A HYBRID SWARM INTELLIGENCE ALGORITHM FOR SIMULTANEOUS FEATURE SELECTION AND CLUSTERING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HASAN GEREN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

JUNE 2022

Approval of the thesis:

A HYBRID SWARM INTELLIGENCE ALGORITHM FOR SIMULTANEOUS FEATURE SELECTION AND CLUSTERING

submitted by **HASAN GEREN** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Esra Karasakal
Head of Department, **Industrial Engineering** ______________

Prof. Dr. Nur Evin Özdemirel
Supervisor, **Industrial Engineering, METU** ______________

Examining Committee Members:

Prof. Dr. Cem İyigün
Industrial Engineering, METU ______________

Prof. Dr. Nur Evin Özdemirel
Industrial Engineering, METU ______________

Prof. Dr. Pınar Karagöz
Computer Engineering, METU ______________

Prof. Dr. Murat Caner Testik
Industrial Engineering, Hacettepe University ______________

Assist. Prof. Dr. Mustafa Kemal Tural
Industrial Engineering, METU ______________

Date: 20.06.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Hasan Geren

Signature :

ABSTRACT

A HYBRID SWARM INTELLIGENCE ALGORITHM FOR SIMULTANEOUS FEATURE SELECTION AND CLUSTERING

Geren, Hasan
M.S., Department of Industrial Engineering
Supervisor: Prof. Dr. Nur Evin Özdemirel

JUNE 2022, 149 pages

In this study, we address the feature selection and clustering problems by using a hybrid swarm intelligence approach. We assume that the number of clusters is known, clusters can be of any shape and have different densities, but there are no outliers or noise. The data set may have high dimensionality and redundant features.

We propose a swarm intelligence algorithm, namely ACOVNS, which is a hybridization of Ant Colony Optimization (ACO) and Variable Neighborhood Search (VNS). We utilize the ACO mechanisms for exploration and enhance its exploitation capability by combining it with VNS. In addition to pheromone values, we make use of some heuristic information to further improve the performance of the algorithm. In the first part of our study, we use our algorithm with an objective function based on the sum of Euclidean distances to solve the clustering problem.

In the second part, we modify the ACOVNS algorithm as F-ACOVNS to perform feature selection and clustering simultaneously. We propose a novel heuristic information that employs the Laplacian Score (LS) and a second pheromone matrix for feature selection. Therefore, the algorithm selects features during clustering by using

distinct pheromone matrices and heuristic information.

Our proposed algorithms are unique in that ACOVNS is the first hybridization of ACO and VNS for clustering and F-ACOVNS is the first algorithm that uses LS as heuristic information. We compared the performance of ACOVNS with some well-known algorithms on nine real-world data sets. For simultaneous feature selection and clustering, we compared F-ACOVNS with known single and multi-objective algorithms using both real and synthetic data sets.

Keywords: Clustering, Feature Selection, Swarm Intelligence, Ant Colony Optimization, Variable Neighborhood Search

ÖZ

EŞZAMANLI ÖZNİTELİK SEÇİMİ VE KÜMELEME İÇİN HİBRİT BİR SÜRÜ ZEKASI ALGORİTMASI

Geren, Hasan
Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Nur Evin Özdemirel

Haziran 2022 , 149 sayfa

Bu çalışmada, bir hibrit sürü zekası yaklaşımı kullanarak kümeleme problemi ele alınmıştır. Yaklaşımımızda, küme sayısının bilindiği, kümelerin herhangi bir şekile ve farklı yoğunluklara sahip olabileceği, ancak aykırı değer veya gürültünün olmadığı varsayılmıştır. Veri seti yüksek boyutlu olabilir ve gereksiz özniteliklere sahip olabilir.

Çalışmada, Karınca Kolonisi Optimizasyonu (ACO) ve Değişken Komşuluk Aramasından (VNS) oluşan ACOVNS adlı bir sürü zekası algoritması önerilmiştir. Keşif yeteneği için ACO'nun mekanizmaları kullanılmış, ACO'nun yerel arama kapasitesi ise VNS ile artırılmıştır. Algoritmanın performansını daha da iyileştirmek için, feromon değerlerine ek olarak bir sezgisel bilgiye başvurulmuştur. Çalışmanın ilk bölümünde, algoritmada amaç fonksiyonu olarak Öklik uzaklıklarının toplamı kullanılarak kümeleme problemi çözülmüştür.

İkinci bölümde, öznitelik seçimi ve kümelemeyi eşzamanlı olarak gerçekleştirmek için ACOVNS algoritması F-ACOVNS olarak modifiye edilmiştir. Öznitelik seçimi için Laplacian Skorunu (LS) kullanan yeni bir sezgisel bilgi ve ikinci bir feromon

matrisi önerilmiştir. Böylece, algoritma kümeleme sırasında farklı feromon matrisleri ve sezgisel bilgiler kullanarak öznitelik de seçer.

ACOVNS kümeleme için ACO ve VNS'nin ilk hibridizasyonu olması, F-ACOVNS ise LS'yi sezgisel bilgi olarak kullanan ilk algoritma olması bakımından benzersizdir. ACOVNS'nin performansı, dokuz gerçek dünya veri setinde bazı bilinen algoritmalarla karşılaştırılmıştır. Eşzamanlı öznitelik seçimi ve kümeleme için F-ACOVNS hem gerçek hem de sentetik veri kümeleri kullanılarak bilinen tek ve çok amaç fonksiyonlu algoritmalarla karşılaştırılmıştır.

Anahtar Kelimeler: Kümeleme, Öznitelik Seçimi, Sürü Zekası, Karınca Kolonisi Optimizasyonu, Değişken Komşuluk Arama

To the future

ACKNOWLEDGMENTS

First of all, I would like to express my special appreciation and thanks to Prof. Nur Evin Özdemirel for her guidance, encouragement, and positive attitude. She has been very kind and helpful. Without her, I wouldn't be able to complete this study.

I would also like to thank to examining committee members Prof. Dr. Cem İyigün, Prof. Dr. Pınar Karagöz, Prof. Dr. Murat Caner Testik, and Assist. Prof. Dr. Mustafa Kemal Tural for their time in reviewing this work and valuable comments.

I would like to specially thank my friends Aydın Emre Yılmaz, Gizem Doğan Yılmaz and Sabri Tufan Erkul for always cheering and motivating me with their presence. I also thank Ersin Telemeci and Çiya Aydoğan for being very kind companions during the courses we took together.

Most importantly, I offer my greatest thanks to Melike Benan Altay to be with me all the time with love and patience. Her contribution to my life is enormous. Melike is my best friend, closest advisor, and life partner. She supported me every step of the way with humor and great wisdom.

Last but not least, special thanks go to my parents Selma and Hüseyin Geren, and my brothers Hakan and Yunus Emre Geren for their endless support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

The rapid increase in the amount of data has been an important issue for the last few decades. This growing data is difficult to handle and analyze. Therefore, data mining has become very important in the processing of data. Data mining is the process of extracting useful information and detecting hidden patterns from data (Arora and Chana, 2014). Data mining has a lot of applications in different areas such as future healthcare, market analysis, intrusion detection, bioinformatics, and computer science. Since each usage area has various needs and these needs have various difficulties, it is very important to develop problem specific techniques.

One of the most studied and important data mining methods is clustering (Zou, 2020). The main aim of clustering is to divide a data set into groups where similar objects are in the same group and dissimilar objects are in different groups. Each object in the data set can be placed in only one cluster. Clustering is an unsupervised learning method, which means that, unlike other data mining techniques, clustering does not require any content or category of the data set to be known in advance. Therefore, clustering techniques are essential for data sets where there is little or no information a priori (Pacheco et al., 2018).

Similarity of objects can be calculated by a variety of methods and the method used for the calculation can significantly affect the outcome of the clustering. Data sets can contain only numerical features, only categorical features, or both of them. Also, data sets may contain different types of object distributions. For example, some data sets may have density differences, arbitrary shaped clusters, outliers, or noise. Therefore, the similarity measure should be selected in accordance with the data type and distribution in the data set.

In addition to the similarity measure, the objective function selection is one of the most important decisions in the clustering problem. The objective function determines how the algorithm performs clustering and measures the quality of the clustering solution. Each cluster in the solution should be compact and connected, and different clusters should be well-separated. However, it is difficult to evaluate the performance of the clustering solution in terms of these properties since clustering is an unsupervised method and true class labels are not available. There exist a lot of objective functions used for clustering, but perhaps the most widely used objective function in the literature is the sum of Euclidean distances between each object and the center of the cluster it belongs to.

The number of clusters in the data set is another challenge in the clustering problem. Some algorithms try to find the number of clusters in the data set by using specific encoding schemes or objective functions while others take the number of clusters as an input parameter. Given the incorrect number of clusters, the algorithm may come up with an undesired solution.

Due to the wide variety of challenging properties of the clustering problem, there is no similarity measure, objective function, or algorithm that has been completely accepted as superior in the literature.

Based on the various features of the clustering problem, in this study, we focus on the clustering problem with the following properties.

- All sample features in the data set are numerical.
- The number of clusters is known.
- Data set may have arbitrary shaped clusters and density differences within and between clusters
- No outliers or noise are considered explicitly.

In reality, the data sets may contain redundant or even misleading features. These features may prevent the clustering algorithm from finding the correct clustering solution. This has brought about the problem of selecting features that are useful in clustering.

The majority of the studies in the literature consider the feature selection and clustering problems separately. They focus on either feature selection or clustering. Some of them first carry out feature selection independently and then use the selected features in their clustering algorithm. Only a limited number of studies perform feature selection and clustering simultaneously.

In this study, we first address the clustering problem. Then, we extend our problem to tackle feature selection and clustering simultaneously.

In the first part, a well-known swarm intelligence algorithm Ant Colony Optimization (ACO) is hybridized with Variable Neighbourhood Search (VNS) and the ACOVNS algorithm is proposed. ACO imitates the foraging behavior of ant colonies by depositing pheromone on the solutions based on their quality. Sample-to-cluster assignments in higher quality solutions receive more pheromone deposit and are favored in solution construction. In addition, some heuristic information on individual sample-to-cluster assignments contributes to selection of these solution components in constructing complete solutions. On the other hand, VNS searches for solution improvements using variable neighborhood sizes. The primary aim of this hybridization is to improve ACO's exploitation ability by using VNS as a local search mechanism. ACOVNS is the first hybridization of ACO and VNS algorithms for the clustering problem. In the algorithm, the similarity measure is the Euclidean distance and the objective function is the sum of Euclidean distances between each sample and the center of the cluster that the sample is assigned to. The heuristic information concerning assignment of a sample to a cluster is based on.

In the second part, the scope of the problem is enlarged to simultaneous clustering and feature selection. The clustering algorithm is modified to solve the two problems together. The clustering part of the algorithm remains the same. However, a new feature pheromone vector and heuristic information for feature selection are introduced to the algorithm to detect the redundant or irrelevant features. The heuristic information for feature selection is calculated by a well-known unsupervised feature scoring technique named Laplacian Score (LS) and given to the algorithm as an input. Moreover, the feature pheromone vector stores the pheromone values of each feature, which are updated according to the objective function values of the solutions where the features are used. Thus, a modified version of ACOVNS (F-ACOVNS) is proposed for simultaneous feature selection and clustering. This solution approach is unique in that it uses an unsupervised filtering method's output as ACO's heuristic information; this is the first use of ACO and LS together.

