ENHANCED NON-PARAMETRIC SEQUENCE LEARNING SHCEME FOR INTERNET OF THINGS SENSORY DATA IN CLOUD INFRASTRUCTURE

EDJE EFETOBOR ABEL

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2021

DEDICATION

This thesis is dedicated to Father who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my Mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

I express my profound gratitude to Whom Honour His due; JEHOVAH the maker of the entire Universe and all that is in it, who gave me the strength and perseverance to embark on this research work.

In the course of this research work, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. Specifically, I would like to convey my sincere appreciation to my main thesis supervisor, **Professor Dr. Muhammad Shafie Abd Latiff**, for his encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor **Dr. Chan Weng Howe** for his guidance, advices and motivation. Without their continued support and interest, this research would not have been the same as presented here. I also appreciate the support and motivation received from lecturers, members and staff of School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia for the series of programmes which aided this research work.

I am also indebted to my Parents, Sir and Lady Johnbest Edje for their financial supports and prayers during the course of this research. I also acknowledge my wife Mrs Edje Gladys, my two Children, Oghenetaga Edje and Gideon Edje for their tremendous support and prayers.

My fellow postgraduate student 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.

ABSTRACT

The Internet of Things (IoT) Cloud is an emerging technology that enables machine-to-machine, human-to-machine and human-to-human interaction through the Internet. IoT sensor devices tend to generate sensory data known for their dynamic and heterogeneous nature. Hence, it makes it elusive to be managed by the sensor devices due to their limited computation power and storage space. However, the Cloud Infrastructure as a Service (IaaS) leverages the limitations of the IoT devices by making its computation power and storage resources available to execute IoT sensory data. In IoT-Cloud IaaS, resource allocation is the process of distributing optimal resources to execute data request tasks that comprise data filtering operations. Recently, machine learning, non-heuristics, multi-objective and hybrid algorithms have been applied for efficient resource allocation to execute IoT sensory data filtering request tasks in IoT-enabled Cloud IaaS. However, the filtering task is still prone to some challenges. These challenges include global search entrapment of event and error outlier detection as the dimension of the dataset increases in size, the inability of missing data recovery for effective redundant data elimination and local search entrapment that leads to unbalanced workloads on available resources required for task execution. In this thesis, the enhancement of Non-Parametric Sequence Learning (N-PSL), Perceptually Important Point (PIP) and Efficient Energy Resource Ranking-Virtual Machine Selection (ERVS) algorithms were proposed. The Non-Parametric Sequence-based Agglomerative Gaussian Mixture Model (NPSAGMM) technique was initially utilized to improve the detection of event and error outliers in the global space as the dimension of the dataset increases in size. Then, Perceptually Important Points K-means-enabled Cosine and Manhattan (PIP-KCM) technique was employed to recover missing data to improve the elimination of duplicate sensed data records. Finally, an Efficient Resource Balance Ranking- based Glow-warm Swarm Optimization (ERBV-GSO) technique was used to resolve the local search entrapment for near-optimal solutions and to reduce workload imbalance on available resources for task execution in the IoT-Cloud IaaS platform. Experiments were carried out using the NetworkX simulator and the results of N-PSAGMM, PIP-KCM and ERBV-GSO techniques with N-PSL, PIP, ERVS and Resource Fragmentation Aware (RF-Aware) algorithms were compared. The experimental results showed that the proposed N-PSAGMM, PIP-KCM, and ERBV-GSO techniques produced a tremendous performance improvement rate based on 3.602%/6.74% Precision, 9.724%/8.77% Recall, 5.350%/4.42% Area under Curve for the detection of event and error outliers. Furthermore, the results indicated an improvement rate of 94.273% F1-score, 0.143 Reduction Ratio, and with minimum 0.149% Root Mean Squared Error for redundant data elimination as well as the minimum number of 608 Virtual Machine migrations, 47.62% Resource Utilization and 41.13% load balancing degree for the allocation of desired resources deployed to execute sensory data filtering tasks respectively. Therefore, the proposed techniques have proven to be effective for improving the load balancing of allocating the desired resources to execute efficient outlier (Event and Error) detection and eliminate redundant data records in the IoT-based Cloud IaaS Infrastructure

ABSTRAK

Awan Internet-Benda (IoT-Cloud) adalah teknologi yang muncul yang membolehkan interaksi mesin-ke-mesin, manusia-ke-mesin dan manusia-ke-manusia melalui Internet. Peranti deria IoT cenderung menghasilkan data deria yang dikenali dengan sifatnya yang dinamik dan heterogen. Oleh itu, sukar untuk dikendalikan oleh peranti deria kerana kuasa dan ruang penyimpanan yang terhad. Walau bagaimanapun, Awan Infrastruktur-sebagai-Perkhidmatan (IaaS-Cloud) memanfaatkan batasan peranti IoT dengan menjadikan kuasa komputeran dan sumber penyimpanannya tersedia untuk melaksanakan data deria IoT. Dalam IaaS IoT-Cloud, peruntukan sumber adalah proses pengagihan sumber yang optimum untuk melaksanakan tugas permintaan data yang terdiri daripada operasi penyaringan data. Baru-baru ini, algoritma pembelajaran mesin, bukan-heuristik, pelbagaiobjektif dan hibrid telah digunakan untuk peruntukan sumber yang cekap untuk melaksanakan tugas permintaan penyaringan data sensori IoT dalam IaaS-Cloud Terpasang-IoT. Walau bagaimanapun tugas penyaringan masih tertakluk kepada beberapa cabaran. Cabaran ini merangkumi perangkap carian global untuk pengesanan pengganjil peristiwa dan kesalahan kerana dimensi set data meningkat dalam ukuran, ketidakupayaan pemulihan data yang hilang untuk penghapusan data bertindih yang berkesan dan perangkap carian tempatan yang membawa kepada beban kerja yang tidak seimbang pada sumber daya yang sedia ada yang diperlukan untuk pelaksanaan tugas. Dalam tesis ini, penambahbaikan algoritma Pembelajaran Jujukan Bukan-Parametrik (N-PSL), Titik Penting Persepsi (PIP) dan Pemeringkatan Sumber Tenaga Efisien–Pemilihan Mesin Maya (ERVS) dicadangkan. Teknik Model Campuran Gaussian Agglomeratif berasaskan-Jujukan Bukan-Parametrik (NPSAGMM) pada awalnya digunakan untuk meningkatkan pengesanan pengganjil peristiwa dan ralat di ruang global apabila dimensi set data meningkat. Setelah itu, teknik Titik Penting Persepsi Terpasang-K-means Cosine dan Manhattan (PIP-KCM) digunakan untuk mengembalikan data yang hilang untuk memperbaiki penghapusan data bertindih yang dikesan. Akhirnya, teknik Baki Sumber yang Cekap Berasaskan Kedudukan Berasaskan Teknik Pengoptimuman Swarm (ERBV-GSO) digunakan untuk menyelesaikan perangkap carian tempatan untuk penyelesaian hampir optimum dan untuk mengurangkan ketidakseimbangan beban kerja pada sumber yang ada untuk pelaksanaan tugas dalam platform Awan-IoT *IaaS*. Eksperimen dijalankan menggunakan simulator NetworkX dan hasil perbandingan teknik N-PSAGMM, PIP-KCM dan ERBV-GSO dengan algoritma NPSL, PIP, ERVS dan Pemecahan sumber sedar (RF-Aware). Hasil eksperimen menunjukkan bahawa teknik N-PSAGMM, PIP-KCM, dan ERBV-GSO yang dicadangkan menghasilkan kadar peningkatan prestasi yang cemerlang berdasarkan kepada ketepatan 3.602%/6.74%, KETAHUI 9.724%/8.77%, Kawasan di bawah Lengkung 5.350%/4.42% untuk mengesan pengganjil peristiwa dan kesalahan. Selanjutnya, hasil menunjukkan kadar peningkatan 94.273% skor F1, Nisbah Pengurangan 0.143 dengan minimum 0.149% Ralat Root Ralat Squared untuk penghapusan data bertindih serta bilangan minimum 608 perpindahan Mesin Maya, 47.62% Penggunaan Sumber dan 41.13% tahap pengimbangan beban untuk peruntukan sumber yang diinginkan digunakan untuk melaksanakan tugas penyaringan data sensori masing-masing. Oleh itu, teknik yang dicadangkan telah terbukti berkesan untuk meningkatkan keseimbangan beban memperuntukkan sumber yang diinginkan untuk pelaksanaan pengesanan pengganjil (Peristiwa dan Kesalahan) yang efisien dan penghapusan catatan data bertindih dalam Infrastruktur Awan *IaaS* yang berasaskan IoT.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACOCS	-	Ant Colony Optimization Cuckko Search
ACO-WSE	-	Ant Colony Web Service Composite
ACS-AR	-	Adaptive Compressive Sensing-based Autoregressive
AFF-WRR	-	Adaptive Firefly-enabled Weighted Round Robin
AKC	-	Adaptive K-means Clustering
AMCO	-	Advance Micro-Cluster-based Continuous Outlier
BBN	-	Bayesian Belief Network
BEATS	-	Block of Eigen Value Algorithm for Time Series Segmentation
BPL	-	Back Propagation Learning
CDCNN	-	Cascading Deep Convolution Neural Network
CEM-KM	-	Cobweb Expectation Maximization K-means
CFSH	-	Correlation Feature Selection-based Heuristic
CNN	-	Convolution Neural Network
COFFGA	-	Combinatorial Ordering First-Fit Genetic Algorithm
CPBO	-	CherryPick-based Bayesian Optimization
CRPCA	-	Clustering-based Recursive Principal Component Analysis
CSLGBP	-	Central-Symmetric Local Gabor Binary Pattern
CSS	-	Chaotic Social Spider
DCNN	-	Deep Convolution Neural Network
DDSS	-	Dynamic Dedication Server Scheduling
DPLB	-	Dynamic Priority Load Balancing
D-SAX	-	Dynamic Symbolic Aggregation Approximation
EH-OFCM	-	Efficient High-order Tensor Fuzzy C-means
EHO-LKSV	-	Elephant Herd Optimization-Linear Kernel Support Vector
EKGC	-	Enhanced Knowledge Granule Clustering
ENOF	-	Ellipsoidal Neighbourhood Outlier Factor

