Decentralized online network performance monitoring for Wireless Sensor/Actuator Networks

Antonios Tzougkarakis

Thesis submitted in partial fulfillment of the requirements for the

Masters' of Science degree in Computer Science and Engineering

University of Crete
School of Sciences and Engineering
Computer Science Department
Voutes University Campus, 700 13 Heraklion, Crete, Greece

Thesis Advisor: Prof. *Panagiotis Tsakalides*

This work has been performed at the University of Crete, School of Sciences and Engineering, Computer Science Department.

The work has been supported by the Foundation for Research and Technology - Hellas (FORTH), Institute of Computer Science (ICS).

University of Crete
Computer Science Department

Decentralized online network performance monitoring for Wireless Sensor/Actuator Networks

Thesis submitted by
Antonios Tzougkarakis
in partial fulfillment of the requirements for the
Masters' of Science degree in Computer Science

THESIS APPROVAL

Author: ΑΤζουγκαράκης
____
Antonios Tzougkarakis

Committee approvals: ________________________
Panagiotis Tsakalides
Professor, Thesis Supervisor

____
Xenophontas Dimitropoulos
Professor, Committee Member

____
Georgios Tzagkarakis
Principal Researcher, Committee Member

Departmental approval: ________________________
Antonios Argyros
Professor, Director of Graduate Studies

Heraklion, July 2019

Decentralized online network performance monitoring for Wireless Sensor/Actuator Networks

Abstract

Wireless Sensor Networks (WSNs) have been introduced to our daily lives in several application domains such as Smart Water, Smart Grids, Smart Homes and Health Monitoring. All these applications require the production of robust and reliable applications which are based on the WSN's optimal performance. Various factors can affect the operation of a WSN ranging from the operational space, the adopted communication protocol, the intra-network dynamics and the status of each individual node.

The main goal is the unattended and the continuous operation of the network in real-life deployments. As such, characterization of the network's high-level performance when it is based exclusively on link-quality estimation can yield episodic snapshots of the performance specific point-to-point links.

The objective of this thesis is to characterize network performance beyond the constraints of 1st hop neighbors and across different layers of a fully functional protocol stack, ranging from the Physical to the Transport and Application layers. Heterogeneous metrics are fused and machine learning methods provide the required means to discover patterns in the data and provide the features that are the most dominant ones by employing feature selection techniques in an unsupervised fashion. Thus, systematic study of end-to-end links' performance could provide the means for understanding the multi-dimensional behavior of an entire network.

Keywords— Wireless sensor networks, Unsupervised learning, Network measurement and analysis, Testbeds and experimental Evaluation

Αποκεντρωμένη επιγραμμική παρακολούθηση επιδόσεων δικτύου για ασύρματα δίκτυα αισθητήρων / επενεργητών

Περίληψη

Τα ασύρματα δίκτυα αισθητήρων έχουν εισέλθει στην καθημερινότητά μας σε ποικίλους τομείς εφαρμογής όπως έξυπνα δίκτυα ύδρευσης, ηλεκτρισμού, σπιτιών όπως και σε συστήματα παρακολούθησης της υγείας. Η παραγωγή ισχυρών και αξιόπιστων εφαρμογών απαιτεί βέλτιστη απόδοση από την πλευρά ενός ασύρματου δίκτυου αισθητήρων. Διάφοροι παράγοντες μπορούν να επηρεάσουν τη λειτουργία ενός ασύρματου δικτύου αισθητήρων και αυτοί κυμαίνοται από τον επιχειρησιακό χώρο, το υιοθετημένο πρωτόκολλο επικοινωνίας μέχρι και τη δυναμική εντός δικτύου και την κατάσταση κάθε μεμονωμένου κόμβου.

Ο κύριος στόχος είναι η μη επιτηρούμενη και η συνεχής λειτουργία του δικτύου σε πραγματικές εφαρμογές. Ως εκ τούτου, ο χαρακτηρισμός των επιδόσεων υψηλού επιπέδου των δικτύων όταν βασίζεται αποκλειστικά στην εκτίμηση ποιότητας συνδέσμου μπορεί να αποδώσει επεισοδιακά στιγμιότυπα των συνδέσμων συγκεκριμένα, από σημείο σε σημείο.

Σκοπός της παρούσας εργασίας είναι να χαρακτηρίσει την απόδοση του δικτύου πέρα από τους περιορισμούς των γειτόνων του πρώτου άλματος και σε διαφορετικά στρώματα μιας πλήρους λειτουργικής στοίβας πρωτοκόλλων, η οποία κυμαίνεται από το επίπεδο συνδέσμου μέχρι και το επίπεδο Εφαρμογής. Οι ετερογενείς μετρήσεις που προέρχονται απο αυτά τα επίπεδα συγχωνεύονται και μέσω των μεθόδων μηχανικής μάθησης ανακαλύπτονται τα μοντέλα μέσα στα δεδομένα αυτά και επιστρέφονται τα χαρακτηριστικά εκείνα που είναι τα πιο κυρίαρχα, χρησιμοποιώντας τεχνικές επιλογής χαρακτηριστικών με μη επιτηρούμενο τρόπο. Επομένως, η συστηματική μελέτη της απόδοσης από το ένα άκρο στο άλλο θα μπορούσε να προσφέρει τα μέσα για την κατανόηση της πολυδιάστατης συμπεριφοράς ολόκληρου του δικτύου.

Λέξεις Κλειδιά— Ασύρματα δίκτυα αισθητήρων, μη εποπτευόμενη μάθηση, μέτρηση και ανάλυση δικτύων, δοκιμαστικές μονάδες και πειραματική αξιολόγηση

Ευχαριστίες

Πρώτα απ ΄όλα, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στον επιβλέποντα καθηγητή Παναγιώτη Τσακαλίδη για τη συνεχή υποστήριξή του στη διάρκεια των μεταπτυχιακών μου σπουδών. Σας ευχαριστώ που πιστέψατε σε μένα και μου δώσατε την ευκαιρία να συνεργαστώ μαζί σας.

Η εκπόνηση της μεταπτυχιακής μου εργασίας δεν θα ήταν δυνατή χωρίς την βοήθεια απο την σύμβουλό μου, Δρ. Αθανασία Πανουσόπουλου. Θα ήθελα να την ευχαριστήσω για τη σταθερή ενθάρρυνσή της, τις εποικοδομητικές ιδέες της και την καθοριστική συνεισφορά της στην ολοκλήρωση αυτής της διατριβής. Επιπλέον θα ήθελα να ευχαριστήσω τον Δρ. Γεώργιο Τζαγκαράκη για τα επικοδομητικά του σχόλια την στήριξη και τις πολύτιμες συζητήσεις που είχαμε κατα την διάρκεια των μεταπτυχιακών μου σπουδών.

Ιδιαίτερες ευχαριστίες απευθύνονται και στα μέλη της επιτροπής εξέτασης της μεταπτυχιακής μου εργασίας, καθηγητή Ξενοφώντα Δημητρόπουλο, και ερευνητή Δρ. Γεώργιο Τζαγκαράκη για τα διορατικά σχόλια και τις ερωτήσεις τους κατά την παρουσίαση αυτής της διατριβής.

Ευχαριστώ επίσης το Ινστιτούτο Πληροφορικής (ΙΤΕ) για την παροχή χρηματοοικονομικής υποστήριξης και για την παροχή του απαραίτητο εξοπλισμού κατά τη διάρκεια αυτής της εργασίας. Θα ήθελα να ευχαριστήσω όλους τους συναδέλφους μου στο Εργαστήριο Επεξεργασίας Σημάτων και για τις στιγμές που περάσαμε μαζί.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά και όλους τους κοντινούς μου ανθρώπους που όλα αυτά τα χρόνια με στηρίζουν και πιστεύουν σε μένα. Σας ευχαριστώ για όλα.

στους γονείς μου

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter is dedicated to introduce some concepts in wireless sensor network (WSN) and some of its applications. The main focus of this Chapter is performance characterization of WSNs and how the scientific community has addressed this issue by introducing various solutions.

1.1 Introduction

Wireless sensor networks (WSNs) have sparked the interest of the scientific community on a variety of issues. Current trends are focusing on automated systems which improve quality of life and can be autonomous and adaptive when problematic scenarios are happening. As a result, WSNs have entered our daily lives via several applications such as Smart Water [4, 5], Smart Grids [6, 7, 8], health monitoring [9, 10] and other smart applications (i.e Smart Homes and others).

Since the backbone of a WSN is the communication between nodes (links) there are several challenges that affect the performance of the network in various ways, such as the adopted communication protocol (i.e IEEE 802.15.4, LoRA) and various environmental factors (i.e humidity, temperature variations). In order to address challenges regarding the quality of service in WSNs, performance characterization is vital for the unattended and continuous operation of the network.

Trends that have occupied the scientific community were focused on characterizing the performance of the network via link quality aspects on point-to-point links. Empirical studies have emerged [1, 11, 12] and tools which analyzes the behavior of the network with the adoption of machine learning techniques [13, 14, 15, 16, 17]. The layout of the bibliography is presented in **Table 1.1**.

Characterization of network performance	
Point-to-point links	
Empirical Studies	[1] [11] [12]
Testbeds	[18] [19] [20]
Learning Techniques	
Supervised Learning	[13] [17]
Other Learning Methods	[14] [16] [15]
End-to-end links	
Feature Selection	[3]

Table 1.1: Bibliography Categorization on WSN performance characterization

Beginning from the empirical studies the authors in [1] conducted experiments with WSNs in controlled environments displayed the implications of common assumptions on the packet delivery performance of WSN with using as means commercial transceivers. The emphasis of the authors was on how observed quantities, such as the Received Signal Strength Indicator (RSSI), the Link Quality Indicator (LQI), the Signal-to-Noise Ratio (SNR), and the Acknowledgment Reception Ratio (ARR) can interpret the observed link behavior. The key finding of this analysis was that the most dominant qualitative characteristics of point-to-point links are the spatial and temporal correlation along with the link asymmetries. Their analysis also showed that the statistical attributes of LQI per packet offer a better correlation with Packet Reception Ratio per link, than one provided by the RSSI. All of their findings are presented on **Table 1.2**.

Similar logic was adopted in [11] were the authors did experimental studies for link quality estimation in controlled environments. The main fact on this work was that different experimental conditions lead to different extracted results. The two main reasons justifying this fact were:

- the lack of standardization in terms of evaluation metrics, assumptions, and approach
- the asymmetry of the hardware

The hardware played an important factor because it introduced antennae irregularities, dependency of radio transceivers on temperature and humidity and finally radio hardware inaccuracy. As a result, there is a inconsistency when link quality is computed by the means of LQI or SNR.

To address this problem the authors in [12] introduced another metric for characterizing link quality which is the triangle metric. This metric combined geometrically the information of PRR, LQI, and SNR into a robust estimator that

Observation	**Implication to the Conceptual Model**
Over short periods, links exhibit either 0% or 100% packet reception ratio need not be applicable to frequent (PRR). Short periods have few links with PRR between 10% and 90% i.e. intermediate links. The portion of intermediate links increases with time.	Estimates from infrequent beacons need not be applicable to frequent data transmissions.
The reception ratio a link observes depends on the channel	Protocol performance can vary over different channels.
Links have temporally correlated reception.	Assuming independent reception over time is not always valid.
External interference from 802.11 can cause losses at multiple nodes.	Assuming that forward and reverse links have different PRRs (direct link) is valid.
Acknowledgement reception ratio (ARR) is usually greater than the packet reception ratio (PRR)	Using PRR in the place of ARR is not valid and can lead to inaccurate link quality estimates

Table 1.2: Key Observations from work [1]

guaranteed a fast and reliable assessment of the point-to-point link quality. The formal description of this metric first denotes a set of n packets are used to sample the channel and m of those packets have been successfully received ($0 < m \leq n$). The LQI and SNR of each succesfully received packet i are denoted by lqi_i and snr_i. Upon reception of the sampling packets, the receiver calculates the window mean SNR and LQI in the following way:

$$\overline{SNR_w} = \frac{\sum_{k=1}^{m} snr_k}{n} \tag{1.1}$$

$$\overline{LQI_w} = \frac{\sum_{k=1}^{m} lqi_k}{n} \tag{1.2}$$

Then, the receiver calculates the distance to the origin(length of hypotenuse):

$$d_{\triangle} = \sqrt{\overline{SNR_w}^2 + \overline{LQI_w}^2} \tag{1.3}$$

Based on the computed distance, the receiver estimates the quality of the send-receiver link according to the rule in which the larger the distance, the higher the quality. Finally the authors have assigned empirical-based thresholds to differentiate the quality of links.

Therefore the problem of link quality performance could be addressed as a prediction problem. A way to resolve it is by using machine learning (ML) to perform the prediction task effectively, without using explicit instructions and by relying on patterns and inference instead. ML is considered to be a subset of artificial intelligence. Therefore ML provides the necessary means for better prediction regarding link-to-link performance.

Shifting to this direction the study in [21] have exploited learning techniques for performance estimation. The efficacy of supervised learning involving two primary phases, namely offline training and online classification was evaluated in [13] again for point-to-point links. The reason behind the supervised learning that they used was the ability to automatically discover relations between readily-available features and the quantity of interest. The general goal was to improve situation-awareness in order to optimize the network communication. Their approach, cast the problem of link quality estimation to a classification problem. Several classification algorithms were tested. The main focus although, was given on decision tree learners [22] and rule learners [23]. As a result their approach lead to similar accuracy compared with traditional batch learners, but with less computational and resource complexity.

A distributed protocol which adopted supervised incremental learning was introduced in [17]. The goal was to estimate wireless link quality based on supervised incremental learning methods. In order to accomplish that they combined Locally Weighted Projection [24] and locally available measures of direct links, such as SNR and traffic rate towards building regression maps between the local network configuration and the expected link quality.

Authors in [14] have adopted a data driven approach which combined the values of PRR and the levels of RSSI, LQI and SNR with logistic regression classifiers. The produced output was the success probability of delivering the next packet. The proposed approach was consisted from three steps:

- data collection
- offline modeling
- online prediction

The link quality prediction that they conducted, used several machine learning methods such as naive Bayes classifier [23], logistic regression [25], and artificial neural networks [26]. Their key finding was that logistic regression works well among the models and it introduced a small computational cost.

Extending the work of [14] the work in [16] employed Stochastic Gradient Descent to address aspects related to the estimation of links with moderate performance. The on-line and unsupervised schemes were able to adapt the wireless dynamics without the need for data collection and model retraining.

Similarly the work in [15] employed machine learning methods to estimate the short evolution of link quality, in order to switch data transmission on a better quality link. LQI metric is predicted by a decision maker called forecaster that can adapt its strategy to predict this metric as close as possible to the real value. The proposed learning and prediction model had great flexibility and could be adapted to different link quality metrics or prediction methods.

As this far, studies have used machine learning for better link quality estimation. Opposed to this fact the work on [27] have used machine learning as a way to increase the efficiency of link sampling. They presented a strategy for link quality monitoring applicable on the Rouging Protocol for Low Power and Lossy Networks (RPL) with minimal overhead and energy waste. To achieve this goal, their system leverages both synchronous and asynchronous monitoring schemes to maintain up-to-date information on link quality and to promptly react to sudden topology changes which can be caused from node mobility.

The characteristic that connects all the above studies is the point-to-point non competitive link analysis that justifies the network behavior. However, there is clear that the literature is lacking on studies that extend the network performance well beyond link quality estimation.

This literature gap was filled in [3]. This work extended the problem of network performance characterization to multi-hop network topologies and introduced a wider range of network parameters that span across all layers of protocol stack (besides SNR, LQI, RSSI). Their focus was the reduction of dimensionality of data by the means of feature selection and thus improve the overall quality of classification. Furthermore dominant factors that impact the behavior of multi-hop links were extracted.

This thesis exploits the framework of [3] and extends the problem of dominant factors that affect the end-to-end links by focusing the problem of feature selection. In this thesis two approaches have been adopted. The first approach (figure 1.1) describes the network dynamics over the network as a graph-based feature selection problem (features: $\Box$). The calculation of dominant features (colored $\Box$) is going to exploited for extracting network-wide behavioral patterns (colored $\bigtriangledown$).

The second approach adopts the feature selection algorithm introduced in [3] were features are described a vector and the feature selection process selects the most dominant ones, based on the representation entropy $H_{\mathbf{A}}$ which is metric that calculates the redundancy of information in the feature matrix $\mathbf{A}$. Both approaches are categorized as unsupervised methods in the machine learning categorization.

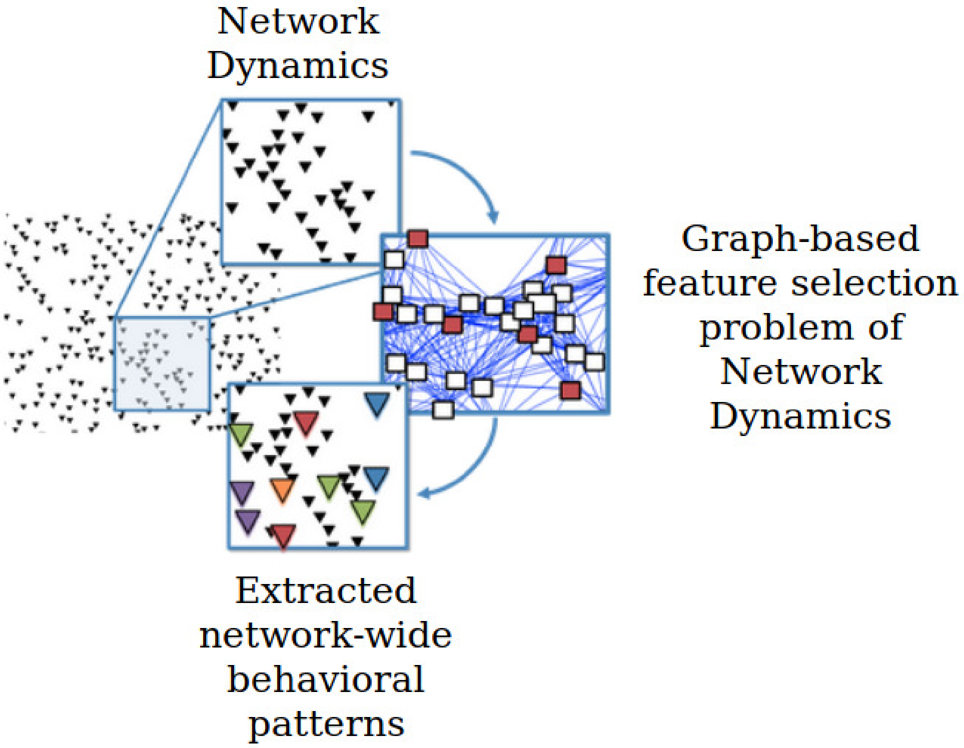
Network Dynamics
Graph-based feature selection problem of Network Dynamics
Extracted network-wide behavioral patterns

Figure 1.1: First approach in this thesis. Network dynamics formulate a graph feature selection problem.

1.2 Motivation

The motivation of this work is based on the fact that performance characterization for WSNs is mainly focused between 1st hop links. This information is not sufficient to characterize the network's high level performance. As a result there is a gap of studies that focus well beyond those links and try to characterize the performance of a multi-hop WSN based on information that goes beyond the Physical Layer (RSSI, LQI). Furthermore, there is a need for studies that explore the performance of a WSN from real-life deployments.

1.3 Contribution

In a nutshell the contributions of this thesis are summarized as follows:

- The combination of network metrics collected from different sides of the network, and corresponding to different layers of the protocol stack to a feature-level fusion mechanism for delivering high-level inference on the dominant network features.
- A design of a framework that is consisted from various unsupervised learning methods and has the ability to analyze different data formats as long as the form is passed as input to it.
- The application of the proposed framework on real-life deployments and the explanation of the findings within the WSN context.
- The designed framework does not introduce computational overhead to the WSN (sensor nodes) and it can be implemented from a low budget single board computer such as the odroid [28].

The rest of this thesis is organized as follows: in Chapter 2 the problem of feature selection is described alongside the categorization of feature selection methods and the algorithms that were used in this thesis. Chapter 3 is focused on the problem formulation and the system displayed in a formal way accompanied with the proposed system architecture. Evaluation studies for the introduced system is presented in Chapter 4. Finally, the conclusions and future work of this thesis are drawn in Chapter 5.

Chapter 2

Background

This Chapter offers information regarding the feature selection background. Classification information is also provided regarding the classifier that was used in this thesis. In particular the basic concept of feature selection is introduced alongside with the categorization of feature selection methods and finally the feature selection methods that were used in this thesis.

2.1 Feature Selection

2.1.1 Problem Formulation And Notation

Patter recognition which works through machine learning is highly affected by the length of data. Nowadays there is a high producing rate of digital data and that leads to a bottleneck when their analysis is held. In order to address the problem of high dimensionality, reduction techniques can be applied. Feature Extraction (FE) and Feature Selection (FS) can reduce the volume of available data and improve the performance of machine learning algorithms.

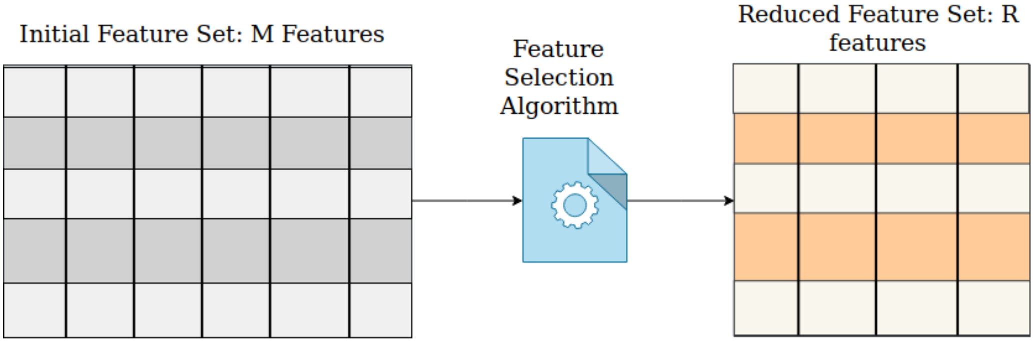
Initial Feature Set: M Features
Feature Selection Algorithm
Reduced Feature Set: R features

Figure 2.1: Feature Selection Procedure

Feature selection is the process which produces a subset of the initial data that is given to it. Figure 2.1 displays that process. Let us consider the initial variable set X of size M, containing feature vectors x_i, where $i = \{1...M\}$. A Feature Selection Algorithm is a function g that extracts a reduced variable set F of R examples f_i, where $i = 1...R$ and $R << M$, from the initial variable set.

2.1.2 Categorization of Feature Selection

Feature selection process can be divided into two main categories. In terms of availability of label information feature selection techniques are classified into 3 categories. Supervised methods [29, 30, 31, 32], semi-supervised [33, 34, 35] and unsupervised methods [36, 37, 38]. The availability of label information offers effective discrimination of features from different classes consisted from samples.

Semi-supervised algorithms are employed when a small portion of data is labeled. As a result these algorithms are exploiting both labeled an unlabeled data. When search criteria for discriminate features is missing, unsupervised learning is the way to do feature selection, but it is a harder problem than the rest methods. Several criteria have been proposed to evaluate feature relevance.

Classification of feature selection can also based on the different strategies of searching. Three methods methods is the result of this classification, i.e filter methods, wrapper methods and embedded methods. In filter methods discrimination of data is done through their character. In general, those methods have as first priority the feature selection process and then perform classification and clustering tasks. As a result, filter methods usually fall into a two-step strategy. First, the whole set of features is ranked according to certain criteria. Afterwards, the features with the highest ranking is selected [39, 40, 41, 42].

Wrapper methods tend to use the learning algorithm itself to evaluate the features. For example the work in [43] has adopted Support Vector Machine methods based on Recursive Feature Elimination (RFE) to select the most relevant gene to cancers. Embedded models perform feature selection in the process of model construction. Overall the categorization of feature selection methods is presented in **Figure 2.2**.

In this work, the exploitation of unsupervised learning via two methods is conducted. In the following section a brief overview will be presented regarding the state of the art in unsupervised feature selection algorithms.

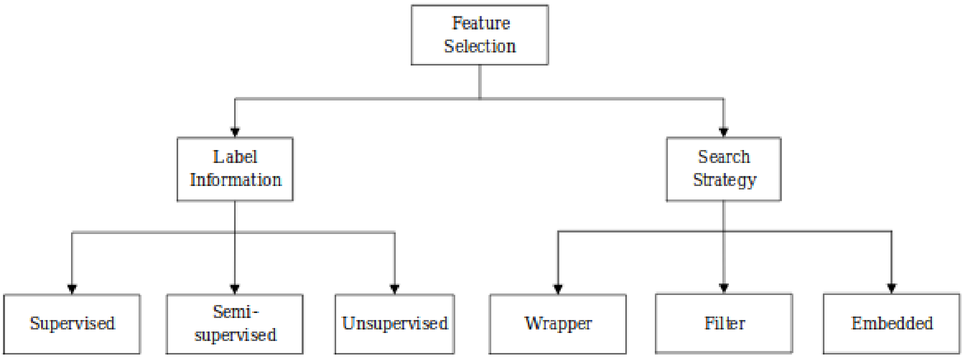

Figure 2.2: Feature Selection Categorization in [2]

2.1.3 Feature Selection Algorithms used in this thesis

2.1.3.1 Graph Clustering with Node Centrality (GCNC)

Morandi and Rostami proposed an unsupervised filter-based approach algorithm which is called Graph Clustering with Node Centrality [44] and it works in three steps:

1. The problem space is represented by a graph $G = (X, E, w_X)$ in which $X = x_1, x_2, ..., F_M$ denotes an original feature set, $E = (x_i, x_j) : x_i, x_j \in F$ denotes the edges of graph and w_{ij} indicates the similarity between two features x_i and x_j connected by the edge (x_i, x_j). In this thesis the Pearson product-moment correlation coefficient [45] it is used as a similarity measure.

2. The graph is divided into several clusters using an efficient community detection algorithm. By exploiting Louvain Community detection algorithm proposed by Blondel et al. [46] detection of communities/clusters is applied to the graph.

3. Selection of the most relevant and influential feature from each cluster is done through a novel iterative search strategy. In particular for each feature x_i in each of the k clusters C_k, an influence value is calculated where

$$Influence(x_i) = TV \times LC$$

TV indicates the normalized term variance of feature x_i and is defined as

$$TV(S, x_i) = \frac{1}{|S|}\sum_{j=1}^{|S|}(A_{ij} - \overline{A})^2$$

where A_{ij} indicates the value of feature x_i for the pattern j, and $|S|$ is the total number of patterns. LC indicates the Laplacian Centrality of vertex u_i

and is defined as

$$LC(u_i, G) = \frac{\Delta E_i}{LE(G)} = \frac{LE(G - LE(G_i)}{LE(G)},$$

where G_i is the graph obtained by deleting u_i from G, and LE is the Laplacian Energy of G. As the computation of influence value for each feature is completed, a comparison is performed with a threshold value δ. If the influence value of the feature is smaller than δ the feature is added into a candidate list for removal. Finally, the features of the candidate list are removed if the difference between its size and the size of cluster is larger than a value ω.

2.1.3.2 Representation Entropy Clustering Feature Selection Algorithm (REC-FSA)

An unsupervised learning technique which combined ranking and clustering techniques was introduced by Panosopoulou et al. in [3]. The algorithm that was introduced relied on the calculation of representation entropy, which is a typical evaluation metric for measuring the amount of redundancy in the feature matrix. The proposed technique relies on a backward search in the feature space and is divided in three main steps:

1. rank the features with respect to the volume uncertainty
2. cluster features that exhibit high redundancy with a top-ranking feature, which is appointed as the head of cluster
3. eliminate all members of the cluster from the feature space as redundant, except from the cluster head

For the first step the procedure that was described in [47] was adopted and the redundancy that each feature contributes to the data set A is calculated as follows:

$$dH_m = H_{A \backslash A_m} - H_A,$$

where A_m is the set of samples corresponding to the mth feature and $H_{A \backslash A_m}$ is the representation entropy of A when the A_m set of samples are not taken into account. Value of dH_m represents the difference in the uncertainty when the mth feature is omited. On the other hand if the mth feature describes information with limited variance or predictable behavior, then the value of $H_{A \backslash A_m}$ will remain similar to the one of H_A. Therefore, the value of $dH_m \to 0$. Ranking of the feature vector is the result of this procedure and the top-ranked feature vector m^* is the one that maximizes the value of $dH_m, m = \{1, 2, ..., M\}$.

While performing the second step of the algorithm the search space is centered around m^*. A k-nearest neighbor technique is employed and the goal is to cluster the features that exhibit the higher value of redundancy with the feature m^*.

Calculation of the pairwise representation entropy $H_{A_{m*,m}}$ between feature m^* and the remaining features m is done, where $m^* \neq m, m \in f_{ij}$ and $H_{A_{m*,m}} = [A_m^*, A_m] \in [0,1]^{D \times 2}$. Values that come as a result from this process are sorted in a descending order. The first k features, along with m^*, form a cluster of features c_{m^*}:

$$c_{m^*} = m^* \bigcup \{m | H_{A_{m^*,m}} \leq h_{m^*}(\kappa)\}$$

where $h_{m^*}(\kappa)$ is the value of the pairwise entropy between the feature m^* and its κth neighbor.