Both of the algorithms are applied to well-known real-life clustering data sets and the data sets randomly generated for feature selection. The experimental results show that the clustering performance of ACOVNS is very promising as compared to the findings of commonly used algorithms in the literature. Additionally, F-ACOVNS detects the relevant features accurately and improves the solution quality of ACOVNS where redundant features exist.

The rest of the thesis is organized as follows. In Chapter 2, definitions of the clustering problem and common clustering methods are given. The properties and challenges of the clustering problem are reviewed. Then, solution approaches of the studies in the literature to the clustering problem are examined. Swarm intelligence approaches for clustering are further analyzed and their properties are given in detail. After clustering, the feature selection problem and related methods are discussed. Finally, simultaneous feature selection and clustering is explained and studies on this subject are reviewed.

In Chapter 3, the proposed clustering algorithm (ACOVNS) is explained in detail. The notation used for clustering, formulations, and pseudo code of the algorithm are given. Then, the experiments are discussed and data set properties are given in detail. Lastly, the performance measures used for clustering are explained and the results of ACOVNS are compared with some known algorithms from the literature.

In Chapter 4, the algorithm designed for simultaneous feature selection and clustering (F-ACOVNS) is described. Detailed parameter tuning of the algorithm and experimental results are provided.

Finally, a summary of work done and major findings of this thesis along with future research suggestions are presented in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Clustering

Clustering is a well-known problem in the literature and has been studied for many years. Clustering aims at partitioning a data set into a collection of groups where more similar objects are in the same group and less similar objects are in different groups (Ezugwu et al., 2021).

Clustering is used in a wide variety of fields in real life. Taib and Bahreininejad (2021) list various fields where clustering has been used in recent years as molecular sciences, Internet of things, geo sciences, text analysis, image analysis, medical diagnosis, energy management, social network analysis, and material sciences.

To address the different and challenging properties of the clustering problem, various clustering methods and algorithms are proposed in the literature. Ramadas and Abraham (2019) classify clustering methods into two categories as hierarchical and partitional methods. In the light of this classification, various algorithms are developed. The most common categories of algorithms used for clustering in the literature are partitional algorithms, hierarchical algorithms, and metaheuristic algorithms (Nanda and Panda, 2014). Furthermore, hybrid use of algorithms belonging to different categories is a common approach. The nature of algorithms cause them to focus on either exploration or exploitation. Therefore, when algorithms are hybridized, the solution quality may increase by focusing on both exploration and exploitation (Mageshkumar et al., 2019).

In addition to these categories, the number and nature of objective functions in the metaheuristic algorithms may also vary. Due to the difficulty of identifying the underlying clusters present in a data set by using a single clustering criterion, multiple optimization criteria can be evaluated simultaneously to increase the performance of the clustering algorithm (José-Garcia and Gómez-Flores, 2016).

Following the above categories, partitional clustering algorithms are reviewed in Section 2.1.1, hierarchical clustering algorithms in Section 2.1.2, and swarm intelligence algorithms for clustering in Section 2.1.3. As the number of metaheuristic clustering algorithms is too large, we restrict our focus on those algorithms based on swarm intelligence.

2.1.1 Partitional Clustering Algorithms

In partitional clustering, different clusters are formed by dividing the entire data set and the clustering solution is improved iteratively by the use of a distance based similarity measure. Moreover, partitional clustering can be soft (fuzzy) or hard (crisp). In soft clustering, data points belong to different clusters with certain weights whereas in hard clustering, each data point can be assigned to only one cluster (Prakash et al., 2019).

K-means is the most often used partitional clustering algorithm in the literature. K-means begin clustering by selecting a pre-determined number of cluster centers at random and assigning objects to the cluster centers that are closest to them. In the following iterations, K-means calculates cluster centroids according to the assigned objects and finds new cluster centroids until it converges to a solution (Krishnasamy et al., 2014). The main disadvantage of the K-means is that it has a tendency to converge to local optima and results are dependent on initial values of cluster centroids (Niknam et al., 2011).

Dinh et al. (2021) proposed a partitional clustering algorithm for numerical and categorical data where the data set may contain missing values. They handle missing values with a decision-tree-based imputation method. Also, they perform clustering using the sum of squared Euclidean distances and an information-theoretic dissimilarity measure. Kuwil et al. (2020) introduced Gravity Center Clustering (GCC), a novel approach for partitional clustering that take outliers and noise in the data set into consideration. GCC also uses the Euclidean distances in the objective function. However, because the Euclidean distance is incapable of detecting outliers and noise, the algorithm uses a threshold value computed using the Critical Distance Clustering Algorithm (Kuwil et al., 2019) to detect outliers and noise. Zhu et al. (2019) modified the K-means and introduced a Grid-K-means algorithm in order to decrease dependency on the initial cluster centers and local optima stagnation. Their algorithm combines the K-means and the idea of meshing in grid clustering. Therefore, they convert the data set to a grid space and group the grids rather than the data points.

Stetco et al. (2015) use Fuzzy C-means (FCM) clustering algorithm, which determines a data point's membership in a cluster using a fuzzy membership function rather than a crisp value. Moreover, they adjust the FCM to produce initial cluster centers that are closer to the clusters' true centers. To accomplish this, they select sample points that are spread out in the data set and utilize them as initial cluster centers.

The approaches used and enhancements made in the studies published in the literature have resulted in a significant improvement in the performance of partitional clustering algorithms. However, initial cluster centers, outliers, and noise continue to be a problem for partitional clustering algorithms (Mittal et al., 2021). Prakash and Singh (2013), Saxena et al. (2017), Mehta et al. (2020), and A. Alam et al. (2021) present in-depth reviews of such algorithms.

2.1.2 Hierarchical Clustering Algorithms

Hierarchical methods divide a data set to form a nested tree. Hierarchical approaches are classified into two types: divisive hierarchical clustering and agglomerative hierarchical clustering. In agglomerative hierarchical clustering, each data point is initially assumed to be a cluster and in every iteration two nearest clusters are merged. Hence, clusters are combined based on single linkage. The algorithm continues until only one cluster remains. In divisive clustering, all data points are assumed to be part of a single cluster, which is then divided into smaller clusters using a distance-based metric until the predefined number of clusters is reached (S. Alam et al., 2015). CURE

(Guha et al., 1998), BIRCH (Zhang et al., 1996), and CHAMELEON (Karypis et al., 1999) are the most frequently used hierarchical clustering algorithms in the literature.

Cilibrasi and Vitányi (2005) proposed a new similarity distance, the Normalized Compression distance (NCD), and applied a hierarchical clustering method using pairwise NCDs given in a distance matrix. They form a dendrogram (ternary tree) from the distance matrix to extract hierarchy of the clusters.

Szekely and Rizzo (2005) developed an agglomerative hierarchical clustering approach based on joint between and within cluster distances, in contrast to commonly used procedures that aim to minimize within-cluster distances or maximize between-cluster distances. Their method is Euclidean distance-based rather than the squared Euclidean distance, and they additionally consider any power of the Euclidean distance in the interval (0,2]. They determine a power of Euclidean distance in the provided interval based on the characterization of the clusters in the data set and merge clusters with the smallest distance iteratively.

Zhao et al. (2005) presented an agglomerative hierarchical clustering algorithm for document clustering. Their methodology integrates characteristics of partitional and agglomerative approaches in order to improve the early-stage performance of the agglomerative methods. As a result, the algorithm creates a hierarchical tree for each partitional cluster and then agglomerate these clusters to build the final hierarchical tree.

Son (2016) proposed a hierarchical fuzzy clustering method named Hierarchical Picture Clustering (HPC). Aside from that, he created a new distance measure that is a combination of the Hamming, Euclidean, and Hausdorff distance measures. The goal of the new distance is to improve the performance and accuracy of the existing standalone distances. The method then produces a distance matrix based on the new distance measure and merges two consecutive clusters to find new cluster centers. It combines two clusters at a time at each stage and repeats the method until the desired number of clusters is reached.

Almeida et al. (2007) enhanced the single linkage algorithm to reduce its sensitivity to outliers and to enable it to work in situations when the number of clusters is unknown a priori. Firstly, by checking nearest-neighbors of each object, the method finds the potential zones of outliers and discards them from the data set. The method then clusters the data in the same way that it is done with the single linkage, using the Euclidean distance matrix. Finally, the algorithm groups the objects that were initially eliminated and re-inserts them into the data set.

Generally, hierarchical approaches are incapable of correcting misclassified objects, and they are not resilient in the presence of noise, outliers and overlapping clusters. Additionally, time complexity is a significant issue when utilizing hierarchical approaches, as the majority of hierarchical methods have a time complexity of at least $O(n^2)$ (Mittal et al., 2021). Detailed review of such algorithms can be found in Lingras and Huang (2005), Saxena et al. (2017), Huang and Ribeiro (2018), Mehta et al. (2020), and A. Alam et al. (2021).

2.1.3 Swarm Intelligence Algorithms for Clustering

Swarm intelligence (SI) is a form of artificial intelligence that is based on the collective qualities of a swarm of agents. A swarm of agents, such as ants, birds, fish, and insects, can perform complicated tasks that a single agent cannot, such as locating food sources and defending against predators (Inkaya et al., 2016). The most commonly used SI algorithms in literature are Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Ant Based Sorting (ABS), and more recently Artificial Bee Colony (ABC).

2.1.3.1 Particle Swarm Optimization (PSO) for Clustering

PSO imitates the movement of birds in search of food. Each particle in the PSO represents a possible solution, and the algorithm traverses the search space utilizing a collection of individuals (Rana et al., 2011). Each particle of the swarm contributes to the environment by adhering to a few simple rules and thereby cooperating and communicating with the other particles. However, the swarm as a whole has a complex global behaviour which is to solve a complex optimization problem (S. Alam et al., 2014).

The fundamental usage of PSO for clustering is to initialize each particle with K random cluster centers and then allocate each data object to the cluster with the nearest centroid using a distance metric. The position and velocity of each particle are then updated in accordance with its fitness value and this procedure is repeated until the algorithm reaches convergence (Mane and Gaikwad, 2014).

There are numerous PSO algorithms with varying properties available for clustering. Between algorithms, the solution representation, the number of objective functions, and the objective functions employed might be quite different. Also, properties of the clustering problem can be very different. A full description of PSO based clustering may be found in the following review articles: Rana et al. (2011), Sarkar et al. (2013), S. Alam et al. (2014), and Esmin et al. (2015).

The main features of some PSO algorithms used for clustering in the literature are given in Table 2.1.

