wolf optimization for IoT botnet detection’, *Journal of Ambient Intelligence and Humanized Computing*. Springer Berlin Heidelberg, 11(7), pp. 2809–2825.

Silva, A. P. R. da, Rocha, B. P. S., Martins, M. H. T., Loureiro, A. A. F., Ruiz, L. B. and Wong, H. C. (2005) ‘Decentralized intrusion detection in wireless sensor networks’, p. 16.

Singh, A., Kotiyal, V., Sharma, S., Nagar, J. and Lee, C. C. (2020) ‘A Machine Learning Approach to Predict the Average Localization Error with Applications to Wireless Sensor Networks’, *IEEE Access*, 8, pp. 208253–208263.

Singh, K. (2015) ‘WSN LEACH Based Protocols : A Structural Analysis’, in *2015 International Conference and Workshop on Computing and Communication (IEMCON)*. IEEE, pp. 1–7.

Singh, S. K., Kumar, P. and Singh, J. P. (2017) ‘A Survey on Successors of LEACH Protocol’, *IEEE Access*. IEEE, 5, pp. 4298–4328.

Siripanadorn, S., Hattagam, W. and Teaumroong, N. (2010) ‘Anomaly Detection using Self-Organizing Map and Wavelets in Wireless Sensor Networks’, in *ACS’10 Proceedings of the 10th WSEAS international conference on Applied computer science*, pp. 291–297.

Soroush, A. D., Kamran, F., Javad, R., Reza, F., Kumbesan, S. and Masih, A. D. (2020) ‘A survey on data aggregation techniques in IoT sensor networks’, *Wireless Networks*. Springer US, 26(2), pp. 1243–1263.

Sudha, L. and Thangaraj, P. (2019) ‘Improving energy utilization using multi hop data aggregation with node switching in wireless sensor network’, *Cluster Computing*. Springer US, 22(s5), pp. 12749–12757.

Sun, B., Shan, X., Wu, K., Member, S. and Xiao, Y. (2013) ‘Anomaly Detection Based Secure In-Network Aggregation for Wireless Sensor Networks’, *IEEE System Journals*, 7(1), pp. 13–25.

Surekha, S. and Jayasuma, G. (2015) ‘Comparison of Feature Selection Techniques for Thyroid Disease’, pp. 20–26.

Taj, M. B. M. and Kbir, M. A. (2016) ‘ICH-LEACH: An enhanced LEACH protocol for Wireless Sensor Network’, in *2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS)*. IEEE, p. 46.

Takianngam, S. and Usaha, W. (2011) ‘Discrete Wavelet Transform and One-Class Support Vector Machines for Anomaly Detection in Wireless Sensor Networks’, in *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*.

Tawhid, M. A. and Ibrahim, A. M. (2020) ‘Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm’, *International Journal of Machine Learning and Cybernetics*. Springer Berlin Heidelberg, 11(3), pp. 573–602.

Tax, D. M. J. and Duin, R. P. W. (1999) ‘Support vector domain description’, *Pattern Recognition Letters*, 20(11–13), pp. 1191–1199.

Tayeh, G. B., Makhoul, A., Demerjian, J., Guyeux, C. and Bahi, J. (2020) ‘Fault tolerant data transmission reduction method for wireless sensor networks’, *World Wide Web*. World Wide Web, 23(2), pp. 1197–1216.

Ullah, I. and Youn, H. Y. (2020) ‘Efficient data aggregation with node clustering and extreme learning machine for WSN’, *Journal of Supercomputing*. Springer US, 76(12), pp. 10009–10035.

Ullah, I., Youn, H. Y. and Han, Y. H. (2021) 'An efficient data aggregation and outlier detection scheme based on radial basis function neural network for WSN', *Journal of Ambient Intelligence and Humanized Computing*. Springer Berlin Heidelberg, (0123456789).

Ur-Rehman, A., Abbasi, A. Z., Islam, N. and Shaikh, Z. A. (2014) 'A review of wireless sensors and networks' applications in agriculture', *Computer Standards & Interfaces*. Elsevier B.V., 36(2), pp. 263–270.

Vijayan, N. and Johny, A. (2016) 'Study on Tradeoffs in Detection of Malicious Data Injections in Wireless Sensor Networks', *Procedia Technology*. Elsevier B.V., 25(Raerest), pp. 378–383.

Vinterbo, S. and Øhrn, A. (2000) 'Minimal approximate hitting sets and rule templates', *International Journal of Approximate Reasoning*, 25(2), pp. 123–143.

Wan, R., Xiong, N., Hu, Q., Wang, H. and Shang, J. (2019) 'Similarity-aware data aggregation using fuzzy c-means approach for wireless sensor networks', *EURASIP Journal on Wireless Communications and Networking (2019)*. EURASIP Journal on Wireless Communications and Networking, 59.

Wang, B. and Mao, Z. (2020) 'Detecting outliers in industrial systems using a hybrid ensemble scheme', *Neural Computing and Applications*. Springer London, 32(12), pp. 8047–8063.