Features that exhibit high redundancy with m^* are accommodated in cluster c_{m^*}. The latter is considered the dominating feature and, therefore appointed as the cluster head. Subsequently, during the third step of the algorithm, the cluster head m^* remains in the search space as the representative feature of cluster c_{m^*}, while all remaining features in c_{m^*} are considered redundant and thus eliminated [48]. The process is repeated until either all features are clustered and discarded, or selected as dominating.

2.2 Classification

The κNN is a a simple but effective method for classification. For a data record t to be classified its κ nearest neighbors are retrieved, and this forms a neighbourhood of t. Majority voting among the data records in the neighbourhood is usually used to decide the classification for t with or without consideration of distance-based weighting. However, to apply κNN successfully is very crucial to choose carefully the value for κ.

Suppose P1 is the point, for which label needs to be predicted and κ equals to 1. First, the nearest point to P1 is calculated and then the label of the nearest point is assigned to P1. This paradigm is explained in **Figure 2.3**.

If the κ is equal to some other value and P1 label needs to predicted, the first step is the calculation of κ close points to P1 and then classify points by majority vote of its neighbors. Each object votes for their class and the class with the most votes is taken as the prediction. In order to estimate the closest similar points, calculation of the distance between points using Euclidean distance is required. The concept of κNN is summarized in three steps:

1. Calculate distance
2. Find closest neighbors
3. Vote for labels

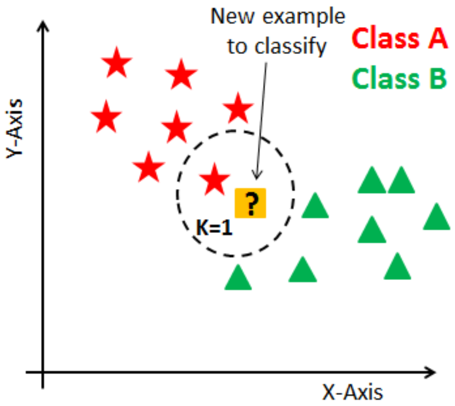

Figure 2.3: Problem statement κNN

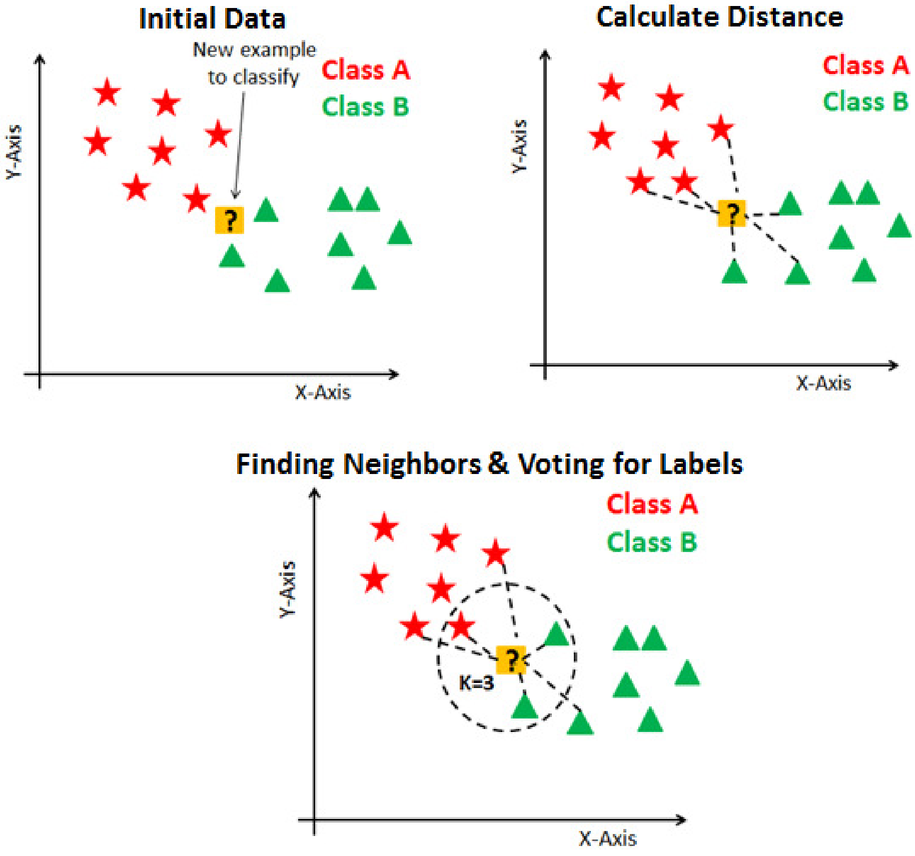

Figure 2.4: Steps of κNN

Chapter 3

Problem Statement, System Architecture

In this Chapter formulation of the problem at hand is presented in a formal way. Afterwards there is a brief introduction of network metrics used in this thesis and the meaning of each one of them. Since this work exploits dominant features extracted from feature selection, an evaluation is necessary through the evaluation metrics. Finally, this Chapter concludes with the implemented system architecture.

3.1 Problem Statement

To state the problem at hand a WSN comprised of energy autonomous, power constrained and IEEE 802.15.4 compliant sensor nodes is needed. The total number of nodes is set N and the main characteristic is that their operation over long periods of time is happening in an unattended fashion. Aspects of network monitoring are implemented via a full protocol stack in each node that extends from the Physical layer to the Application layer. Lifetime of each sensor node is described through the adopted network policy, the hardware characteristics of the transceiver chip, and the input voltage supply.

The in-network operation is characterized by the establishment of end-to-end links which expresses the unicast connections between different sensor nodes at the Application Layer. An end-to-end link between a sensor node i and a node j is described as $i \rightarrow j$ and it is constructed over a network path $P_{ij} = \{i..., k, ...j\}$ and k node in a relay node between the to end-to-end nodes i and j (**Figure 3.1**).

During network operation each node $k \in P_{ij}$ can monitor network metrics which are relevant to the functionality of the node and the quality of the link $i \rightarrow j$. Those metrics span across the protocol stack and furthermore remain independent of the specific solution that is adopted by each layer. Parameters that are being monitored and represent the network functionality are the RSSI, LQI per received

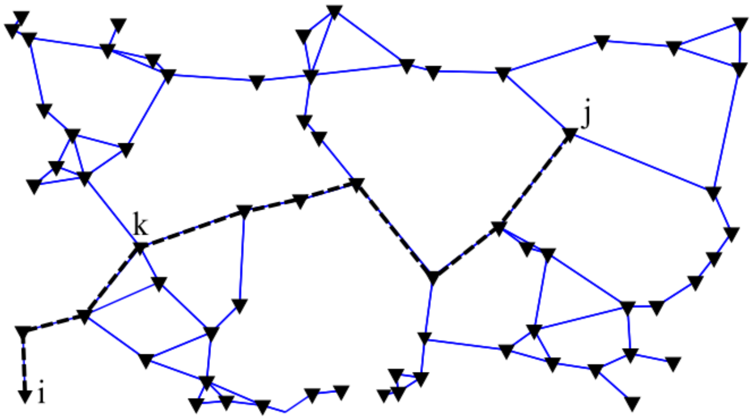

Figure 3.1: An example of an end-to-end link $i \to j$ and the corresponding path P_{ij} (dashed line), established over a multi-hop network topology (solid line) [3]

packet, the Noise Floor (NF), the unicast network activity at the MAC layer, the battery level and the on-board temperature, humidity. Packets received by the jth node and the packets transmitted by the ith node define the PRR_{ij} which define the performance of the link. PRR set of values are set between 0 and 1 ($PRR \in [0, 1]$). Those values can be classified into discrete, user-defined labels l_{ij}.

The automated calculation of the network factors that are relevant and have an impact on the classification of performance for the link $i \to j$ is the problem at hand to different values of l_{ij}. As the deployments of a WSN shift towards more complex non single-hop links, PRR_{ij} can be affected by different factors that vary with respect to the operation space, the ambient conditions and intra-network behavior. All related metrics are summarized in **Table 3.1**.

The result is a feature vector $\boldsymbol{f}_{ij}$ of size M, which is consisted from different network metrics available at different layers and different sides of P_{ij}. The question raised is whether the resulting vector conveys the dominating attributes can characterize the network performance of the $i \to j$ link expressed in terms of a class labels l_{ij}. Characterization of attributes as dominant implies that they have sufficient information to predict value of class l_{ij}. The redundant information is eliminated and thus the objective is to exploit the contents of $\boldsymbol{f}_{ij}$ to automatically calculate the subset of R features ($R \leq M$) $\boldsymbol{f}^*_{ij} \subseteq \boldsymbol{f}_{ij}$ that are most relevant to define the performance of each link $i \to j$.

$$\boldsymbol{f}_{ij} \to \;(PRX^*_{ij},\; LQI^*_{ij},\; NF_{ij},\; |P_{ij}|,\; T_i,\; H_i,\; V_i)$$

Network Metric	Description
PRX^*_{ij}	Receive Power over path P_{ij} (dBm)
LQI^*_{ij}	Link Quality Indicator over path P_{ij}
NF^*_{ij}	Noise Floor Quality Indicator over path P_{ij}
$\|P_{ij}\|$	Length of path P_{ij}
T_i	On-board temperature of the inth node (^oC)
H_i	On-board humidity of the inth node (%)
V_i	Input power level for the ith node (Volt)

Table 3.1: The network metrics employed for forming the feature vector $\boldsymbol{f}_{ij}$

3.2 Metrics

As mentioned before metrics that were used in this work are **Received Signal Strength Indicator** (RSSI), **Link Quality Indicator** (LQI), **Noise Floor** (NF), **Path length** ,**Temperature**, **Humidity**, **Battery voltage**. Each one of them denotes something different about the WSN and it's operation. The following paragraphs are dedicated to present each metric and what it represents.

RSSI is a measurement in telecommunications that presents the power of a received signal strength. It is implemented and widely-used in 802.11 standards [49]. **LQI** stands for Link Quality Indicator. LQI estimates how easily the received signal can be modulated when considering noise in the channel [49]. Some practical examples are presented in **Figure 3.2**.

Noise floor signal theory is the measure of the signal created from the sum of all the noise sources and unwanted signals within a measurement system, where noise is defined as any signal other than the one being monitored. In radio communication and electronics, this may include thermal noise, black body, cosmic noise as well as atmospheric noise from distant thunderstorms and similar and any other unwanted man-made signals, sometimes referred to as incidental noise. If the dominant noise is generated within the measuring equipment (for example by a receiver with a poor noise figure) then this is an example of an instrumentation noise floor, as opposed to a physical noise floor. These terms are not always clearly defined, and are sometimes confused [50].

Path length denotes how many sensor nodes a packet needed to go through in order to arrive to it's final destination; the sink node. Overall, path length express the end-to-end hop count from a source to a destination over the network. Routing path performance can be evaluated as well as the reliability of the system.

Temperature in the proposed system is measured in degree Celsius. Each sensor measures the temperature of the environment in which is deployed. As a

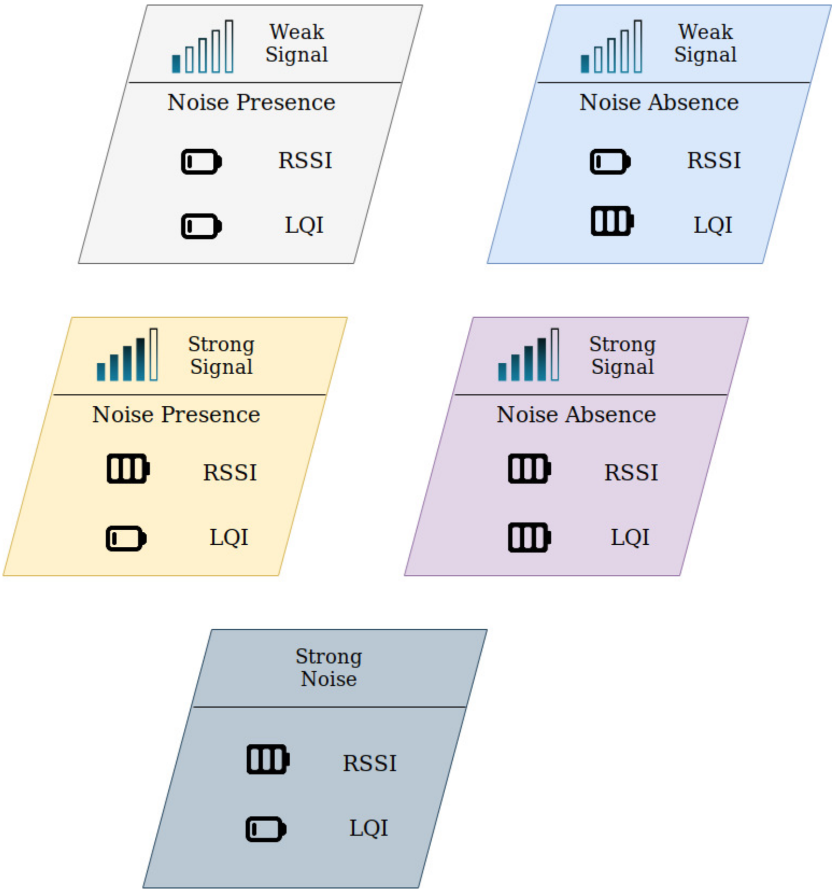

Figure 3.2: Examples of RSSI and LQI

metric is very useful because it provides information about the environment and if it is combined with other metrics such as LQI and RSSI can deliver useful insights about the WSN operation.

Humidity in the proposed system is measured by a percentage. Its value denotes the amount of water vapour present in the air. It ranges from 0% to 100%.

Battery voltage is measured in volts and it defines the operation of wireless sensors. As the WSN operates the battery voltage in each mote fades up to the point that the mote cannot operate anymore. It's important to keep that information and notice how it affects the network operation throughout it's lifetime.

Packet loss rate (PLR) is a crucial and popular link quality metric for wireless sensor networks (WSNs) [51]. A proposed model was introduced that connects PLR to link quality indicator (LQI). Specifically the proposed PLR model for 802.15.4 links, is a function of LQI and packet payload size.

$$PLR = \frac{1}{1 + (\alpha/L) * exp(\beta * LQI)}$$

LQI denotes the value of LQI, L is the packet payload size in bytes, and α and β are two model parameters.

3.3 System Architecture

In this section the proposed system architecture is going to by analyzed thoroughly. The proposed system operates in two modes. Either data are parsed through a file, either the system operates as a unit with a real time operating wireless sensor network and performs analysis on the fly. **Figure 3.3** presents flow of data inside the system's implementation.

First step is the data segmentation. Data is divided into time segments $S_1, S_2, ..S_i$ and its size can be defined as S_{size}. The computation of each segment is based on the timestamp that data had been collected. Next step includes the division of each segment S_i into observation windows $W_1, W_2, ...W_j$ with size W_{size}. For example if segmentation size is set to x hours the observation window will be set in y minutes. For every observation window there is an overlapping value set to it intentionally to improve the efficiency of the next steps (i.e feature extraction). As the computation of observation windows is completed, the feature extraction process begins. For each metric that can be found in the data for each observation window, its feature is computed (for example for RSSI measurements the possible features would be the mean value, the standard deviation etc). Features which have been extracted from each observation window are then fused into a feature matrix where the columns represent each metric's feature. This feature matrix is then parsed to the Bucket of Feature Selection Algorithms (BFSA) which outputs reduced feature matrices that are equal to the number of feature selection algorithms that are being used.

Once the information from the BFSA is extracted, the evaluation process of this output is initiated. The employed mechanism is presented in **Figure 3.4**. Reduced feature matrix alongside the initial feature matrix is passed through this evaluation process in order to compute the representation entropy of the reduced matrix, and the compressing ratio. For the classification accuracy process there is an additional step where the computation of labels for Packet Reception Ratio is needed. For a specific segment for each observation window that constitutes it the computation of Packet Reception Ratio is held in order to form a matrix with the segment's packet ratio values. These values are then labeled and parsed to the classification accuracy process along with the dominant features in order for the classifier to predict some randomly chosen missing values. Results of this process are then stored for post-processing. The process of evaluation is repeated for each segment S_i.

3.3.1 Offline Analysis

Offline analysis uses as input data that came from a file. It's form needs to be defined in order for the system to understand the data that is being given to it. Computation of each segment timestamp and each observation window with the overlapping is essential. The first (or starting) timestamp comes from the first line of file. All segment timestamps are being computed till the last line of the file. Data for each observation window are imported and the computation of features begins. After features are computed for each observation window are then imported into a matrix for a corresponding segment. This process is repeated until the last segment. Finally data are parsed to BFSA and as a result reduced feature matrices are extracted.

Next step is the evaluation of dominant attributes that were extracted from the feature selection process. Compression ratio,representation entropy and execution times is the three evaluation steps. The final and fourth step is to evaluate the extracted information (dominant features) via classification accuracy process. As the evaluation step is completed the results are written to a file for post processing.

3.3.2 Online Analysis

Online analysis operates differently than offline. Specifically, the main difference is the way that analysis is being held. It is done for the data that have been collected for a segment length from a real time operating WSN. For example if data collection starts at 17:00 and the segment size is set to 2 hours, the first analysis will be conducted at 19:00 for the data that have been collected from 17:00-19:00. The next analysis will happen at 21:00 and the data window will be from 19:00 to 21:00.

As data are gathered for the feature selection process the data collection process is still operating without intervention. Data are divided into observation windows with the user-defined overlapping pattern. For each observation window matrix, computation of features is conducted. The features from each observation window are then concatenated into one matrix which represents the features for a whole segment. The feature selection process is called in order to extract the most dominant ones. Finally the last step is the evaluation of this information via classification, representation entropy, compression ratio and execution times.

This process will be repeated every segmentation size hours which is defined by the user. As long as the WSN provide the system with sufficient data, it will continue to analyze and evaluate them. The analysis is called online because the data analyzed have gathered at a similar close time to the operating WSN. The focus of online analysis is the central processing unit. The offline analysis was held by a Personal Computer equipped with an Intel Core i7-6700 at 3.4 GHz and RAM size of 16 GB RAM. On the other hand, online analysis (data collection, feature computation, evaluation) was conducted by an Odroid U3+ single board computer. This computer has computational limitations due to its 1.7 GHz quad core processor and 2 GB LPDDR2 RAM. Overall, the proposed system is efficient enough to run on those constraints.

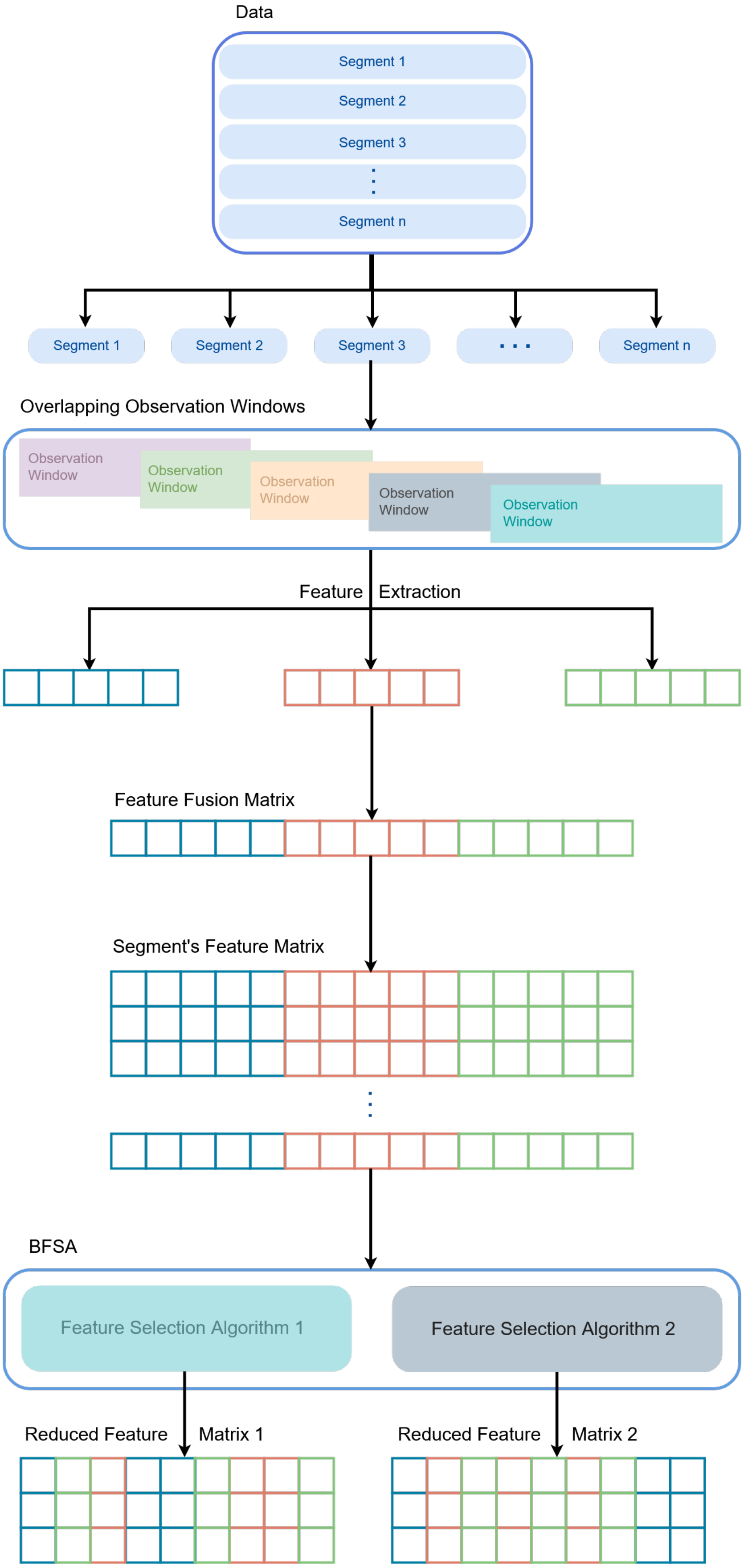

Figure 3.3: System Architecture Flow of Information

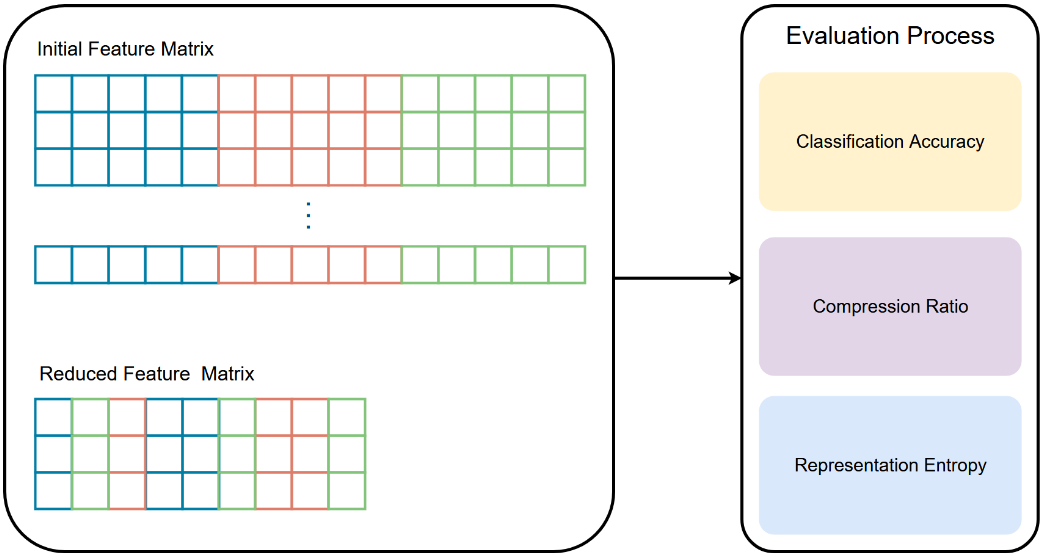

Figure 3.4: System Architecture For Evaluation on Extracted Information

Chapter 4

Evaluation Studies

This Chapter provides information regarding the evaluation metrics for better comprehension of the results. Furthermore experimental setup information for the system's supported modes is provided (offline/online). Data used in offline analysis was gathered from a WSN that operated in a desalination plant in the framework of Hydrobionets project. In online analysis data was gathered from a real-time working WSN deployed inside the University of Crete specifically in the classrooms of Computer Science's Department. The extracted information was evaluated on various perspectives.

4.1 Evaluation Metrics

In the following sections some concepts regarding the evaluation of the feature selection process are presented. This subsection is dedicated to present those concepts for better coherence. Specifically, concepts such as **compression ratio, representation entropy** and **classification accuracy**. Since the system has two feature selection algorithms, provided information is going to be compressed. Compression ratio denotes the rate of information after the feature selection process. For example if a total number of 100 features are given to feature selection process and 50 were returned, the compression ratio is 50%.

Compressed information could contain sufficient data that could represent the initial ones. In order to define how good the reduced information, **representation entropy** computation is needed. As a typical evaluation metric, it defines the amount of redundancy in a matrix [48]. In order to interpret this information better it's value is normalized between 0 to 1. The closer the value it is to 1, the better quantity of information is provided in contrast to the initial data. Representation entropy is computed through the following formula:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

In order to produce more understanding results, representation entropy values are being normalized from 0 to 1 where 0 denotes that the reduced information cannot represent the initial data well and 1 denotes that the information can represent well the initial data.

Classification accuracy is a metric which evaluates classification models [52]. Labels are an essential part for the data representation that are going to be predicted. Next step of the process, is the separation of data between training and testing. Testing data are then copied and labels are being subtracted from the data. When training process is completed, the data are being used to predict the missing labels from the testing data. A comparison is being done after the prediction of the labels with the actual data before the subtraction. Formally, classification accuracy has the following definition:

$$classification\ accuracy = \frac{Number\ of\ correct\ predictions}{Total\ Predictions}$$

4.2 Offline Experimental Setup

As mentioned in previous section the proposed system has two modes; offline and online. For the offline mode data comes from a dataset that was developed under Hydrobionets project. The WSN was deployed at an industrial environment during the period 05-06/06/2014. Sensor nodes that were employed were AdvanticSyS XM1000 and CM5000-SMA which are IEEE 802.15.4-compliant devices. The sink node that was used was a pandaboard. All described devices are presented on **Figure 4.1**

The WSN, comprised $N = 10$ nodes was part of an industrial smart water network which was deployed at a fully functional pilot desalination plant ($40\ x\ 12\ s.m$). Its purpose was to monitor and control the phenomenon of fouling process. This is related to the concentration of unwanted bacterial matter on the surface of the reverse osmosis membranes.

The sensors were deployed on key factor locations. Namely the sea water intake, the pre-treatment, the security filters and the reverse osmosis. The deployment as a result was a challenging one since various factors affected the communication. For example the metallic environment, bulky water tanks, heavy machinery and operating devices such as pumps and valves and finally the human presence of technical stuff.

Sensors operated with non-rechargeable batteries. Information regarding the network metrics and other data (such as temperature, humidity etc) were implemented through a customized protocol stack. The transmission period was set to 6 seconds. Deployment is presented on **Figure 4.2**.

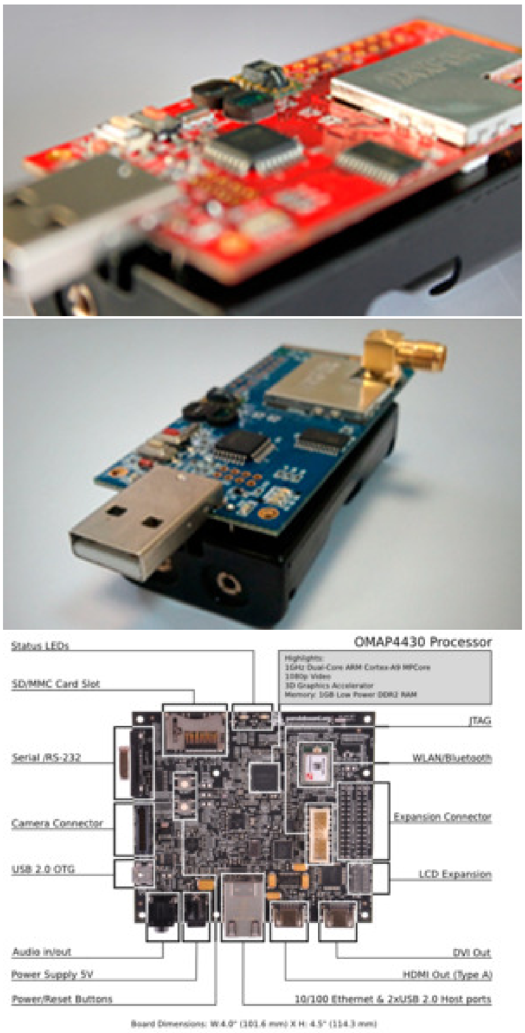

Figure 4.1: Hardware Used in Hydrobionets project

In order for this analysis to take place data is parsed from the Hydrobionets project which took place in an industrial environment with 10 IEEE 802.15.4 compliant sensor nodes and sampling rate set to 6 seconds. Data has been analyzed entirely and divided into various observation window and segment sizes. Various parameters were examined. Behavior of representation entropy, compression ratio, classification accuracy, dominant statistics, dominant metrics and execution times are presented in the follow sections.