2.1.3.2 Ant Colony Optimization (ACO) for Clustering

ACO is a population based approach which mimics the collective behaviour of the ants. A traveling ant deposits pheromone on the ground, leaving a trail of this substance in its wake. The more ants that follow a route, the more appealing it becomes for others to follow. Thus, the process is characterized by a positive feedback loop, in which the probability of one ant choosing a path grows in direct proportion to the number of ants that have previously chosen that way (Dorigo et al., 1996).

In its basic form, ACO is a constructive heuristic. In the literature, various methods were used to improve the weaknesses of ACO and to obtain better results. For instance, Shelokar et al. (2004) note that because typical ACO's local search capabilities is limited, the algorithm has a propensity to become trapped at local optima. Therefore, they incorporate a mechanism for performing local search into the algorithm. Another common improvement is to use some heuristic information for choosing solution components in solution construction (Dorigo et al., 1996). Apart from the pheromone values, heuristic information is utilized to guide the search process in ACO algorithms, as it provides prior knowledge about the given problem (Abd-

Alsabour et al., 2013). For example, Pacheco et al. (2018) use the similarity between two data objects as heuristic information and, as a result, consider the similarity of objects to one another when assigning them to clusters.

Because each ant in the colony represents a clustering solution, the solution may be represented in a variety of ways. Solution representations utilized in ACO algorithms are divided into two types, according to Inkaya et al. (2016). In the first, each ant displays the cluster assignments of all of the points in the data set, while in the second, each ant connects pairs of points as it moves.

Detailed specifications of ACO algorithms used in the literature for clustering may be found in the following review articles: Nanda and Panda (2014), Inkaya et al. (2016), Jabbar et al. (2018), Ezugwu et al. (2021), and Abualigah et al. (2020).

Table 2.2 summarizes the ACO algorithms that have been used in the literature for clustering, including the number of objective functions, the objective function(s) used, whether the number of clusters known or not, the shapes of the clusters in the data set, the type of data in the data set, and the solution representation used.

2.1.3.3 Ant Based Sorting (ABS) for Clustering

ABS is inspired by ant collection and sorting behaviors such as corpse clustering, brood sorting, and nest construction. The data pieces are dispersed in a two-dimensional grid, and each ant walks randomly through it, picking up and dropping off data items. Choosing to pick up or drop off an item is a probabilistic decision impacted by the data items in the ant's immediate environment. If ants are surrounded by similar data objects, the probability of dropping an object increases; conversely, if they are surrounded by dissimilar data objects, the probability of choosing an object increases. As a result of implementing these processes, clustering of the objects on the two-dimensional grid can be achieved. (Boryczka, 2009).

Nanda and Panda (2014), and Jabbar et al. (2018) contain in-depth reviews of such algorithms.

The details of some ABS algorithms used for clustering in the literature are given in Table 2.3.

2.1.3.4 Artificial Bee Colony (ABC) for Clustering

The ABC algorithm is an optimization algorithm that is based on the intelligent behavior of a swarm of honey bees. The ABC algorithm divides the artificial bee colony into three distinct groups: employed bees, onlookers, and scouts. A bee waiting on the dancing area to decide on a food source is referred to as an onlooker, whereas a bee traveling to a previously visited food source is referred to as an employed bee. A scout bee is one that conducts random searches (Karaboga and Basturk, 2007).

In the ABC algorithm, the first half of the colony is made up of employed bees, and the second half is made up of the onlookers. The ABC generates a randomly distributed initial population. Then, an employed bee modifies its position (the solution it represents) based on local information and calculates a new fitness value using an objective function. If the new position is better than the previous one, she discards the previous one and places nectar proportional to the new fitness value at the new position. The onlookers then collect nectar information and new positions (solutions), choose a solution based on the probability calculated using the fitness value, and determine whether or not the incumbent solution has been improved. If the employed bees are unable to improve a solution after a set number of iterations, the scouts replace it with a random one (Karaboga and Ozturk, 2011). The algorithm repeats these steps until it converges to a solution.

ABC algorithms for clustering are discussed in detail in the following review papers Karaboga et al. (2014), Mane and Gaikwad (2014), Nanda and Panda (2014), and A. Kumar et al. (2017).

The details of some ABC algorithms used for clustering in the literature are given in Table 2.4.

2.1.3.5 Hybrid Swarm Intelligence (SI) Algorithms for Clustering

In the literature, hybrid algorithms are frequently employed to overcome the shortcomings of standalone algorithms and improve their performance. Because each algorithm has its own capabilities and may focus on either exploration or exploitation, hybrid algorithms were invented to balance exploration and exploitation (Mageshkumar et al., 2019). Thus, the primary goal of hybrid algorithms is to combine the advantageous features of at least two methods (Hatamlou et al., 2012).

Table 2.5 contains information on different hybrid methods involving swarm intelligence used for clustering in the literature.

Table 2.1: PSO Algorithms for Clustering in the Literature

Article	**Multi-objective**	**# of clusters**	**Arbitrary shapes**	**Data type**	**Clustering objective**	**Solution representation**	**Application**
Alam et al., 2015	No	Given and not given	No	Numeric	Euclidean distance	Each agent represents the cluster centroids	-
Ali et al., 2012	Yes	Not given	Yes	Numeric	Mobile ad-hoc networks environment functions	Each agent represents the cluster centroids	Mobile ad-hoc networks
Pham et al., 2018	No	Not given	Yes	Image	Modified fuzzy entropy	Each agent represents the cluster centroids	MRI brain image segmentation
Li et al., 2017	Yes	Not given	Yes	Numeric	Inter-cluster density and intra cluster density	An agent represents a graph that connecting points	Complex network clustering
Tsai et al., 2015	No	Given	No	Numeric and image	Euclidean distance	An agent shows cluster assignments of all points	-
Netjinda et al., 2015	No	Given	No	Numeric	Euclidean distance	An agent shows cluster assignments of all points	-
Kuo et al., 2011	No	Not given	No	Numeric	Euclidean distance	Each agent represents the cluster centroids	Order clustering
Jiang and Wang, 2014	No	Given	No	Numeric	Euclidean distance	Each agent represents the cluster centroids	-
Fornarelli and Giaquinto, 2013	No	Not given	Yes	Image	Euclidean distance	Each agent represents the cluster centroids	-

Table 2.1 continued from previous page

Article	**Multi-objective**	**# of clusters**	**Arbitrary shapes**	**Data type**	**Clustering objective**	**Solution representation**	**Application**
Zhang et al., 2014	No	Given	No	Numeric and image	Euclidean distance	Each agent represents the cluster centroids	-
Song et al., 2017	No	Given	No	Text	Cosine measure	An agent shows cluster assignments of all points	-
Jarboui et al., 2007	No	Given	No	Numeric	Euclidean distance	An agent shows cluster assignments of all points	-
Cura, 2012	No	Not given	No	Numeric	Euclidean distance	Each agent shows number of clusters and cluster centroids	-
Das et al., 2008	No	Not given	Yes	Numeric	Kernelized CS measure	Activation function values and cluster centroids	-
Masoud et al., 2013	Yes	Not given	No	Numeric	Euclidean distance, VRC and DBI	An agent shows cluster assignments of all points	-

Table 2.2: ACO Algorithms for Clustering in the Literature

Article	Multi-objective	# of clusters	Arbitrary shapes	Data type	Clustering objective	Solution representation	Application
Inkaya et al., 2015	Yes	Not given	Yes	Numeric	Adjusted compactness and relative separation	An ant connects pairs of points as it moves	-
Ramos et al., 2009	No	Not given	Yes	Numeric	Euclidean distance	An ant shows cluster assignments of all points	Identifying disease risk
Chowdhury and Das, 2012	No	Not given	Yes	Numeric	Euclidean distance	An ant shows cluster assignments of all points	-
Menèndez et al., 2016	No	Given and not given	No	Numeric	Euclidean distance and Silhoutte index	An ant shows cluster assignments and medoids	-
Kao and Cheng, 2006	No	Given	No	Numeric	Euclidean distance	An ant shows cluster assignments of all points	-
Shelokar et al., 2004	No	Given	No	Numeric	Euclidean distance	An ant shows cluster assignments of all points	-
Halder et al., 2011	No	Not given	No	Image	Density measure	An ant shows cluster assignments of all points	Landuse map generation
Yang and Kamel, 2006	No	Not given	No	Numeric	Cosine similarity	An ant shows cluster assignments of all points	-
Salama and Freitas, 2014	No	Given	No	Numeric	Euclidean distance	Medoid based representation	Classification

Table 2.2 continued from previous page

Article	**Multi-objective**	**# of clusters**	**Arbitrary shapes**	**Data type**	**Clustering objective**	**Solution representation**	**Application**
Hu et al., 2015	No	Given	No	Numeric	Euclidean distance	An ant shows cluster assignments of all points	-
Yang et al., 2018	No	Given	No	Numeric	Euclidean distance	An ant shows cluster assignments of all points	-
Pacheco et al., 2018	No	Not given	No	Numeric	Silhouette index	An ant connects pairs of points as it moves	-

Table 2.3: ABS Algorithms for Clustering in the Literature

Article	**Multi-objective**	**# of clusters**	**Arbitrary shapes**	**Data type**	**Clustering objective**	**Solution representation**	**Application**
Boryczka, 2009	No	Not given	Yes	Numeric	Euclidean distance	Ant carries data points	-
Dziwi‘nski et al., 2012	No	Given	No	Numeric and text	Cosine similarity	Ant carries data points	Clustering text documents
Elkamel et al., 2015	No	Not given	Yes	Numeric	Euclidean distance	Ant carries data points	Image indexing
Ghosh et al., 2008	No	Given	Yes	Numeric	Within cluster similarity	Each point is represented by an ant	-
Zhang et al., 2013	No	Not given	Yes	Numeric	Euclidean distance	Each point is represented by an ant	-

Table 2.4: ABC Algorithms for Clustering in the Literature

Article	**Multi-objective**	**# of clusters**	**Arbitrary shapes**	**Data type**	**Clustering objective**	**Solution representation**	**Application**
Dilmac and Korurek, 2015	No	Given	Yes	Numeric	Euclidean distance	Each agent represents the cluster centers	ECG heartbeat classification
Karaboga and Ozturk, 2011	No	Given	No	Numeric	Euclidean distance	Each agent represents the cluster centers	-
Ozturk et al., 2015	No	Not given	No	Numeric and image	VI index	Each agent represents the cluster centers	-
Su et al., 2012	No	Not given	No	Numeric	Fuzzy C-means	Each agent represents the cluster centers	-
Zabihi and Nasiri, 2018	No	Given	No	Numeric	Euclidean distance	Each agent represents the cluster centers	-
Das et al., 2018	No	Given	No	Numeric	Euclidean distance	Each agent represents the cluster centers	-