EODSP	-	Entropy Outlier Detection Semi-supervised
ERBV-GSO	-	Efficient Resource Balance Virtual Machine-Glowworm Swarm Optimization
ERVS	-	Energy Efficient Resource Ranking Virtual Machine Selection
FCBF	-	Fast Correlation-Based Filter
FEM	-	Fair Energy-efficient Virtual Machine Placement
FFNN	-	Feed-forward Neural Networks
GACO	-	Genetic Ant Colony Optimization
GA-OCSTuM	-	Genetic-based One-class Support Tucker Machine
GSO	-	Glowworm Swarm Optimization
GVN	-	Genetic Variable Neighbourhood
HDBN	-	Hybrid Deep Belief Network
HMLT	-	Homoscedasticity Measurement Leven’s Test
IaaS	-	Infrastructure as a Service
IEDCA-IRE	-	Intra-class Extra-class Discriminative Correlation Analysis
IFSC-KM	-	Incremental Fast Searching Clustering-based K-mediods
IoT	-	Internet of Things
IPCA	-	Iterative Principal Component Analysis
LDA	-	Linear Discriminant Analysis
L-HOSVD	-	Lanczos-based High Order Singular Value Decomposition
LPT	-	Laplace Pyramid Transform
LRP	-	Logistic Regression-based Prediction
MGPCA	-	Multivariate Gaussian-based Principal Component Analysis
MLPC	-	Multilayer Perceptron Classifier
MMAS	-	Max-Min Ant System
MWPCA	-	Moving Window Principal Component Analysis
N-PSAGMM	-	Non-Parametric Sequence Learning Agglomerative Gaussian Mixture Model
OSTM	-	One-class Tensor Machine

OSVM	-	One-class Support Vector Machine
PAA	-	Piecewise Aggregation Approximation
PaaS	-	Platform as a Service
PCA	-	Principal Component Analysis
PIP	-	Perceptually Important Points
PIP-KCM	-	Perceptually Important K-means Cosine and Manhattan
PPSO	-	Parallel Particle Swarm Optimization
PSO	-	Particle Swarm Optimization
RF-A	-	Resource Fragmentation- Aware
RIPCA	-	Robust Incremental Principal Component Analysis
ROF	-	Resolution Outlier Factor
SaaS	-	Software as a Service
SIFT	-	Scale Invariant Feature Transform
SLB	-	Stochastic Load Balancing
STDD-KSTDD	-	Support High-order Tensor Data Description-based Kernel Support High-order Tensor Data Description
SVDD	-	Support Vector Data Description
SVM	-	Support Vector Machine
SVMP	-	Service-oriented Virtual Machine Placement
SVMRFE+CBR	-	Support Vector Machine Recursive Function Elimination-based Correlation Bias Reduction
TKMC	-	Two-step K-means Clustering
WA-VMC	-	Workload Aware-Virtual Machine Consolidation
WHAR	-	Weighted Component Human Activity Recognition
WSN	-	Wireless Sensor Network

LIST OF SYMBOLS

Rels	-	Relative Sequences
Refs	-	Reference Sequences
ϴ	-	Sequence Operator
β	-	Influence Relative Grade Point
RMF	-	Relative Mass Factor
μ	-	Median
z	-	Data Points
D	-	Dimension
Ƒ	-	Probability
Ϙ	-	Reverse Parameters
MR	-	Missing Data
ꝏi	-	Unknown Data Records
Fm	-	Number of Features
KIC	-	K-means Imputation Clustering
Wr	-	Weighted Distance
Ȣg	-	Energy Usage
Ԏ	-	Task processing Time
IDTg	-	Idle State
Ȝ	-	Resource Utilization Rate
Ȝc	-	CPU Utilization
$Ȝ_m$	-	Memory Utilization
lf_p	-	Luciferin Value
ss	-	Size Moving Steps
ꝱ	-	Luciferin Decay Coefficient
ꝏ	-	Luciferin Improvement Coefficient
ꝶ	-	Neighbourhood Range
Rs	-	Sensor Range
n_r	-	Number of Suitable Neighbours
PB	-	Probability of predicted movement direction to neighbour

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

The integration of the Internet of Things (IoT)-based Cloud computing is an emerging ubiquitous network infrastructure, offering diverse distributed and transparent services in most computing applications. Smart devices such as sensors, micro-controllers, mobile phones, servers designated at local sites, and the cloud can interact with one another due to this new integrated technology. As the popularity and extensive use of IoT-cloud increases over the years, IoT sensory data are generated, and various IoT-cloud applications are implemented to deliver desired services to end-users across the globe. IoT sensors can sense, pre-process, store and transmit sensed data directly to the Internet without any human intervention. IoT sensor nodes do not interact with each other or Internetwork to transmit their sensed data to a connected sink node, unlike Wireless Sensor Network (WSN) nodes. However, IoT sensors do not have the computation power and storage required to process and store IoT sensing data characterized by dynamicity and heterogeneity. This challenge is resolved with the support of the cloud`s heterogeneous processing power and storage resources, which led to integrating IoT and Cloud computing.

Cloud resources can be accessed at anytime, anywhere for the storing and processing of IoT sensed IoT sensed data and applications over the Internet. Three service models are accessible on the cloud, namely, Infrastructure-as-a-Service (IaaS), Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS). The IaaS is made of enormous computing resources managed by IaaS providers that allocate resources to clients on a pay-as-you-go basis. IoT sensing data management, application systems to high-performance computing applications. Resources such as Servers, Virtual Machine (VM), CPU, networks, and storage systems (hard disks and memory) are available and accessible on IaaS to manage workloads, ranging from IoT sensing data management to application systems high-performance computing applications. Figure 1.1 illustrates the architecture of IoT-based Cloud computing infrastructure. It comprises three layers beginning from the IoT sensor device layer that generates various sensor data to be transmitted directly to the cloud. The network layer devices are introduced to minimize the occurrences of latency delay during data transmission from IoT sensors to the cloud data-center. The upper layer consists of the cloud platform where sensed data retrieved are stored and processed.

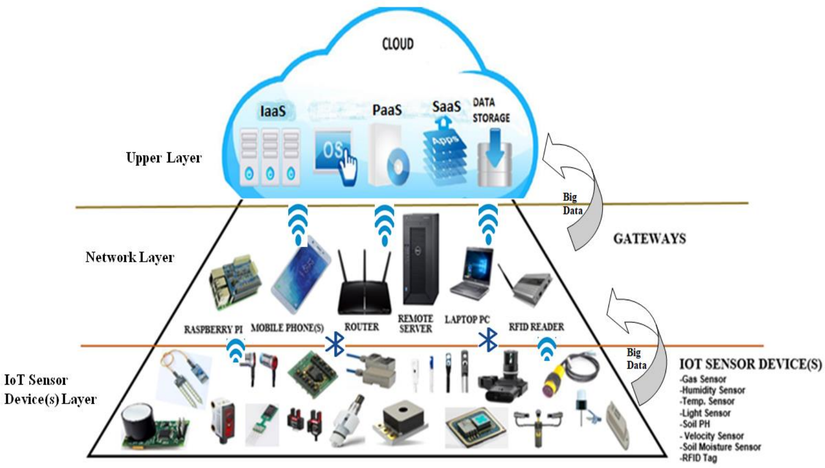

Figure 1.1 IoT-based Cloud Computing Architecture

The allocation of resources to execute IoT sensory data's filtering process is a critical challenge for Infrastructure-as-a-Service (IaaS) in cloud computing. Also, the filtering processes of IoT sensed data entails the utilization of algorithms. Implementing an algorithm required for practical IoT sensed data filtering is of great importance. Efficient resource allocation is the "optimal selection of desired resources (e.g., Servers and Virtual Machines) from the resource pool, required to execute the data filtering or application request tasks/cloudlets on the IaaS cloud platform.

Data filtering is known as the process of defining, detecting and correcting or eliminating errors or duplicates in raw data samples to reduce the impact on succeeding analysis. Filtering algorithms are the techniques and procedures to tackle the filtering issues in applications to obtain its effectiveness, although it might not always be attainable. In the last few decades, machine and deep learning algorithms are highly recognized algorithms for effective filtering of IoT sensed data in various areas of engineering, health, computing, geo-informatics, bio-informatics and signal processing. Furthermore, an efficient resource allocation algorithm is required to select optimal resources in IaaS cloud to execute the data filtering requests tasks for better performance.

Non-parametric Sequence Learning (N-PSL) and Perceptually Important Point (PIP) algorithms are unsupervised machine learning techniques capable of data filtering. Existing studies confirmed that N-PSL and PIP are reliable and performs better in terms of outlier detection and elimination of data redundancy than current machine learning algorithms, including the Linear Discriminate Analysis (LDA), Support Vector Machine (SVM), Wavelet Transform (WT) and Principal Component Analysis (PCA). Also, N-PSL is deployed in multi-discipline researches such as Zhou *et al.,*(2018), Mallor *et al.*,(2017), Andersson *et al.*, (2016), Neely *et al.*,(2012), Timm and Barth (2011) and Chen *et al.*, (2010). On the other hand, the Energy-efficient Resource Ranking-based Virtual Machine Selection (ERVS) algorithm is adapted to allocate resources to execute tasks. Its adaption is attributed to its ability to balance workloads on running physical machines (PMs). Thus, VMs (tasks) can be migrated from overloaded PMs to under-loaded PMs to minimize the workload imbalance on running PMs.

1.2 Problem Background

Sensed data is mainly generated by multiple IoT sensor devices embedded in the environment which they sense. IoT sensory data's main characteristic is inaccurate sensing due to several limitations such as unreliable reading that leads to data anomalies or outliers. Also, the data's dynamic nature is a result of a large amount of data generated from multiple sensors in real-time, which leads to redundant data records. Conversely, the increase in demand for IaaS resources in the cloud leads to load balancing challenges due to inefficient resource allocation to execute IoT sensed data filtering or application request tasks, both in static and real-time.

Data outliers are mainly of two types, namely event and error outliers, which are common in sensing data readings. An event outlier is when there is a sudden change in data readings triggered by unexpected circumstances. On the other hand, an error outlier is regarded as the data point that diverges significantly away from other data points caused by faulty sensor readings, as depicted in Figure 1.2.

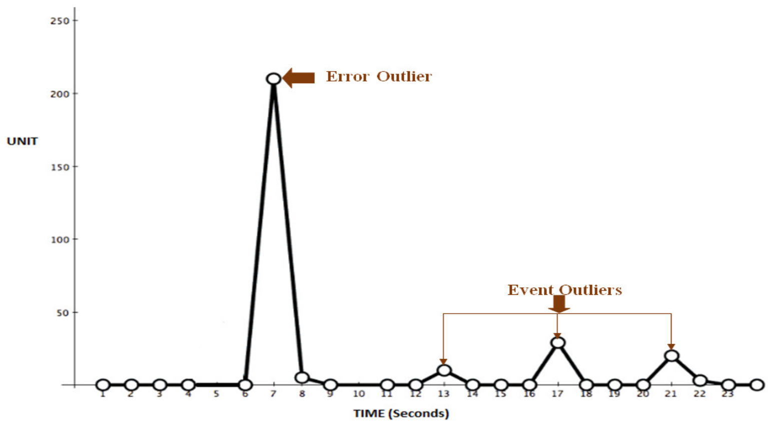

Figure 1.2 Example of Event and Error Outliers

An example of an event outlier can be discovering sudden gas leakage from a cylinder container, the discovery of unauthorized persons gaining access to prohibited company's files, and the detection of abnormality in a patient's health status. Therefore, event outliers are very important in all areas of research field due to its capability to identify unforeseen events that may be disastrous to human existence on earth. There are several filtering techniques deployed for detecting outliers in sensory data, generated by IoT sensor devices for IaaS cloud computing infrastructure. These techniques range from machine learning to deep learning algorithms, implemented in the cloud to identify outliers in vast amounts of sensed datasets.