Figure 4.2: Industrial Deployment of [3]

As a first step there was an analysis of data for the κ parameter of the REC-FSA feature selection algorithm. There was a formulation of experiments which could display the results as far the representation entropy values. The main goal was to achieve high results from REC-FSA regarding the representation entropy. The value of κ was expressed as a percentage of rows of the initial feature matrix. Those results are presented in **Figures 4.3, 4.4, 4.5, 4.6**.

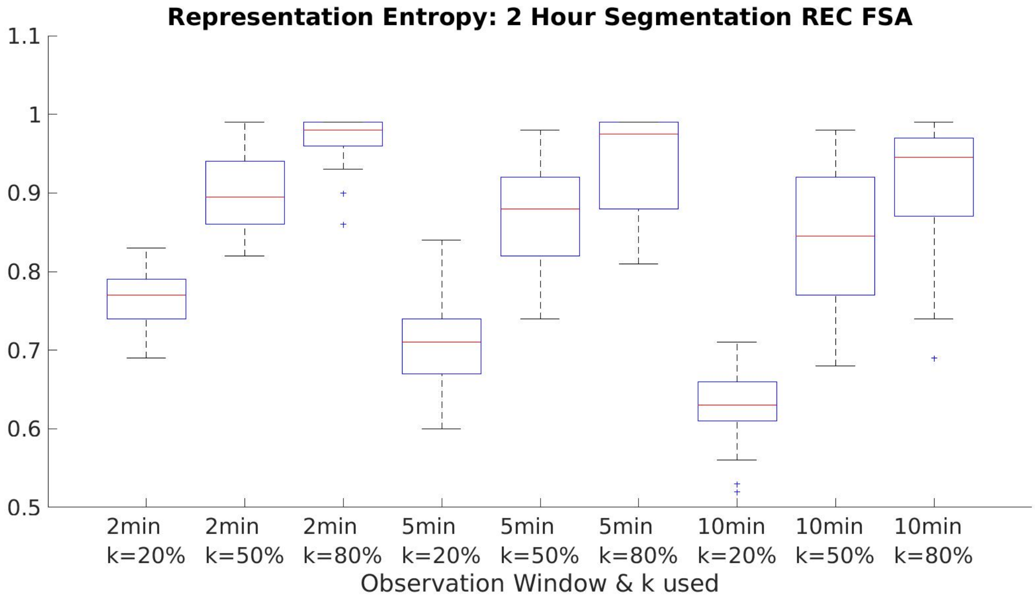

Figure 4.3: Representation Entropy of REC-FSA in various κ sizes 2 hour segments

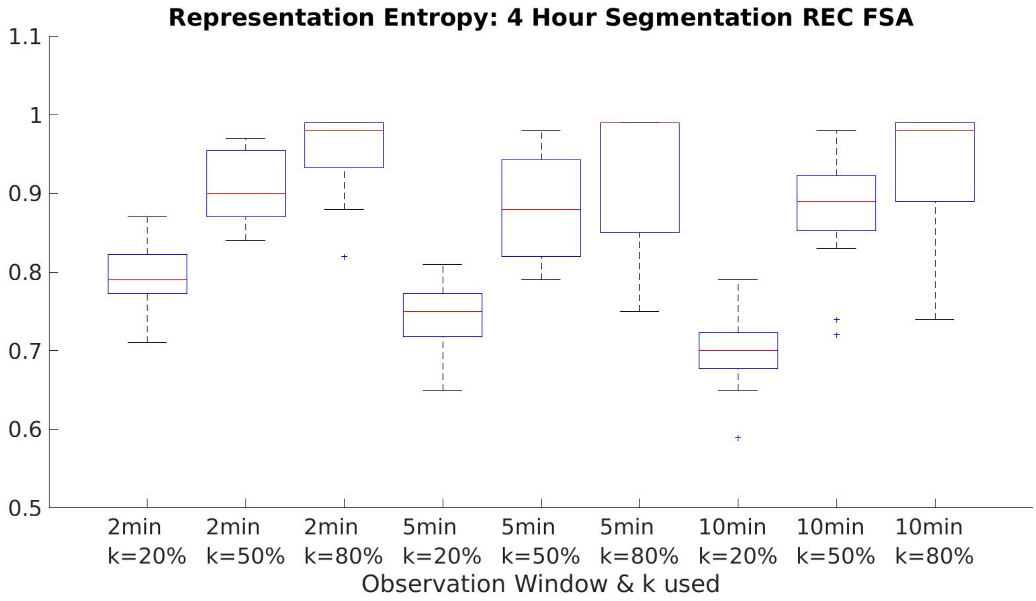

Figure 4.4: Representation Entropy of REC-FSA in various κ sizes 4 hour segments

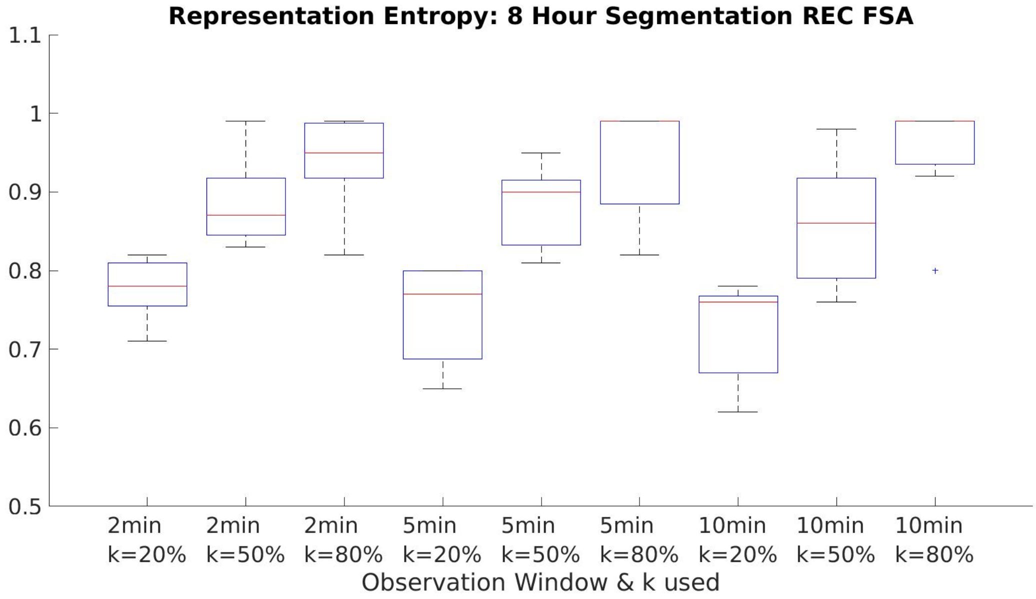

Figure 4.5: Representation Entropy of REC-FSA in various κ sizes 8 hour segments

As the size of κ is increases the extracted results from REC-FSA provide higher Representation Entropy results. This due to the increased size of dominant/reduced attributes that the feature selection algorithm outputs. Driven by these results the κ value for the experiments was set $80\% \times size\ Of\ Rows\ Of\ Feature\ Matrix$.

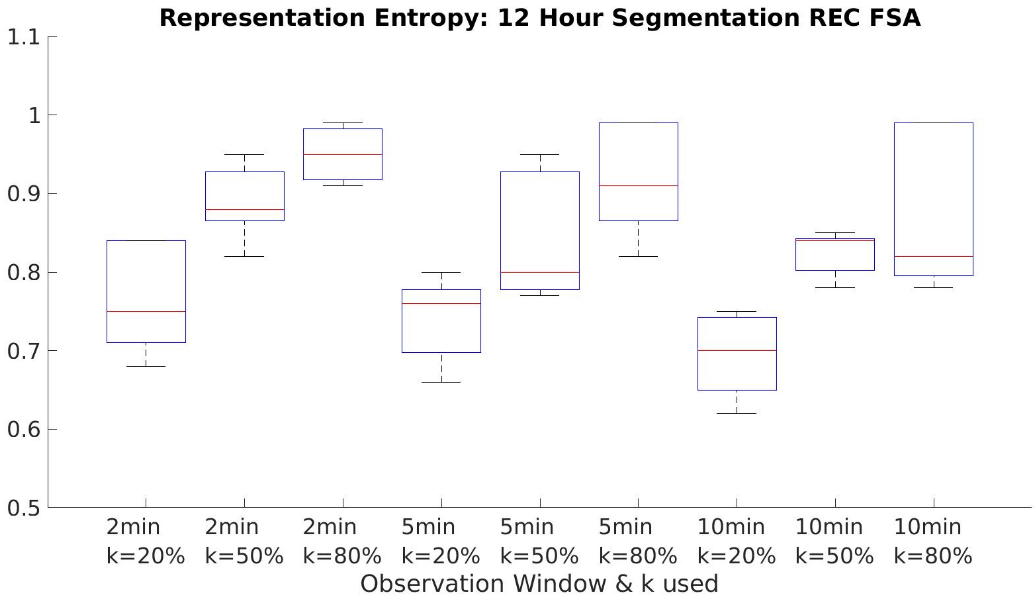

Figure 4.6: Representation Entropy of REC-FSA in various κ sizes 12 hour segments

4.3 Representation Entropy

Figures 4.7, 4.8, 4.9, 4.10 shows the results for representation entropy in different segmentation and observation sizes. For segmentation size set to 2 hours is noticeable that REC-FSA algorithm has a more dense distribution of values than GCNC through various observation window sizes. Mean value of representation entropy for observation window size of 2 and 10 minutes is the same in both algorithms. The minimum value of representation entropy for the REC-FSA algorithm is lower than GCNC for observation window sizes of 2 and 5 minutes. Another observation in 2,4,8,12 hour segmentation plot is that the increased size of observation windows do not necessary lead to better representation results. For example 2 minute observation window for GCNC has the best representation entropy distribution than other observation window sizes for the same algorithm. Finally the GCNC representation entropy value distribution become more dense as the observation window size increases.

Similar observations can be made also in 4 hour segmentation size figure. Again, REC-FSA has a more dense distribution between the various observation window sizes than GCNC. For segmentation sizes set to 8 and 12 hour respectively, a different behavior can be detected. Bigger segmentation in the data lead to smaller density of values. Comparatively, smaller segmentation sizes imply a bigger density in representation entropy values. An advantage of those values for GCNC algorithm is becoming more intense for segment size of 12 hours.

Observation window size is an important factor that affects representation entropy. If it is increased, fewer data consist the feature matrix in each segment. As a result, low observation window sizes such as 2,5 minutes perform better than the larger ones. Bigger segmentation sizes (more data in each segment) do not guarantee better results. Although there is a lower distribution of values in 12 hour segmentation than in 4 hour segmentation.

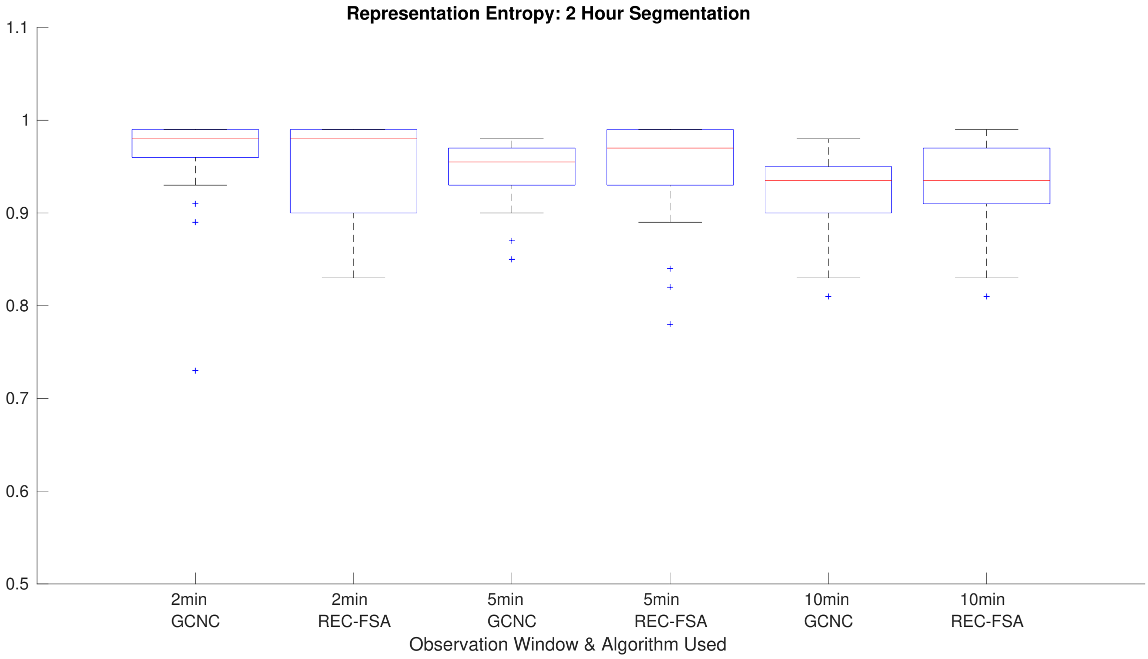

Figure 4.7: Representation Entropy of system from Hydrobionets dataset for 2 hour segments and various observation window sizes

4.4 Compression Ratio

In **Figures 4.11, 4.12, 4.13, 4.14** compression ratio values for each feature selection algorithm is presented in box plots for each segment and observation window sizes. It is distinct, that GCNC outperforms REC-FSA in any segment size and observation window with better compression ratio which exceeds 94%. On the other hand, there are cases that this compressed information cannot represent the initial data as good as the REC-FSA, so vital information is being lost. For example this behavior is encountered for segment size of 4, 8 hours and observation window size of 10, 5 minutes respectively. Overall though, GCNC has performed better for both representation entropy values and compression ratio.

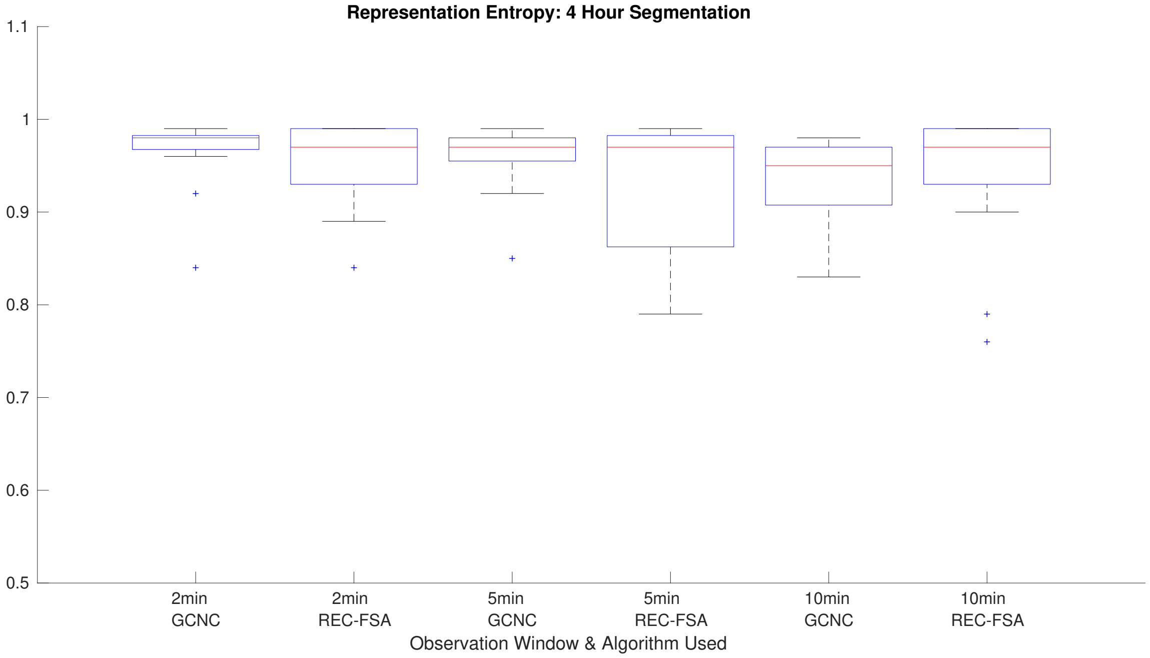

Figure 4.8: Representation Entropy of system from Hydrobionets dataset for 4 hour segments and various observation window sizes

Compression ratio of both algorithms is very high, but the important observation is that even though the returned information is less than 30% of the initial one, it can still represent it pretty accurate. This conclusion can be extracted if information from figures **(4.7, 4.8, 4.9, 4.10 and 4.11, 4.12, 4.13, 4.14)** is combined. As a behavior seems to be common for every segmentation and observation window size.

4.5 Classification Accuracy

In order to test how well the compressed information can reconstruct the initial one, the system uses k-nearest neighbors classification algorithm (**KNN**). By predicting the packet reception ratio labels and compare them with the actual ones, a percentage of accuracy is produced. **Figures 4.15, 4.16, 4.17, 4.18** shows those results for various segment and observation window sizes. Occasionally both unsupervised feature selection algorithms can achieve 100% accuracy. As a result, compressed information is enough to represent missing values accurate. Generally, both algorithms produce very accurate results which exceed 92% for every segmentation and observation window sizes. Nevertheless, for segmentation size of 2 hours and observation window of 5 and 10 minutes the classification accuracy results approached 100%.

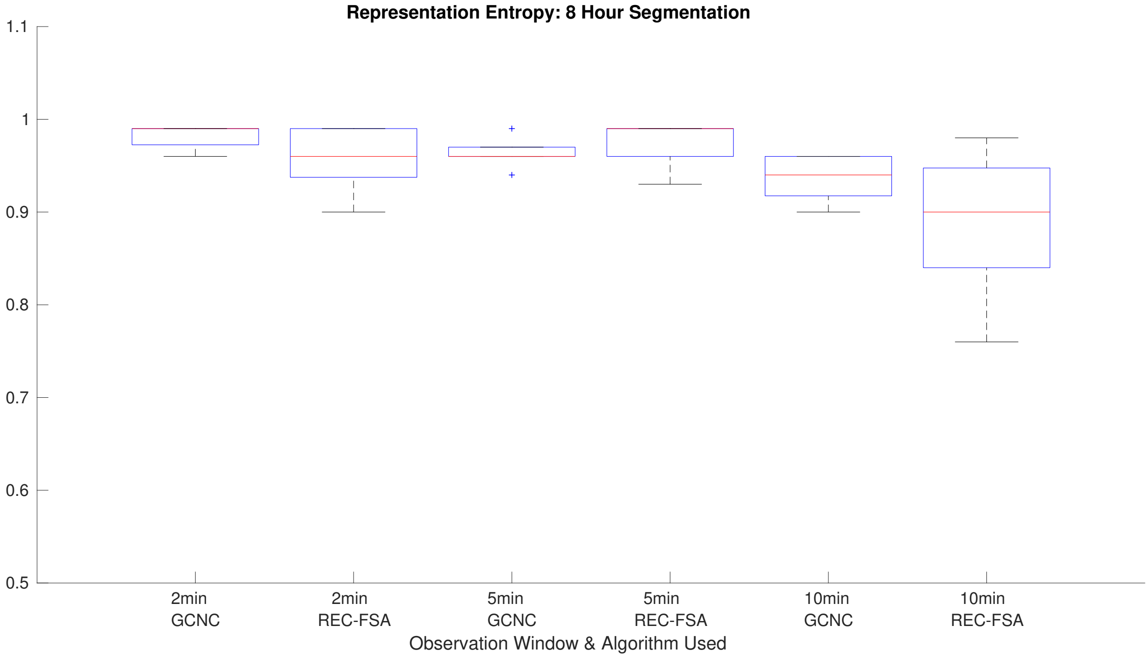

Figure 4.9: Representation Entropy of system from Hydrobionets dataset for 8 hour segments and various observation window sizes

4.6 Execution Time

Evaluation of both feature selection algorithms can be done though measurement of execution times for each one of them. Clearly in **figure 4.19, 4.20, 4.21, 4.22**, GCNC outperforms REC-FSA. For both algorithms execution times are below 5 seconds which is fast, although GCNC most of the times runs below 0.5 seconds. Behavior of REC-FSA has a lower performance especially in the 2 minute observation window size. Data size towards the feature selection algorithms increases, as observation window size shrinks. Due to that increased number of data REC-FSA operates slower than GCNC. The reason behind this number is the implementation of GCNC though graphs which are faster than the linked lists that are used for the implementation of REC-FSA.

As segmentation window size increases, GCNC algorithm execution times remains intact and below 0.5 seconds. On the other hand REC-FSA does not have the same behavior. Clearly, as segmentation size increases the execution time of REC-FSA increases as well. For example in the minute observation window the max execution for 2 hour segments was 1 second and for 8 hours segments was 3 seconds. This means that as the data are increased due to larger segmentation sizes so does the execution time of the algorithm. In addition, observation window sizes are also an important factor that affect the performance of REC-FSA. For example

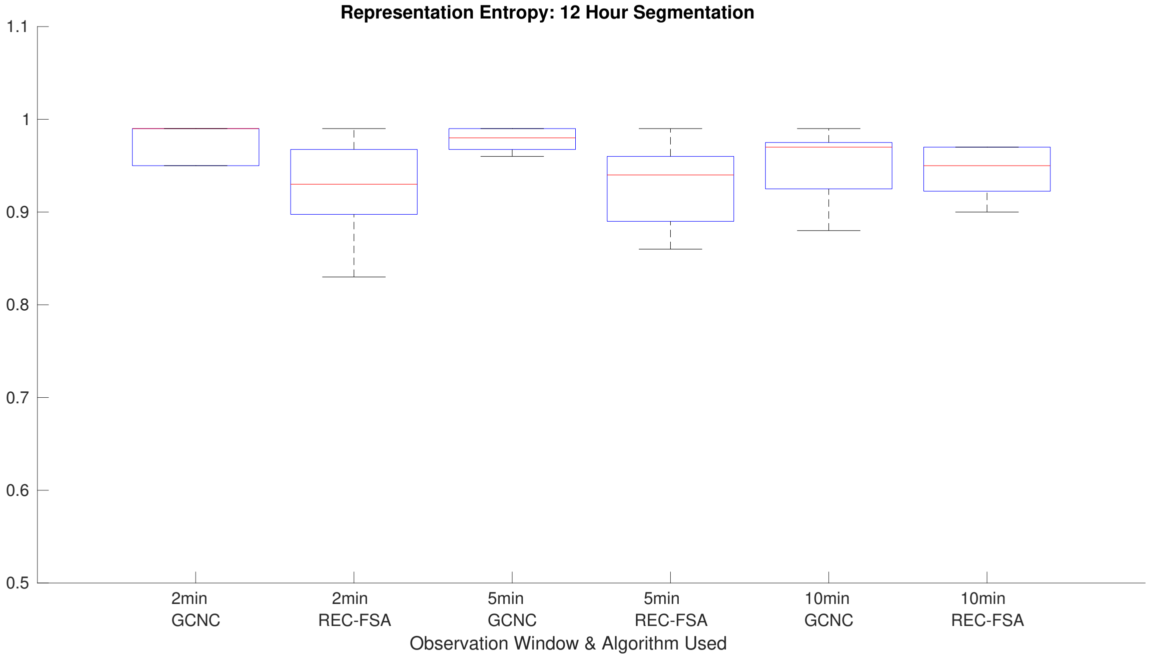

Figure 4.10: Representation Entropy of system from Hydrobionets dataset for 12 hour segments and various observation window sizes

in 8 hour segmentation size execution times of REC-FSA are lower as observation window size increases. This phenomenon is logical because as observation window size increases fewer attributes are inserted in the dominant feature matrix of this segment.

In the following section an temporal aspect analysis was conducted as far the dominant attributes produced by both algorithms in various segment and observation window sizes.

4.7 Temporal Aspect Analysis

First the behavior of both FSAs was evaluated through a fixed set of segment size (i.e 2 hours). **Figure 4.23** displays the number of dominant attributes per observation window size of 2, 5 and 10 minutes. As the observation window increases the sum of extracted dominant features either decreases (in 5 minutes) or increases (between 5 minute and 10 minute observation windows). Since the algorithms does not display a standard behavior for the various window sizes, the observation of the algorithm behavior in small observation window size (2 minutes) was selected.

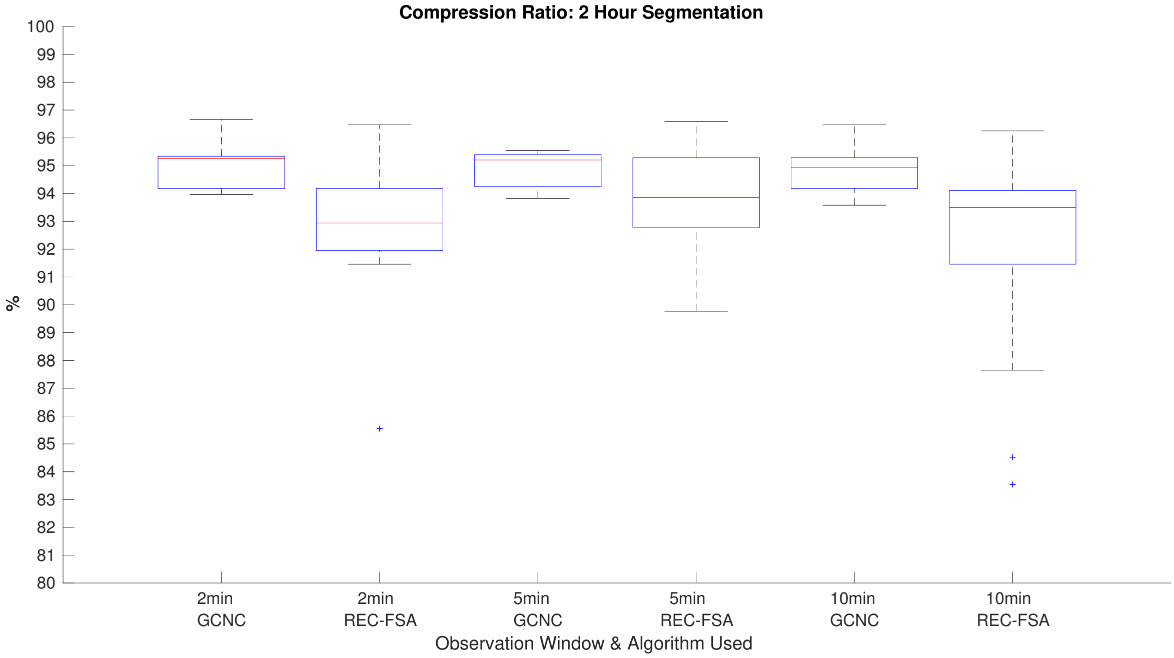

Figure 4.11: Compression Ratio of system from Hydrobionets dataset for 2 hours segment and various observation window sizes

An analysis was held for observation window size of 2 minutes and various segment sizes (2,4,8,12 hours) regarding the behavior of dominant attributes that the Feature Selection Algorithms (FSAs) produced.

At first the goal was to observe the number of dominant attributes that each algorithm outputs when the segment size is increasing. **Figure 4.24** displays the aforementioned behavior. Clearly, the number of total produced dominant features is declining as the segment size increases because the division of the dataset is different. The duration of the network in the collected data is 48 hours and the division of that value by 2 returns a number of 24 segments. However, if the division is done with a value of 12 (hours) the segments are only 2.

Driven by the aforementioned results the analysis was extended to the perspective of the most dominant attribute that each algorithm produced for a segment size of 2 minutes and various segment sizes. Starting from the GCNC FSA the **figures 4.25 4.26, 4.27, 4.28** present the number of appearances of dominant attributes for different segment sizes (2,4,8,12 hours).