Wang, D., Yeung, D. S., Tsang, E. C. C., Member, S., Yeung, D. S., Tsang, E. C. C. and Member, A. (2006) 'Structured One-Class Classification', *IEEE Trans Syst Man Cybern B Cybern*, 36(6), pp. 1283–1295.

Wang, N. and Zhu, H. (2012) 'An Energy Efficient Algrithm Based on LEACH Protocol', *2012 International Conference on Computer Science and Electronics Engineering*. Ieee, pp. 339–342.

Weng, J., Zhang, Y. and Hwang, W.-S. (2003) 'Candid Covariance-Free Incremental Principal Component Analysis', *IEEE transactions on pattern analysis and machine intelligence*, 25(8), pp. 1034–1040.

Widhalm, D., Goeschka, K. M. and Kastner, W. (2021) 'An open-source wireless sensor node platform with active node-level reliability for monitoring applications', *Sensors*, 21(22), pp. 1–44.

Wong, Y. L. and Shamsuddin, S. M. (2019) 'Biometric identification based scores discretization on multi forms of image structures', *International Journal of*

Advances in Soft Computing and its Applications, 11(3), pp. 1–23.

Xie, M., Han, S. and Tian, B. (2011) 'Highly Efficient Distance-Based Anomaly Detection through Univariate with PCA in Wireless Sensor Networks', *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*. Ieee, pp. 564–571.

Xie, M., Han, S., Tian, B. and Parvin, S. (2011) 'Anomaly detection in wireless sensor networks: A survey', *Journal of Network and Computer Applications*. Elsevier, 34(4), pp. 1302–1325.

Xie, M., Hu, J., Han, S. and Chen, H.-H. (2013) 'Scalable Hypergrid k-NN-Based Online Anomaly Detection in Wireless Sensor Networks', *IEEE Transactions on Parallel and Distributed Systems*, 24(8), pp. 1661–1670.

Xie, M., Hu, J., Han, S. and Chen, H. (2012) 'Scalable Hyper-Grid k-NN-based Online Anomaly Detection in Wireless Sensor Networks', pp. 1–11.

Xue, Y., Jia, W., Zhao, X. and Pang, W. (2018) 'An Evolutionary Computation Based Feature Selection Method for Intrusion Detection', *Security and Communication Networks*, 2018.

Yahyaoui, A., Abdellatif, T. and Attia, R. (2019) 'Hierarchical anomaly based intrusion detection and localization in IoT', *2019 15th International Wireless Communications and Mobile Computing Conference, IWCMC 2019*. IEEE, pp. 108–113.

Yazirli, Y. and Kan-Kilinç, B. (2019) 'Comparison of Algorithms Based on Rough Set Theory for a 3-Class Classification', *International Journal of Research - Granthaalayah*, 7(8), pp. 394–401.

Yektaparast, A., Nabavi, F.-H. and Sarmast, A. (2012) 'An Improvement on LEACH Protocol (Cell-LEACH)', in *2012 14th International Conference on Advanced Communication Technology (ICACT)*. IEEE, pp. 992–996.

Yu, Y., Li, K., Zhou, W. and Li, P. (2012) 'Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures', *Journal of Network and Computer Applications*. Elsevier, 35(3), pp. 867–880.

Yuan, F., Zhan, Y. and Wang, Y. (2014) 'Data Density Correlation Degree Clustering Method for Data Aggregation in WSN', *IEEE Sensors Journal*, 14(4), pp. 1089–1098.

Yue, H., Jiang, Q., Yin, C. and Wilson, J. (2018) 'Research on data aggregation and transmission planning with Internet of Things technology in WSN multi-channel aware network', *The Journal of Supercomputing*. Springer US.

Zain, N., Seman, I. A. B. U., Kushairi, A. and Ramli, U. M. I. S. (2013) 'Metabolite Profiling Of Oil Palm Towards Understanding Basal Stem Rot (BSR) Disease', *Journal of Oil Palm Research*, 25(1), pp. 58–71.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J. (2020) 'A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction', *Journal of Applied Science and Technology Trends*, 1(2), pp. 56–70.

Zhang, J., Hu, P., Xie, F., Long, J. and He, A. (2018) 'An Energy Efficient and Reliable In-Network Data Aggregation Scheme for WSN', *IEEE Access*, 6, pp. 71857–71870.

Zhang, L., Zhang, G., Yu, L., Zhang, J. and Bai, Y. (2004) 'Intrusion detection using rough set classification', *Journal of Zhejiang University-SCIENCE A*, 5(9), pp. 1076–1086.

Zhang, P., Wang, J., Guo, K., Wu, F. and Min, G. (2018) 'Multi-functional secure data aggregation schemes for WSNs', *Ad Hoc Networks*, pp. 86–99.

Zhang, R., Nie, F., Li, X. and Wei, X. (2019) 'Feature selection with multi-view data: A survey', *Information Fusion*. Elsevier B.V., 50(November 2018), pp. 158–167.

Zhang, Y. (2010) *Observing the Unobservable Distributed Online Outlier Detection in Wireless Sensor Networks*.