The detection of event and error outliers is a challenging process due to sensory data characteristic (Souza and Amazonas, 2015; Santos et al., 2018). Most of these algorithms mainly focus on the generality of outlier detection without considering either event or error type of outliers. However, some of the algorithms implemented are able to detect both error and event outliers in a reasonable amount of data sample such as the research work of Zhang *et al*. (2013), Kamal *et al*. (2015), Yu *et al*. (2017), Santamaria *et al*. (2018) and Nesa *et al*. (2018), without considering the heterogeneity and dimensionality size of the sensory datasets. Hence, further research is required to develop an effective and reliable algorithm for discovering event and error outliers in IoT sensing datasets.

The redundant sensed data records are prevalent in IoT sensory dataset need to be filtered out or eliminated, to obtain the actual data required for onward processing. The process of eliminating redundant data records from a given data sample involves removing all but one of the replica records in the data sample. Figure 1.3 depicts a typical redundant data records scenario for more insight.

Report Number	Report Time	Outside Air Temp	Fuel Consumption	Id. Number	Course	Speed	Latitude	Longitude	Create Time
20									
2523464	16:03:57	6,2500	16000,0000	467190001004975	17,0000	91,7500	57,5403	12,5322	12:14:12
2523513	16:03:57	6,2500	16000,0000	467190001004975	17,0000	91,7500	57,5403	12,5322	12:14:13
2523873	16:03:57	6,2500	16000,0000	467190001004975	17,0000	91,7500	57,5403	12,5322	12:14:19
2523399	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:08
2523425	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:08
2523427	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:10
2523507	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:13
2523572	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:14
2523870	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:19
2523378	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:07
2523416	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:09
2523456	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:12
2523470	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:12
2523478	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:13
2523580	16:04:27	6,2500	49600,0000	467190001004975	353,0000	66,7500	57,5456	12,5344	12:14:14
2523376	16:04:57	6,2500	10000,0000	467190001004975	329,0000	74,2500	57,5504	12,5304	12:14:07
2523400	16:04:57	6,2500	10000,0000	467190001004975	329,0000	74,2500	57,5504	12,5304	12:14:08
2523421	16:04:57	6,2500	10000,0000	467190001004975	329,0000	74,2500	57,5504	12,5304	12:14:08

Figure 1.3 Examples of Duplicate Records in dataset

A critical observation of the dataset records indicates that the dataset is retrieved from a single-car due to the ID Number column (i.e., the fifth column highlighted). The Report Time column indicates the time at which the contents of the report where obtained. Furthermore, the rows marked with yellow have the same Report Time and the records appeared to be identical. Hence, records in the yellow background are regarded as redundant or duplicate records. The solution is to eliminate all but leaving one of the records with a similar report time.

An effective filtering algorithm is required to eliminate the redundant sensed data records with optimality solutions. Most of the algorithms developed by existing researchers mainly focus on discovering redundant data records by comparing the features in a dataset. Redundant data is identified when features with similar records are discovered. Therefore, identical features are eliminated from the dataset. This type of procedure is not very effective in IoT sensing-based IaaS Cloud computing environment due to the unstructured pattern of sensed data generated by IoT sensors on real-time basis (Xie and Chen 2017; Doghman *et al*., 2017). Therefore, the filtering process that considers eliminating identical sensed data records or points from a given dataset is required for onward research.

Due to the high demand of IaaS resource usage and challenges of allocating them in cloud data-center, scheduling of resources for the execution of IoT sensed data filtering operations in IoT-based cloud computing has attracted researchers' attention in recent times. Load balancing is one of the significant issues in IoT-based IaaS Cloud computing infrastructure (Ghomi *et al.,* 2017). It is the process of allocating and reallocating the workload among available resources to maximize throughput while minimizing the cost and response time, improving performance and resource utilization as well as the reduction of energy consumption (Priyanka *et al*., 2016; Goyal *et al*., 2016; Ghomi *et al*., 2017). Figure 1.4 illustrates a general picture of the loading balancing procedure in the cloud data center.

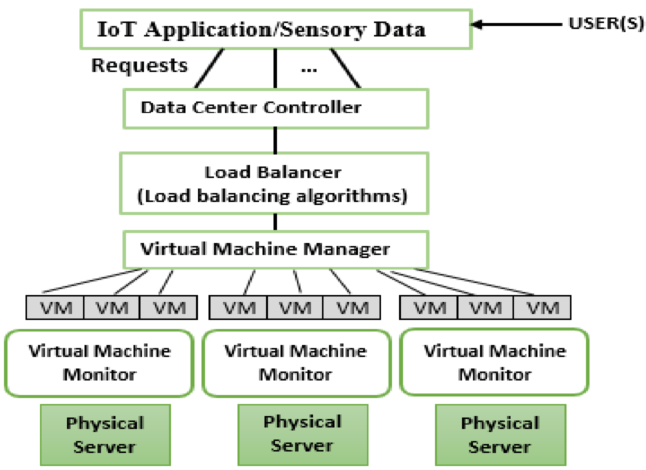

Figure 1.4 An Example of Load Balancing Model

Various Non-heuristics, hybrid and multi-objective algorithms have been deployed to resolve the load balancing problem, which produces optimal solutions for large size problems in the research work of Abed and Yunis (2019), Muhiudin and

Almogren (2019), Makala and Viswanathan (2019), Xavier and Annadura (2018), Jing *et al.*, (2018), Hallawi *et al.,* (2017), Chen and Chen (2017) and Yu *et al.* (2016). However, the performance level of algorithms reduces as the problem size increases. Therefore, further research is required to improve the performance of existing algorithms for optimal solutions for large size problems due to local space entrapment in IoT-based Cloud IaaS environment.

In IoT-based Cloud IaaS computing, effective and efficient data filtering with the support of optimized resource allocation algorithms is necessary to retrieve useful information with minimum resource utilization and computation time. Ways to filter sensing data during acquisition depend on various aspects such as the dynamicity of data, the volume of data generated over time, the issues related to the filtering process, the availability of storage and computation resources to store and execute the data filtering operation. Other issues are either the communication distance between the IoT sensing environmental coverage and the cloud IaaS location or cloud resource demand variation by end-users.

The problem leading to the research, as demonstrated in Figure 1.5, highlights the significant requirements that should be given thorough attention. The existing algorithms used in IoT-based IaaS Cloud computing infrastructure cannot guarantee the global search of detecting event and error outlier as the dimension of the sensory dataset increases in size, not considering the missing data effect that impedes the elimination of redundant sensory data records to obtain relevant data ones. They also suffer from inefficient load balancing of available resource allocation due to local entrapment. These challenges affect the users' expectations in terms of recall, precision and area under curve for event/error outlier predictions, retrieval of useful data that is free from redundancies that ensure improved reduction ratio, f1-score with minimum root mean squared error. Also, satisfying the cloud service provider's objectives, such as optimal workload balancing of active resources with a minimum number of virtual machine migration and resource utilization.

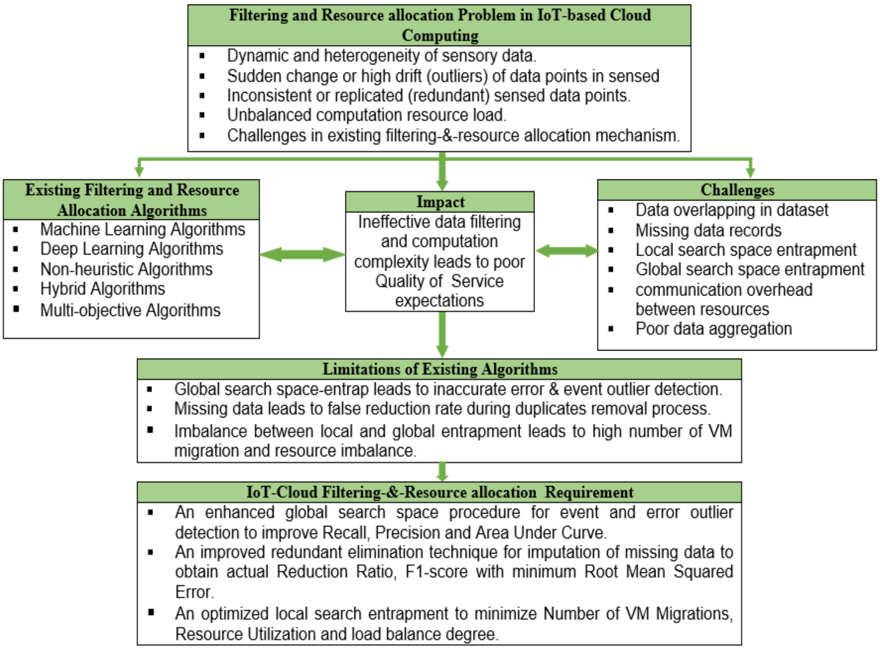

Figure 1.5 Problem Leading to Research

Global search is the ability to explore the search to cover as much distance as possible to obtain new or best solutions within the search space. Finding the efficient algorithm with high scalability to converge in the discovering of both event and error outliers in IoT sensed datasets is challenging. Existing researches in this field are yet to address the challenge of global entrapment that often leads to ineffective outliers (event/error) detection in IoT sensed datasets. The unsupervised clustering and outlier factor algorithms can help resolve this challenge and provide better performance at the global optimal IoT sensed dataset region.

In IoT-based cloud computing, redundant data records in IoT sensed datasets are mostly considered irrelevant data that utilizes huge storage space. Consequently, “without the recovery of missing data records in the datasets before embarking on redundant records elimination leads to removing some relevant data together with redundant records” (Gong et al., 2016; Feng et al., 2017). Missing data causes similarity data record mismatched subsequences during the redundant identification process. Therefore, a technique that can recover missing data before the elimination of redundant data records is required. The unsupervised imputation algorithm can be deployed for missing data recovery to eliminate redundant data records.

Local search space is searching for a neighborhood solution, continuously finding it challenging to obtain a near-optimal solution and deciding to attain its iteration limits. Local search often leads to poor convergence with high computation time. Therefore, a meta-heuristic algorithm that can optimize the local search for a near-optimal solution is highly needed for better resource load balancing performance on the cloud datacenter.