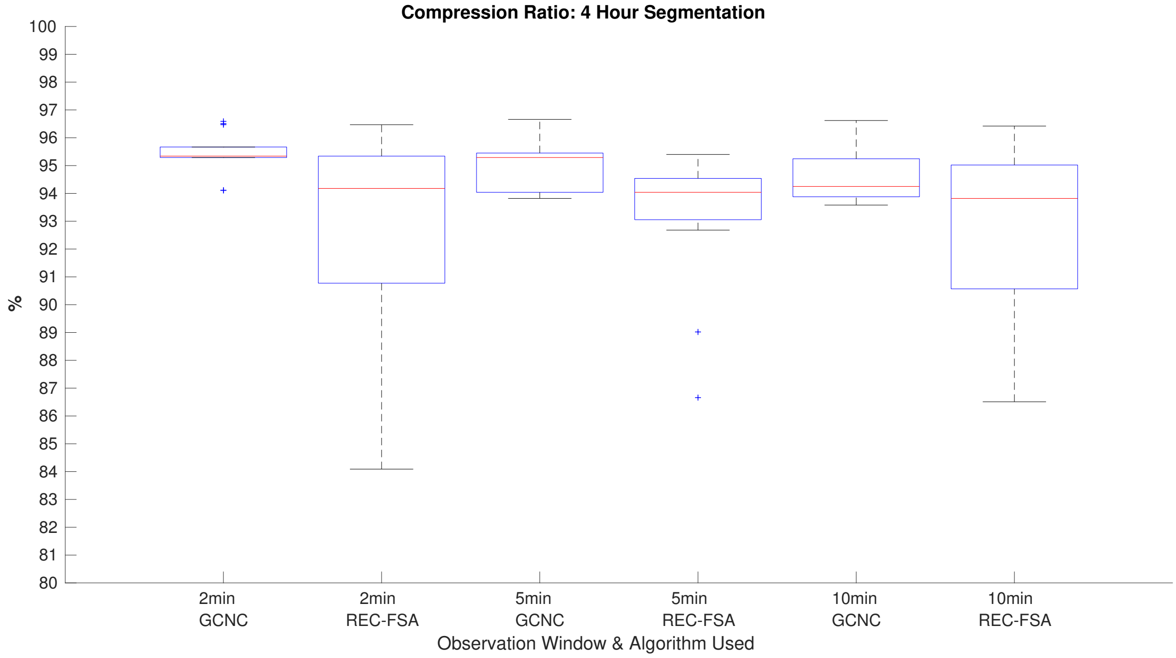

Figure 4.12: Compression Ratio of system from Hydrobionets dataset for 4 hours segment and various observation window sizes

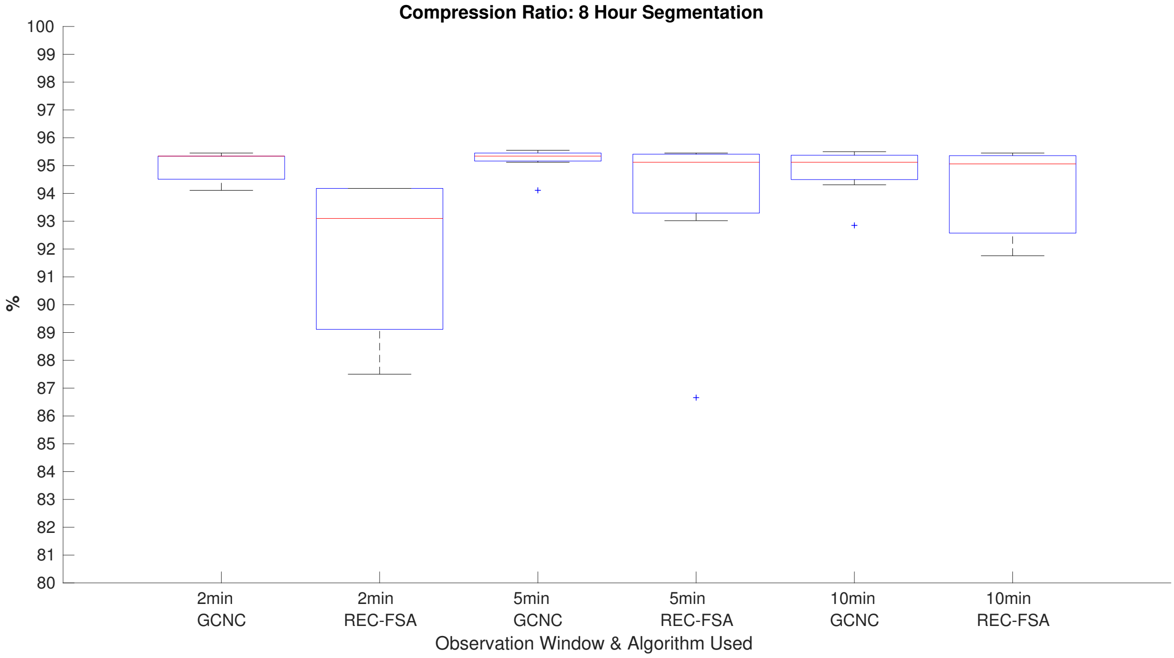

Figure 4.13: Compression Ratio of system from Hydrobionets dataset for 8 hours segment and various observation window sizes

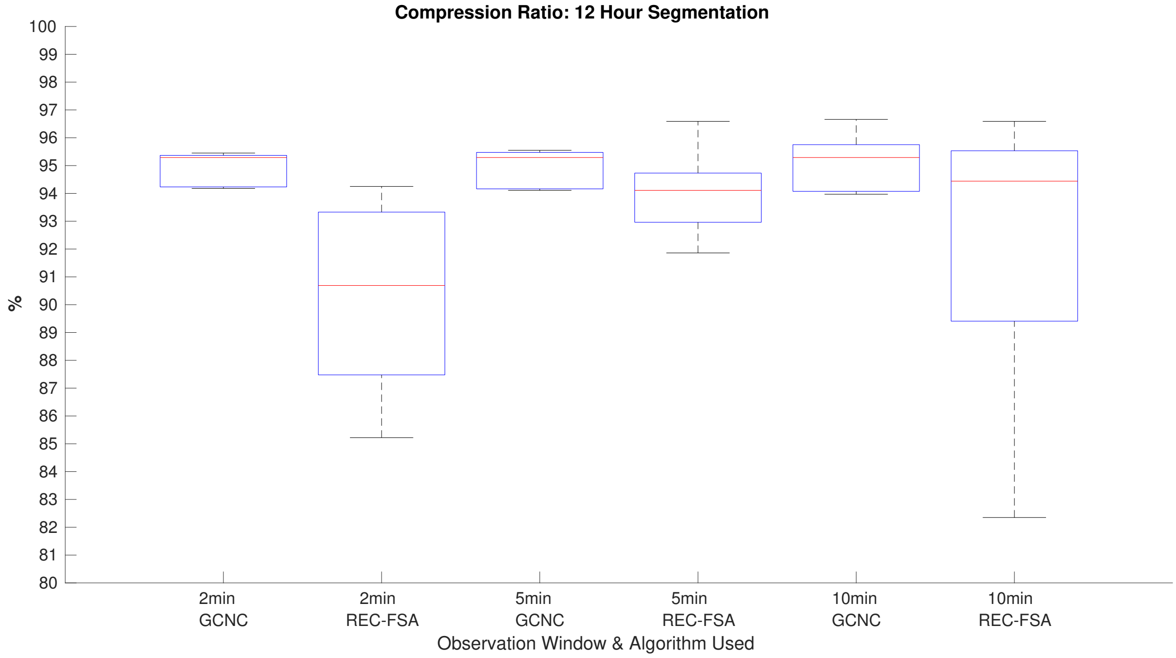

Figure 4.14: Compression Ratio of system from Hydrobionets dataset for 12 hours segment and various observation window sizes

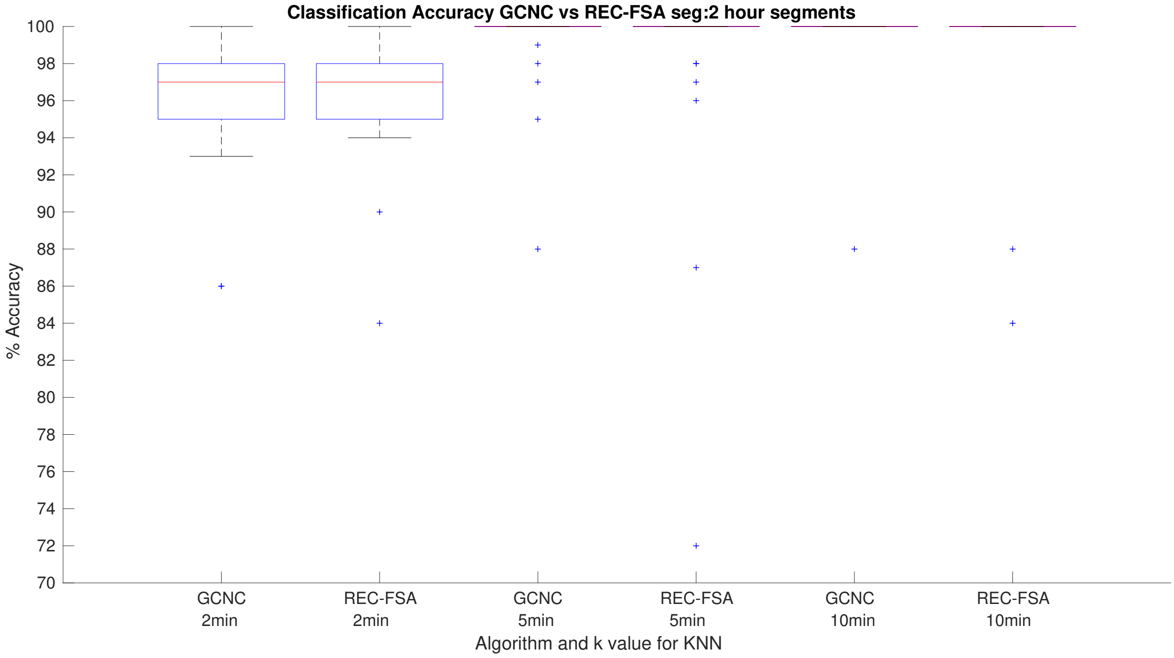

Figure 4.15: Classification Accuracy of system from Hydrobionets dataset for 2 hour segment and various observation window sizes

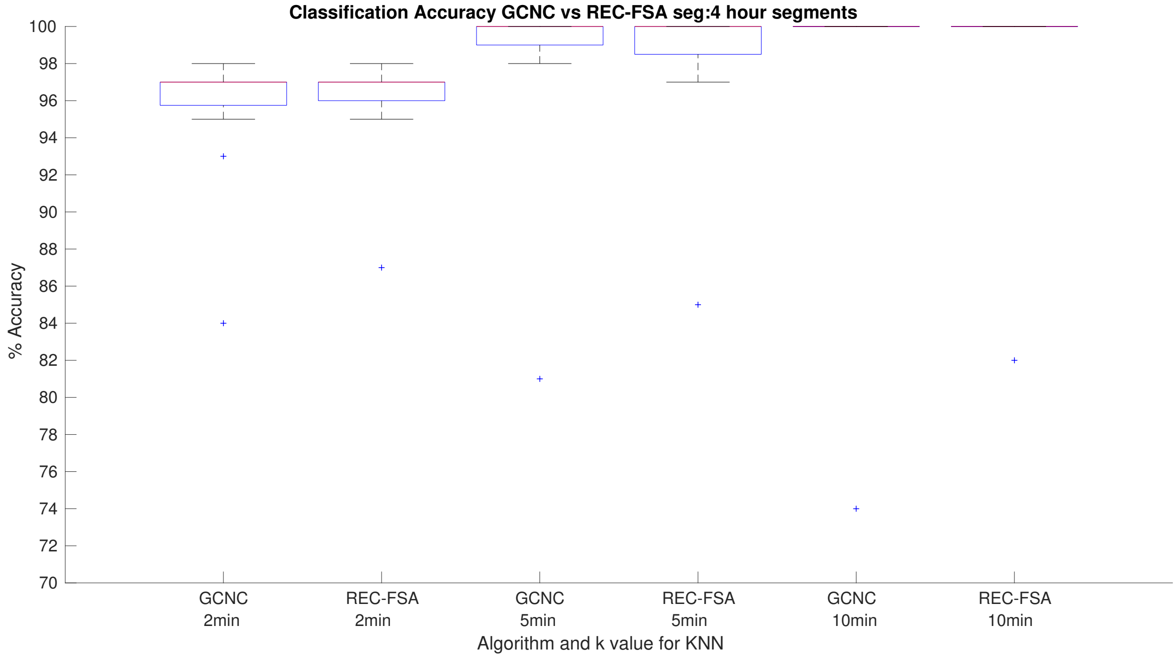

Figure 4.16: Classification Accuracy of system from Hydrobionets dataset for 4 hour segment and various observation window sizes

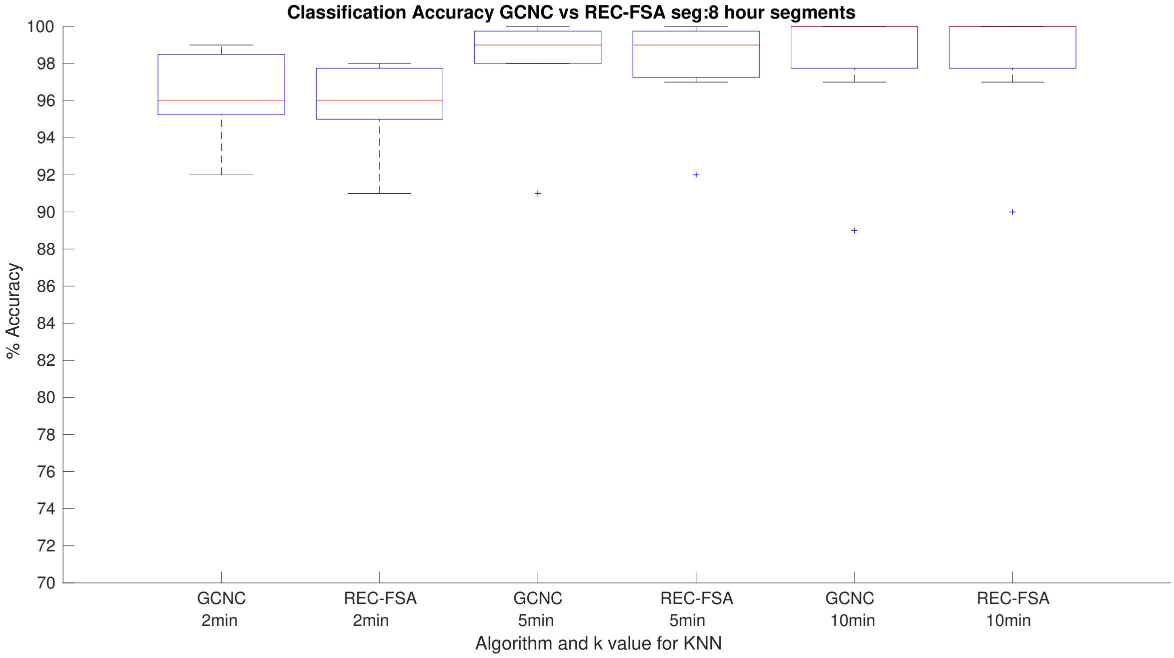

Figure 4.17: Classification Accuracy of system from Hydrobionets dataset for 8 hour segment and various observation window sizes

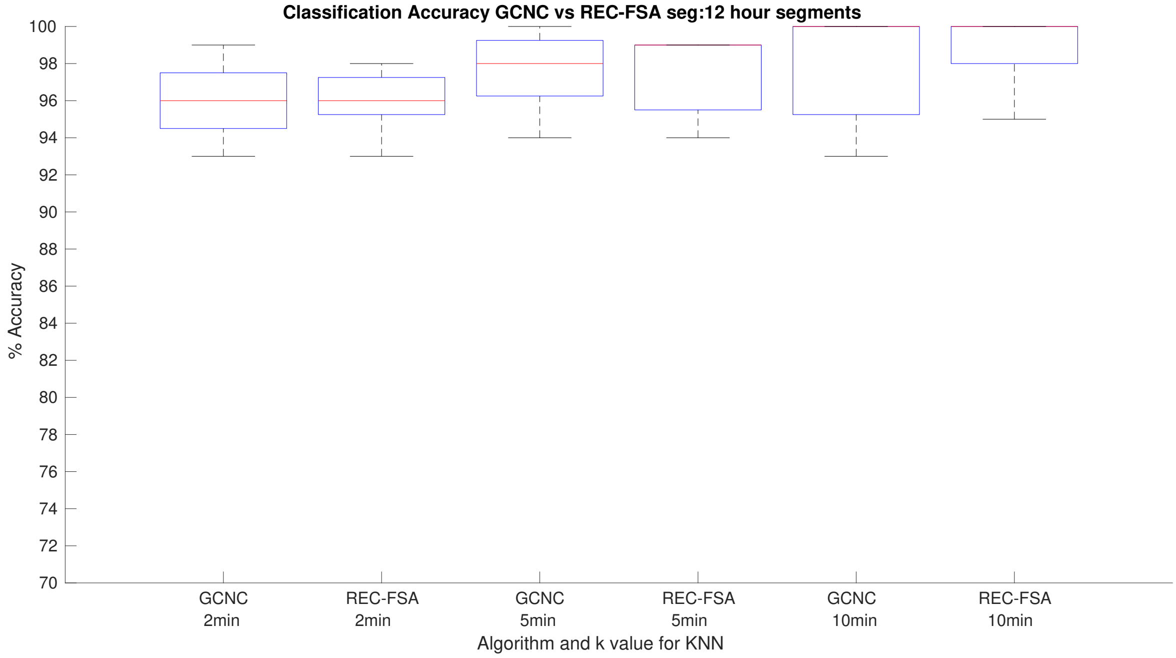

Figure 4.18: Classification Accuracy of system from Hydrobionets dataset for 12 hour segment and various observation window sizes

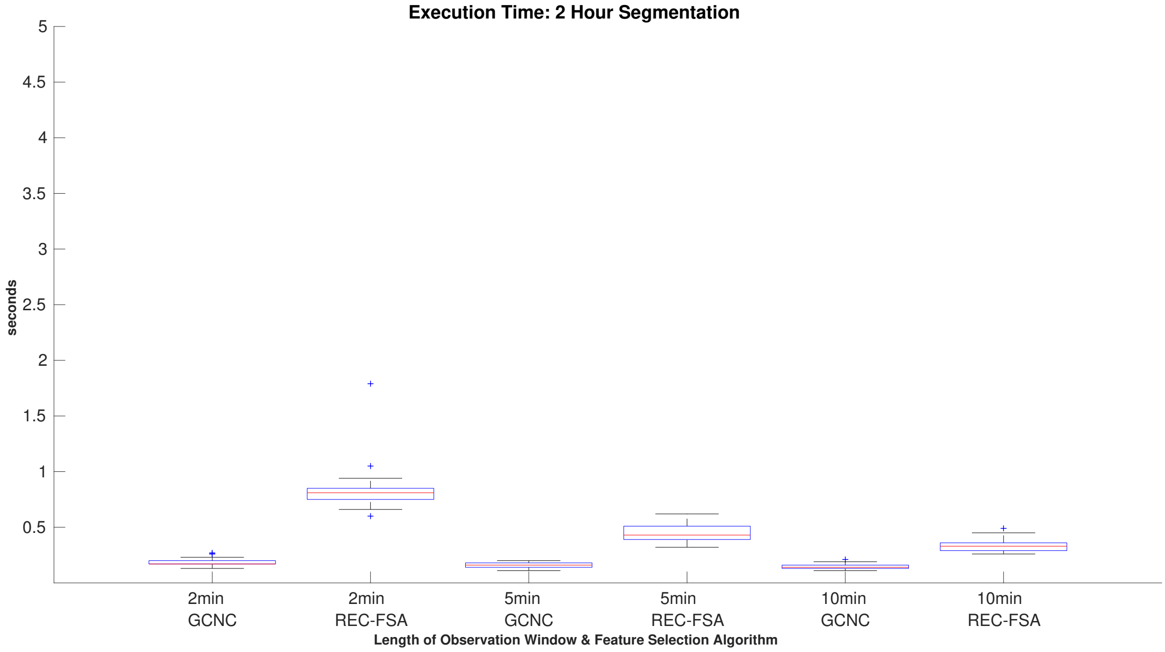

Figure 4.19: Execution Time of each feature selection algorithm for 2 hour segment and various observation window sizes

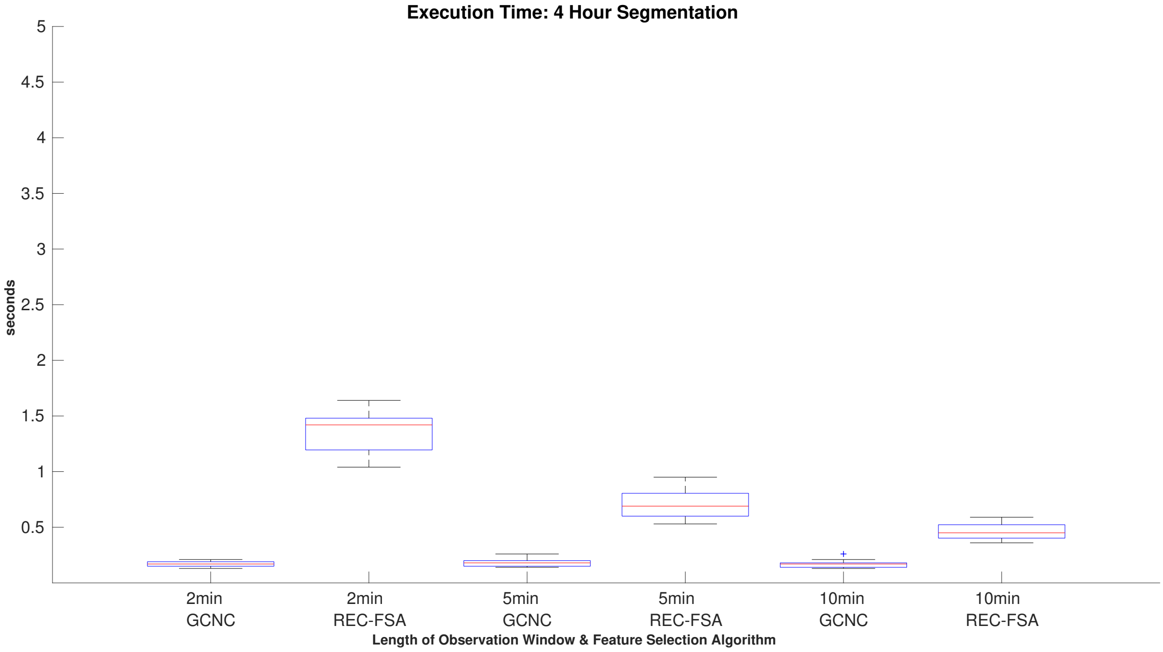

Figure 4.20: Execution Time of each feature selection algorithm for 4 hour segment and various observation window sizes

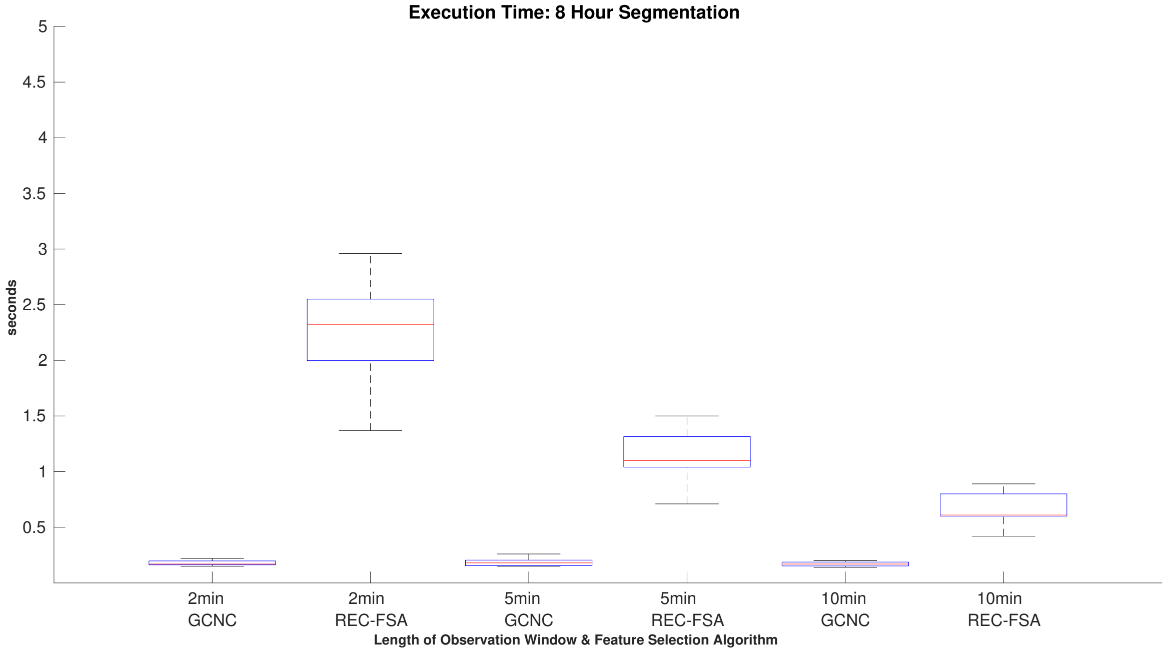

Figure 4.21: Execution Time of each feature selection algorithm for 8 hour segment and various observation window sizes

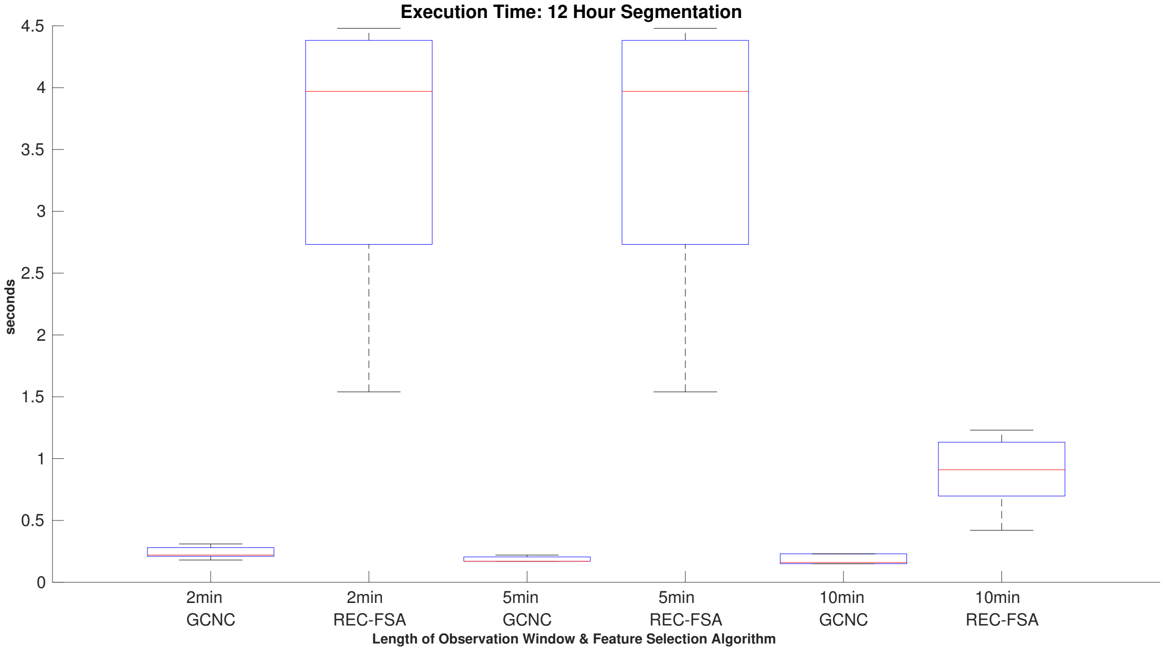

Figure 4.22: Execution Time of each feature selection algorithm for 12 hour segment and various observation window sizes

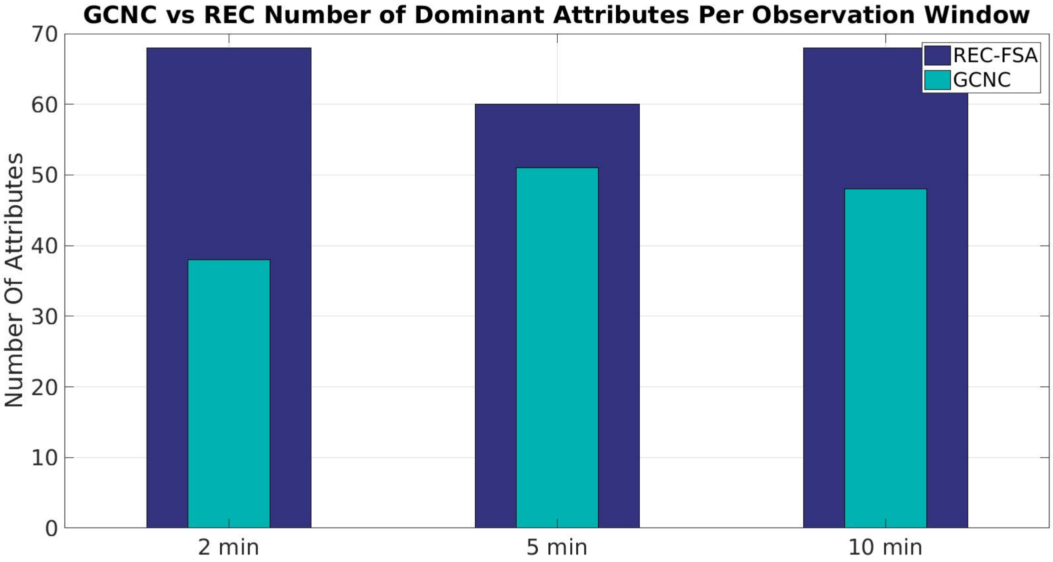

Figure 4.23: Number of Attributes for segment size of 2 hours and various observation window sizes