Table 2.5: Hybrid Algorithms for Clustering in the Literature

Article	Algorithms	Multi-objective	# of clusters	Arbitrary shapes	Data type	Clustering objective	Solution representation	Application
Niknam and Amiri, 2010	PSO, ACO and k-means	No	Given	Yes	Numeric	Euclidean distance	Each agent represents cluster centroids	Market segmentation
Hatamlou et al., 2012	GSA and k-means	No	Given	No	Numeric	Euclidean distance	Each agent represents cluster centroids	-
C.-L. Huang et al., 2013	ACO and PSO	No	Given	No	Numeric	Modified fuzzy entropy	Each agent represents cluster centroids	-
Y. Kumar and Sahoo, 2015	MCSS and PSO	No	Given	No	Numeric	Euclidean distance	An agent shows cluster assignments of all points	-
Wu et al., 2020	ACO and DE	No	Given	No	Numeric	Euclidean distance	Each agent represents cluster centroids	-
Chen et al., 2020	QALO and k-means	No	Given	No	Numeric	Euclidean distance	Each agent represents cluster centroids	Intrusion detection
L. M. Abuligah et al., 2018	KHA and GA	Yes	Given	No	Text	Cosine similarity and Euclidean distance	An agent shows cluster assignments of all documents	Text clustering
X. Zhang et al., 2018	BBO and GWO	No	Given	No	Numeric	Euclidean distance	An agent shows cluster assignments of all points	-

Table 2.5 continued from previous page

Article	Algorithms	Multi-objective	# of clusters	Arbitrary shapes	Data type	Clustering objective	Solution representation	Application
Thong et al., 2016	FC-PFS and PSO	Yes	Not given	Yes	Numeric and image	Picture Fuzzy clustering and alternative Silhouette index	Each agent represents cluster centroids	-
Elyasigomari et al., 2015	COA and GA	No	Given	No	Numeric	Euclidean distance	An agent shows cluster assignments of all points and centroids	Gene selection in cancer classification
Bouyer and Hatamlou, 2018	ICS and PSO	No	Given	No	Numeric	Euclidean distance	Each agent represents cluster centroids	-
Hatamlou, 2017	PSO and BBBC	No	Not given	No	Numeric	Euclidean distance	Each agent represents cluster centroids	-

GSA: Gravitational Search Algorithm, MCSS : Magnetic Charged System Search Algorithm,
DE : Differential Evolution, QALO : Quantum-inspired Ant Lion Algorithm, KHA : Krill Herd Algorithm,
BBO : Biogeography-Based Optimization, GWO : Grey Wolf Optimization Algorithm,
FC-PFS : Picture Fuzzy Clustering, COA : Cuckoo Optimization Algorithm, GA : Genetic Algorithm,
ICS : Improved Cuckoo Search Algorithm, BBBC = Big Bang-Big Crunch Algorithm

2.2 Feature Selection

Data preparation is critical for effective data mining. Feature selection is a critical and commonly used approach for data preprocessing in preparation for data mining. It eliminates irrelevant or redundant features, and immediately benefits applications by speeding up data mining algorithms and boosting mining performance metrics such as predicted accuracy and result comprehension (Liu and Yu, 2005). While data mining frequently involves huge and multidimensional data, the majority of clustering techniques in the literature are sensitive to either the number of samples in the data set or multidimensionality, or both. Different features have varying effects on clusters; some are beneficial to detect clusters, while others may be useless or even decrease the clustering accuracy (Dash and Liu, 2000).

In supervised learning, algorithms for feature selection optimize some function of predicted accuracy. Due to the existence of predefined class labels, it is logical to retain just those features that are related to or lead to these classes (Dy and Brodley, 2004). However, as a result of recent technological advancements and the massive amount of unlabeled data generated in a variety of applications such as text mining, bioinformatics, image processing, social media, and intrusion detection, unsupervised feature selection methods have garnered considerable interest in the scientific community (Solorio-Fernández et al., 2020).

The methods for selecting features are generally classified into two broad categories: filtering and wrapper methods. Filtering approaches attempt to identify an appropriate feature subset based on the general properties of the data instead of a learning algorithm. Typically, filters compute a score for each feature based on specific evaluation criteria. Then, the decision maker can select the features with the highest scores. Wrappers, on the other hand, necessitate the use of a clustering technique to determine the goodness of potential feature subsets. Wrappers gather a subset of features via search strategies and then evaluate the subset's quality using a clustering algorithm. This method is carried out repeatedly until the stopping criteria are satisfied (Hancer et al., 2020). The wrapper method's feature subset clustering performance is typically superior to the filter method's feature subset clustering performance. However, because the clustering algorithm must evaluate each feature subset separately, this method has a high computational complexity, which may be an issue when working with large-scale data sets (Cai et al., 2018).

Fisher score is one of the most extensively used filtering approaches for selecting features in supervised learning. Fisher score's primary principle is to discover a collection of features such that the distances between data points belonging to different classes are as large as possible, while the distances between data points belonging to the same class are as small as possible (Gu et al., 2012). However, because Fisher score is a supervised approach, it requires a priori knowledge of the class label. As a result, it is unsuitable for unsupervised tasks such as clustering. However, there are also unsupervised filtering methods such as data variance and Laplacian score (LS). Data variance criterion is good at finding ways to represent the data, but it does not explicitly determine which features are good at telling which sample belongs to which cluster. However, LS is based on the fact that two data points that are near to one another are most likely related to the same cluster. LS accomplishes this by utilizing a nearest neighbor graph and searching for features that relate to the graph structure (He et al., 2005).

F-ACOVNS algorithm makes use of Laplacian Score as heuristic information. Therefore, it is described in detail below, as proposed by He et al. (2005).

Let L_r denote the Laplacian Score of the r-th feature. Let f_{ri} denote the i-th sample of the r-th feature, $i = 1, ..., m$. The LS algorithm can be stated as follows:

1. Construct a nearest neighbor graph G with m nodes. The i-th node corresponds to $\mathbf{x}_i$. We put an edge between nodes i and j if $\mathbf{x}_i$ and $\mathbf{x}_j$ are "close", i.e. $\mathbf{x}_i$ is among k nearest neighbors of $\mathbf{x}_j$ or $\mathbf{x}_j$ is among k nearest neighbors of $\mathbf{x}_i$. When the label information is available, one can put an edge between two nodes sharing the same label.
2. If nodes i and j are connected, put $S_{ij} = e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}$, where t is a suitable constant. Otherwise, put $S_{ij} = 0$. The weight matrix S of the graph models the local structure of the data space.
3. For the r-th feature, we define:

$$\mathbf{f}_r = [f_{r1}, f_{r2}, ..., f_{rm}]^T, D = diag(S\mathbf{1}), \mathbf{1} = [1, ..., 1]^T, L = D - S$$

where the matrix L is often called graph Laplacian. Let

$$\tilde{\mathbf{f}}_r = \mathbf{f}_r - \frac{\mathbf{f}_r^T D\mathbf{1}}{\mathbf{1}^T D\mathbf{1}}\mathbf{1}$$

4. Compute the Laplacian Score of the r-th feature as follows:

$$L_r = \frac{\tilde{\mathbf{f}}_r^T L\tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^T D\tilde{\mathbf{f}}_r} \tag{2.1}$$

He et al. (2005) provide insight into LS as follows. "For each feature, its Laplacian score is computed to reflect its locality preserving power. LS is based on the observation that, two data points are probably related to the same topic (belong to the same cluster) if they are close to each other. In fact, in many learning problems such as classification, the local structure of the data space is more important than the global structure. In order to model the local geometric structure, we construct a nearest neighbor graph. LS seeks those features that respect this graph structure."

Jović et al. (2015), Bolón-Canedo et al. (2013), Ferreira and Figueiredo (2012), Vipin and Minz (2014), and Ang et al. (2015) discuss filtering approaches in detail.

Wrapper approaches are mainly divided into two categories which are sequential and heuristic search methods. The sequential selection methods begin with an empty set or complete set and gradually add or remove features until an objective function is optimized (Chandrashekar and Sahin, 2014). Typically, a clustering algorithm is used to evaluate the feature subset, and the best feature subset is selected according to clustering accuracy.

Evolutionary Local Selection Algorithm (ELSA) is perhaps the most commonly used bio-inspired wrapper method which searches feature subsets by using k-means and Gaussian Mixture clustering algorithms. Each clustering solution is coupled with a vector whose elements represent the quality of the assessment criteria, which are based on the cohesion of clusters, inter-class separation, and maximum likelihood. The evaluation stage selects the features that optimize the objective functions (Solorio-Fernández et al., 2020).

Binary PSO (BPSO) is a swarm intelligence technique that has been utilized in the literature for feature selection. Each particle in BPSO is given in binary form, which accurately depicts the simple yes or no decision regarding whether a feature should be selected or not. Then, in each iteration, the k-nearest neighbor method is employed to clustering the data for selected features. (Chuang et al., 2008).

Expectation-Maximization (EM) clustering is another extensively used wrapping approach. The fundamental concept is to search across the space of feature subsets, assessing each subset using EM clustering. It identifies relevant features by utilizing a greedy search technique such as sequential forward or backward elimination (Dy and Brodley, 2000).

Detailed information about wrapper approaches are given in the following research papers by El Aboudi and Benhlima (2016), Venkatesh and Anuradha (2019), Kumari and Swarnkar (2011), and Visalakshi and Radha (2014).

2.3 Simultaneous Clustering and Feature Selection

Decomposing feature selection and clustering into two distinct problems and solving them independently frequently results in information loss. As a result, solving them concurrently in a single integrated method results in superior solutions.

One of the ways used to address these two problems simultaneously is to perform clustering wherein the features are selected on a random basis (Swetha and Susheela Devi, 2012). However, without a mechanism to guide the search, finding the target feature subset at random can be highly computationally expensive. Including both cluster and feature information in the solution representation, on the other hand, is a method that has been employed in the literature. The most frequently used representation is centroids for clusters and binary coding for features where each feature has a value of 1 when selected and 0 when not selected (Saha et al., 2014). An alternative representation for this problem is to define the decision variables as cluster centroids but to compute the centroids using only the selected features. As a result, feature information is retained in an indirect manner in the solution representation (Dutta et al., 2013). Another option is to use a multi-objective framework with clustering and feature selection objective functions. Algorithms taking this approach attempt to optimize both objective functions for detecting the right subset of features and partitioning the data (Saha et al., 2015).

Dutta et al. (2014) study feature selection and clustering simultaneously by employing multiple objective functions. In their work, they depict solutions using cluster centroids but using only selected features. Saha et al. (2015) use the multi-objective Simulated Annealing technique for simultaneous feature selection and clustering. They use the following objective functions: Euclidean distance, point symmetry based distance, and the number of features. In addition, to represent solutions, they use cluster centroids and binary coding of features.

Prakash and Singh (2015) combine the BPSO and k-means algorithms to perform feature selection and clustering simultaneously. BPSO employs binary coding of features to represent feature selection solutions, whereas the k-means algorithm does the clustering using selected features. In their study, BPSO uses a common clustering objective function, the Silhoutte index. The Silhouette Index calculates the "optimal" number of clusters by dividing the average distance within a cluster by the minimum distance between clusters, hence it tries to reflect the optimal clustering effect (Wang and Xu, 2019). Prakash and Singh (2019) combine GSA and k-means for simultaneous feature selection and clustering, however unlike their prior work, they hybridize a swarm intelligence method using two objective functions. They use the Silhouette index to determine the number of clusters, and the number of features to determine the relevant features. They represent solutions using binary coding of features.