1.3 Problem Statement

IoT Sensory data filtering processes with the support of resource allocation strategy for IaaS in cloud computing infrastructure is considered a critical challenge. The filtering process is composed of two scenarios known as outlier detection and elimination of redundant data. Outliers (event and error) and redundant data records are usually classified as unwanted data. However, outliers are often predicted due to their usefulness, whereas redundant ones are eliminated out rightly from a given dataset. The detection of outliers (event and error) and eliminating redundant data are addressed by utilizing suitable algorithms, including machine and deep learning algorithms. Also, suitable algorithms are required to address the unbalanced workload of IaaS in the cloud, which includes non-heuristics and meta-heuristic algorithms. The results generated by the filtering and resource allocation algorithms depict the computation complexities of the input size objective functions. Therefore, effective filtering and efficient resource allocation techniques are necessary to resolve the abovementioned challenges in IoT-based Cloud IaaS infrastructure.

Some of the existing researches consider the detection of error and event outliers in a given dataset instead of the majority of them that mainly focus on the generality of outlier detection. The Non-parametric Sequence Learning (N-PSL) algorithm is deployed for detection event and error outliers with some significant performance improvement rate in terms of Accuracy, Specificity, Recall, Precision, False Positive Rate and Area under Curve. However, as the dimension of the dataset increases in size (i.e. 16 attributes and 51,500 objects), the performance rate of N-PSL algorithm degrades or becomes sub-optimal based on outlier (event/error) detection with the output of 69.849%/78.719% Recall, 57.590%/75.895% Precision and 81.633%/89.082% Area under Curve. This is due to its weaknesses in global space search for event and error outlier detection as the dimension of the dataset increases with numerous data irregularities. Therefore, a machine learning technique is critically needed to resolve the above-mentioned challenges to enhance the N-PSL algorithm's performance rate based on Recall, Precision and Area under Curve.

It is imperative to establish the correlation between sensed data records according to their timestamp to eliminate redundant ones from the whole dataset effectively. Due to the unstructured nature of IoT sensing data records generated overtime. The Perceptually Important Points (PIP) algorithm considers the similarity between data records and produces substantial performance results before and after eliminating redundant sensed data records. Nevertheless, as a weakness of eliminating relevant or useful data points together with redundant ones due to missing data in a given dataset. This leads to its sub-optimal performance rate based on F1-score, Reduction ratio and Root mean squared error. Therefore, an imputation technique is highly needed to handle missing data to improve the PIP algorithm's performance rate in terms F1-score, Reduction ratio Root mean squared error.

Conversely, the computation complexity of IoT sensed data filtering operations increases exponentially as their workload size increases. However, the existing non-heuristic, hybrid and multi-objective algorithms have produced better performance in exploring global and local search space in the available IaaS computation resources. However, the Energy-efficient Resource Ranking- based Virtual Machine Selection (ERVS) algorithm is still caught up in local search entrapment resulting in sub-optimal performance results. Therefore, the technique that can improve the local search space is highly needed to enhance the performance rate of ERVS algorithm in terms of Resource utilization, VM migration and Load balancing degree.

1.4 Research Questions

In view of the above discussion, the following research questions are presented towards attaining the research objectives of this research work:

i. How to avoid global optima for the performance improvement rates of event and error outlier detection as the dimension of sensed dataset increases in size?

ii. How to recover missing data before and after elimination of redundant sensed data records for optimal solutions?

iii. How to avoid the local optima to minimize workload imbalance of resources and enhance optimality solutions for resource allocation optimization?

1.5 Research Goal

This research aims to achieve optimal sensed data filtering-enabled resource allocation techniques that are efficient to explore global search for the detection of error and event outliers, recovery of missing data for redundant sensed data record elimination, and avoid local optima for optimal resource solution.

1.6 Research Objectives

This study mainly focuses on implementing IoT sensed data filtering and resource allocation techniques for IaaS Cloud computing infrastructure. Three objectives are established as:

i. To develop a Non-parametric Sequence Agglomerative Gaussian Mixture Model (N-PSAGMM) technique to minimize global search for improving event and outlier detection based on precision, recall and area under curve.

ii. To develop a Perceptually Important Point-enabled K-means Imputation (PIP-KCM) technique to recover missing data for improving the elimination of redundant data records in terms of reduction ratio, f1-score with minimum root mean squared error.

iii. To develop an Efficient Resource Balanced Ranking Virtual Machine Selection-based Glowworm Swarm Optimization (ERBV-GSO) technique for avoiding local optima to minimize the unbalanced workload, number of VM migration and resource utilization of available IaaS resources.

1.7 Research Scope

The scope of this research work is as follows:

i. This research considers the performance improvement rates based on precision, recall and area under curve to detect event and error outliers. Also, the performance improvement rates in terms of f1-score, reduction ratio with minimum root mean squared error for effective elimination of redundant sensory data records as well as load balancing degree, number of virtual machine migrations and resource utilization for optimal workload balancing on available resources for the execution tasks in IoT sensing-based IaaS Cloud computing infrastructure.

ii. All benchmarks and datasets are generated from the University of California (UCI) machine learning repository, except one that was retrieved from a laboratory.

iii. It also considers dependent and independent workload using a data-center Aware Round Robin load balancing Policy for resource allocation.

The physical and virtualization of IoT sensor devices, gateways and communication network bandwidth, Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) issue are out of the scope of this research.

1.8 Thesis Organization

The structure of the remaining thesis is outlined as follows:

Chapter 2 presents a detailed description of previous and related research works. It explicitly shows the position of the research from previous literature. This chapter also briefly discusses the literature that has deployed the Non-parametric Sequence Learning (N-PSL) algorithm in resolving several challenges in other research fields. Chapter 3 describes the research methodology of the work. It is categorized into three research phases. The first phase comprise of the research background, preparation and planning. The second phase presents the design and development of effective filtering and efficient resource allocation techniques according to the set objectives of this research work. The third phase is made up of testing, performance evaluation and validation of the proposed techniques of this research. Chapter 4 presents the implementation of the filtering techniques, namely N-PSAGMM and PIP-KCM. The proposed filtering techniques are formulated based on the Non-parametric Sequence Learning (N-PSL) and Perceptually Important Points algorithms. Also, experimentation and analysis of results for the proposed techniques and comparing them with related existing filtering techniques. Chapter 5 presents the implementation of the resource allocation technique namely ERVB-GSO. A detailed description of how the proposed technique is designed and developed is depicted in this chapter. Experimentation and discussion of results, as well as comparing its performance with other existing resource allocation algorithms, are also presented. Chapter 6 presents the conclusion, a brief discussion of this research's contributions and suggests future directions. It also presents the achievements of the set objectives and the comparative performance evaluati

REFERENCES

Abawajy, J. H. and Hassan, M. M. (2017). Federated Internet of things and cloud computing pervasive patient health monitoring system. *IEEE Communications Magazine*, 55(1), pp.48-53.

Adhikari M. and Srirama, S., N. (2019). Multi-objecitve Accelerated Particle Swarm Optimization with a Container-based Scheduling for Internet of Things in Cloud Environment. *Journal Of Network and Computer Applications (Elsevier)*, 137,pp.35-61.

Aggarwal, C. C. (2015). Outlier Analysis, *Data Mining Journal (Springer),*pp.237-263.

Ahmad, A., Khan, M., Paul, A., Din, S., Rathore, M. M., Jeon, G., and Choi, G. S. (2018). Toward modeling and optimization of features selection in Huge volumeData based social Internet of Things. *Future Generation Computer Systems (Elsevier)*, 82(2018), pp.715-726.

Akbar, A., Khan, A., Carrez, F., and Moessner, K. (2017). Predictive analytics for complex IoT data streams. *IEEE Internet of Things Journal*, 4(5), pp.1571-1582.

Ali, Z., Muhammad, G. and Alhamid, M. F. (2017). An automatic health monitoring system for patients suffering from voice complications in smart cities. *IEEE Access*, 5(2017), pp.3900-3908.

Alipourfard, O., Liu, H. H., Chen, J., Venkataraman, S., Yu, M., and Zhang, M. (2017). Cherrypick: Adaptively unearthing the best cloud configurations for huge volumedata analytics. *Proceedings of 2017 Symposium on Networked Systems Design and Implementation*, pp.469-482.

Aljawarneh, S. A. and Vangipuram, R. (2018). GARUDA: Gaussian dissimilarity measure for feature representation and anomaly detection in Internet of Things, *The Journal of Supercomputing (Springer Link)*, Pages 1-38

Amir, G. and Murtaza, H. (2015). Byond the hype: Huge volumedata concepts, methods and analytics, *International Journal of Information Management (Elsevier)*, 35, pp.137-144.

Anderson, J., L., R., Graham, M., S., Zsoldos, E. and Sotiropoulos, S., N. (2016). Incorprating outlier detection and replacement into a non-parametric framework for movement and distortion corrction of diffusion MR, *Neurolmage (Elseveir)*, 141, pp.556-572.

Andersson, J., L., R. and Sotiropoulos (2015). Non-parametric representation and prediction of single-and multi-shell diffusion-weighted MRI data using Gaussain processes, *Neurolmage (Elsevier)*, 122, pp.166-176.

Arijeet Satapathy, Seva Mahapatra and Riteshi K. Pandey (2018). Optimization Algorithm: Glowworm Swarm Optimization, *System Programming Project Report, International Institute of Information Technology*, pp.1-18.

Arunachalam, D., and Kumar, N. (2018). Benefit-based Comsumer Segmentation and Perforamnce Evaluation of Clustering Approaches: An Evidence of Data-driven Deciosion-making, *Expert Systems with Applications (Elsevier)*, 111, pp.11-34.

Aya, A., Oussama, G., Abdulfattah, M., O. and Mohamed A. (2017), Outlier detection approaches for wireless sensor networks: A survey.*Computer Networks (Elsevier*), 129(2017), pp.319-333.

Babu, D., L., D. and Krishna P., V. (2013). Honey bee Behaviour Inspired Load Balancing of Tasks in Cloud Computing Environments, *Applied Soft Computing (Elsevier)*, 13, pp.2292-2303.

Babukartik, R., G. and Dhavachelvan, P. (2012). Hybrid Algorithm using the Advantage of ACO and Cuckoo Search for Job Scheduling, International *Journal of Information Technolgoy Convergence and Services (IJITCS)*, 2(4), pp.25-34.

Bagnall, A., Lines, J., Hills, J., Bostrom, A. (2015). Time-Seres Classifcation with COTE: The Collective of Transformation-based Ensemles for IoT Huge volumeData, *IEEE Transactions on Knowledge and Data Engineering*, 27(9), pp.2522-2535.

Bandyopadhyay, S., Ukil, A., Puri, C., Singh, R., Bose, T. and Pal, A. (2016). SensIPro: Smart Sensor Analytics for Internet of Things, *Proceedings of 2016 IEEE Symposium on Computers and Communications (ISCC)*, pp.1-7.