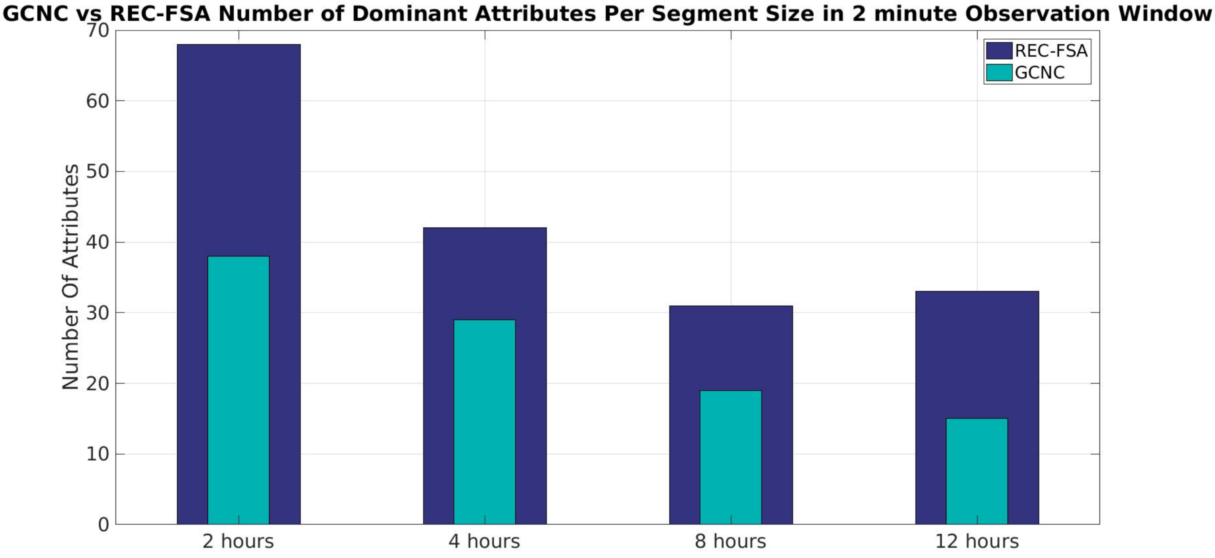

Figure 4.24: Number of Attributes for observation window size set to 2 minutes and various segment sizes

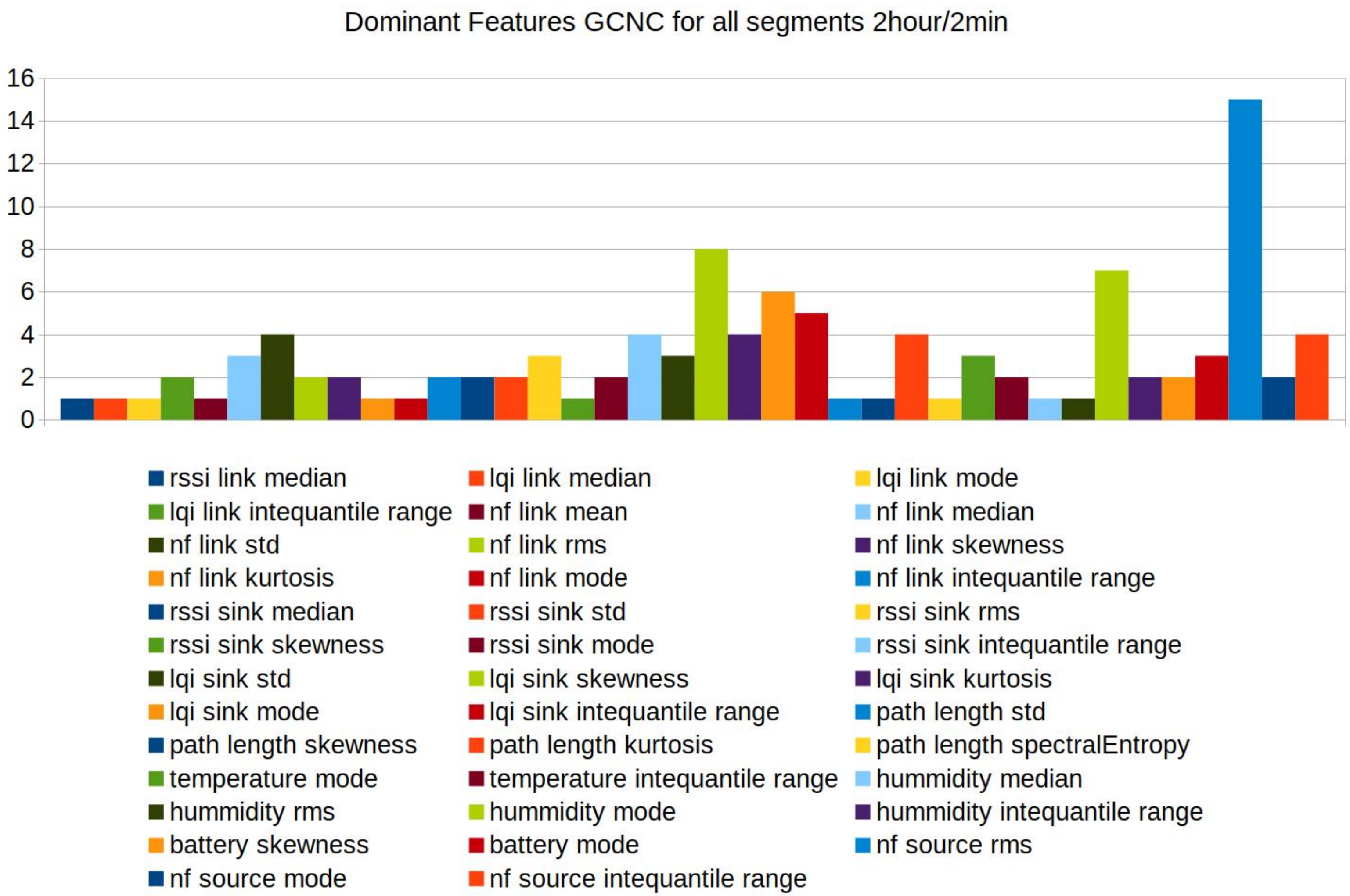

Figure 4.25: Number of feature appearances for GCNC in 2 hour segmentation for 2 minute observation window

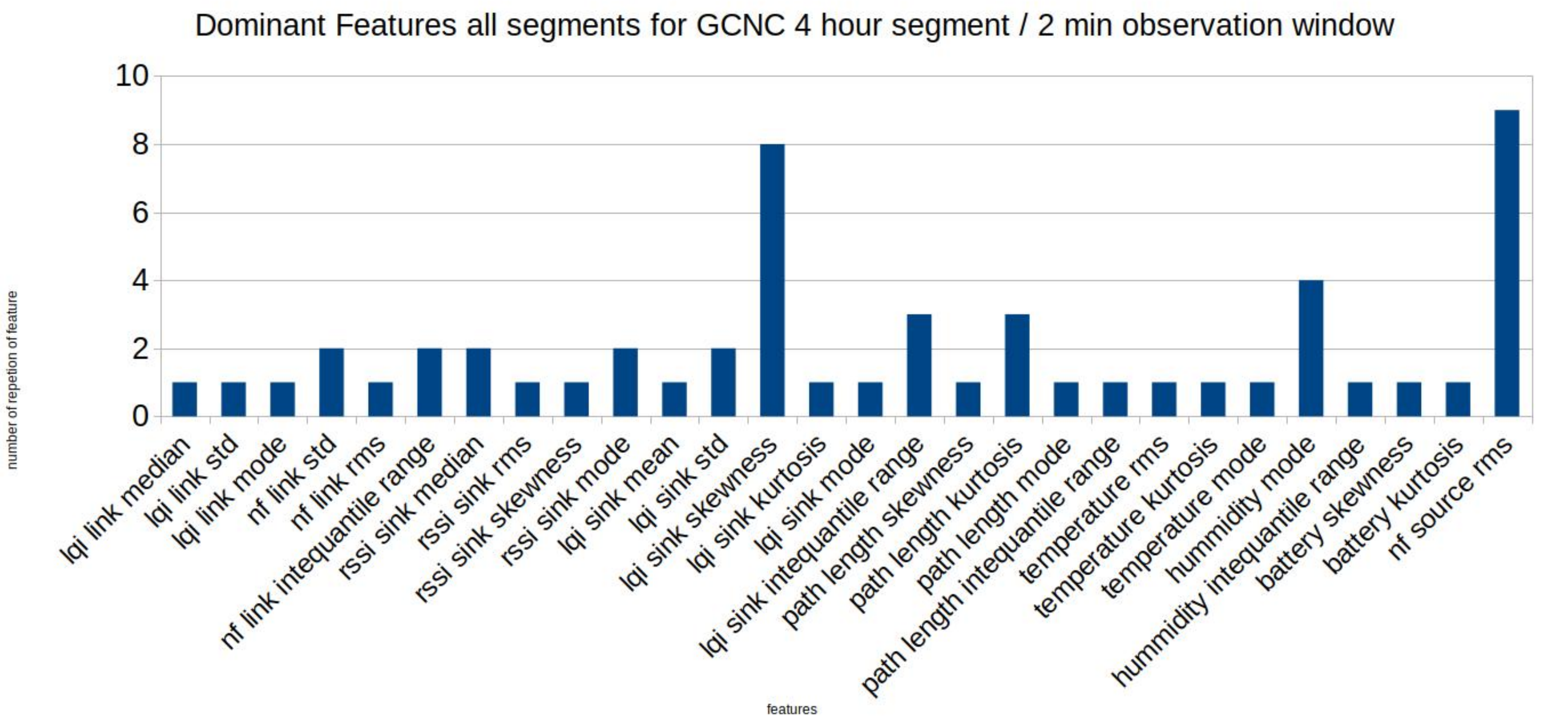

Figure 4.26: Number of feature appearances for GCNC in 4 hour segmentation for 2 minute observation window

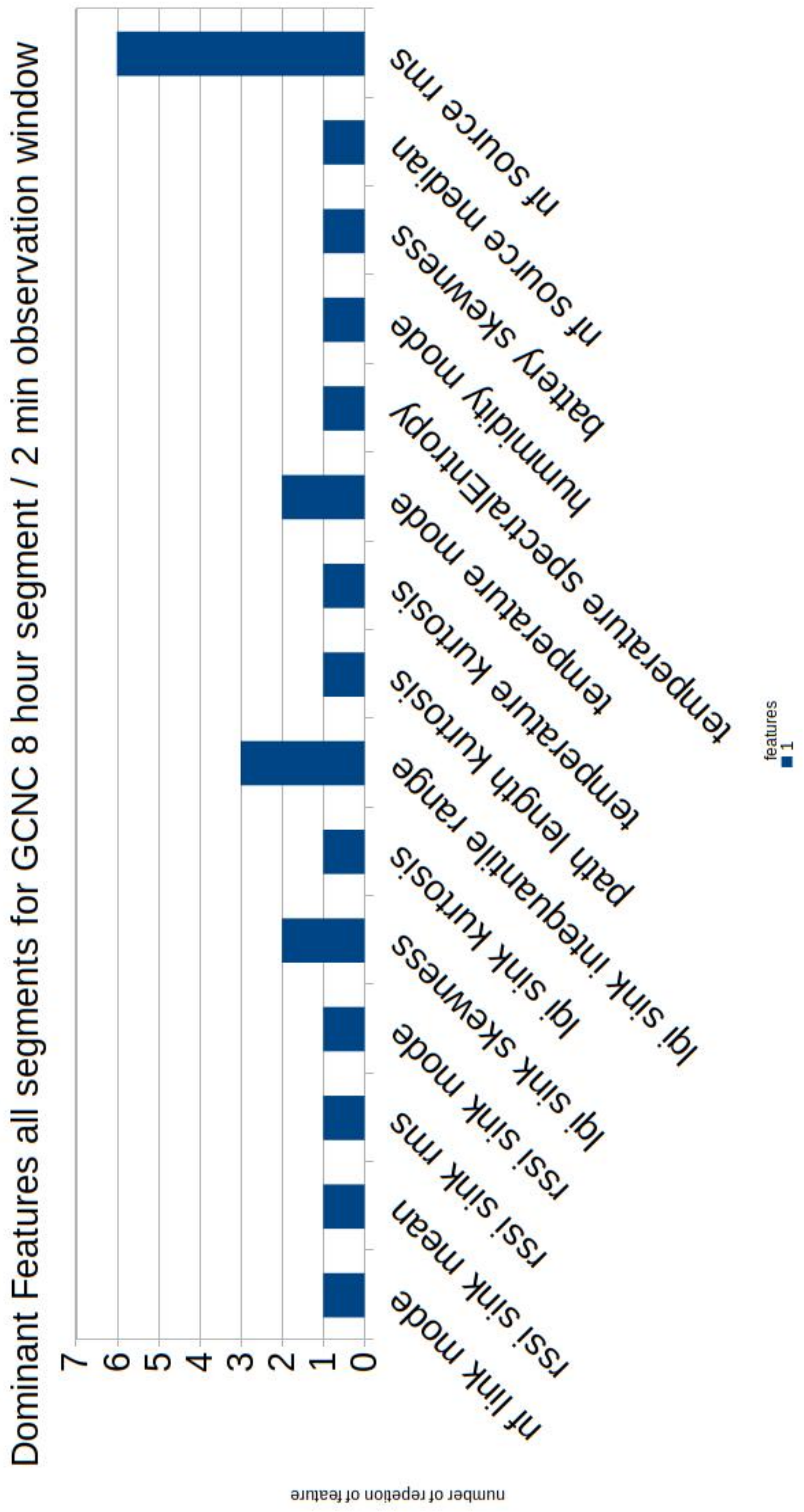

Figure 4.27: Number of feature appearances for GCNC in 8 hour segmentation for 2 minute observation window

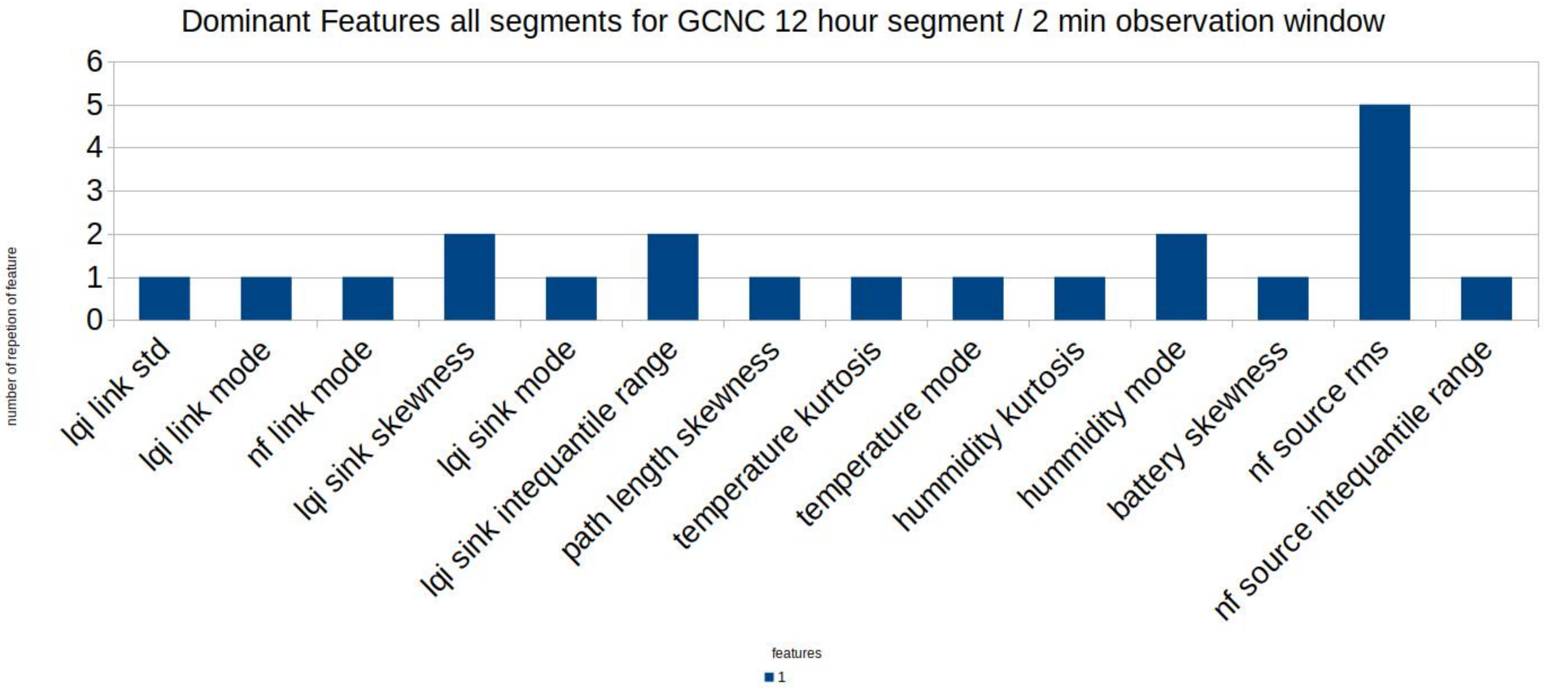

Figure 4.28: Number of feature appearances for GCNC in 12 hour segmentation for 2 minute observation window

Analyzing the figures, it is clear that the most dominant feature was the Root Mean Square (RMS) of Noise Floor for the sink node. This observation is logical since the WSN that was in operation was deployed near heavy industrial machinery so the Noise was present in the environment. The second most common feature varied as the segment size changes. For example in **figures 4.25, 4.26**, the skewness for LQI of sink node appeared as the second most dominant attribute. In **figure 4.27** GCNC ranked second the interquartile range of LQI for sink node. **Figure 4.28** displays that several attributes ranked second as the most dominant. Humidity measurements described by mode, skewness of LQI for sink node and interquartile range of LQI for sink node were all ranked second. The same type of analysis was also conducted for REC-FSA feature selection algorithm.

Figures 4.29, 4.30, 4.31, 4.32 display the dominant attributes along with the number of total appearances. The first ranked feature in **figures 4.29, 4.30**, was the RMS of LQI for sink node. In **figures 4.31 4.32** there isn't a feature that outranked the others. In 8 hour segment size the most dominant features with three appearances were the spectral entropy of Noise Floor for the source node and the skewness of LQI for a link in the WSN. In 12 hour segments various features appeared as most dominant. This behavior led to the observation that for REC-FSA algorithm as the segment size is increasing the most dominant attributes are increasing as well.

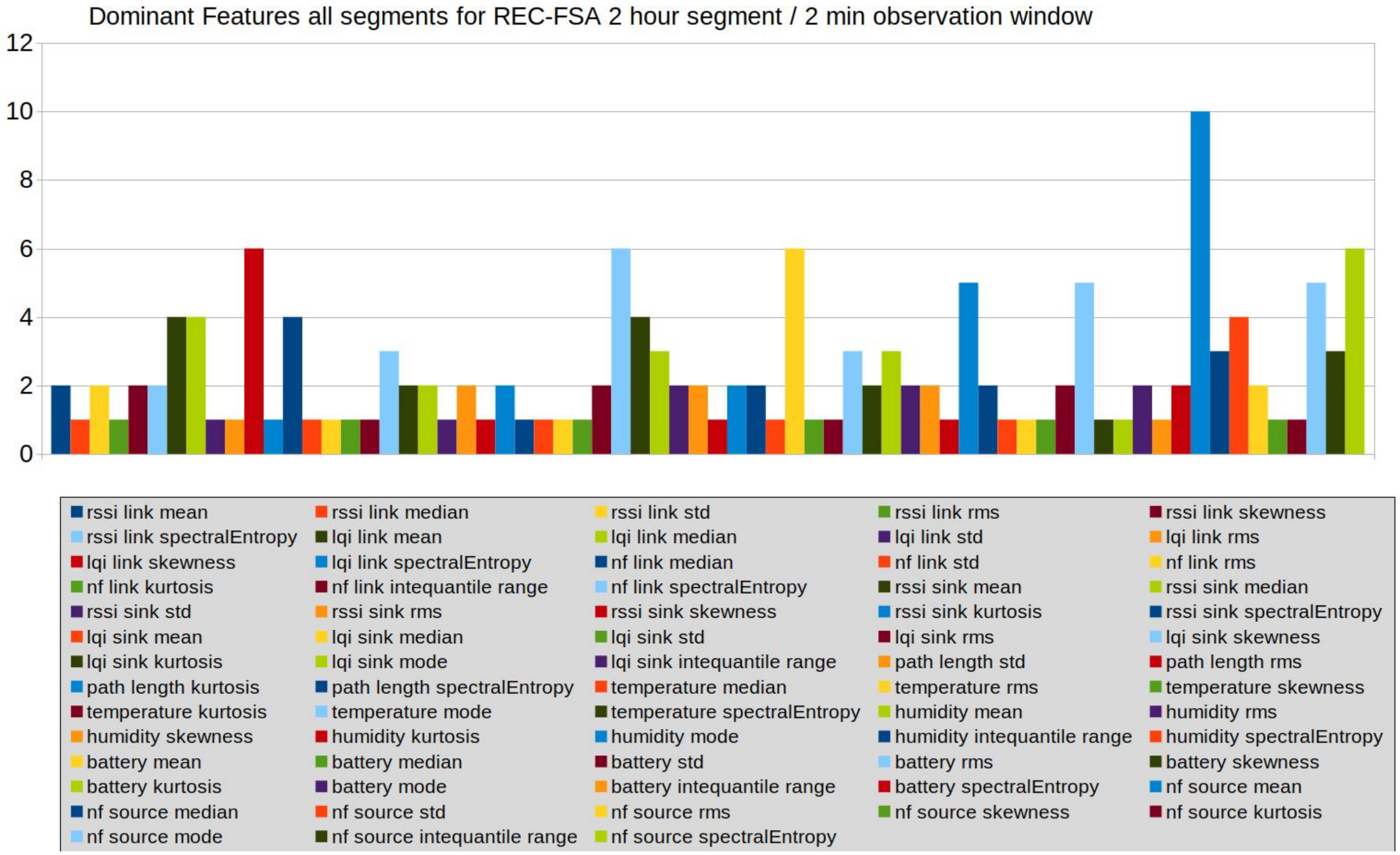

Figure 4.29: Number of feature appearances for REC-FSA in 2 hour segmentation for 2 minute observation window

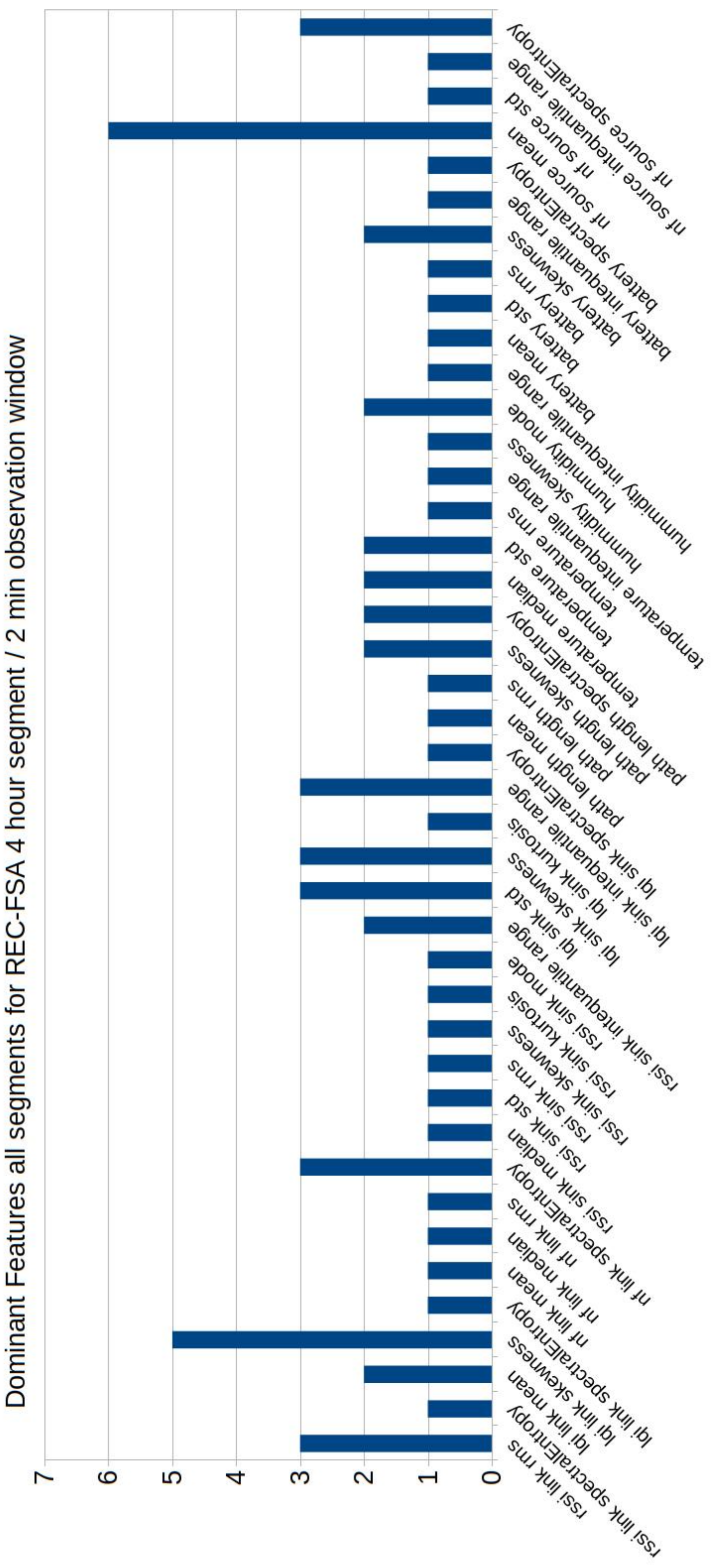

Figure 4.30: Number of feature appearances for REC-FSA in 4 hour segmentation for 2 minute observation window

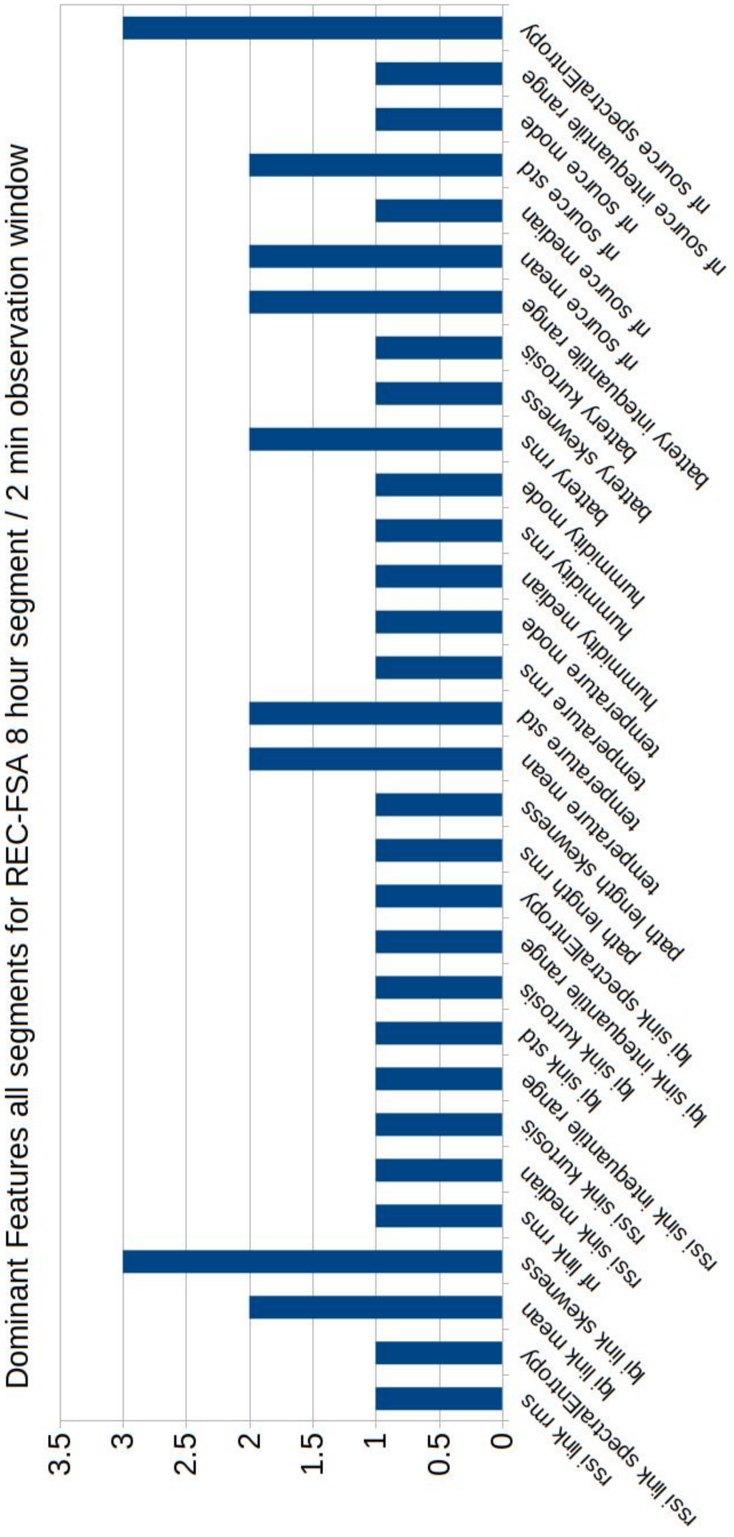

Figure 4.31: Number of feature appearances for REC-FSA in 8 hour segmentation for 2 minute observation window

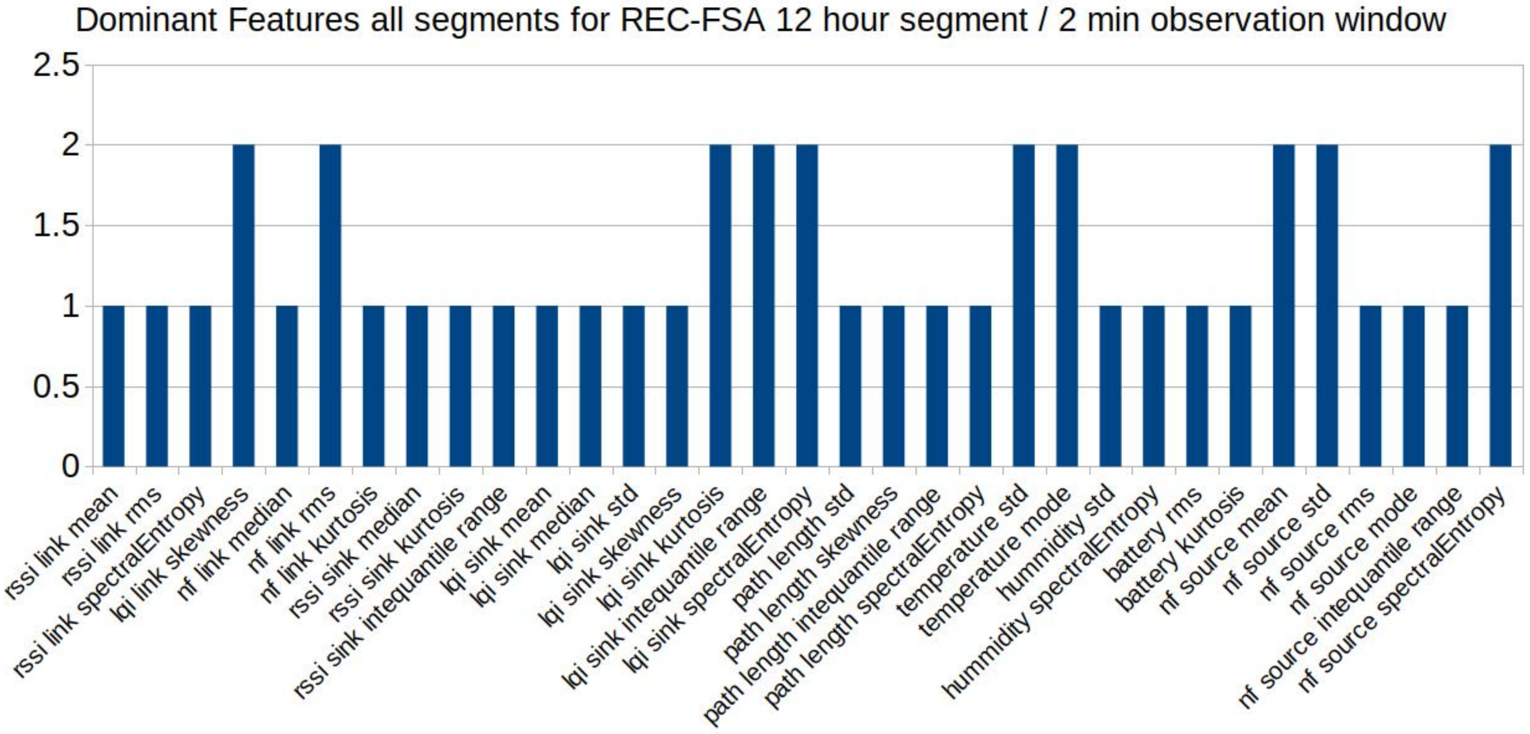

Figure 4.32: Number of feature appearances for REC-FSA in 12 hour segmentation for 2 minute observation window

4.8 Conclusions for Offline Analysis

Overall, offline analysis produced very high results regarding the representation entropy and classification accuracy. As a result, the information that was produced from the feature selection algorithms was enough and capable of representing the initial one. This behavior of good results could come from the distribution of packets in observation windows. **Figure 4.33** presents the time difference between two packets for each sensor node (1-10). From the figure it could be implied that there is a significant distribution of packets below 6 seconds that was the initial information of the Hydrobionets dataset.