Hancer (2020), in the multi-objective Differential Evolution algorithm, employ the Silhouette index, the WB index, and the number of features. Differential Evolution represents solutions using variable-length strings containing cluster codes and feature activation codes. On the other hand, the Niching Memetic algorithm by Sheng et al. (2008) employs the same representation approach, but it uses Euclidean distance as an objective function. The same representation scheme is also used by Sarvari et al. (2010) and Hancer (2018).

O'Neill et al. (2018) employ the PSO algorithm for simultaneous feature selection and clustering using the Silhouette index with connectivity. In addition, to represent particles, they use cluster centroids with only selected features. Lensen et al. (2016) also propose a PSO technique where they describe particles using medoids and binary coding of features. They use cluster compactness and separability as objective functions. Javani et al. (2011) is another work that employs PSO for simultaneous feature selection and clustering. They represent solutions in their work using cluster centroids, feature weights, and activation thresholds. As a result, if the weight of a feature is above the threshold, that feature is selected. Additionally, they use the Conn index and modified kernel PBM validation functions to determine the objective function values.

Alakuş (2018) proposes a multi-objective clustering algorithm (MOCNC) and extend this algorithm for simultaneous feature selection and clustering (MOCNC-F). MOCNC algorithm is a combination of Neighborhood Construction (NC) and a multi-objective evolutionary algorithm (NSGA-II). Three objective functions are used in the algorithm: connectivity, compactness, and separation. The connectivity objective is achieved by the NC algorithm in the preprocessing step and then NSGA-II takes output of the NC algorithm as input. Compactness is defined as the average of the distances between data points in clusters and their respective cluster centroids, whilst separation is defined as the single-link linkage metric. When feature selection and clustering are performed simultaneously, MOCNC identifies nondominated clustering solutions for each feature subset, and then MOCNC-F determines the dominance of distinct clustering solutions arising from different feature subsets. MOCNC-F produces a set of nondominated clustering solutions, each with a unique compactness and separation value, also perhaps with a unique feature subset. The algorithm uses adjacency-based representation where the genes correspond to the links between pairs of NC closures. The mth closure is linked (or connected) to the nth closure if mth gene takes the value of n.

Additionally, Alakuş (2018) combines her feature selection approach with the ΔMOCK clustering algorithm proposed by Garza-Fabre et al. (2017). ΔMOCK uses two objective functions: intra-cluster variance and cluster connectivity. In Alakuş's implementation, the MOCNC-F algorithm uses ΔMOCK instead of MOCNC for clustering so the new algorithm is called ΔMOCK-F.

Table 2.6 summarizes simultaneous feature selection and clustering algorithms found in the literature, many of which are based on SI.

Table 2.6: Simultaneous Feature Selection and Clustering Algorithms in the Literature

Article	**Algorithm(s)**	**Multi-objective**	**# of clusters**	**Arbitrary shapes**	**Data type**	**Objective Function(s)**	**Solution representation**
Dutta el al., 2014	MOGA	Yes	Given	Yes	Numeric	Intra-cluster distance, and inter-cluster distances	Cluster centroids with only selected features
Saha et al., 2015	Simulated Annealing	Yes	Not given	Yes	Numeric	Euclidean distance, point symmetry based distance, and the number of features	Cluster centroids and binary coding of features
Prakash and Singh, 2015	BPSO and k-means	No	Given	Yes	Numeric	Silhouette index	Binary coding of features
Prakash and Singh, 2019	GSA and k-means	Yes	Not given	Yes	Numeric	Silhouette index and the number of features	Binary coding of features
Hancer, 2020	Differential Evolution	Yes	Not given	Yes	Numeric	Silhouette index, WB index, and the number of features	Varible length string with cluster centroids and feature activation codes
Sheng el al., 2008	NMA	No	Not given	Yes	Numeric	Euclidean distance	Variable length string with cluster centers and binary coding of features
O'Neill et al., 2018	PSO	No	Not given	No	Numeric	Combined Silhouette index and Connectedness	Cluster centroids with only selected features

Table 2.6 continued from previous page

Article	Algorithm(s)	Multi-objective	# of clusters	Arbitrary shapes	Data type	Objective Function(s)	Solution representation
Sarvari et al., 2010	Harmony Search	No	Not given	Yes	Numeric	k-means	Variable length string with cluster centers and binary coding of features
Lensen et al., 2016	PSO	No	Given	Yes	Numeric	Cluster compactness and separability	Medoids and binary coding of features
Hancer, 2018	Differential Evolution	No	Not given	Yes	Numeric	VI index	Varible length string with cluster centroids and feature activation codes
Javani et al., 2011	PSO	Yes	Given	Yes	Numeric	Conn and modified kernel PBM validation functions	Cluster centroids, weights of features, and activation threshold
Alakuş, 2018	MOCNC-F	Yes	Not given	Yes	Numeric	Connectivity, compactness and separation	Adjacency-based representation
Garza-Fabre et al., 2017	ΔMOCK	Yes	Not given	Yes	Numeric	Intra-cluster variance and cluster connectivity	Locus-based adjacency representation

MOGA: Multi Objective Genetic Algorithm, BPSO : Binary Particle Swarm Optimization,
GSA : Gravitational Search Algorithm, NMA : Niching Memetic Algorithm

CHAPTER 3

SOLUTION APPROACH FOR CLUSTERING: ACOVNS

We present our clustering algorithm, a hybridization of ACO and VNS (ACOVNS), in this chapter. ACOVNS primarily utilize ACO mechanisms for exploration and VNS mechanisms for exploitation. The overall structure of the ACOVNS algorithm is represented in Figure 3.1.

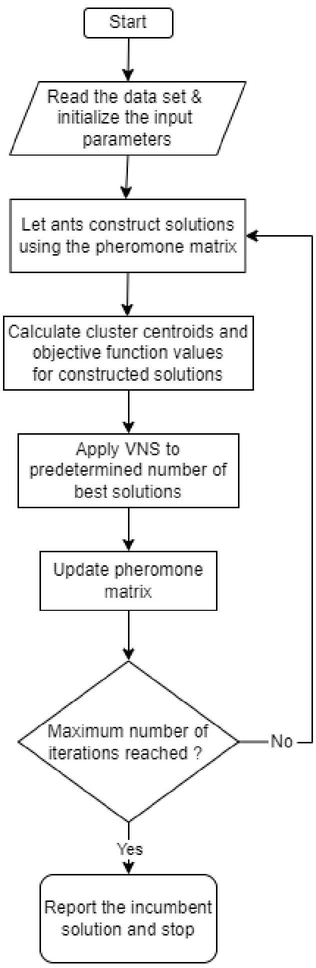

Figure 3.1: Flowchart of the ACOVNS algorithm

The main flow of the algorithm is as shown in Figure 3.1. Prior to describing the specifics of ACOVNS, we discuss some ACO features of the algorithm in Section 3.1. The notation and formulations used in the algorithm are given in Sections 3.2 and 3.3, respectively. In Section 3.4, VNS and its use within ACO are described. Then, we give the pseudocode of our proposed algorithm ACOVNS in Section 3.5. Following that, the data sets, parameter settings, and performance measures used in the experiments are described in Sections 3.6.1 and 3.6.2. Finally, we compare our algorithm with the algorithms in the literature in Section 3.6.3.

3.1 Ant Colony Optimization Features

3.1.1 Solution Representation

In our ACO algorithm, each ant represents a clustering solution. A solution is represented by a vector of length N where N is the number of samples. Each element of the vector corresponds to one of the objects in the data set. The value assigned to an element in the vector is the cluster number to which the object is assigned. Hence, in this representation, each ant shows cluster assignments of all samples in the form of a vector. Figure 3.2 illustrates an example vector where there are 7 objects and 3 clusters. The solution in the illustration states that objects 1, 4, and 5 are assigned to cluster 3, objects 2 and 3 are assigned to cluster 2, and objects 6 and 7 are assigned to cluster 1.

Object 1	Object 2	Object 3	Object 4	Object 5	Object 6	Object 7
3	2	2	3	3	1	1

Figure 3.2: Solution Representation Example

3.1.2 Solution Construction and Pheromone Update

In each iteration, each ant uses the pheromone matrix to construct a solution. The pheromone matrix stores pheromone values for object-cluster pairs. Ants determine probabilities for object-cluster assignments based on the pheromone values, and then construct solutions in accordance with those probabilities. In general, the higher the pheromone value is, the larger the probability of respective object-cluster assignment is. Hence, an ant assigns each sample to a cluster using the roulette wheel selection with these probabilities in constructing a solution.

The quality of a solution is determined by its objective function value. Following the construction of a solution by each ant, the value of the objective function associated with that solution is calculated. The solutions constructed by all ants are then sorted based on their objective function values, and a predetermined number of solutions with the best objective function values can be selected.

In the first iteration, each ant constructs a solution by randomly assigning objects to clusters, as the pheromone values are set to be equal at the start. Then, using object-cluster assignments, the cluster centroids are determined and the objective function values are calculated. The objective function values are then used to update the pheromone values for the predetermined number of best solutions. The better the objective function value is, the higher the pheromone deposits are to the object-cluster pairs assigned in the respective solution.

At the end of each iteration, pheromone values for all object-cluster pairs are also decreased proportionally to a rate of evaporation. However, the overall pheromone values of those object-cluster pairs in the predetermined number of best solutions are still increased due to relatively higher levels of pheromone deposits. The algorithm continues with these updated pheromone values in the following iteration, where increased pheromone values result in higher object-cluster assignment probabilities. By repeating this procedure for each iteration, the ant colony eventually converges to a clustering solution with the best objective function value.

The exact formulations for probability of selection and pheromone update will be given in Section 3.3.

3.1.3 Heuristic Information

In addition to the pheromone values, some heuristic information is also used in determining the object-cluster assignment probabilities. The contribution of heuristic information can be explained as follows. In each iteration, pheromone values are increased for all object-cluster pairs found in the predetermined number of best solutions by the same amount, using objective function values of those solutions. However, when taken as a solution component, not every object-cluster assignment included in a solution has the same quality. Hence, heuristic information in ACO is defined to differentiate the quality of individual solution components (Dorigo et al., 1996). After an ant constructs a solution, heuristic information is found based on the distance between an object and centroid of the cluster it is assigned to in that solution. Hence, with the inclusion of this heuristic information, object-cluster assignment probabilities differ for different solution components found in the same solution. They also differ from one ant's solution to another.

When heuristic information is used in the algorithm, the probability of an object-cluster assignment is calculated by combining heuristic information values together with the pheromone values. Heuristic information defined in this manner also takes advantage of each ant's solution constructed in the previous iteration. Therefore, while the pheromone matrix contains the collective memory of the colony, heuristic information contains the memory of an individual ant.

3.1.4 Additional Exploration Mechanism

To increase the exploration capability of the algorithm and to avoid premature convergence, we incorporate a mechanism that allows probabilistic object-cluster assignment with the lowest pheromone value. This mechanism enables the algorithm to seek out distinct solutions in a manner similar to the mutation operation in a genetic algorithm. As the ACO algorithm typically converges to a few distinct solutions and their associated pheromone values become much higher than the values for the other object-cluster assignments, it is possible to find an improvement in the solutions produced by assigning a sample to a cluster with the lowest pheromone value.