Bastiani, M., Cottaar, M., Fitzgibbon, S., P., Suri, S., Alfaro-Almagro, F., Sotiropoulos., S., N., Jbaddi, S. and Andersson., J. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction, *NeuroImage (Elsevier)*, 184, pp.801-812.

Basu S., Karuppiah, M., Selvakumar, K., Li, K-C., Islam, S., K., H., Hassan, M., M. and Bhuiyan, Md., Z., A. (2018). An Intelligent/Cognitive Model of Task Scheduling for IoT Applications In Cloud Computing Environment. *Future Generation Computer Systems (Elsevier),* 88, pp.254-261.

Bedzek, J., C., Ehrlich, R., Full, W. (2016). FCM: The Fuzzy c-Means Clustering Algorithm, *Computers and Geosciences*, 10(2-3), pp.191-203.

Beloglazov A., Abawajy J. and Buyya R. (2012). Energy-awere resource allocation heuristics for efficient management of datacenters for Cloud Computing, *Future Generation Computer Systems*, 28(5), pp. 755-768.

Bhargava, A. (2019). Grouping of Medical Drugs Used for Similar Symptoms by Mining Clusters from drug Benefits Reviews, *Proceedings of 2019 International Conference on Sustainable Computing in Science*, pp.1-10.

Bikmukhamedov, R., F. and Nadeev, A., F. (2019). Lightweight Machine Learning Classifiers of IoT Traffic Flows. *Proceedings of 2019 IEEE Conference on Systems of Signal Synchronization, Generating and Processing in Telecommunications*, pp.1-5.

Boveiri, H., R., Khayami., R., Elhoseny, M. and Gunasekaran, M. (2018). An Efficient Swarm-Intelligent Approach for Task Scheduling in Cloud-based Internet of Things Applications. *Journal of Ambient Intelligence and Humanized Computing (Springer)*, 10, pp.3469-3479.

Branch, J., W., Giannella, C., Szymanski, B., Wolff, R. and Kargupta, H. (2013). In-network outlier detection in wireless sensor networks, *Knowledge and Information Systems (Springer)*, 34, pp.23-54.

Bu, F. (2018). An efficient fuzzy c-means approach based on canonical polyadic decomposition for clustering huge volumedata in IoT. *Future Generation Computer Systems (Elsevier)*, 88(2018), pp.675-682.

Burgues, J., Jimenez-Soto,J., M. and Marco,S. (2018). Estimation of limit of detection in semiconductor gas sensors through linearized calibration models, *Analytica Chemical Acta (Elsevier)*, 1013, pp.13-25.

Cabrieto, J., Tuerlincks, F., Kuppens, P., Grassmann, M. and Ceulemans, E. (2016). Detecting Correlation changes in multivariate time series: A comparsion of

Four non-parametric change point detection methods, *Journal Behavior Research Methods (Springer)*, 49, 988-1005.

Cai, H., Xu, B., Jiang, L. and Athanasios, V. (2017). IoT-Based Huge volumeData Storage Systems in Cloud Computing: Perspecives and Challenges. *IEEE Internet of Things Journal*, 4(1), pp. 75-87.

Chang, H.-T., Mishra, N., & Lin, C.-C. (2015). IoT big-data centred knowledge granule analytic and cluster framework for BI applications: a case base analysis. *PloS one*, 10(11), pp.1-23.

Chang, K-H. (2015). e-Design: Computer-Aided Engineering Design, *Elsevier Book Series*, pp.1-1234.

Chavez, R., Rooco, R., Gutierrez, A., G., Dorner, M. and Estay, S., A. (2019). A Self-Calibrated Non-Parametric Time Series Analysis Approach for Assessing Insect Defoliation of Broad-Levaed Deciduous Nothogagus pumilio Forests, *MDPI Remote Sensing*, 11(2), pp. 1-20.

Chen, Y., C., Sun, W., C. and Juang, J., C. (2010). Outleir detection technique for RSS-based localization problems in wireless sensor networks, *Proceedings of 2010 on SICE Annual Conference*, pp. 657-662.

Chen, Y., Wang, K., Zhong, P. (2016). One Class Supprt Tensor Machine, *Knowledge-Based Systems*, 96(2016), pp.14-28.

Chen, Y.-H., and Chen, C.-Y. (2017). Service oriented cloud VM placement strategy for Internet of Things. *IEEE Access*, 5, 25396-25407.

Cheng, S., Cai, Z., Li, J. and Gao, H. (2016). Extracting kernel dataset from huge volumesensory data in wireless sensor networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(4), pp.813-827.

Daneshpazhouh, A. and Sami, A. (2014). Entropy-based outlier detection using semi-supervised approach with few positive examples. *Pattern Recognition Letters*, 49(2014), pp.77-84.

Das, A., Acharya, U., R., Panda, S., S. and Sabut, S. (2019). Deep Learning based Liver Cancer Detection using Watershed Transform and Gaussian Mixture Model Techniques, *Cognitive Systems Research (Elsevier)*, 54, pp.165-175.

Delicato., F., C., Pires P., F. and Batista, T. (2017). Resource Management for Internet of Things, *Springer Briefs in Computer Scienc*e, pp.1-5.

Delimargas, A., Skevakis, E., Halabian, H., Lambadaris, I., Seddigh, N., Nandy, B. and Makkar, R. (2015). IPCA for network anomaly detection. *Proccedings of 2015 Conference at IEEE Military Communications Conference*, pp.617-622.

Deng, X., Jiang, P., Peng, X., Mi, C. (2019). An Intelligent Outlier Detection Method with One-class Support Tucker Machine and Genetic Algorithm Toward Huge volumeSensor Data in Internet of Things, *IEEE Transactions on Industrial Electronics*, 66(6), pp.4672-6683.

Deng, X., Jiang, P., Peng., X. and Mi, C. (2017). Support High-order Tensor Data Description for Outleir Detection in High-dimensional Huge volumesensor Data, *Future Generation Computer Systems (Elsevier)*. 81(2018), pp.177-187.

Diaz-Rozo, J., Bielza, C., and Larrañaga, P. (2018). Clustering of Data Streams With Dynamic Gaussian Mixture Models: An IoT Application in Industrial Processes. *IEEE Internet of Things Journal*, 5(5), pp.3533-3547.

Doghman, F., Chaczko, Z. and Jiang, J. (2017). A Review of Aggregation Algorithms for the Internet of Things, *Proceedings of 2017 IEEE Internationl Conference on Systems Engineering*, pp.481-487.

Elhoseny, M., Abdelaziz, A., Salama, A., S., Riad, A. M., Muhammad, K. and Sangaiah, A., K. (2018). A Hybrid model of Internet of Things and Cloud Computing to Management huge volumedata in Health Services Applications. *Future Generation Computer Systems (Elsevier)*, 86, pp.1383-1394.

Erfani, S., M., Rajasegarar, S., Karunasekera, S. and Leckie, C. (2016). High-dimensional and Large-scale anomaly Detection using a Linear One-class SVEM with Deep Learning, *Pattern Recognition (Elsevier)*. 58(2016), pp.121-134.

Falk, E., Camino, R., State, R. and Gurbani, V., K. (2017). On Non-parametric models for detecting outages in the mobile network, *Proceedings of 2017 IFIP/IEEE International Symposium on Integrated Network and Service Management (MI)*, pp. 1139-1142.

Fekade, B., Maksymyuk, T., Kuryk, M. and Jo, M. (2018). Probabilistic Recovery of Incomplete Sensed Data in IoT, *IEEE Internet of Things Journal*, 5(4), pp. 2282-2292.

Feng, L., Kortoci, P. and Liu, Y. (2017). A Multi-tier Data Reduction Mechanism for IoT Sensors. *Proceedings of 2017 ACM International Conference on Internet of Things*, pp.1-8.

Fernandez, L., Yan., J., Fonollosa, J., Burgues, J., Gutierrez, A. and Marco, S. (2018). A Practical Method to Estimate the Resolving Power of a Chemical Sensor Array: Application toFeature Selection, *Frontiers in Chemistery*, 6, pp.1-18.

Fonollosa, J., Sheik, S., Huerta, R. and Marco, S. (2015). Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring, *Sensors and Actuators (Elsevier)*, 215, pp.618-629.

Ganz, F., Barnaghi, P. and Carrez, F. (2013). Information abstraction for heterogeneous real world Internet data. *.IEEE Sensors Journal*, 13(10), pp.3793-3805.

Ghanbari, Z., Navimipour, N., J., Hosseinzadeh, M. and Darswesh, A. (2019). Resource Allocation mechanisms and approaches on the Interent of Things, *Cluster Computing (Springer)*, 22, pp.1253-1282.

Ghomi, E. J., Rahmani, A. M, Qader, N. N. (2017). Load-balancing Algorithms in Cloud Computing: A survey, *Journal of Network and Computing Applications*, 88, pp.50-71.

Gong, X., Yain-Whar, S., Fong S. and Biuk-Aghai, R., P. (2016). Financial time series pattern matching with extended UCR Suite and Support Vector Machine. *Expert Systems With Applications (Elsevier)*, 55, pp. 284-296.

Gonzalez-Vidal, A., Barnaghi, P. and Skarmeta, A., F.(2018). BEATS: Blocks of Eigenvalues Algorithm for Time Series Segmentation, *IEEE Transactions On Knowledge and Data Engineering*, 30(11), pp.2051-2064.

Goyal R., Shipra, and Verma, M. K. (2016). "Load balancing techniques in cloud computing environment: a review." *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(4), pp.583-588.

Gupta H. and Sahu K. (2016). Honey Bee Behavior Based Load Balancing of Task in Cloud Computing, *International Journal of Science and Research*, 3(6), pp.842-846.

Ha, J., Seok, S. and Lee, J-S. (2015). Robust Outlier Detection Using Instability Factor, *Knowledge-Based Ststems (Elsevier)*, 63, pp.15-23.

Hagberg, A., Schult, D. and Swart., P. (2019). Refence Networkx, *In networkx.github*, pp. 1-772.

Haghighat, M., Abdel-Mottaleb, M., and Alhalabi, W. (2016). Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Transactions on Information Forensics and Security*, 11(9), pp.1984-1996.

Hallawi, H., Mehnen, J. and He, H. (2017). Multi-Capacity Combinatorial Ordering GA in IoT Application to Cloud resources allocation and efficient virtual machines consolidation. *Future Generation Computer Systems (Elsevier),* 69, pp.1-10.

Han, Q., Meng, F., Hu, T. Chu, F. (2017). Non-parametric hybrid models for wind speed forecasting, *Energy Conversion and Management (Elsevier)*, 184, pp.554-568.

Havet, A., Schiavoni, V., Felber, P., Comat, M., Rouvoy, R. and Fetzer, C. (2017). GENPACK: A Generational Schedular for Cloud Datacenters, *Proceedings of 2017 IEEE Internetional Conference on Cloud Engineering (IC2E)*, pp. 95-104.