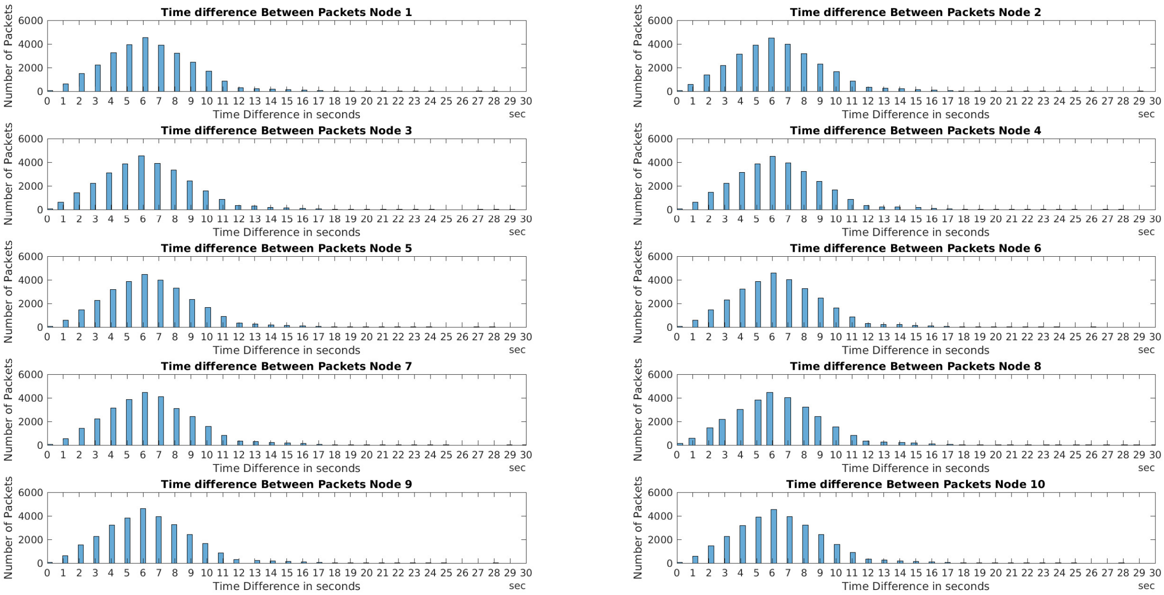

Figure 4.33: Time difference of packets for each sensor from Hydrobionets Dataset

The temporal aspect analysis displayed several behaviors regarding the FSAs. First an increase of observation window does not necessarily lead to an increase to dominant features. There is either an increase or decrease. GCNC algorithm could denote a most dominant feature as the segment size increases. On the other hand REC-FSA had several most dominant features especially in 12 hour segments. Driven by the aforementioned results, a choice was made regarding the real-life WSN deployment. Segment size would be set to 2 hours and the observation window size will be equal to 2 minutes. Furthermore, another aspect should be considered and it involves the sampling rate of sensors. As a result, two sampling rates were chosen. First was 6 seconds (same as the dataset) and the other was 18 seconds. The following sections describe the analysis that was conducted.

4.9 Online Experimental Setup

Online experimental deployment is presented in **figure 4.35**. Experiments were conducted inside Computer Science Department (CSD) at University of Crete. Two classrooms hosted the deployment of sensor nodes. Motes compliant with IEEE 802.15.4 were used and specifically Zolertia's Z1 devices[53]. The device is presented in **figure 4.34**. As sink node the same type of mote was used in order to pass information to Odroid single board computer [28] which stores the information and performs the feature selection process.

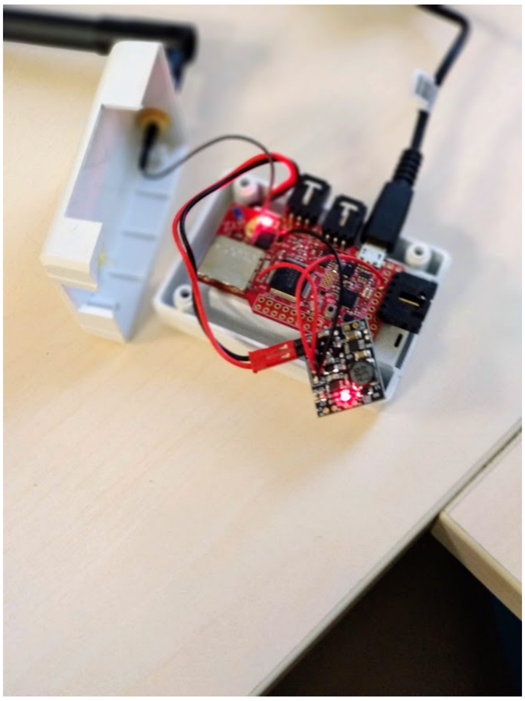

Figure 4.34: Zolertia's Z1 sensor mote

Actual deployment of nodes is also displayed in the following figure 4.36. Sensor nodes were deployed at an indoor environment with various factors to interfere with their communication. Some of them were operating personal computers, obstacles such as walls, doors and interference from the University's Wi-Fi and other electronic devices.

Results from online system are presented below. Analysis is divided into two parts. Packets were sampled and sent every 6 seconds for one set of experiments and 18 seconds for another set. The set of experiments includes feature selection process being called every 2 hours (segment size) with 2 minute observation windows. Data are collected at real time from a WSN gathered at sink node and then parsed for

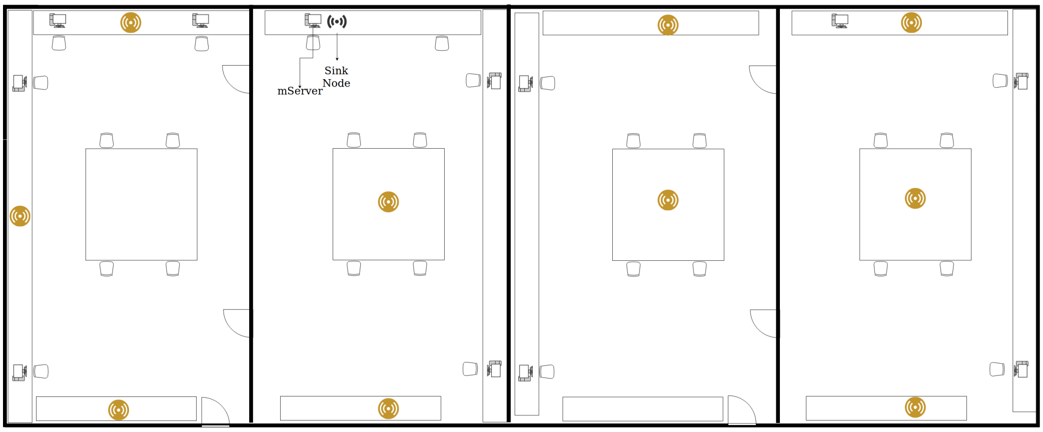

Figure 4.35: Deployment of sensor nodes inside CSD rooms

computation. Data are divided to the defined size of observation window size, features are computed and the feature matrix of a segment is completed. Then this matrix is parsed to the BFSA for feature selection and the reduced feature matrix is extracted. In order to evaluate the produced result a focus has been made on representation entropy, compression ratio, classification accuracy, packet reception ratio, metric and statistic with the greatest appearance.

nd every 6 hours for 6 minute observation windows.

4.10 6 Second Sampling Rate Analysis

4.10.1 Representation Entropy

Representation entropy values for online system when sampling rate is set to 6 seconds has a big distribution. **Figure 4.37** displays that distribution for both feature selection algorithms. Max value of GCNC is below 1 and for REC-FSA is 1. Representation entropy values range between 0.05 and 0.6 for GCNC and 0.15 to 0.85. Mean value of representation entropy of REC-FSA algorithm is better than GCNC's so overall, from representation entropy's perspective REC-FSA performs better.

4.10.2 Classification Accuracy

Figure 4.38 present the performance of each algorithm extracted information for prediction of packet reception ratio labels. At fist glance, both feature selection algorithms seems to have similar performance. Although, minimum value of classification accuracy is spotted for the second features selection algorithm. Mean value of REC-FSA is higher than GCNC and the distribution of values are higher in REC-FSA than GCNC. On the other hand GCNC has achieved classification accuracy of 100% than REC-FSA which is a little lower than that. Considered that REC-FSA performed also better in representation entropy value is logical that better information can lead to better label prediction by the KNN algorithm.

4.10.3 Compression Ratio

Compression ratio performance is showed in **figure 4.39**. GCNC outperforms REC-FSA with compression ratio values range from 93 to 96%. Mean values for both feature selection algorithms have a difference of 16% which has an impact in the extracted information. GCNC returns more reduced information than REC-FSA so it affects both representation entropy values and classification accuracy. Previous conclusions can be verified, by the behavior of compression ratio for both feature selection algorithms.

4.10.4 Classification Accuracy And Packet Reception Ratio

Classification accuracy in respect to packet reception ratio values is showed in **figure 4.40, 4.41**. In figure 4.40 values of classification accuracy varies in different packet reception ratios. There isn't a certain behavior between day and night time. High packet reception ratios do not guarantee high classification accuracy values. For example during time interval of 18:04-20:04 even though there is packer reception ratio of 90% packet reception ratio is 50%. Another example is at the end of the experiment where classification accuracy for GCNC is 100% even though the packet reception ratio is 55%. This can only mean that the packet reception ratio labels for that particular segment is similar.

For REC-FSA behavior of classification accuracy and packet reception ratio seems to be the same as GCNC. Both algorithms seems to be very close to the classification accuracy results with small differences such as in time interval of 4:04-6:04 where classification accuracy of REC-FSA exceeds GCNC by 12%.

4.10.5 Dominant Statistical Aspect

Statistical with most appearance for experiment of sampling 6 seconds is presented in **figures 4.42, 4.43**. Mean and median statistical have the most appearances in segments for GCNC algorithm (figure 4.42) of 4 times each. On the other hand REC-FSA algorithm gave mean statistical in 5 segments and kurtosis appeared in 3 segments.

4.10.6 Dominant Metric Aspect

Display of the most common metric inside each segment is presented in **figures 4.44, 4.45** for each feature selection algorithm. For both feature selection algorithms was the most common metric that appeared as dominant for most of the segments. Although packet loss rate, path length,RSSI also appeared as dominant in some segments but Noise Floor surpassed them in times of occurrence. REC-FSA algorithm denotes in segment time of 2:04 - 22:04 different dominant metric than GCNC. Those metrics that appear is RSSI, LQI, path length and battery. RSSI and LQI as dominant attribute means that there are value fluctuations for these metrics which leads to the conclusion that noise interfered to the received signal and distorts it. Path length as dominant feature explains that packets either travels through a short path or a long path inside the WSN. Certainly there is an important variation of values for path length in order to appear as dominant feature.

4.10.7 Conclusions for 6 second Sampling Rate

Analysis of 6 second sampling rate produced moderate results regarding the representation entropy and classification accuracy. For the particular experiment Packet Reception Ratio values varied throughout the operation of WSN and so did the classification accuracy. The implemented feature selection algorithms performed similar with small value differences. An exception though still existed regarding the compression ratio where GCNC achieved higher values.

Figure 4.46 displays the distribution of packets throughout time. Packets for each node are gathered round 6 seconds.

4.11 18 Second Sampling Rate Analysis

4.11.1 Representation Entropy

Representation entropy values inside segments is presented in **figure 4.47**. REC-FSA excels over GCNC in value distribution. Mean value of REC-FSA is below 0.25 and GCNC is 0.1. Max value for each feature selection algorithm is above 0.35 for REC-FSA and below 0.2 for GCNC. For both algorithms value of representation entropy is low, so the returned information cannot reconstruct the initial information so well.

4.11.2 Classification Accuracy

Classification accuracy distribution values is presented in figure **4.48**. Mean value for REC-FSA is approximately 55% and for GCNC is above 50%. Overall, GCNC algorithm performed worse than REC-FSA with larger distribution values. Minimum value of GCNC is lower than REC-FSA, so it is clear that REC-FSA outperformed GCNC and as a conclusion the returned information from that algorithm gave enough information for better prediction of packet reception ratio values.

4.11.3 Compression Ratio

Compression ratio performance is showed in **figure 4.49**. GCNC outperforms REC-FSA with compression ratio values range from 93 to 96%. Mean values for both feature selection algorithms have a difference of 16% which has an impact in the extracted information. GCNC returns more reduced information than REC-FSA so it affects both representation entropy values and classification accuracy. Previous conclusions can be verified, by the behavior of compression ratio for both feature selection algorithms.

4.11.4 Classification Accuracy And Packet Reception Ratio

Classification accuracy in respect to packet reception ratio is displayed in **figures 4.50,4.51**. For segments of time 17:41 - 11:41 is obvious that even though packet reception ratio values are between 95-100% classification accuracy does not reach the same values. As mentioned before representation entropy values from the returned information is low so this has an impact on the prediction of packet reception labels. During date time classification accuracy values seem to perform better than night time. Both feature selection algorithms perform similarly when it comes to classification accuracy with small value differences (1-13%).

4.11.5 Dominant Statistical Aspect

Statistical with most appearance for experiment of sampling 18 seconds is presented in **figures 4.52, 4.53**. Mean and kurtosis statisticals have the most appearances in segments for GCNC algorithm (figure 4.42) with 6 and 5 segments each. On the other hand, REC-FSA algorithm gave mean and spectral entropy statistical in 5 segments.

4.11.6 Dominant Metric Aspect

Metrics that appeared most as dominant attributes are presented in **figures 4.54, 4.55** for each feature selection algorithm. Metric that appeared most for several segments was packet loss rate (figure 4.54. Computation of this metric is depended from LQI. Packet loss rate formula also depends from parameters a,b and payload size which are constant. As a conclusion this metric was affected by distortions of the LQI value and as a result GCNC algorithm. Even though it appears as dominant metric packet reception ratio for those segments is very high except in the segment 7:41 - 9:41 (figure 4.50) where packet reception ratio is between 85% - 90%.

REC-FSA algorithm showed different behavior regarding the metric with the most appearance. Figure 4.55 displays that the most common metric in segments was Noise Floor. Interference of noise affected the network and it's value in a way that the feature selection algorithm extract it as dominant for each segment. In the meantime where noise floor weren't the most dominant feature, LQI and RSSI appeared two times. As a conclusion variation in these metric values has happened in order for REC-FSA algorithm to choose it as dominant. Noise interference is the reason behind this variation.

4.11.7 Conclusions for 18 second Sampling Rate

Analysis of 18 second sampling rate produced moderate results regarding the representation entropy and classification accuracy. Even though the packet reception ratio was above 70% the information extracted from the feature selection process couldn't represent the initial one that well.

4.12 Conclusions Between the different Sampling Rate Results

Various conclusions were implied above from the displayed results. Different sampling rate lead to different system behavior. For example representation entropy values are better when sampling rate is set to 6 seconds rather than 18 seconds. Even though the extracted information represents better the initial one when sampling rate is set to 6 seconds, classification accuracy values seems to be similar.

Compression ratio also has the same behavior with similar results for GCNC algorithm, although for REC-FSA the values of compression ratio have a bigger distribution in 6 seconds rather than 8 seconds sampling. Reason behind this behavior is data size that the feature selection algorithm receives. Maybe for REC-FSA as data size decreases the distribution of compression ratio increases.

GCNC algorithm from the statistical perspective gave different statistics as dominant for each sampling rate. In 6 second sampling the greatest appearance of a statistic was mean and median in 4 segments each and in 18 second sampling median and kurtosis appeared in 6 and 5 segments respectively. REC-FSA on the other hand had mean statistic as the most common statistic in both sampling rates. The second greatest statistic for this feature selection algorithm was kurtosis in 6 second sampling and spectral entropy in 18 second sampling.

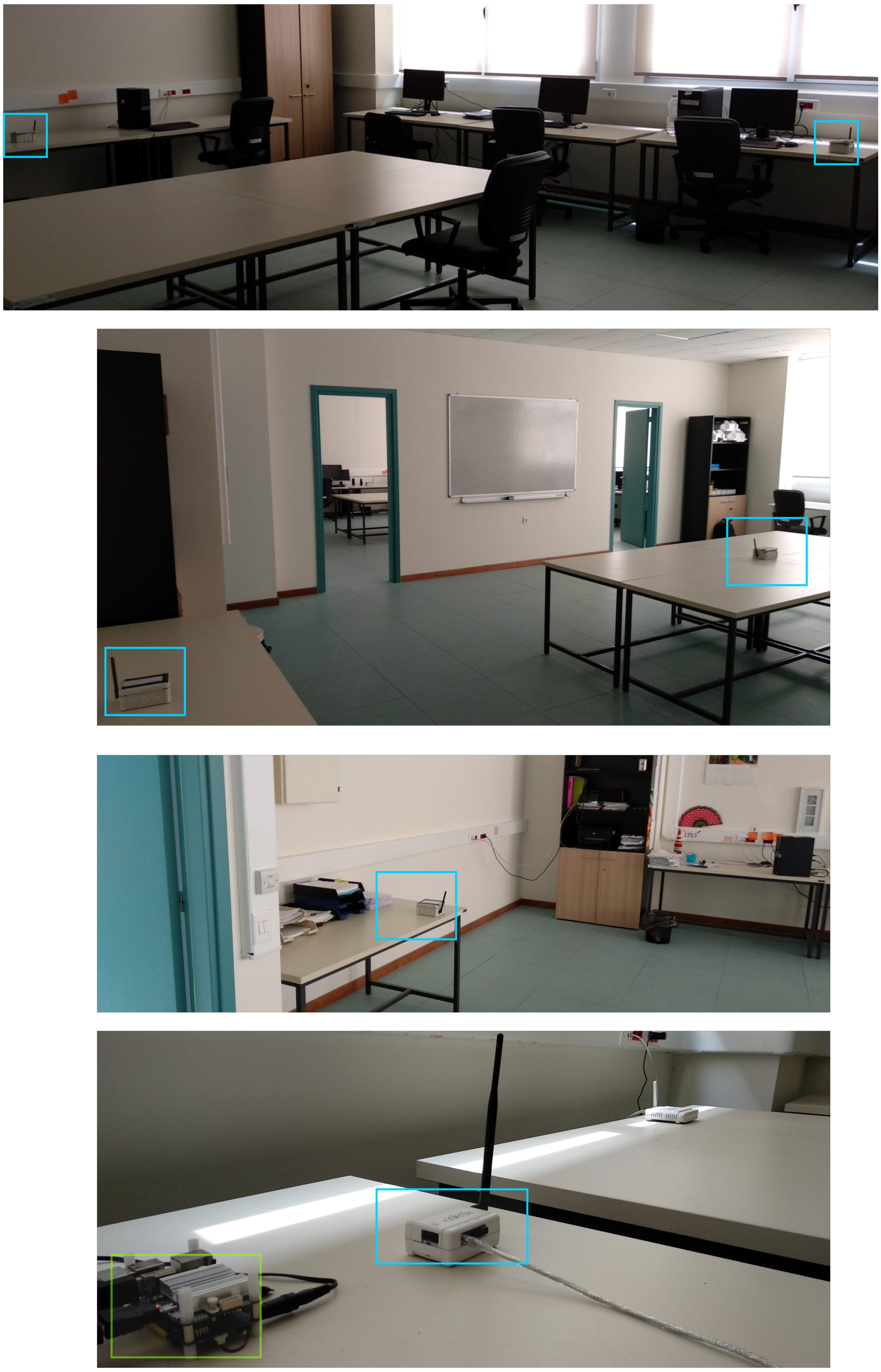

Figure 4.36: Experimental Setup Display. Node positions inside the classroom on top 3 pictures, sink and odroid in the bottom one

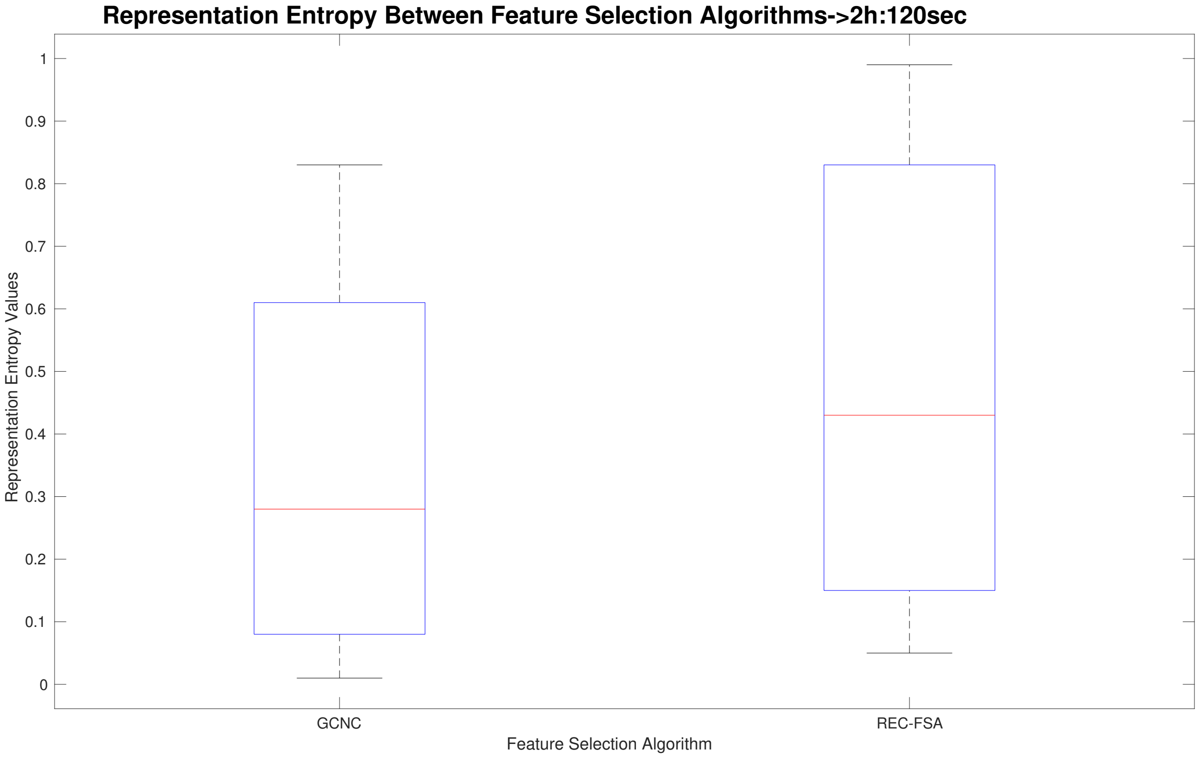

Figure 4.37: Representation Entropy box plot for each Feature Selection Algorithm

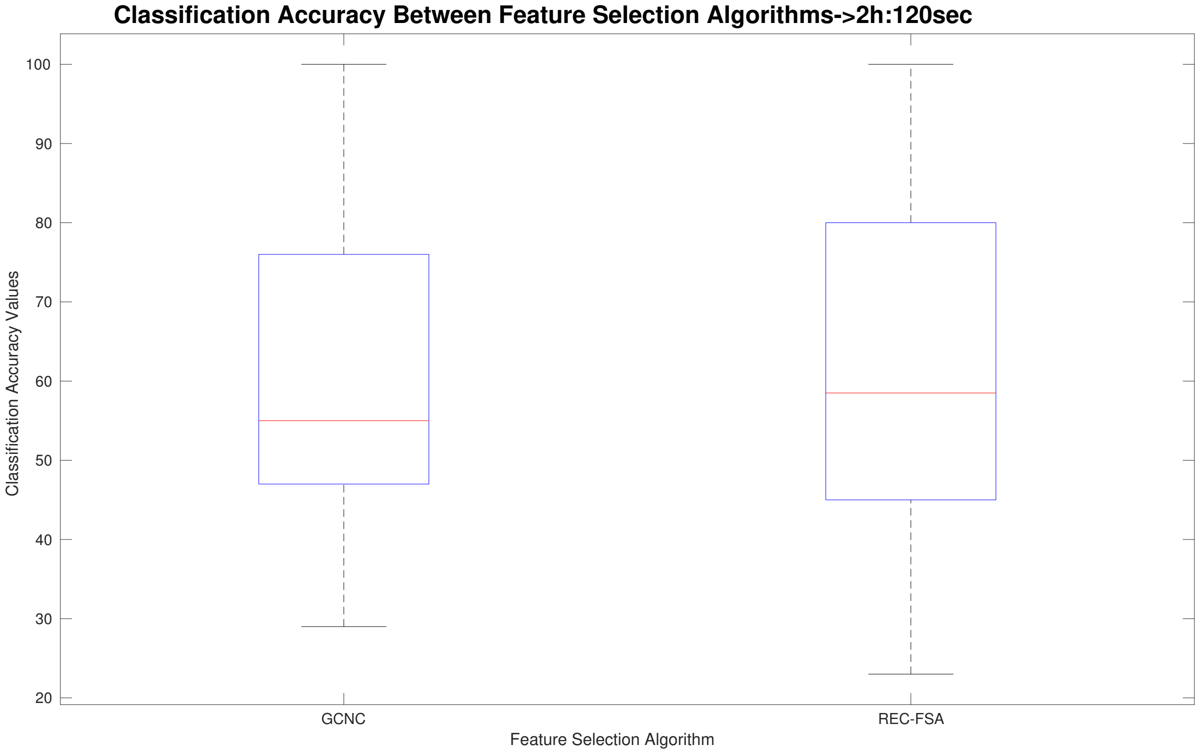

Figure 4.38: Classification Accuracy box plot for each Feature Selection Algorithm