3.2 Notation

The following notation is used in the forthcoming formulations and pseudocode of the algorithm.

N_{ants}	Number of ants
N	Number of samples (given)
N_c	Number of clusters (given)
N_f	Number of features (given)
N_{iter}	Number of iterations
N_{ik}	Number of samples in cluster k in ant i's solution
i	Ant index, $i = 1, ..., N_{ants}$
j	Sample index, $j = 1, ..., N$
k	Cluster index, $k = 1, ..., N_c$
f	Feature index, $f = 1, ..., N_f$
S_i	Selected feature subset in ant i's solution (in ACOVNS, S_i contains all features available in the data set and it is the same for all i; it is used for compatibility with F-ACOVNS)
x_j	Location of sample j (given)
x_{jf}	Location of sample j in terms of feature f (given)
c_{ik}	Location of cluster centroid k in ant i's solution
c_{ikf}	Location of cluster centroid k in terms of feature f in ant i's solution
$d_i(x_j, c_{ik})$	Euclidean distance between x_j and c_{ik} in ant i's solution
T_{jk}	Pheromone value of assigning sample j to cluster k
τ	Initial pheromone matrix where $T_{jk} = 1$ for all j, k
ρ	Evaporation rate
h_{ijk}	Heuristic information value of assigning sample j to cluster k in ant i's solution
P_{jk}	Probability of assigning sample j to cluster k
P_{ijk}	Probability of assigning sample j to cluster k in ant i's solution

P_{lowest}	Probability of assigning a sample to the cluster that has the lowest pheromone value
y_{ijk}	1 if sample j is assigned to cluster k in ant i's solution, 0 otherwise
F_i	Objective function value of ant i's solution
L	Predetermined number of best solutions
V_{NS}	Neighborhood size
V_{MNS}	Maximum neighborhood size
r	a uniform random number between 0 and 1
$HeuristicInfoUsed$	1 if heuristic information is used, 0 otherwise
$NormalizationUsed$	1 if the data set is normalized, 0 otherwise

3.3 Formulations

Normalization of Feature Values

Numerical features taking values from drastically different ranges can have a negative effect on clustering performance. In real-world data sets, one feature may take values from the range (0,1) whereas values of another may be in the order of thousands. In such circumstances, a single feature might dominate the others and have a significant impact on the outcome of clustering algorithms due to its extremely large values in comparison to the others. Normalizing all features to values between 0 and 1 can improve the algorithm's performance and ensure that it is not misled. Therefore, we use Equation (3.1) for normalization.

Let x'_{jf} be the original value of feature f of sample j. If the data set is to be normalized then

$$x_{jf} = \frac{x'_{jf} - \min_j\{x'_{jf}\}}{\max_j\{x'_{jf}\} - \min_j\{x'_{jf}\}} \quad \forall j, f, \tag{3.1}$$

otherwise $x_{jf} = x'_{jf}$.

Cluster Centroids

Cluster centroids are calculated in the algorithm by averaging the feature values of samples assigned to the associated cluster.

$$c_{ikf} = \frac{1}{N_{ik}} \sum_{j=1}^{N} y_{ijk} x_{jf} \quad \forall i, k, f \in S_i \tag{3.2}$$

Euclidean Distances

The Euclidean distance between a sample and centroid of the cluster to which the sample is assigned is calculated as in Equation (3.3).

$$d_i(x_j, c_{ik}) = \sqrt{\sum_{f \in S_i} (x_{jf} - c_{ikf})^2} \quad \forall i, j, k \tag{3.3}$$

Objective Function

ACOVNS uses a single objective function based on the Euclidean distances. Objective function value for each solution is calculated by summing the Euclidean distances between each sample and the cluster centroid which the sample is assigned to. The lower the objective function value is, the higher the solution quality is. The formulation of the objective function is given in Equation (3.4).

$$F_i = \sum_{j=1}^{N} \sum_{k=1}^{N_c} y_{ijk} d_i(x_j, c_{ik}) \quad \forall i \tag{3.4}$$

Heuristic Information

The distance between each sample and its cluster centroid provides heuristic information for the respective object-cluster assignment. Each ant utilizes the preceding iteration's heuristic information from the solution it constructed, when assigning samples to clusters in the current iteration. The formulation of the heuristic information is given in Equation (3.5).

$$h_{ijk} = \frac{1}{d_i(x_j, c_{ik})} \quad \forall i, j, k \tag{3.5}$$

Probability of Cluster Selection

Probability of cluster selection shows the probability of each sample being assigned to each cluster. If heuristic information is not used, the probability of selection is determined using Equation (3.6a); otherwise, Equation (3.6b) is used. Note that, in the former case, selection probabilities in an iteration are the same for all ants. In the latter, however, they are different for different ants due to the heuristic information reflecting quality of individual solution components.

$$P_{jk} = \frac{T_{jk}}{\sum_{l=1}^{N_c} T_{jl}} \quad \forall j, k \tag{3.6a}$$

$$P_{ijk} = \frac{T_{jk} h_{ijk}}{\sum_{l=1}^{N_c} T_{jl} h_{ijl}} \quad \forall i, j, k \tag{3.6b}$$

If $\sum_{l=0}^{k-1} P_{ijl} < r \leq \sum_{l=0}^{k} P_{ijl}$ (where $P_{ij0} = 0$), then sample j is assigned to cluster k in ant i's solution.

Pheromone Update

Pheromone values are updated based on the solutions constructed by the colony. All pheromone values are reduced proportionally to the evaporation rate at the end of each iteration. Furthermore, for object-cluster assignments made in the L best solutions, the pheromone values are increased by the inverse of the objective function values. The formulation of pheromone update is given in Equation (3.7).

$$T_{jk} = (1 - \rho) T_{jk} + \sum_{i=1}^{L} y_{ijk} \frac{1}{F_i} \quad \forall j, k \tag{3.7}$$

Note that, according to the general framework of a typical ACO algorithm, two opposing forces are at play here. On one hand, pheromone values are decreased to avoid premature convergence and improve exploration capability of the algorithm. On the other hand, they are increased for better solutions to introduce eliticism and improve the exploitation capability.

3.4 Variable Neighborhood Search

Local search methods for combinatorial optimization work by making a series of small changes to an initial solution that incrementally enhance the value of the objective function until a local optimum is discovered. As a result, each iteration produces an improved solution that is in the neighborhood of the current solution. However, local search heuristics usually use one neighborhood structure and size. In contrast to the majority of other local search methods, VNS does not follow a linear path, but instead investigates more distant neighborhoods of the current incumbent solution by systematically changing the neighborhood and jumps to a new one if and only if an improvement is made (Mladenović and Hansen, 1997). The details of the general VNS is given in Figure 3.3.

Because ACO is a constructive heuristic, local search strategies have been demonstrated to considerably improve the performance of ACO (Mavrovouniotis et al., 2016). As the ants in the colony are prone to choosing cluster assignments with high pheromone values, it is beneficial to use a local search mechanism to find solutions in the neighborhood of these high pheromone value solutions. For this purpose, we include VNS as a local search method in our ACO algorithm and propose ACOVNS.

Following the construction of a solution by each ant in the colony, ACOVNS applies VNS to the L best solutions. The steps of VNS are listed as follows.

Step 1. Set the current neighborhood size V_{NS} = 1
Step 2. Select V_{NS} objects at random and assign each object to the cluster whose centroid is closest to it.
Step 3. If an improvement is discovered or V_{NS} reaches its maximum value of V_{MNS}, terminate the local search and replace the solution with the new one. Otherwise, increase V_{NS} by one and go to Step 2.

An example of the VNS implementation is given in Figure 3.4. The figure illustrates

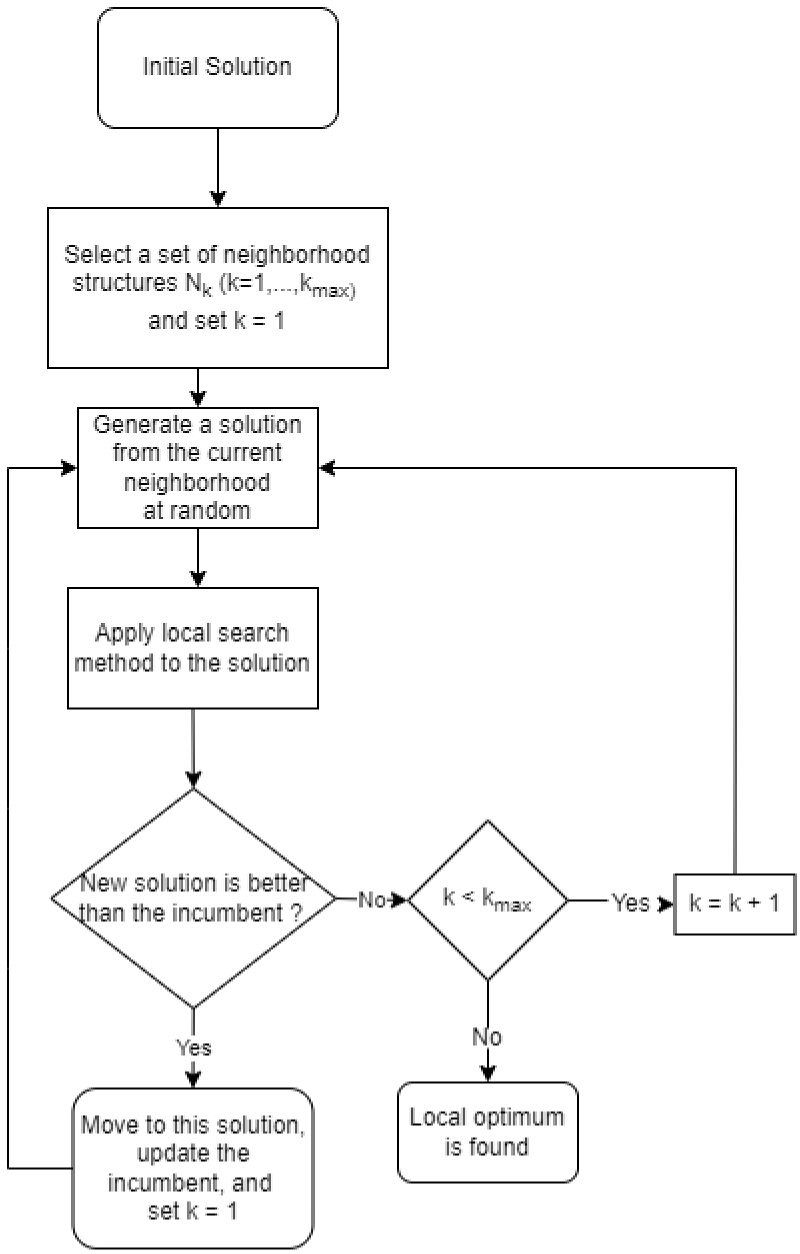

Figure 3.3: The General VNS algorithm

the local search process when the neighborhood size is one and two. When the neighborhood size is one, object 6 is selected and assigned to cluster 1, which is assumed to be the nearest cluster. Assuming that there is no improvement found by the new assignment, the neighborhood size is increased by one. When the neighborhood size is two, objects 3 and 6 are selected and their cluster assignments are changed from 2 and 3 to 3 and 1, respectively. Assuming an improvement is discovered, the local search stops and the new solution replaces the previous one.