Hieu, N. T., Di Francesco, M., & Jääski, A. Y. (2014). A virtual machine placement algorithm for balanced resource utilization in cloud datacenters. *Proccedings of 2014 IEEE International Conference on Cloud Computing*, pp.474-481.

Huang, J., Zhu, Q., Yang, L. and Feng, J. (2016). A non-parametric outlier detection algorithm based on Natural Neighbor, *Knowledge-based Systems (Elsevier)*, 92, pp.71077.

Idrees, F., Rajarajan, M., Conti, M., Chen, T., M. and Rahulamathavan, Y. (2017). Pindroid: A novel Android Malware detection system using ensemble learning methods, *Computers and Security (Elsevier)*,68, pp.36-46.

Islam, T., Mukhopadhyay, S. C., and Suryadevara, N. K. (2017). Smart sensors and Internetof things: a postgraduate paper, *IEEE Sensors Journal,* 17(3), 577-584.

Jarwan, A., Sabbah, A. and Ibnkahla, M. (2019). Data Transmission Reduction Schemes in WSNs for Efficient IoT Systems, *IEEE Journal on Elected Areas in Communications*, 37(6), pp.1307-1324.

Jing, W., Miao, Q., and Chen., G. (2018). An Open Scheduling Framework for QoS Resource Managemnt in the Internet of Things, *KSII Transactions on Internet and Information Systems*, 12(9), pp.4103-4121.

Kalaiselvi T., Nagaraja P. and Bashith Z. Abdul Bashit (2017). A Review on Glowworm Swarm Optimizaton, *International Journal of Information Technology*, 3(2), pp.49-56.

Kamal, S., Ramadan, R. A. and Fawzy, E., R. (2015). Smart outlier detection of wireless sensor network. *Facta Universitatis, Series: Electronics and Energetics*, 29(3), pp.383-393.

Karkhanis, B. and Sharma, S. (2019). A Review On: Finding Outlier Points On Real Dimensional Data Sets, *Engineering Journal*, 7, pp. 45-58.

Kaur, D., Aujla, G., S., Kumar, N., Zomaya, A., Y., Perera, C. and Ranjan, R. (2018). Tensor-Based Huge volumeData Management Scheme for Dimensionality Reduction Problem in Smart Grid System: SDN Perspective, *IEEE Transactions on Knowledge and Data Engineering*, 30(10), pp.1985-1998.

Kaur, P., D. (2015). A survey on huge volume data storage strategies, *Proceedings of 2015 International Conference on Green Computing, Internet of Things*, pp.2209-2212.

Khadkari, H. and Maghrebi, S., G. (2016). Necessity of the Integration Internet of Things and Cloud Services with Quality of Service Assurance Approach, *Bulltin De La Societe Royale Des Sceicnes De Liege*, 85, pp.434-445.

Khamparia, A. and Pandey, B. (2015). Heuristics and Non-heurisitcs Based Service Composition Methods, *Proceedings of 2015 National Conference on Management and Technology for Skill Development: Innovative approaches*, pp. 1-9.

Klein, B. and Bodenson, R. (2019). Advanced Topics with Python: Software for Complex Networks, *Python-course*, pp.1-2.

Kolozali, Ş., Puschmann, D., Bermudez-Edo, M. and Barnaghi, P. (2016). On the effect of adaptive and nonadaptive analysis of time-series sensory data. *IEEE Internet of Things Journal*. 3(6), pp.1084-1098.

Kondo, Y., Salibian-Barrera, M. and Zamar, R. (2016). RSKC: An R Package for a Robust and Sparse K-Means Clustering Algorithm, *Journal of Statistical Software*, 72(5), pp.1-24.

Kong, X., Chang, J., Niu, M., Huang, X., Wang, J. and Chang, S. I. (2018). Research on real time feature extraction method for complex manufacturing huge volumedata. *The International Journal of Advanced Manufacturing Technology*, 99(5-8), pp.1101-1108.

Kontaki, M., Gounaris, A., Papadopoulos, A. N., Tsichlas, K. and Manolopoulos, Y. (2016). Efficient and flexible algorithms for monitoring distance-based outliers over data streams. *Information systems*, 5(5), pp.37-53.

Kuang, L., Yang, L., T., Chen, J., Hao, F., and Chanqing, L. (2018). A Holistic Approach for Distributed Dimensionality Reduction of Huge volumeData, I*EEE Transactions on Cloud Computing*, 6(2), pp.506-518.

Kuhn , D., Probst, P., Thomas, J. and Bischl, B. (2018) Automatic Exploration of Machine Learning Experiments on OpenML, *Cornell University Website Portal*, pp.1-6.

Kumar, S. and Krishnan, S. (2018). Internet of Things Based Clinical Decsion Support System Using Data Mining Techniques, J*ournal of Advanced Research in Dynamic and Control Systems*, 10(4), pp.132-139.

Kundu, S. (2019). Various Types of Distances Metrics in MAchine Learning, *Analytics*. pp.2-7.

Kurdi, H., Ezzat, F., Altoaimy, L., Ahmed, S., H. and Youcef-Toumi, K. (2018). MultiCuckoo: Multi-Cloud Service Composition Using a Cuckoo-Inspiere Algorithm for the Internet of Things Applications, *IEEE Accesss*, 6, pp.56737-56749.

Lakshmanaprabu, S., Shankar, K., Khanna, A., Gupta, D., Rodrigues, J. J., Pinheiro, P. R., and De Albuquerque, V. H. C. (2018). Effective features to classify huge volumedata using social Internet of things. *IEEE Access*, 6 (2018), pp.24196-24204.

Leroux, S., Bohez, S., De Coninck, E., Verbelen, T., Vankeirsbilck, B., Simoens, P. and Dhoedt, B. (2017). The cascading neural network: building the Internet of Smart Things. *Knowledge and Information Systems*, 52(3), pp.791-814.

Li, C., Wei, W., Li, J. and Song, W. (2017). A cloud-based monitoring system via face recognition using Gabor and CS-LBP features. *The Journal of Supercomputing*, 73(4), pp.1532-1546.

Li, H., Li, W., Wang, H. and Wang, J. (2018). An optimization of virtual machine selection and placement by using memory content similarity for server consolidation in cloud. *Future Generation Computer Systems (Elsevier)*, 84, pp.98-107.

Li, H., Ota, K. and Dong, M. (2018). Learning IoT in edge: deep learning for the Internet of things with edge computing. *IEEE Network*, 32(1), pp.6-101.

Li, J., Lu, Z., Zhang, W., Wu, J., Qiang, H., Li, B. and Hung, P. C. (2018). SERAC3: Smart and economical resource allocation for huge volumedata clusters in community clouds. *Future Generation Computer Systems (Elsevier).* 85, pp.210-221.

Li, J., Song, S., Zhang, Y. and Zhou, Z. (2016). Robust K-means-Median and K-means Clustering Algorithms for Incomplete Data, *Mathematical Problems in Engineering*, 2016, pp.1-8.

Li, P., Chen, Z., Yang, L. T., Zhang, Q. and Deen, M. J. (2018). Deep convolutional computation model for feature learning on huge volumedata in Internet of Things. *IEEE Transactions on Industrial Informatics*, 14(2), pp.790-798.

Liberatore F. and Camacho-Collados M. (2016). A Comparison of Local Search Methods for the Multicriteria Police Districting Problem on Graph, *Mathematical Problems in Engineering (Hindawi)*, 2016, pp.1-14.

Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Yunsheng, M. and Hou, P. (2018). A new deep learningbased food recognition system for dietary assessment on an edge computing service infrastructure. *IEEE transactions on services Computing*, 11(2), pp.249-261.

Luo, T. and Nagarajan, S. G. (2018). Distributed Anomaly Detection Using Autoencoder Neural Networks in WNS for IoT, *Proceedings of 2018 IEEE International Conference on Communications,* pp.1-6.

Lyu, L., Jin, J., Rajasegarar, S., He, X. and Palaniswami, M. (2017). Fog-empowered anomaly detection in IoT using hyperellipsoidal clustering. *IEEE Internet of Things Journal*, 4(5), pp.1174-1184.

Ma, J., Jiang, J., Liu, C. and Li, Y. (2017). Feature Guided Gaussian Mixture Model with Semi-supervised EM and Local Geometric Constraint for Retinal Image Registration, *Infmration Sciences (Elsevier)*. 427, pp.128-142.

Mallor, F., Leon, T., Boeck, L., D., Gulck, S., V., Meulders, M. and Meerssche, B., V., D. (2017). A methods for detecting malfunctions in PV solar panels based on electricity production monitoring, *Solar Energy (Elsevier)*, 153, pp.51-63.

Manogaran, G., Vijayakumar, V., Varatharajan, R., Kumar, P., M., Sundarasekar, R. and Hsu, C-H. (2018). Machine Learning Based Huge volumeData Processing Framework for Cancer Diagnosis Using Hidden Markov Model and GM Clustering, *Wireless Personal Communications (Springer Link),*102(3), pp.2099-2116.

Marjani, M., Fariza, N., Gani, A., Karim, A., Hashem, I., A., T. and Saddiqa, A. (2017). Huge volumeData Analytics: Achitecture, Opportunities, and Open Research Challenges. *IEEE Access*, 5(2017), pp.5247-5261.

Marwa M. Abed and Manal F. Younis(2019). Developing Load Balancing for IoT-Cloud Computing Based on Advanced Firefly and Wieghted Round Robin Algorithms, *Baghdad Science Journal,* 16(1), pp.130-139.

Masdari, M., Nabavi, S., S. and Ahmadi, V. (2016). An Overview of machine placement schemes in cloud computing, J*ournal of Network and Computer Applications (Elsevier)*, 66, pp.106-127,

McCormick J. Patrick, Lin Hung-mo, Deiner G. Stacie and Levin A. Mathew (2018). Validation of the All Patient Refined Diagnosis Related Group (APR-DRG) Risk of Mortality and Severity of Illness Modifiers as a Measure of Perioperative Risk, *Journal of Medical Systems (Springer),* 42 (81), pp. 2-6.

Meer D. W., Munkhammar J. and Widen J. (2018). Probabiistic Forecasting of solar power, Electricity Consumption and Net Load: Investigating the effect of Seasons, Aggregation and Penetration on Prediction Intervals, *Solar Energy (Elsevier)*, 171, pp. 397-413.

Mekala, M. S., and Viswanathan, P. (2019). Energy-efficient virtual machine selection based on resource ranking and utilization factor approach in cloud computing for IoT. *Computers & Electrical Engineering (Elsevier)*, 73, pp.227-244.

Modi, K. and Oza, B. (2016), Outlier Analysis Approaches in Data Mining,Internat*ional Journal of Innovative Research in Technology,* 3(7), pp.6-12.