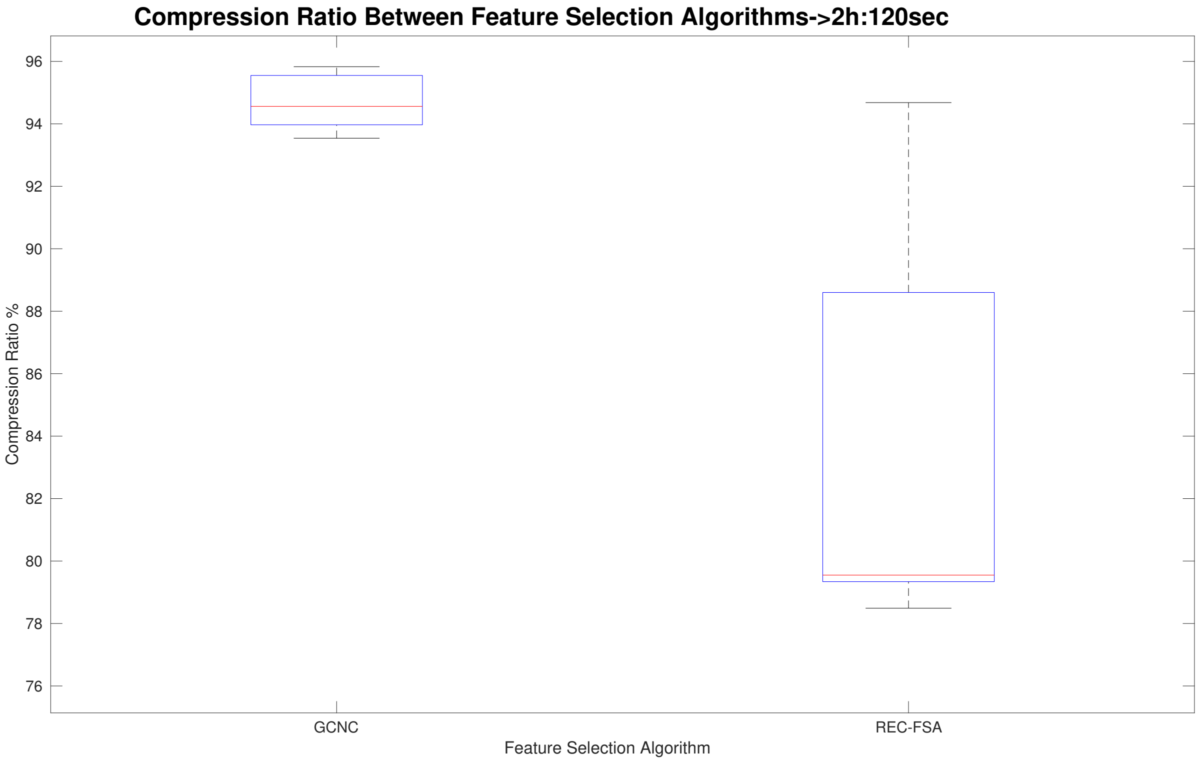

Figure 4.39: Compression Ratio for each Feature Selection Algorithm

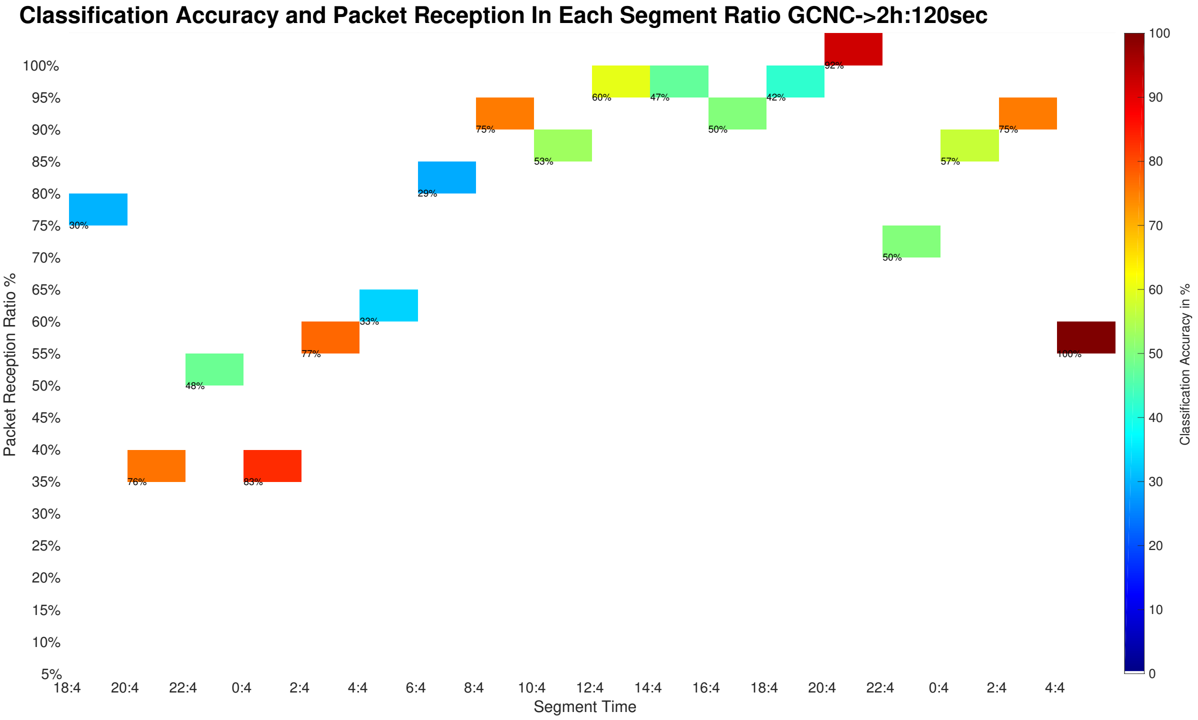

Figure 4.40: Classification Accuracy and Packet Reception Ratio in each segment for GCNC feature selection algorithm

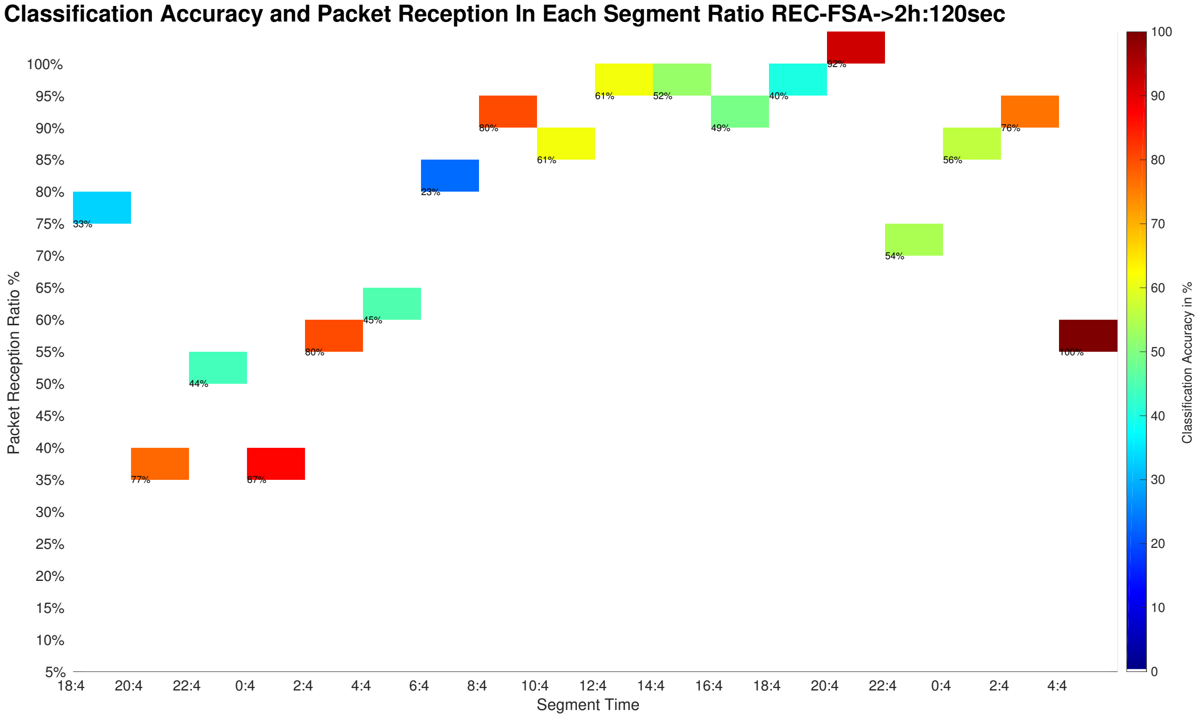

Figure 4.41: Classification Accuracy and Packet Reception Ratio in each segment for REC-FSA feature selection algorithm

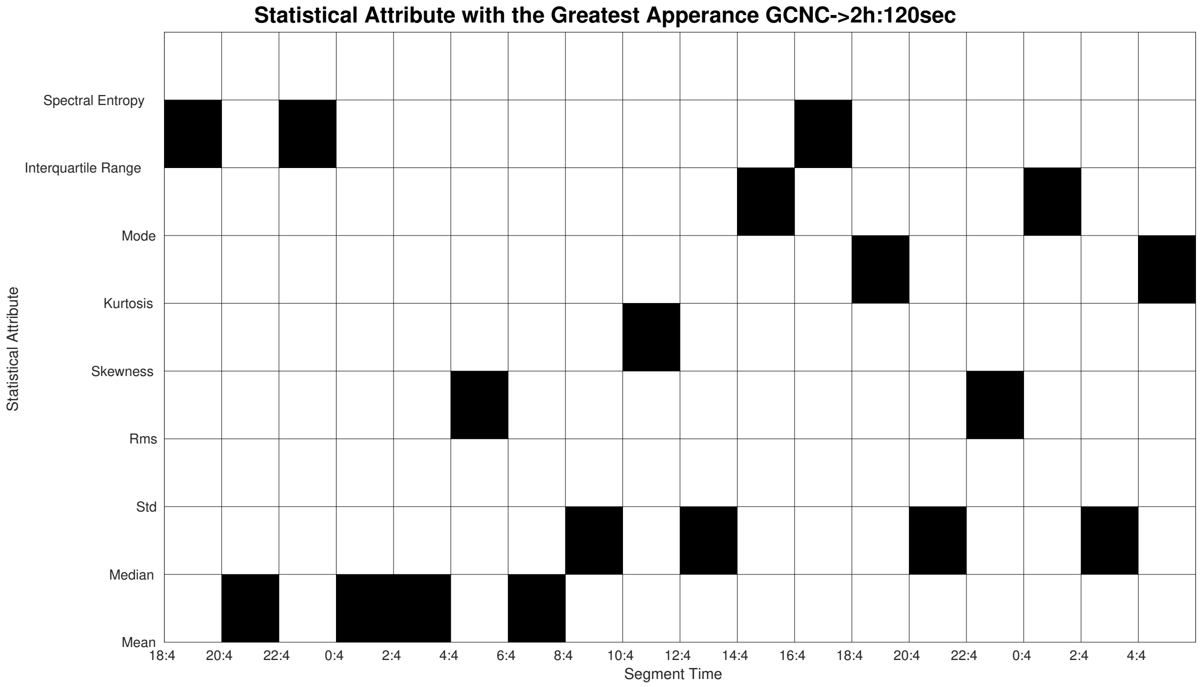

Figure 4.42: Statistical with the highest appearance for GCNC feature selection algorithm

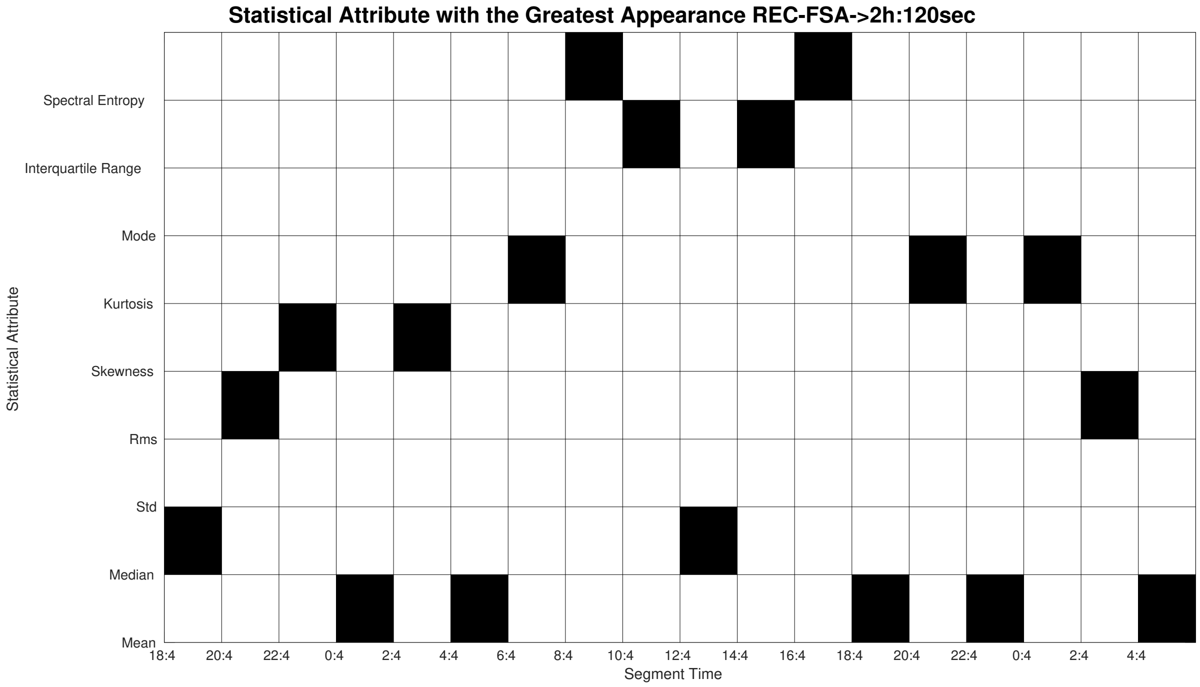

Figure 4.43: Statistical with the highest appearance for REC-FSA feature selection algorithm

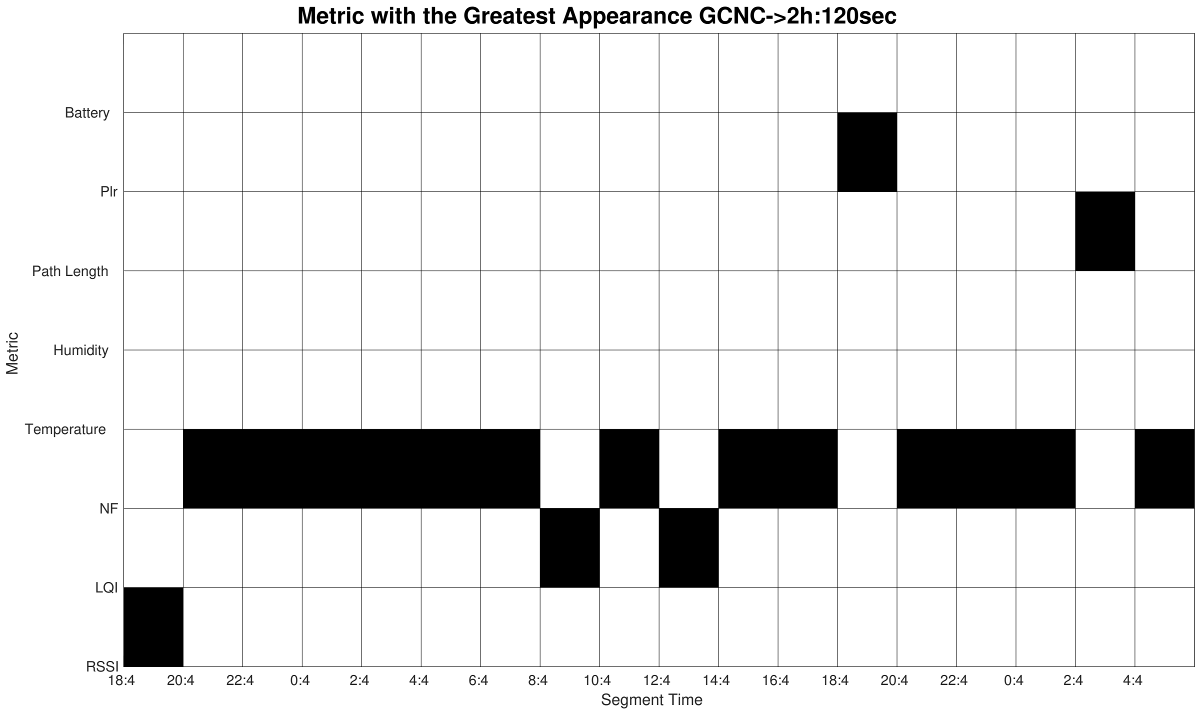

Figure 4.44: Metric with the highest appearance for each feature selection algorithm

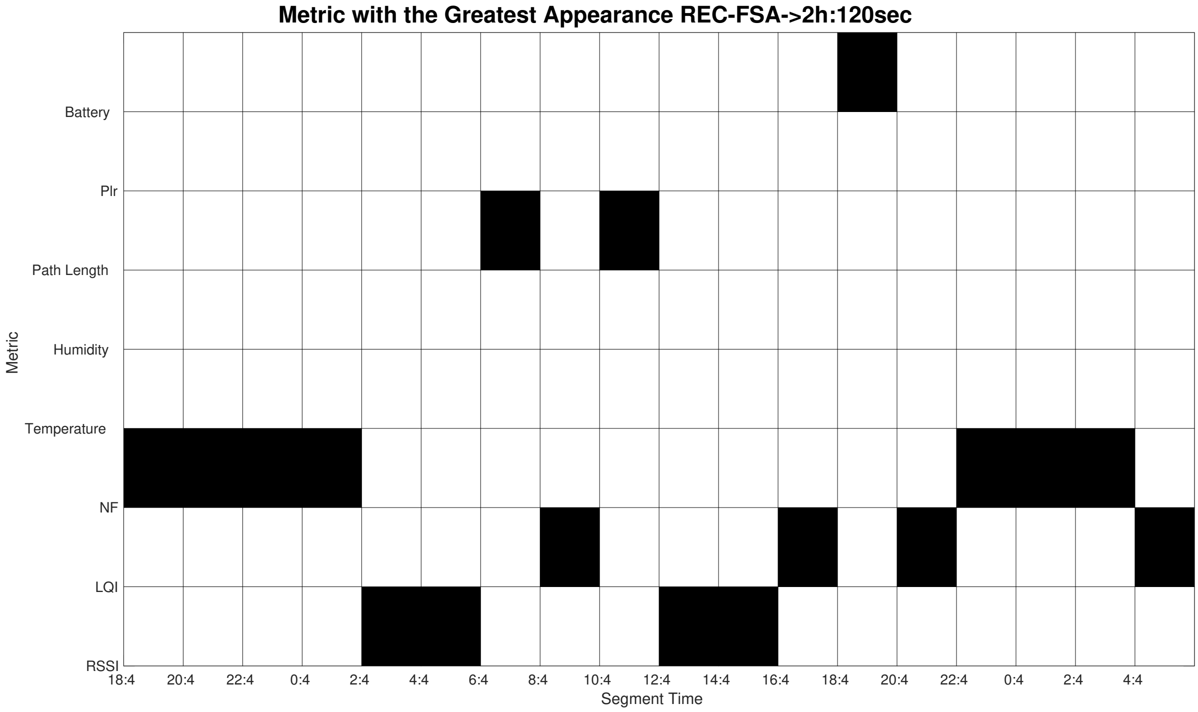

Figure 4.45: Metric with the highest appearance for each feature selection algorithm

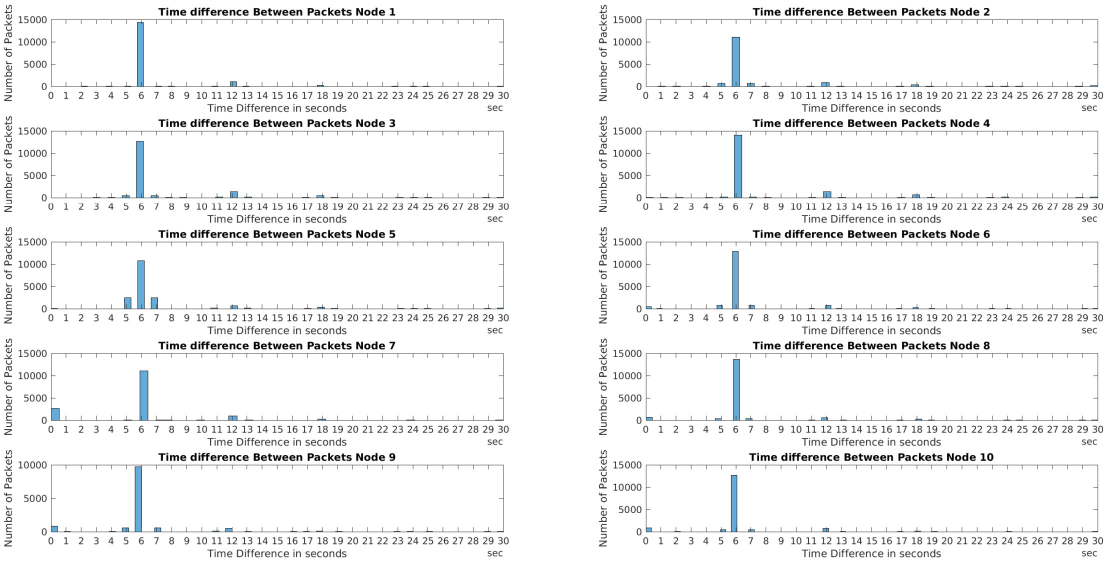

Figure 4.46: Time difference of packets for each sensor from University's Experiment

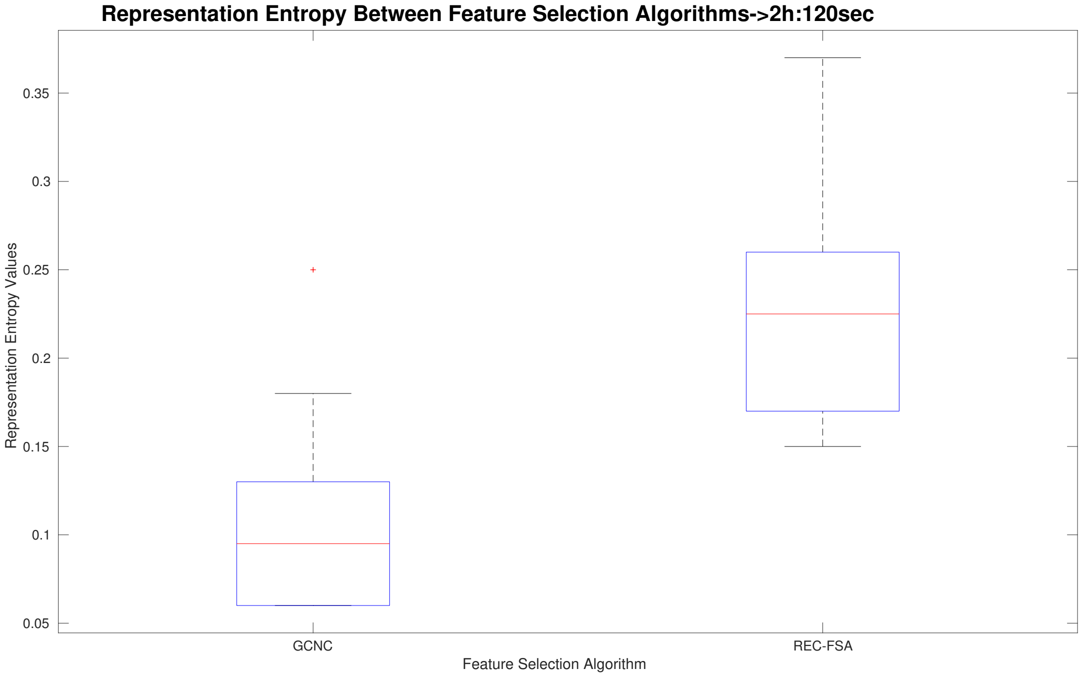

Figure 4.47: Representation Entropy box plot for each feature selection algorithm

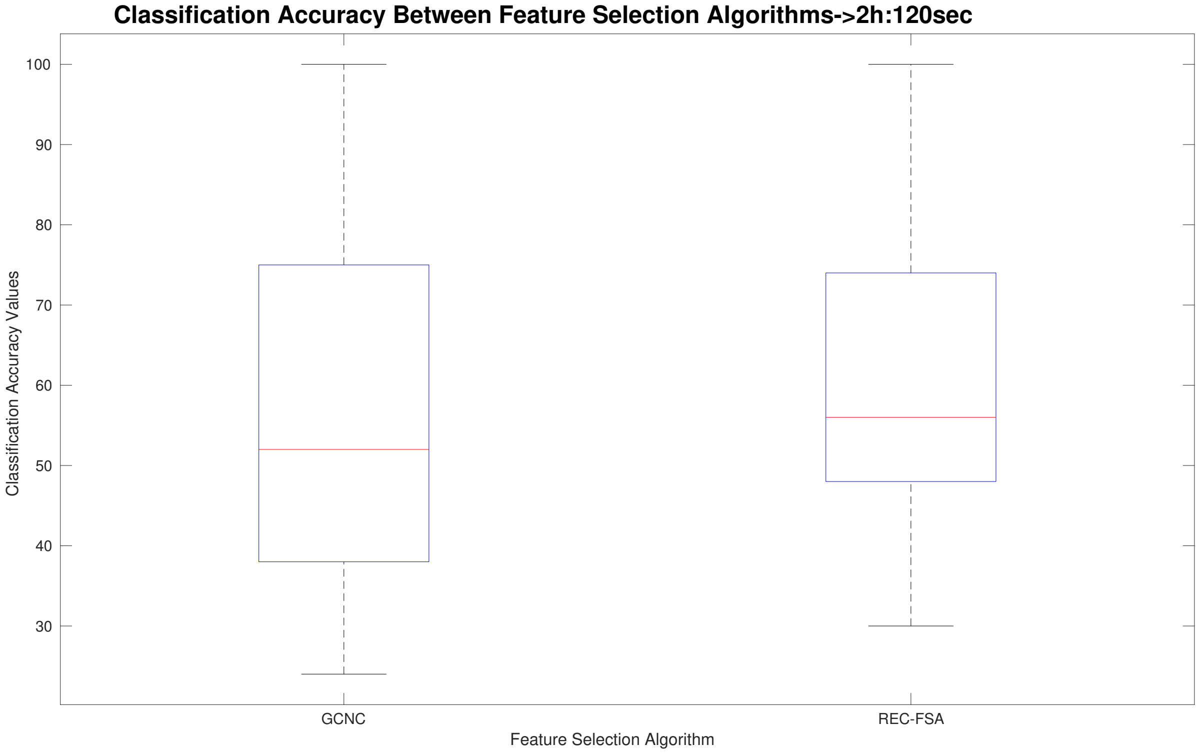

Figure 4.48: Classification Accuracy box plot for each feature selection algorithm