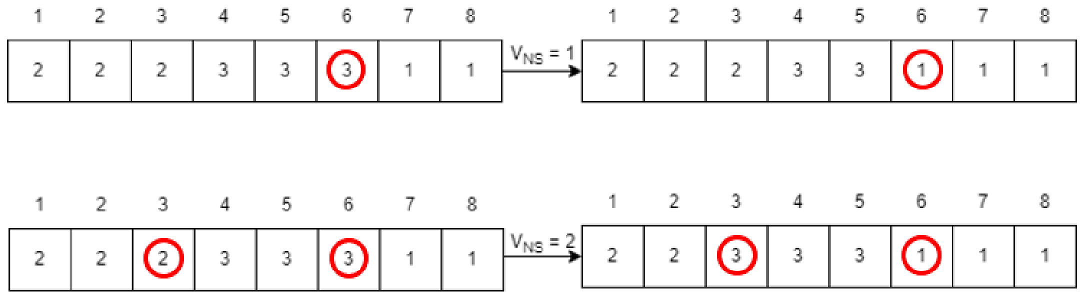

Figure 3.4: VNS example

3.5 Description of the ACOVNS Algorithm

Pseudocode of ACOVNS is given in Algorithm 1.

Initialization step of the algorithm is given in lines 1-3. Algorithm parameters are initialized in line 2. Problem parameters are initialized and the data set is read in line 3. Lines 4-8 check if normalization is to be used.

From line 9 to 54, the algorithm's main loop is executed. The colony's solution construction, cluster centroid computation, and objective function value calculation are completed between lines 10-34. Between lines 11 and 26, a single solution is constructed by an ant, by assigning each of the samples to a cluster. Cluster centroids for the solution are calculated in lines 27-30. Using the cluster centroids, the objective function value is calculated in lines 31-33.

In line 35, the solutions are sorted and between lines 36-51 VNS is applied to L best solutions. Then, pheromone values are updated in lines 52-53. Finally, after N_{iter} iterations are performed in lines 9-54, the algorithm terminates in line 55.

Algorithm 1 Pseudocode of the ACOVNS Algorithm

```
Initialization:
Input algorithm parameters N_ants, N_iter, ρ, τ, P_lowest, V_MNS, L,
    NormalizationUsed, HeuristicInfoUsed
Read problem parameters N, N_c, N_f, and feature values of samples in the data
    set, x'_jf for all j and f; initialize S_i for all i to contain all N_f features in the data
    set
if NormalizationUsed = 1 then
    Normalize the feature values using Equation (3.1)
else
    Set x_jf = x'_jf for all j and f
end if
for iter = 1 to N_iter do
    for i = 1 to N_ants do
        Construct a solution:
        for j = 1 to N do
            Generate a uniform random number r between 0 and 1
            if r ≥ P_lowest then
                if HeuristicInfoUsed = 1 then
                    Calculate heuristic information for all k using Equation (3.5)
                    Calculate probabilities of selection for all k using Equation (3.6b)
                    Assign sample j to a cluster using roulette wheel selection
                else
                    Calculate probabilities of selection for all k using Equation (3.6a)
                    Assign sample j to a cluster using roulette wheel selection
                end if
            else
                Assign sample j to the cluster that has the lowest pheromone value
            end if
        end for j
        Calculate cluster centroids:
        for k = 1 to N_c do
```

Calculate centroid of cluster k using Equation (3.2)
end for
Calculate the objective function value:
Calculate Euclidean distances for all j and k using Equation (3.3)
Calculate the objective function value using Equation (3.4)
end for *i*
Sort constructed solutions in ascending order of objective function values and keep L best solutions
Perform local search on L **best solutions:**
for $i = 1$ to L **do**
Set $V_{NS} = 1$
$NewObjectiveValue$ = A very large value
while $V_{NS} < V_{MNS}$ and $NewObjectiveValue \geq F_i$ **do**
Generate V_{NS} random integers between 1 and N
Assign samples corresponding to random integers to their closest clusters in solution i
Calculate new cluster centroids for all k using Equation (3.2)
Calculate new Euclidean distances for all j and k using Equation (3.3)
Calculate $NewObjectiveValue$ using Equation (3.4)
V_{NS} = $V_{NS} + 1$
end while
if $NewObjectiveValue \leq F_i$ **then**
Set F_i = $NewObjectiveValue$ and replace original solution i with the improved one
end if
end for *i*
Pheromone update:
Update pheromone matrix for all j and k using Equation (3.7)
end for *iter*
return N_{ants} Clustering Solutions

3.6 Computational Results for ACOVNS

ACOVNS is implemented in Matlab R2021b. All runs where made on personal computers having 8GB RAM and i7-11800H processor. In this section, properties of real-world data sets, parameter settings, performance measures and results are given. Section 3.6.1 gives the information about data sets used in clustering experiments. Section 3.6.2 gives the parameter settings of the algorithm and describes performance measures used. Finally, Section 3.6.3 summarizes the experimental results and comparisons with the algorithms from the literature.

3.6.1 Data Sets

Clustering tests are conducted on nine real-world data sets. These data sets exhibit a variety of difficult properties and are frequently utilized in the clustering literature. The data sets have the characteristics presented in Table 3.2.

Table 3.2: Properties of Real-World Data Sets

Data Set	**Sample Size**	**Number of Features**	**Number of Clusters**	**Data Range**
Iris	150	4	3	0.1 - 7.9
Breast Cancer	683	9	2	1 - 10
Wine	178	13	3	0.13 - 1680
Glass	214	9	6	0 - 75.41
Thyroid	215	5	3	-0.7 - 144
Liver Disease	345	6	2	0 - 297
CMC	1473	9	3	0 - 49
Zoo	101	17	7	0 - 8
Wdbc	569	30	2	0 - 4254

In order to evaluate performance of the algorithm, data sets with different properties are selected. Except for Zoo, all of the data sets include exclusively numeric feature values. Zoo also contains some categorical features. The sample sizes in the selected data sets range from 101 to 1473, the number of features is between 4 and 30, and the number of clusters is between 2 and 7.

3.6.2 Parameter Settings and Performance Measures

ACOVNS algorithm takes the number of ants, the number of iterations, initial pheromone values, evaporation rate, probability of selecting the cluster with the lowest pheromone, the number of best solutions to perform VNS, and maximum neighborhood size of VNS as initial parameters. Experiments on these parameters are many in the literature, and the majority of research used 0.05 as the evaporation rate and 50 as the number of ants. Pheromone values are often initialized as 1. Also, the number of best solutions to perform a local search is usually taken as 5 in the studies using ACO (Shelokar et al., 2004). For these parameters, we use the same settings as in the literature. However, with the aid of some pilot runs we decided to use 0.005 as the probability of selecting the cluster with the lowest pheromone. It acts as a mutation operator in the genetic algorithms and this probability is a commonly used mutation rate in the literature (Haupt, 2000). Finally, we limit the maximum size of the VNS neighborhood to 10. Thus, it may attempt to alter the cluster assignments of no more than ten objects. The number of iterations is determined for each data set by examining the convergence behaviour.

We did, however, conduct experiments with four main alternative algorithm settings. In the first setting, the data set is not normalized and no heuristic information is employed. The data set is normalized in the second setting, but no heuristic information is used. The third setting employs heuristic information but does not normalize the data set. Finally, both heuristic information is applied and the data set is normalized in the fourth setting. For each data set, the algorithm was replicated 10 times with each configuration.

The performance of the algorithm is measured by two performance metrics, namely F-measure and Rand Index. "The geometric mean of the proportion of pairs that belong to the same cluster in both the algorithm's solution and the true partition, relative to the proportion of pairs that belong to the same cluster in each partition, is used to calculate the F-measure" (Brun et al., 2007). The formulation of F-measure is given in Equation (3.10), following the notation.

a_{jl}: 1 if samples j and l are assigned to the same cluster in true clustering solution, 0 otherwise

b_{jl}: 1 if samples j and l are assigned to the same cluster in the algorithm's solution, 0 otherwise

If $a_{jl} = b_{jl} = 1$, then True Positive (TP)
If $a_{jl} = b_{jl} = 0$, then True Negative (TN)
If $a_{jl} = 0$ and $b_{jl} = 1$, then False Positive (FP)
If $a_{jl} = 1$ and $b_{jl} = 0$, then False Negative (FN)

$$precision = \frac{TP}{TP + FP} \tag{3.8}$$

$$recall = \frac{TP}{TP + FN} \tag{3.9}$$

$$Fmeasure = 2\frac{(precision)(recall)}{precision + recall} \tag{3.10}$$

The Rand Index, on the other hand, "calculates the proportion of pairs of objects that agree by belonging to the same or distinct clusters in both the algorithm's solution and the true partition" (Brun et al., 2007). The formulation of Rand Index is given in Equation (3.11).

$$RI = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.11}$$

Summary of the experimental results are given in Table 3.3. A column titled best in the table displays the best result obtained out of 10 replications, whereas a column titled average displays the average value acquired from these 10 replications. The largest average in each row is shown in bold face. Detailed results for each replication of each data set are given in Table A.1 (see Appendix A).

Table 3.3 demonstrates that the algorithm yields the best average results (or tied for the best) in eight of the nine data sets in terms of F-measure and seven of the nine data

Table 3.3: Results for the Best and Average of 10 replications with Four Algorithm Settings

		Without Normalization and Heuristic Information		With Normalization only		With Heuristic Information only		With Both Normalization and Heuristic Information	
Data set	**Measure**	**Best**	**Average**	**Best**	**Average**	**Best**	**Average**	**Best**	**Average**
Iris	F-measure	0.870	**0.849**	0.880	0.848	0.869	0.842	0.880	0.835
	Rand Index	0.913	**0.899**	0.920	0.898	0.913	0.894	0.920	0.889
Wine	F-measure	0.593	0.591	0.934	**0.926**	0.604	0.593	0.934	**0.926**
	Rand Index	0.723	0.721	0.955	**0.950**	0.727	0.721	0.955	**0.950**
Breast Cancer	F-measure	0.945	**0.943**	0.943	0.942	0.943	**0.943**	0.943	**0.943**
	Rand Index	0.940	**0.938**	0.938	0.937	0.938	0.937	0.938	**0.938**
Glass	F-measure	0.378	0.351	0.407	0.360	0.481	**0.425**	0.410	0.389
	Rand Index	0.726	**0.711**	0.735	0.700	0.724	0.709	0.730	0.696
Thyroid	F-measure	0.700	0.642	0.591	0.580	0.712	0.650	0.850	**0.717**
	Rand Index	0.710	0.661	0.602	0.592	0.720	0.670	0.819	**0.708**
Liver Disease	F-measure	0.587	**0.587**	0.552	0.543	0.587	**0.587**	0.552	0.549
	Rand Index	0.500	**0.500**	0.501	**0.500**	0.500	**0.500**	0.501	**0.500**
CMC	F-measure	0.369	0.365	0.369	0.366	0.371	0.367	0.379	**0.368**
	Rand Index	0.562	**0.560**	0.560	0.557	0.561	**0.560**	0.559	0.556
Zoo	F-measure	0.769	0.643	0.871	0.753	0.803	0.745	0.966	**0.795**
	Rand Index	0.898	0.854	0.942	0.898	0.915	0.890	0.983	**0.914**
Wdbc	F-measure	0.809	0.806	0.876	0.876	0.809	0.804	0.879	**0.877**
	Rand Index	0.780	0.776	0.866	0.866	0.779	0.773	0.869	**0.867**

sets in terms of Rand Index when heuristic information is used. When both heuristic information and normalization are applied, the algorithm produces the best average results (or tied for the best) for six of the nine data sets. The algorithm achieves the best average results in fewer data sets using only normalization or only heuristic information. In general, we chose the setting in which both heuristic information and normalization were used for our further computational experiments for comparison because the algorithm found the best average results the most.