Mohammadi, M., Al-Fuqaha, A., Guizani, M. and Oh, J.-S. (2018). Semisupervised deep reinforcement learning in support of IoT and smart city services. *IEEE Internet of Things Journal*, 5(2), pp.624-635.

Mohiuddin, I. and Almogren, A. (2019). Workload aware VM consolidation method in edge/cloud computing for IoT applications. *Journal of Parallel and Distributed Computing (Elsevier),* 123, pp.204-214.

Moura J. and Hutchison, D. (2016). Review and analysis of networking challenges in cloud computing, *Journal of Network and Computer Applications (Elsevier)*, 60(2016), pp.113-129.

Muhammad, G., Alhamid, M. F., Alsulaiman, M. and Gupta, B. (2018). Edge computing with cloud for voice disorder assessment and treatment. *IEEE Communications Magazine*. 56(4), pp.60-65.

Munandar, T., A., Azhari, Musdholifah A. and Arsyad, L. (2019). Modified Agglomerative Clustering with Location Quotent for Identification of Regional Potential Sector, *Journal of Theoretical and Applied Information Technology*, 95(5), pp.1191-1199.

Narman, H. S., Hossain, M. S., Atiquzzaman, M. and Shen, H. (2017). Scheduling Internet of things applications in cloud computing. *Annals of Telecommunications (Springer)*, 72(1-2), pp.79-93.

Nasser, T. and Tariq, R., S. (2015). Huge volumeData Challenges. *Journal of Computer Engineering and Information Technology*, 4(3), pp.1-10.

Neely, M., Guilder M., V., Yamada., W., Schumitzky., A. and Jellife, R. (2012). Accurate detection of outliers and subpopulations with Pmetrics, a non-parametric and parametric pharmacometric modeling and simulation pacakge for R, *PMC Journal*, 34(4), pp.467-476.

Nekouie, A. and Moattar, M., H. (2019). Missing Value Imputation for breast cancer diagnosis data using tensor factorization improved by enhanced reduced adaptive particle swarm optimization, *Ournal of King Saud University-Computer and Information Sciences*, 31, pp.287-294.

Nesa, N., Ghosh, T. and Banerjee, I. (2018). Non-parametric sequence-based learning approach for outlier detection in IoT. *Future Generation Computer Systems (Elsevier)*, 82(2018), pp.412-421.

Nguyen, G., Dlugolinsky, S., Bobak, M., Tran, V., Garcia, A., L., Heredia, I., Malik, P. and Hluchy, L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey, *Artificial Intelligence Review*, 52, pp.77-124.

Nurezayana Zainal, Azlan Mohd Zain, Nor Haizan Mohamed Radzi andMuhamad Razib Othman (2016). Glowworm swarm optimization (GSO) for optimzation of machining parameters, *Journal of Intelligent Manufacturing*, 27, pp.797-804.

Osanaiye, O., Chen, S., Yan, Z., Lu, R., Choo, K., R. and Dlodlo, M. (2017). Frome Cloud to Fog Computing: A Review and a Concepual Live VM Mitigation Framework, *IEEE Access*, 5, pp.8284-8300.

Paloma Ferrando-Vivas, Jones Andrew, Rowan M. Kathryn and Harrison A. David (2017). Development and Validation of the new ICNARC model for Prediction of Acute Hospital Mortality in Adult critical care, *Journal of Critical Care (Elsevier)*, 38, pp. 335-339.

Papageorgiou, A, Cheng, B. and Kovacs, E. (2015), Real-Time Data Reduction at the Network Edge of Internet-of-Things Systems, *Proceedings of 2015 IEEE International Conference on Network and Service Management*, pp.284-291.

Paulraj, G. J. L., Francis, S. A. J., Peter, J. D. and Jebadurai, I., J. (2018). Resource-aware virtual machine migration in IoT cloud. *Future Generation Computer Systems (Elsevier)*, 85, pp.173-183.

Posegga Oliver (2018). Network Analysis in Python: A Brief Introuction, *Proceedings of 2018 International workshop of MMB*, pp.1-8.

Priyanka, S., Baaga, P. and Gupta, S. (2016). Assorted load balancing algorithms in cloud computing: A survey."*International Journal of Computer Applications*, 143(7), pp.1-8.

Puschmann, Daniel., Barnaghi, P. and Tafazolli, R. (2017). Adaptive clustering for dynamic IoT data streams. *IEEE Internet of Things Journal*, 4(1), pp.64-74.

Raafat, H. M., Hossain, M. S., Essa, E., Elmougy, S., Tolba, A. S., Muhammad, G. and Ghoneim, A. (2017). Fog intelligence for real-time IoT sensor data analytics. *IEEE Access*, 5(2017), pp.24062-24069.

Raffat, H., M., Hossain, M., S., SamirElmougy, E., E., Tolba, A., S., Muhammad, G. and Ghoniem, A. (2017). Fog Intelligence for Real-Time IoT Sensor Data Analytics, *IEEE Access*, 05, pp.24062-240069.

Rafferty, M., Liu, X., Laverty, D. M. and McLoone, S. (2016). Real-time multiple event detection and classification using moving window PCA. *IEEE Transactions on Smart Grid*. 7(5), pp.2537-2548.

Rao, K., S. and Thilagam P., S. (2016). Heuristics based Server Consolidation with Residual Defragmentation in Cloud Datacenter, *Future Generation Computer Systems (Elsevier)*, 50, pp.87-98.

Rehman, M. H., Ahmed, E., Yaqoob, I., Hashem, I. A. T., Imran, M. and Ahmad, S. (2018). Huge volumeData Analytics in Industrial IoT Using a Concentric Computing Model. *IEEE Communications Magazine, 56*(2), pp.37-43.

RocketStats, R. (2017). An Introduction to Gaussian Mixture Modeling, *R news and Tutorials COntributed by Hundereds of R bloggers*, pp.1-4.

Samie, F., Bauer, L. and Henkel, J. (2019). From Cloud Down to Things: AN Overview of Machine Learning in Internet of Things, *IEEE Internet of Things Journal*, 6(3), pp.4921-4934.

Santamaria, A. F., De Rango, F., Serianni, A.and Raimondo, P. (2018). A real IoT device deployment for eHealth applications under lightweight communication protocols, activity classifier and edge data filtering. *Computer Communications (Elsevier)*, 128 (2018), pp.60-73.

Santos, J., Leroux, P., Wauters, T., Volckaert, B. and Turck, F. D. (2018). Anomaly detection for Smart City applications over 5G low power wide area networks, *Proceedings of 2018 IEEE/IFIP Conference of Nework Operations and Management Symposium*, pp.1-9.

Sezer, O., B., Dogdu, E. and Ozbayoglu, M., A. (2018). Context-Aware Computing. Learning, and Huge volumeData in Internet of Things: A Survey, *IEEE Internet of Things Journal*. 5(1), pp.1-27.

Shahid, N., Naqvi, I. H.and Qaisar, S. B. (2015), Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey. *Artificial Intelligence Review (Springer)*, 43 (2), pp.193-228.

Shakhari, S. and Banerjee, I. (2019). A Multi-class Classifiction System forContinous Water Quality Monitoring, *Heyliyon (Elsevier)*, 5(5), pp.1-11.

Sharma, N. and Guddeti, R., M. (2016). Multi-objective energy efficient Virtual Machines Allocations at the cloud datacenter. *IEEE Transactions on Services Computing*, 1(99), pp.1-2.

Sharma, V., You, I. and Kumar, R. (2017). Intelligent Sensing Model for Anomalies Detection in Cross Platform OSNs with a Case Study on IoT, *IEEE Access*, 5, pp.3284-3301.

Shidik, G., Azhari, A. and Mustofa, K. (2016). Improvement of energy efficiency at cloud datacenter based on fuzzy markov normal algorithm vm selection in dynamic vm consolidation. I*nternational Review on Computers and Software,* 11(6), pp.511-520.

Silas , P., S-S., Marques, W., S., Rossi, F., D., Redrigues, G., C. and Calherios, R., N. (2017). Performance and accuracy trade-off analysis of techniques for anomaly detection in IoT sensors, *Proceedings of 2017 International Conference on Information Networking (ICOIN),* pp.486- 491.

Simar, L. and Wilson, P., W. (2015). Statistical Approaches for Non-parametric Frontier Models: A Guided Tour, *International Statistical Review*, 83(1), pp.77-110.

Skakun, S., Franch, B., Vermote, E., Roger, J-C., Becker-Reshef, I., Justice, C. and Kussul, N. (2017). Early Season Large-area winter Crop Mapping Using MODIS NDVI data, growing degree days Information and a Gaussian Mixture Model, *Remote Sensing of Environemnt (Elsevier)*, 195, pp.244-258.

Souto, M. C. P., Jaskowiak, P. A., and Costa, I. G. (2015). Impact of Missing Data Imputation methods on gene Expression Clustering and Classfication, *BMC Bioinformatics*, pp.1-9.

Souza, A, M., C. and Amazonas, J. R. A. (2015). An Outlier Detection Algorithm using Huge volumeData Processing and Internet if Things Architecture, *Procedia (Elsevier)*, 52, pp.1010-1015.

Souza, P. S. S., Marques, W. D. S., Rossi, F. D., Rodrigues, G. C. and Calheiros, R. N. (2017). Performance and accuracy trade-off analysis of techniques for anomaly detection in IoT sensors, *Proceedings of 2017 International Conference on Information Networking (ICOIN),* pp. 486- 491.

Sperati Simone, Alessandrini Stefano and Monache Delle Luca (2016). An Application of the ECMWF Ensemble Prediction System for short-term solar Power Forecasting, *Solar Energy (Elsevier)*, 133, pp. 437-450.

Stocker, M., Rönkkö, M., and Kolehmainen, M. (2012). Making sense of sensor data using ontology: A discussion for residential building monitoring. I*FIP International Conference on Artificial Intelligence Applications and Innovations*, pp.341-350.

Surender Reddy Salkkuti and Seong-Cheol Kim (2019). Congestion Management Using Multi-objective Glowworm Swarm Optimization Algorithm, *Journal of Electrical Engineering and Technology*, 14, pp.1565-1575.

Swain, S., K., Prasad, S. and Krishna M., V. (2019). Fault Detection using Unsupervised Feature Learning on Huge volumeData Based on Improved Weighted SoftMax Regression with Jaya Optimization, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(9), pp.1757-1767.

Takeuchi, K. (2016). Nonparametric Stastistics: Asymptotics, *International Encyclopedia of the Social and Behavioral Sciences*, 2nd Edition, pp.885-890.

Tambe, G., U. and Kumar, S. (2016). A Review on Heterogenous Internet of Things Cloud Federation, *International Journal of Recent Trends in Engineering and Research (IJRTER)*, 2(12), pp.324-329.

Tareghian, S. and Bornaee, Z. (2015). Algorithm to improve job scheduling problem in cloud computing environment, *Proceeding of 2015 International Conference on Knowledge-Based Engineering and Innovation (KMEI)*, pp.684-688.