Zhang, Y., Hamm, N. a. S., Meratnia, N., Stein, a., van de Voort, M. and Havinga, P. J. M. (2012) 'Statistics-based outlier detection for wireless sensor networks', *International Journal of Geographical Information Science*, 26(8), pp. 1373–1392.

Zhang, Y., Meratnia, N. and Havinga, P. (2009a) 'Adaptive and Online One-Class Support Vector Machine-Based Outlier Detection Techniques for Wireless Sensor Networks', *2009 International Conference on Advanced Information Networking and Applications Workshops*. Ieee, pp. 990–995.

Zhang, Y., Meratnia, N. and Havinga, P. (2009b) 'Hyperellipsoidal SVM-Based Outlier Detection Technique for Geosensor Networks', pp. 31–41.

Zhang, Yang, Meratnia, N. and Havinga, P. (2010) 'Outlier Detection Techniques for Wireless Sensor Networks : A Survey', *IEEE Communications Surveys & Tutorials*, 12(2), pp. 159–170.

Zhang, Y., Meratnia, N. and Havinga, P. J. M. (2008) 'An Online Outlier Detection Technique for Wireless Sensor Networks using Unsupervised Quarter-Sphere Support Vector Machine', in *International Conference on Intelligent Sensors, Sensor Networks and Information Processin*, pp. 151–156.

Zhang, Y., Meratnia, N. and Havinga, P. J. M. (2010) 'Ensuring high sensor data quality through use of online outlier detection techniques', *Int. J. Sensor Networks*, 7(3).

Zhang, Y., Meratnia, N. and Havinga, P. J. M. (2013) 'Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine', *Ad Hoc Networks*. Elsevier B.V., 11(3), pp. 1062–1074.

Zhong, C., Chen, Y. and Peng, J. (2020) 'Feature selection based on a novel improved tree growth algorithm', *International Journal of Computational Intelligence Systems*, 13(1), pp. 247–258.

Zhu, B., Bedeer, E., Nguyen, H. H., Barton, R. and Henry, J. (2021) 'Improved Soft-k-Means Clustering Algorithm for Balancing Energy Consumption in Wireless Sensor Networks', *IEEE Internet of Things Journal*, 8(6), pp. 4868–4881.

Zou, Z., Tseng, T. L., Sohn, H., Song, G. and Gutierrez, R. (2011) 'A rough set based approach to distributor selection in supply chain management', *Expert Systems with Applications*. Elsevier Ltd, 38(1), pp. 106–115.

LIST OF PUBLICATIONS

Journal Article

Indexed Journal

1. Nurfazrina Mohd Zamry, Anazida Zainal Murad A Rassam (2018), “Unsupervised Anomaly Detection for Unlabelled Wireless Sensor Networks Data”, *Int. J. Advance Soft Compu. Appl* 10 (2), 172-192. **(Scopus Indexed)**
2. Zamry, N.M.; Zainal, A.; Rassam, M.A.; Alkhammash, E.H.; Ghaleb, F.A.; Saeed, F. Lightweight Anomaly Detection Scheme Using Incremental Principal Component Analysis and Support Vector Machine. Sensors 2021, 21, 8017. https://doi.org/10.3390/s21238017 **(Q2, IF: 3.576)**

Non-Indexed Journal

1. Nurfazrina M Zamry, Anazida Zainal, Murad A Rassam (2018) “Lightweight anomaly detection for wireless sensor networks based on CCIPCA and One-Class SVM” Advanced Science Letters 24 (10), 7371-7376.

2. Nurfazrina Mohd Zamry, Anazida Zainal Murad A Rassam (2019), “Distributed CESVM-DR anomaly detection for Wireless Sensor Network”, *International Journal of Innovative Computing* 9 (1).

International Conference

1. Nurfazrina Mohd Zamry, Anazida Zainal Murad A Rassam (2021), “LEACH-CR: Energy Saving Hierarchical Network Protocol Based on Low-Energy Adaptive Clustering Hierarchy for Wireless Sensor Networks” 2021 3rd International Cyber Resilience Conference (CRC), 1-6. **(Scopus Indexed)**

2. Nurfazrina M Zamry, Anazida Zainal, Murad A Rassam (2018) “Lightweight anomaly detection for wireless sensor networks based on CCIPCA and One-Class SVM” *Advanced Science Letters* 24 (10), 7371-7376. **(Web of Science Conference)**

3. Nurfazrina M Zamry, Anazida Zainal, Murad A Rassam, Majid Bakhtiari, Mohd A Maarof (2015), “Selection of Soil Features for Detection of Ganoderma Using Rough Set Theory”, Pattern Analysis, Intelligent Security and the Internet of Things, 303-314. **(Scopus Indexed)**

4. Nurfazrina M Zamry, Mohd A Maarof, Anazida Zainal (2015), “Islamic Web content filtering and categorization on deviant teaching”, Advances in Intelligent Systems and Computing, 287, pp. 667–678. **(Scopus Indexed)**