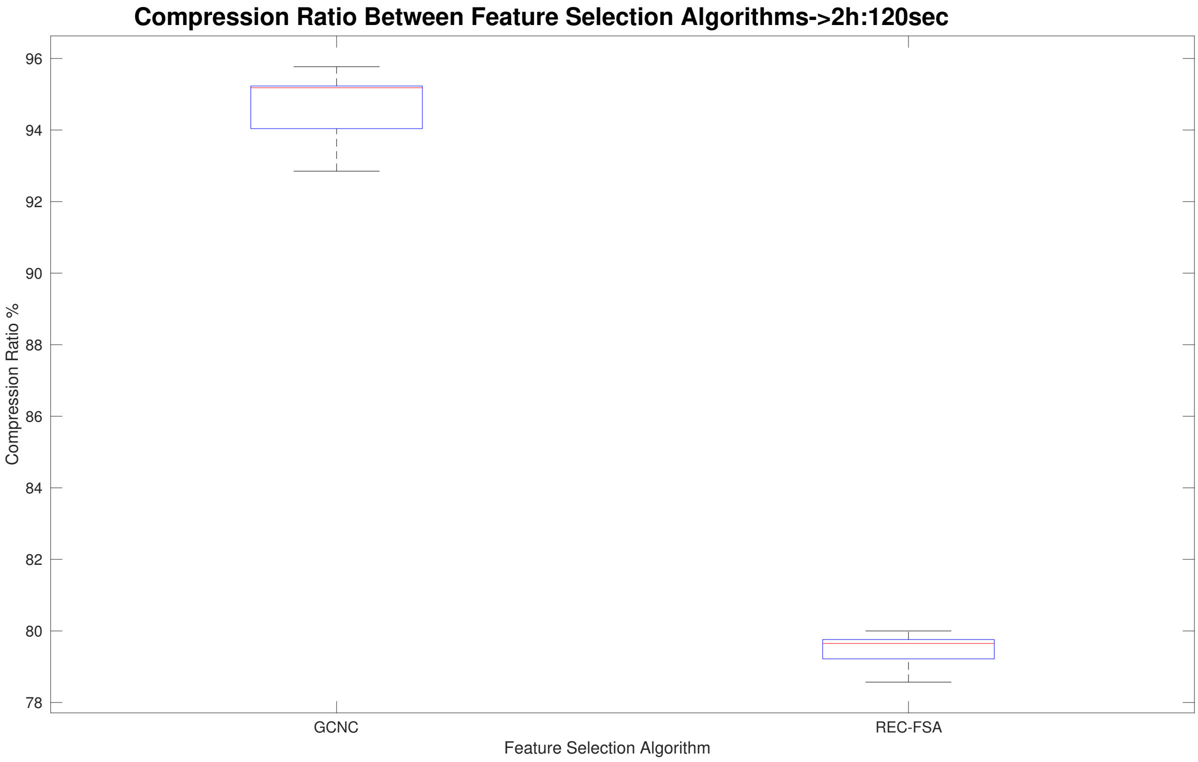

Figure 4.49: Compression Ratio for each feature selection algorithm

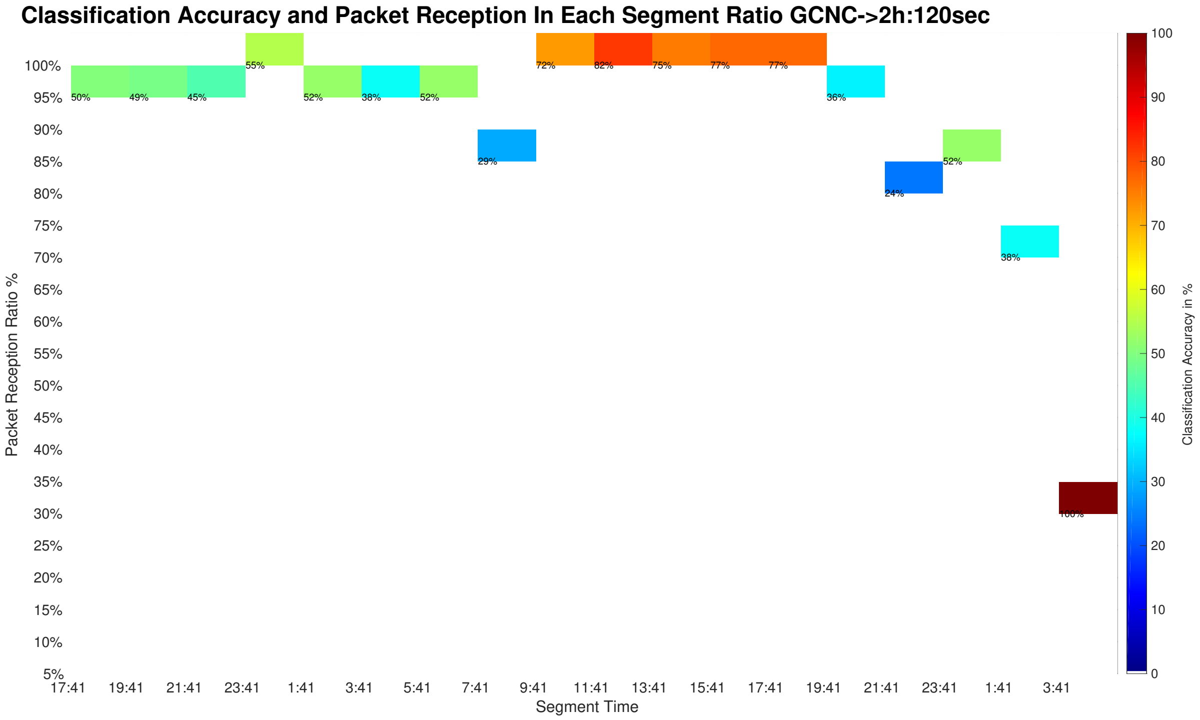

Figure 4.50: Classification Accuracy and Packet Reception Ratio in each segment for GCNC feature selection algorithm

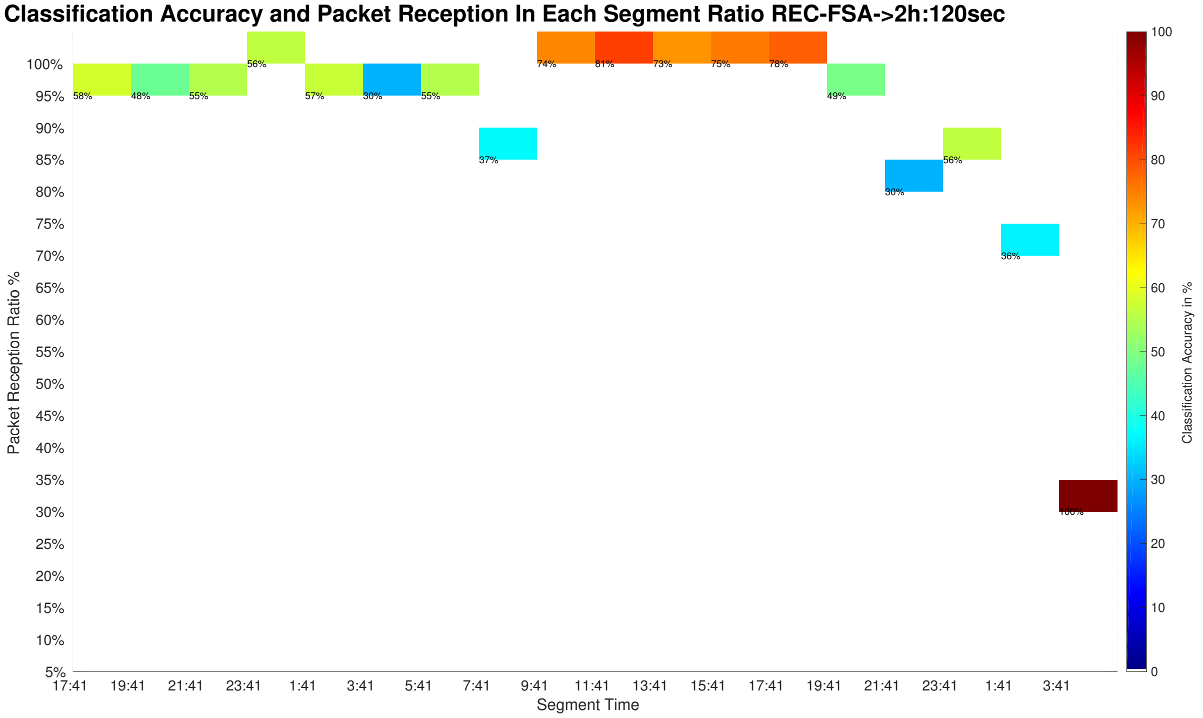

Figure 4.51: Classification Accuracy and Packet Reception Ratio in each segment for REC-FSA feature selection algorithm

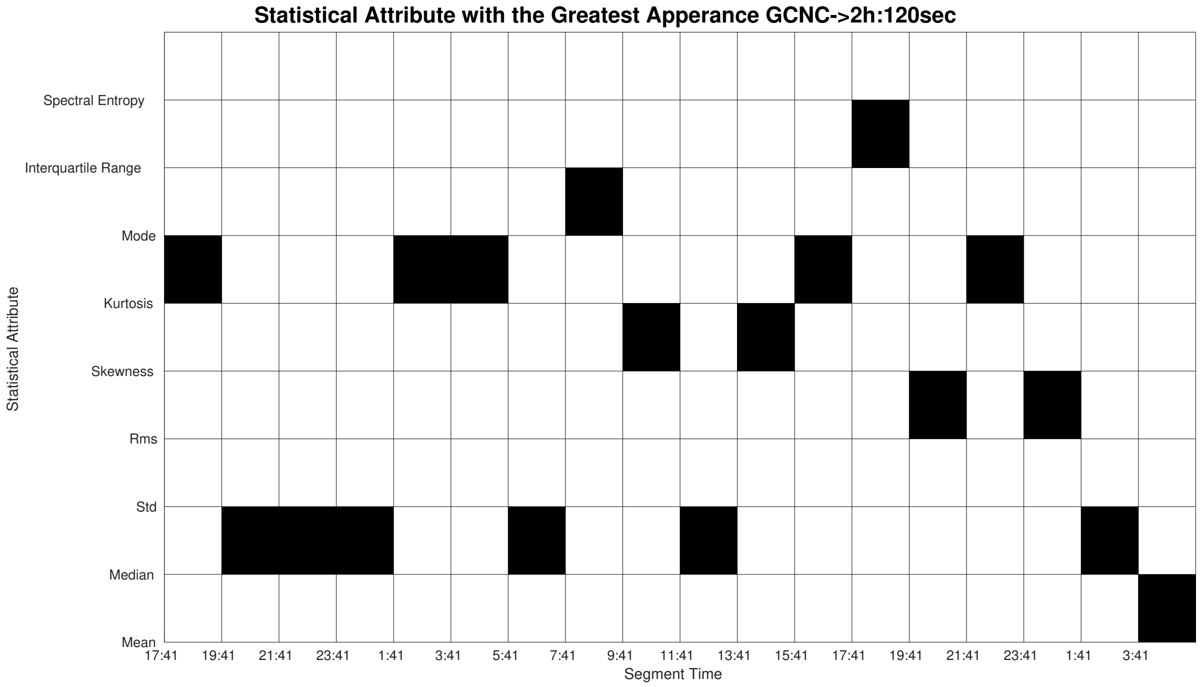

Figure 4.52: Statistical with the highest appearance for GCNC feature selection algorithm

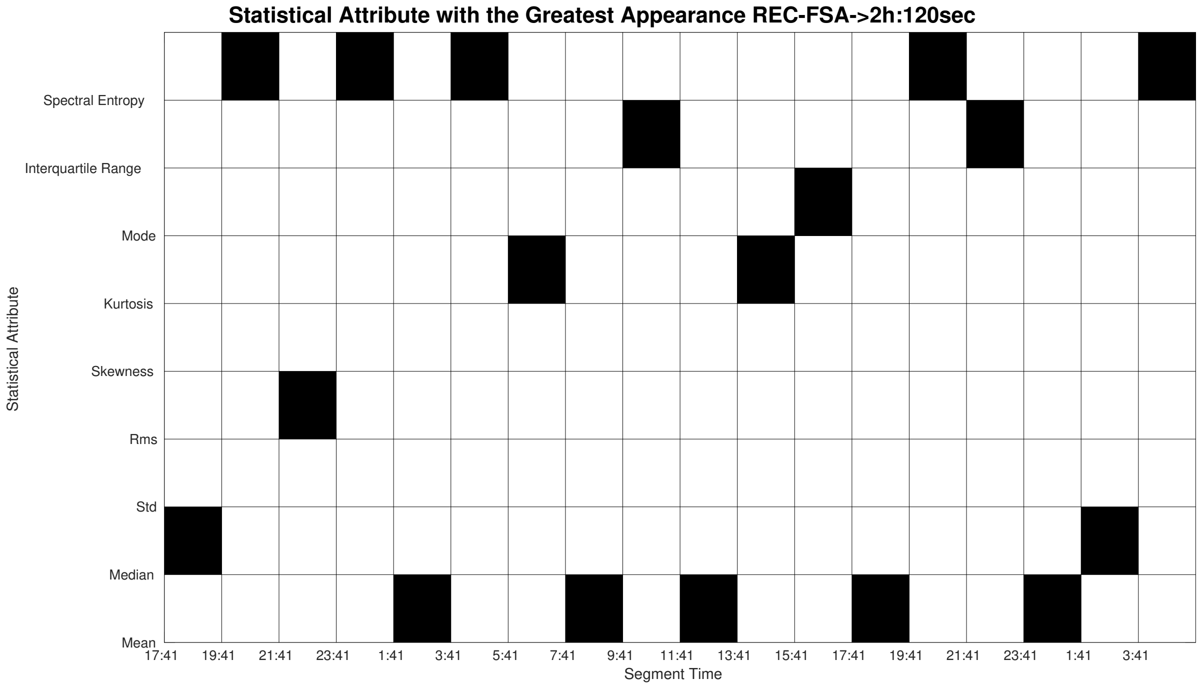

Figure 4.53: Statistical with the highest appearance for REC-FSA feature selection algorithm

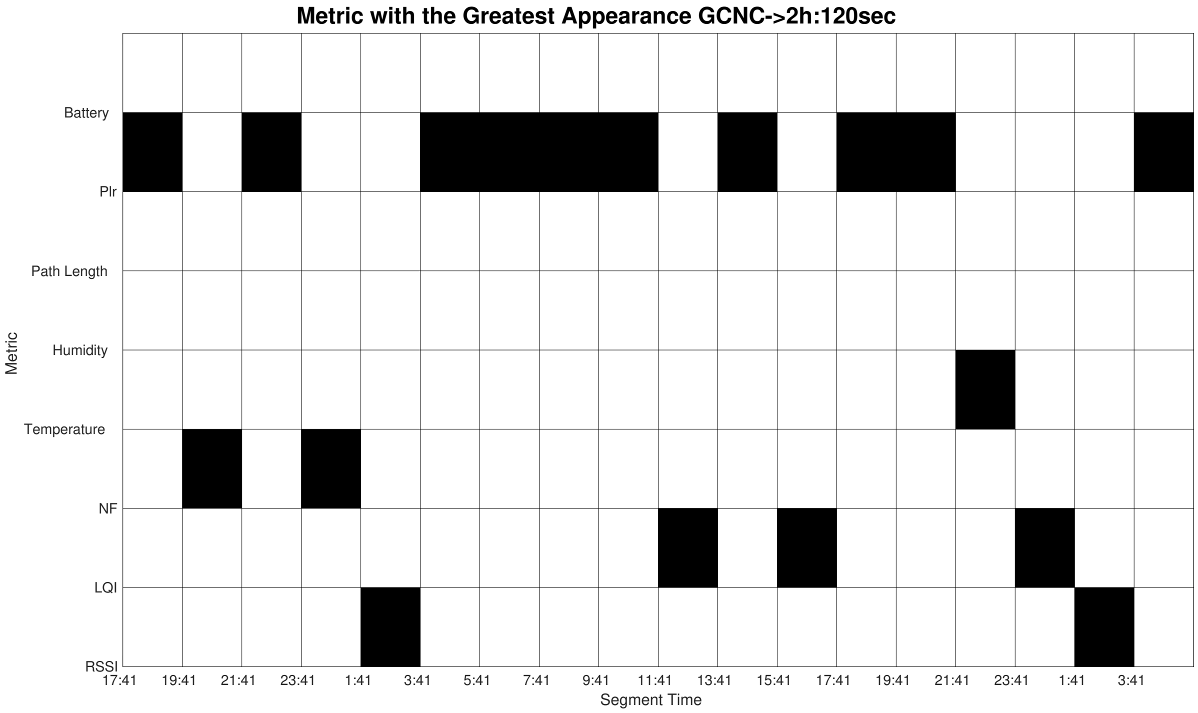

Figure 4.54: Metric with the highest appearance for each feature selection algorithm

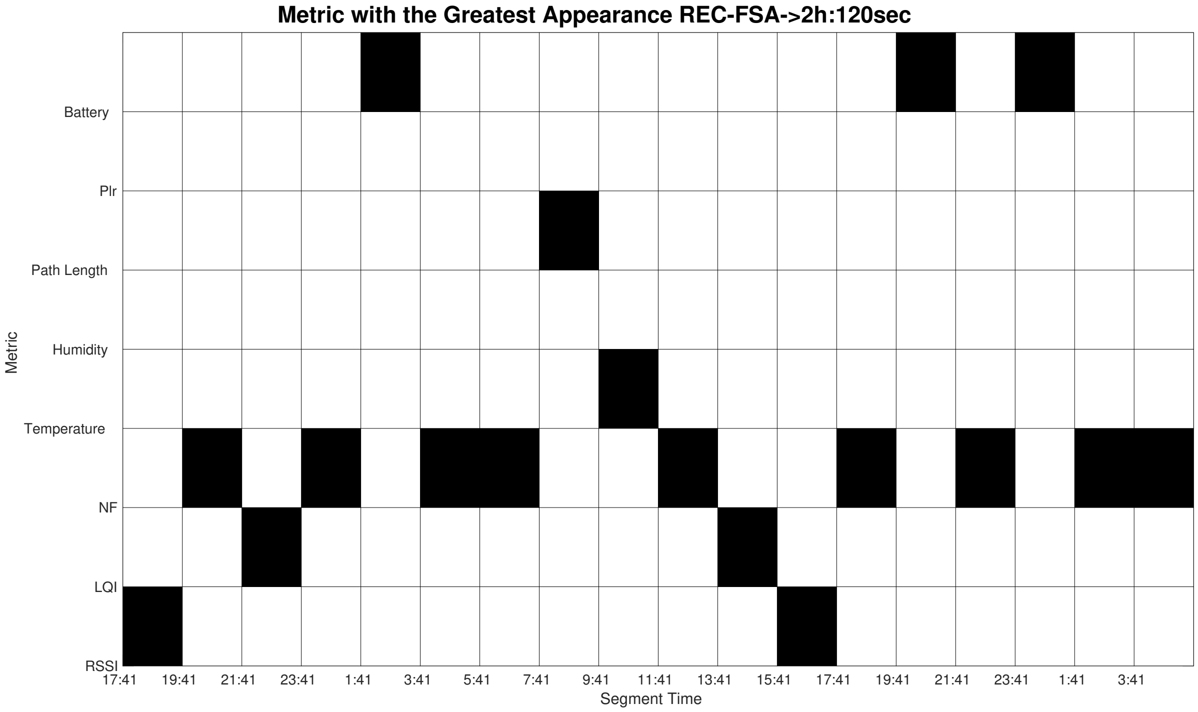

Figure 4.55: Metric with the highest appearance for each feature selection algorithm

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This thesis focused on the problem of unsupervised feature selection for performance characterization of a WSN without human intervention. Previous works have tried different approaches which were based exclusively on information of point-to-point links. Those studies can provide only partial representation of the network's performance. Machine learning approaches that were adopted by those studies demand the training of a classifier in order to predict link-quality performance which leads to network overhead. Our proposed system adopted an unsupervised approach where performance characterization is done via robust and low complexity learning methods which don't require a classifier to be trained. As a result, an efficient mechanism was developed which can run on an Odroid U3+ single board computer [28]. Our objective was to demonstrate the performance of end-to-end links over a multi-hop network via unsupervised selection of dominant features that have crucial impact. The system was tested on data gathered from a WSN that operated in a desalination plant in the framework of the Hydrobionets Project (dataset) for the offline mode. In online mode data was parsed to the system from a real-time operating WSN that was deployed inside the University's of Crete classrooms. Both experiments lead to several important observations. The deployment of our real-time WSN has shown that sampling rate of sensors is a factor that affects the performance of dominant feature selection. Specifically performance was increased when the sampling rate was set to 6 seconds instead of 18. In offline mode the extracted information had better quality in terms of representation entropy than the online, however the deployments can't be compared due to different hardware and environmental setup.

5.2 Future Work

The extensions of this work are numerous. Deployments of the WSN in this work could be done in various environments (indoor or outdoor) to check the performance characterization for various conditions. Since an analysis was held to the dominant characteristics of a mutli-hop sensor network there could be actions that lead to better network performance. Therefore a decision making component could provide the mean to change the network characteristics by adapting the WSN routing protocol and communication.

The system is based on unsupervised feature selection methods. A way to improve the network's dominant feature extraction is based on the increment of unsupervised feature selection methods by 1 or 2. The results for each feature selection algorithm will be compared for each segment and the best ones are parsed to the decision making component in order to adjust the network accordingly.

This work has focused on a low cost and extendable implementation. Certainly, since there is a growth of technology in various parts of human life, from Internet of Things to Smart Cities and Smart Water implementations, it could be a part of those applications. It could constitute a Smart Water implementation and monitor the network operation and how it is affected though several conditions (i.e soil, temperature and humidity parameters) especially in remote places where the electricity infrastructure is incomplete and only battery operated network could be deployed.

Bibliography

[1] Kannan Srinivasan, Prabal Dutta, Arsalan Tavakoli, and Philip Levis. An empirical study of low-power wireless. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):16, 2010.

[2] Jianyu Miao and Lingfeng Niu. A survey on feature selection. *Procedia Computer Science*, 91:919–926, 2016.

[3] Athanasia Panousopoulou, Mikel Azkune, and Panagiotis Tsakalides. Feature selection for performance characterization in multi-hop wireless sensor networks. *Ad Hoc Networks*, 49:70–89, 2016.

[4] Susanna Spinsante, Mirco Pizzichini, Matteo Mencarelli, Stefano Squartini, and Ennio Gambi. Evaluation of the wireless m-bus standard for future smart water grids. In *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1382–1387. IEEE, 2013.

[5] Brendan O'Flynn, Rafael Martinez-Catala, Sean Harte, C O'Mathuna, John Cleary, C Slater, F Regan, Dermot Diamond, and Heather Murphy. Smartcoast: a wireless sensor network for water quality monitoring. In *32nd IEEE Conference on Local Computer Networks (LCN 2007)*, pages 815–816. IEEE, 2007.

[6] Vehbi C Gungor, Bin Lu, and Gerhard P Hancke. Opportunities and challenges of wireless sensor networks in smart grid. *IEEE transactions on industrial electronics*, 57(10):3557–3564, 2010.

[7] Melike Erol-Kantarci and Hussein T Mouftah. Wireless sensor networks for cost-efficient residential energy management in the smart grid. *IEEE Transactions on Smart Grid*, 2(2):314–325, 2011.

[8] Melike Erol-Kantarci and Hussein T Mouftah. Suresense: sustainable wireless rechargeable sensor networks for the smart grid. *IEEE Wireless Communications*, 19(3):30–36, 2012.

[9] Ashraf Darwish and Aboul Ella Hassanien. Wearable and implantable wireless sensor network solutions for healthcare monitoring. *Sensors*, 11(6):5561–5595, 2011.

[10] Gilles Virone, A Wood, Leo Selavo, Quihua Cao, Lei Fang, Thao Doan, Zhimin He, and J Stankovic. An advanced wireless sensor network for health monitoring. In *Transdisciplinary conference on distributed diagnosis and home healthcare (D2H2)*, pages 2–4. Citeseer, 2006.

[11] Nouha Baccour, Anis Koubâa, Claro Noda, Hossein Fotouhi, Mário Alves, Habib Youssef, Marco Antonio Zúñiga, Carlo Alberto Boano, Kay Roemer, Daniele Puccinelli, et al. *Radio link quality estimation in low-power wireless networks.* Springer, 2013.

[12] Carlo Alberto Boano, Marco Antonio Zúniga, Thiemo Voigt, Andreas Willig, and Kay Romer. The triangle metric: Fast link quality estimation for mobile wireless sensor networks. In *2010 Proceedings of 19th International Conference on Computer Communications and Networks*, pages 1–7. IEEE, 2010.

[13] Yong Wang, Margaret Martonosi, and Li-Shiuan Peh. Predicting link quality using supervised learning in wireless sensor networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 11(3):71–83, 2007.

[14] Tao Liu and Alberto E Cerpa. Data-driven link quality prediction using link features. *ACM Transactions on Sensor Networks (TOSN)*, 10(2):37, 2014.

[15] Dana Marinca and Pascale Minet. On-line learning and prediction of link quality in wireless sensor networks. In *2014 IEEE Global Communications Conference*, pages 1245–1251. IEEE, 2014.

[16] Tao Liu and Alberto E Cerpa. Temporal adaptive link quality prediction with online learning. *ACM Transactions on Sensor Networks (TOSN)*, 10(3):46, 2014.

[17] Gianni A Di Caro, Michal Kudelski, Eduardo Feo Flushing, Jawad Nagi, Imran Ahmed, and Luca M Gambardella. Online supervised incremental learning of link quality estimates in wireless networks. In *2013 12th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, pages 133–140. IEEE, 2013.

[18] Geoffrey Werner-Allen, Patrick Swieskowski, and Matt Welsh. Motelab: A wireless sensor network testbed. In *Proceedings of the 4th international symposium on Information processing in sensor networks*, page 68. IEEE Press, 2005.

[19] Nouha Baccour, Anis Koubâa, Maissa Ben Jamâa, Denis Do Rosario, Habib Youssef, Mário Alves, and Leandro B Becker. Radiale: A framework for designing and assessing link quality estimators in wireless sensor networks. *Ad Hoc Networks*, 9(7):1165–1185, 2011.

[20] Manjunath Doddavenkatappa, Mun Choon Chan, and Akkihebbal L Ananda. Indriya: A low-cost, 3d wireless sensor network testbed. In *International*

conference on testbeds and research infrastructures, pages 302–316. Springer, 2011.

[21] Tom Mitchell, Bruce Buchanan, Gerald DeJong, Thomas Dietterich, Paul Rosenbloom, and Alex Waibel. Machine learning. *Annual review of computer science*, 4(1):417–433, 1990.

[22] Patrick Knab, Martin Pinzger, and Abraham Bernstein. Predicting defect densities in source code files with decision tree learners. In *Proceedings of the 2006 international workshop on Mining software repositories*, pages 119–125. ACM, 2006.

[23] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24, 2007.

[24] Sethu Vijayakumar and Stefan Schaal. Locally weighted projection regression: An o (n) algorithm for incremental real time learning in high dimensional space. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, volume 1, pages 288–293, 2000.

[25] David G Kleinbaum, K Dietz, M Gail, Mitchel Klein, and Mitchell Klein. *Logistic regression*. Springer, 2002.

[26] Robert J Schalkoff. *Artificial neural networks*, volume 1. McGraw-Hill New York, 1997.

[27] Xuedong Liang, Ilangko Balasingham, and Sang-Seon Byun. A reinforcement learning based routing protocol with qos support for biomedical sensor networks. In *2008 First International Symposium on Applied Sciences on Biomedical and Communication Technologies*, pages 1–5. IEEE, 2008.

[28] Odroid U3+. `https://www.hardkernel.com/`.

[29] Lior Wolf and Amnon Shashua. Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach. *Journal of Machine Learning Research*, 6(Nov):1855–1887, 2005.

[30] Zheng Zhao and Huan Liu. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th international conference on Machine learning*, pages 1151–1157. ACM, 2007.

[31] Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding. Efficient and robust feature selection via joint 2, 1-norms minimization. In *Advances in neural information processing systems*, pages 1813–1821, 2010.

[32] Jianping Li, Zhenyu Chen, Liwei Wei, Weixuan Xu, and Gang Kou. Feature selection via least squares support feature machine. *International Journal of Information Technology & Decision Making*, 6(04):671–686, 2007.

[33] Zheng Zhao and Huan Liu. Semi-supervised feature selection via spectral analysis. In *Proceedings of the 2007 SIAM international conference on data mining*, pages 641–646. SIAM, 2007.

[34] Zenglin Xu, Irwin King, Michael Rung-Tsong Lyu, and Rong Jin. Discriminative semi-supervised feature selection via manifold regularization. *IEEE Transactions on Neural networks*, 21(7):1033–1047, 2010.

[35] Pin Wang, Yongming Li, Bohan Chen, Xianling Hu, Jin Yan, Yu Xia, and Jie Yang. Proportional hybrid mechanism for population based feature selection algorithm. *International Journal of Information Technology & Decision Making*, 16(05):1309–1338, 2017.

[36] Martin HC Law, Mario AT Figueiredo, and Anil K Jain. Simultaneous feature selection and clustering using mixture models. *IEEE transactions on pattern analysis and machine intelligence*, 26(9):1154–1166, 2004.

[37] Eric P Xing and Richard M Karp. Cliff: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts. *Bioinformatics*, 17(suppl_1):S306–S315, 2001.

[38] Pavel Pudil, J Novovičová, N Choakjarernwanit, and Josef Kittler. Feature selection based on the approximation of class densities by finite mixtures of special type. *Pattern Recognition*, 28(9):1389–1398, 1995.

[39] Kenji Kira and Larry A Rendell. A practical approach to feature selection. In *Machine Learning Proceedings 1992*, pages 249–256. Elsevier, 1992.

[40] Laura Elena Raileanu and Kilian Stoffel. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41(1):77–93, 2004.

[41] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.

[42] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (8):1226–1238, 2005.

[43] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[44] Parham Moradi and Mehrdad Rostami. A graph theoretic approach for unsupervised feature selection. *Engineering Applications of Artificial Intelligence*, 44:33–45, 2015.

[45] Md Monirul Kabir, Md Shahjahan, and Kazuyuki Murase. A new local search based hybrid genetic algorithm for feature selection. *Neurocomputing*, 74(17):2914–2928, 2011.

[46] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[47] V Madhusudan Rao and VN Sastry. Unsupervised feature ranking based on representation entropy. In *2012 1st International Conference on Recent Advances in Information Technology (RAIT)*, pages 421–425. IEEE, 2012.

[48] Pabitra Mitra, CA Murthy, and Sankar K. Pal. Unsupervised feature selection using feature similarity. *IEEE transactions on pattern analysis and machine intelligence*, 24(3):301–312, 2002.

[49] Karl Benkic, Marko Malajner, P Planinsic, and Z Cucej. Using rssi value for distance estimation in wireless sensor networks based on zigbee. In *2008 15th International Conference on Systems, Signals and Image Processing*, pages 303–306. IEEE, 2008.

[50] RM Page-Jones. Notes on the rsgb observations of the hf ambient noise floor. *Radio Soc. Great Britain, Bedford, UK*, 2003.

[51] Yan Zhang, Songwei Fu, Yuming Jiang, Matteo Ceriotti, Markus Packeiser, and Pedro José Marrón. An lqi-based packet loss rate model for ieee 802.15. 4 links. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2018.

[52] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 986–996. Springer, 2003.

[53] Zolertia Z1 DataSheet. `http://zolertia.sourceforge.net/wiki/images/e/e8/Z1_RevC_Datasheet.pdf`.