Tie, J., Chen, W., Sun, C., Mao, T. and Xing, G. (2018). The Application of Agglomerative Hierarchical Spatial Clustering algorithm in Tea blending, *Cluster Computing (Springer)*, 22, pp.6059-6068.

Timm, F. and Barth, E. (2011). Non-parametric texture defect detection using weibull features, *Proceedings of 2011 Workshop on Image Processing: Machine Vision Applications*, 7877, pp. 1-12.

Tripathy, A., Agrawal, A. and Ruth, S., K. (2016). Classification of sentiment reviews using n-gram machine learning approach, *Expert Systems with Applications (Elsevier)*, 57, pp.117-126.

Turabieh, H., Salem, A., A. and Abu-El-Rub, A. (2018). Dynamic L-Rn recovery of missing Data in IoMT applications, *Future Generation Computer Systems (Elsieveir)*, 89, pp.575-583.

Ukil, A., Bandyoapdhyay, S., Puri, C., Pal, A. (2016). IoT healthcare analytics: The importance of anomaly detection (2016). *Proceedings of 2016 IEEE International Conference on Advanced Information Networking and Applications (AINA)*, pp. 994-997.

Usman, M., J. (2018). Energy-efficient resource Allocation Schemes based on Ehhanced Flower Polinitation Algorithm for Cloud Computing Datacenter, *Thesis Reprot*, pp. 1-183.

Vantanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-aho, M., Oresic, M., Honkela, T. and Lahdesmaki, H. (2016). Self-organizing and Missing values in SOM and GTM, *Neurocomputing (Elsevier)*, 147, pp.60-70.

Varatharajan, R., Manogaran, G. and Priyan, M. (2018). A huge volumedata classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing. *Multimedia Tools and Applications*, 77(8), pp.10195-10215.

Vasconcelos, I., Vasconcelos, R. O., Olivieri, B., Roriz, M., Endler, M. and Junior, M. C. (2017). Smartphonebased outlier detection: a complex event processing approach for driving behavior detection. J*ournal of Internet Services and Applications*, 8(1), pp.1-30.

Verma, P.and Sood, S. K. (2018). Fog assisted-IoT enabled patient health monitoring in smart homes. *IEEE Internet of Things Journal*, 5(3), pp.1789-1796.

Wan, X., Guan, X., Wang., T., Bai, G. and Choi B-Y. (2018). Application deployment using Microservice and Docker containers: Framework and Optimization. *Journal of Network and Computer Applications (Elsevier),* 119, pp.97-109.

Wang, J., Tang, S., Yin, B., and Li, X., Y. (2012). Data gathering in wireless sensor networks through intelligent compressive sensing. *Proceedings of 2012 IEEE INFOCOM*, pp. 603-611.

Weyenberg, G., Huggins, P., M., Schardi, C., L., Howe, D., K. and Yoshida, R. (2014). KDETREES: Non-parametric Estimation of phylogenetic tree distributions, *Bioinformatics*, 30(16), pp.2280-2287.

Wijaya, D. R., Sarno, R. and Zulaika, E. (2016). Sensor array optimization for mobile electronic nose: Wavelet transforms and filters based feature selection approach. *International Review on Computers and Software*, 11(8), pp.659-671.

Wong, S. L., Ooi B., Y. and Liew S., Y. (2017). A Novel Data Reduction Technique with Fualt-Tolerance for Internet-of-Things, *Association for Computing Machinery (ACM)*, 2(1), pp.214-221.

Wu, H., Xian, J., Wang, J., Kandge, S. and Mohapatra, P. (2018). Missing data Recovery Reconstruction in Ocean Wireless Sensor Networks, *Computer Communications (Elsevier)*, 132, pp.1-9.

Wu, J., Lin, W., Shi, G., Li, L. and Fang, Y. (2016). Orientation Selectivity Based Visual Pattern for Reduced-reference image assessment, *Information Sciences (Elsevier)*, 351, pp.18-29.

Wu, M. Z., Wang, Y. T. and Liao, Z. C. (2014). A new shelf life prediction method for farm products based on an agricultural IOT. *Advanced Materials Research*, 846-847, pp.1830-1835.

Wu, Z., Mao, K., and Ng, G.-W. (2019). Enhanced feature fusion through irrelevant redundancy elimination in intra-class and extra-class discriminative correlation analysis. *Neurocomputing (Elsevier)*, 335(2019), pp.105-118.

Xavier, V., M., A. and Annadurai, S. (2018). Chaotic Social Spider algorithm for Load Balancing aware Task Scheduling in Cloud computing, *Cluster Computing Journal (Springer)*, 22, pp.287-297.

Xie, S. and Chen, Z. (2017). Anomaly Detection and Redundancy Elimination of Huge volumeSensor Data in Internet of Things, *Computer Science: Distributed, Parallel and Cluster Computing*, pp.1-22.

Xing., G., Xu., X., Xiang., H., Xue, S., Ji, S. and Yang, J. (2017). Fair energy-efficent virtual machine scheduling for Interent of Things Applications in Cloud environment, *International Journal of Distributed Sensor Networks (Sage)*, 13(2), pp.1-11.

Yan, H., Wan, J., Zhang, C., Tang, S., Hua, Q. and Wang, Z. (2018). Industrial huge volumedata analytics for prediction of remaining useful life based on deep learning. *IEEE Access*, 6, pp.17190-17197.

Yan, K. and Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical (Elsevier)*, 212, pp.353-363.

Yang, M-S., Lai, C-Y. and Lin, C-Y. (2012). A Robust EM clustering Algorithm for Gaussian mixture models, *Pattern Recognition (Elsevier)*, 45(11), pp.3950-3961.

Yang, X-S. (2016). Nature-Inspired Optimization Algorithms, *Elsevier Book Series,* pp.1-300

Yeongho C. and Yujin Lim (2016).Optimization Appraoch for Resource Allocation on Cloud Computing for IoT, *International Journal of Distributed Sensor Networks*, 2016, pp.1-6.

Yu Tianqi., Wang, X.and Shami, A. (2017). Recursive principal component analysis-based data outlier detection and sensor data aggregation in iot systems. I*EEE Internet of Things Journal*, 4(6), pp.2207-2216.

Yu, J., Kim, M., Bang, H.-C., Bae, S.-H. and Kim, S.-J. (2016). IoT as a applications: cloud-based building management systems for the Internet of things. *Multimedia Tools and Applications*, 75(22), pp.14583-14596.

Yu, L., and Lan, Z. (2015). A Scalable, Non-parametric Method for Detecting Performance Anomaly in Large Scale Computing, *IEEE Transcations on Parallel and Distributed Systems*, 27(7), 1902-1914.

Yu, L., Chen, L., Cai, Z., Shen, H., Liang, Y. and Pan, Y. (2016). Stochastic load balancing for virtual resource management in datacenters. *IEEE transactions on Cloud Computing*, pp.1-14.

Yu, Q., Chen, L. and Li, B. (2015). Ant colony Optimization Applied to Web Service Compositions in Cloud computing, *Computer and Electrical Engineering (Elsevier)*, 41, pp.18-27.

Yuan, M., and Sheng, H. (2017). Research on the fusion method of spatial data and multimedia information of multimedia sensor networks in cloud computing environment. *Multimedia Tools and Applications*, 76(16), pp.17037-17054.

Zhang, Q., Zhu, C., Yang, L. T., Chen, Z., Zhao, L., and Li, P. (2017). An incremental CFS algorithm for clustering large data in industrial Internet of things. *IEEE Transactions on Industrial Informatics*, 13(3), pp.1193-1201.

Zhang, Rui., Ji, P., Mylaraswamy, D., Srivastava, M., and Zahedi, S. (2013). Cooperative sensor anomaly detection using global information. *Tsinghua Science and Technology*, 18(3), pp.209-219.

Zhang, S., Wu, X., and You, Z. (2017). Jaccard DIstance based weigthed sparse representation for coarse-to-fine plant species Recognition, *Plos One*, 12(6), 1-14.

Zhang, W., Tan, S., Lu, Q., Liu, X. and Gong, W. (2015). A Genetic-Algorithm-Based Approach for Task Migration in Pervasive Clouds, *International Journal of Distributed Sensor Networks (Sage)*, 2015, pp.1-9.

Zhao, Z., Chen, W., Wu, X., Chen, P., C., Y. and Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast, *IET Intelligent Transport Systems Journal*, 11(2), pp.68-75.

Zhong, P. and Fukushima, M. (2007). Regularized Nonsmooth Newton Method for Multi-class Support Vector Machines, *Journal of Optimization Methods and Software*, 22(1), pp.225-236.

Zhou, Y., Zou, H., Arghandeh, R., Gu, W. and Spanos, C., J. (2018). Non-Parametric Outliers Detection in Multiple Time Series, A Case Study: Power Grid Data Analysis, *Proceedings of 2018 AAAI Conference on Artificial Intelligence,* pp.4605-4612.

Zong, B., Song, Q., Min, M., R., Cheng, W., Lumezanu, C., Cho, D. and Chen, H. (2018). Deep Autoencoding Gaussian Mixture Model for Unsuupervised

Anomaly Detection, *Proceedings of the 2018 International Conference on Learning Representations*, pp.1-19.

LIST OF PUBLICATIONS

Journal with Impact Factor

1. **Edje** E. Abel, Abd Latiff Muhammad Shaffie and Chan Weng Howe (2021). Enhanced Non-parametric Sequence-based Learning Algorithm for Outlier Detection in the Internet of Things, *Applied Intelligence (Springer).* 53, pp. 1889-1919. (**Q2, IF:3.325**)

2. **Abel** E. Edje and Muhammad Shafie Abd Latiff (2021). The Utilization of Algorithms for Cloud Internet of Things Application Domains: A review. *Frontiers of Computer Science (Springer).* 15(3), pp. 1-27. (**Q3, IF:1.275**).

Indexed Journal

1. **Edje** Efetobor Abel and Abd Latiff Shafie Muhammad (2021). Management of WSN-enabled Cloud Internet of Things: A Review, *International Journal of Computing and Digital Systems*. 10(1), pp. 353-372. (**Index by SCOPUS**).

2. **Edje** E. Abel, Muhammad Shaffie Abd Latiff and Weng Howe Chan (2021). Deployment of Internet of Things-based Cloudlet-Cloud for Surveillance Operations, *IAES International Journal of Aetificial Intelligence*. 10(1), pp. 24-34. (**Index by SCOPUS**).

3. **Edje** Efetobor Abel and Abd Latiff Shafie Muhammad (2019). Cloud Computing Enabled Datacenter Infrastructure Development and Deployment by IT Firms. *International Journal of Computing and Digital Systems*, 9(1), pp. 37-50. (**Index by SCOPUS**).