Table 3.4 summarizes the algorithm's computation times where normalization and heuristic information are unused. Table 3.5 summarizes the algorithm's computation times for normalized data sets. The algorithm's computation times are summarized in Table 3.6 when heuristic information is used but data sets are not normalized. Table 3.7 summarizes the algorithm's computation times where both normalization and heuristic information are used.

Table 3.4: Computation Times Without Normalization and Heuristic Information

Data Set	Number of Iterations	Objective Function Calculation (seconds)	Construction (Including Objective Calculation) (seconds)	Improvement (Local Search with VNS) (seconds)	Total Time per Replication (seconds)
Iris	400	4.69	9.50	1.11	10.61
Breast Cancer	1000	48.54	90.75	11.93	102.68
Wine	550	7.71	15.37	2.34	17.71
Glass	1100	17.72	35.06	6.12	41.18
Thyroid	900	14.12	28.20	4.36	32.56
Liver Disease	1100	27.78	53.94	7.72	61.66
CMC	3000	303.83	562.20	65.38	627.58
Zoo	600	5.70	11.69	3.33	15.03
Wdbc	1000	43.76	82.30	19.12	101.43

Table 3.5: Computation Times With Normalization Only

Data Set	Number of Iterations	Objective Function Calculation (seconds)	Construction (Including Objective Calculation) (seconds)	Improvement (Local Search with VNS) (seconds)	Total Time per Replication (seconds)
Iris	400	4.73	9.54	1.31	10.86
Breast Cancer	1000	48.52	90.37	12.97	103.34
Wine	550	7.79	15.40	3.19	18.59
Glass	1100	17.88	35.01	6.87	41.88
Thyroid	900	14.19	28.05	4.77	32.82
Liver Disease	1100	27.03	51.87	8.28	60.14
CMC	3000	308.24	568.60	76.46	645.06
Zoo	600	5.23	10.30	3.01	13.31
Wdbc	1000	42.40	77.25	20.55	97.79

Table 3.6: Computation Times With Heuristic Information Only

Data Set	Number of Iterations	Objective Function Calculation (seconds)	Construction (Including Objective Calculation) (seconds)	Improvement (Local Search with VNS) (seconds)	Total Time per Replication (seconds)
Iris	150	1.89	9.85	0.59	10.45
Breast Cancer	400	20.60	79.49	5.30	84.80
Wine	200	3.10	15.01	0.84	15.84
Glass	450	8.13	60.83	3.04	63.87
Thyroid	450	7.59	38.77	2.46	41.23
Liver Disease	400	10.23	40.72	2.68	43.39
CMC	500	55.16	265.47	11.80	277.27
Zoo	250	2.68	19.64	1.31	20.96
Wdbc	400	19.09	68.70	8.01	76.71

Table 3.7: Computation Times With Both Normalization and Heuristic Information

Data Set	Number of Iterations	Objective Function Calculation (seconds)	Construction (Including Objective Calculation) (seconds)	Improvement (Local Search with VNS) (seconds)	Total Time per Replication (seconds)
Iris	150	1.88	9.48	0.62	10.10
Breast Cancer	400	20.86	81.45	6.01	87.47
Wine	250	3.74	18.37	1.23	19.59
Glass	450	8.26	62.22	3.16	65.37
Thyroid	600	10.00	50.70	3.64	54.34
Liver Disease	400	10.46	42.29	2.92	45.21
CMC	500	55.91	272.09	9.41	281.50
Zoo	250	2.52	18.72	1.06	19.79
Wdbc	400	18.61	66.77	8.48	75.25

Computation time tables demonstrate that when heuristic information is used, the algorithm converges in fewer iterations. Thus, even though using heuristic information increases the computing time for a single iteration, the total time is reduced dramatically when the data set contains more than 250 samples. Most of the total time is spent on solution construction whereas improvement time with VNS is fairly low.

3.6.3 Computational Results and Comparisons

We compared ACOVNS with algorithms that are commonly used in the literature for comparisons. Most studies in the literature use F-measure, therefore we compared the results in terms of F-measure. Computing times are takes as they are reported in respective articles and given only for informative purposes. Because each data set was solved in a different number of studies, comparisons were made in separate tables for each data set. In the studies utilized for comparison, the findings are provided as the average of several algorithm replications. As a result, we used the average F-measure and computing times obtained across replications of algorithms.

The comparisons for the Iris data set are given in Table 3.8. ACOVNS outperforms 19 of the 21 algorithms in terms of F-measure.

Table 3.8: Comparisons on Iris Data Set

Algorithm	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.782	0.4
GA	Singh and Kumar (2020)	0.778	140
PSO	Singh and Kumar (2020)	0.782	30
ACO	Singh and Kumar (2020)	0.779	75
CSS	Y. Kumar and Sahoo (2015)	0.787	56
CCSSA	Y. Kumar and Sahoo (2015)	0.791	48
MCSS	Y. Kumar and Sahoo (2015)	0.79	53
MCSS-PSO	Y. Kumar and Sahoo (2015)	0.794	41
CSO	Singh and Kumar (2020)	0.776	-
P-CSO	Singh and Kumar (2020)	0.784	-
ACO-SA	Niknam et al. (2008)	0.786	25
PSO-SA	Niknam et al. (2008)	0.785	17
SA	Fathian and Amiri (2008)	0.776	32
TS	Fathian and Amiri (2008)	0.777	135
HBMO	Fathian and Amiri (2008)	0.781	82
PSO-ACO	Niknam et al. (2008)	0.787	17
PSO-ACO-k	Niknam and Amiri (2010)	0.788	16
EM	Mehdizadeh et al. (2017)	0.81	-
K-EM	Mehdizadeh et al. (2017)	0.832	-
MEM	Mehdizadeh et al. (2017)	0.862	-
K-MEM	Mehdizadeh et al. (2017)	0.898	-
ACOVNS	Proposed Algorithm	0.835	10

The results of the algorithms for the Breast Cancer data set are reported in Table 3.9. ACOVNS finds better solutions than 20 of the 21 algorithms. However, in the Breast Cancer data set ACOVNS is slower than 5 of the algorithms including k-means.

Table 3.10 presents the performance of the algorithms for the Wine data set. ACOVNS shows superior performance among all of the algorithms and performs 0.19 better than its closest competitor in terms of F-measure. However, ACOVNS's F-measure

Table 3.9: Comparisons on Breast Cancer Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.829	0.5
GA	Singh and Kumar (2020)	0.819	135
PSO	Singh and Kumar (2020)	0.819	123
ACO	Singh and Kumar (2020)	0.821	123
CSS	Y. Kumar and Sahoo (2015)	0.847	134
CCSSA	Y. Kumar and Sahoo (2015)	0.856	130
MCSS	Y. Kumar and Sahoo (2015)	0.859	126
MCSS-PSO	Y. Kumar and Sahoo (2015)	0.863	112
CSO	Singh and Kumar (2020)	0.831	-
P-CSO	Singh and Kumar (2020)	0.833	-
ACO-SA	Niknam et al. (2008)	0.829	40
PSO-SA	Niknam et al. (2008)	0.829	28
SA	Fathian and Amiri (2008)	0.818	126
TS	Fathian and Amiri (2008)	0.818	130
HBMO	Fathian and Amiri (2008)	0.825	136
PSO-ACO	Niknam et al. (2008)	0.83	17
PSO-ACO-k	Niknam and Amiri (2010)	0.83	16
EM	Mehdizadeh et al. (2017)	0.814	-
K-EM	Mehdizadeh et al. (2017)	0.866	-
MEM	Mehdizadeh et al. (2017)	0.902	-
K-MEM	Mehdizadeh et al. (2017)	0.964	-
ACOVNS	Proposed Algorithm	0.943	87

is 0.40-0.45 higher than those of the majority of algorithms.

Results for the Glass data set are given in Table 3.11. Because the Glass data set is one of the most difficult in the literature, the majority of algorithms produce F-measure values less than 0.5. ACOVNS outperforms only five of the algorithms on the Glass data set, while the remaining algorithms outperform ACOVNS. In the Glass data set, it scores poorly in terms of solution quality.

Results for the Thyroid data set are summarized in Table 3.12. ACOVNS produces the lowest outcome of all the algorithms in the Thyroid data set.

The results for the Liver Disease data set are given in Table 3.13. The table demonstrates that ACOVNS outperforms the other algorithms in terms of F-measure. While the other algorithms may achieve a maximum F-measure of 0.496, ACOVNS achieves

Table 3.10: Comparisons on Wine Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.521	0.9
GA	Singh and Kumar (2020)	0.515	170
PSO	Singh and Kumar (2020)	0.518	123
ACO	Singh and Kumar (2020)	0.519	121
CSS	Y. Kumar and Sahoo (2015)	0.529	126
CCSSA	Y. Kumar and Sahoo (2015)	0.535	124
MCSS	Y. Kumar and Sahoo (2015)	0.537	120
MCSS-PSO	Y. Kumar and Sahoo (2015)	0.542	114
CSO	Singh and Kumar (2020)	0.521	-
P-CSO	Singh and Kumar (2020)	0.523	-
ACO-SA	Niknam et al. (2008)	0.52	84
PSO-SA	Niknam et al. (2008)	0.52	38
SA	Fathian and Amiri (2008)	0.515	129
TS	Fathian and Amiri (2008)	0.516	140
HBMO	Fathian and Amiri (2008)	0.518	121
PSO-ACO	Niknam et al. (2008)	0.519	33
PSO-ACO-k	Niknam and Amiri (2010)	0.521	30
EM	Mehdizadeh et al. (2017)	0.724	-
K-EM	Mehdizadeh et al. (2017)	0.694	-
MEM	Mehdizadeh et al. (2017)	0.702	-
K-MEM	Mehdizadeh et al. (2017)	0.729	-
ACOVNS	Proposed Algorithm	0.926	19

a value of 0.549, which is more than 5 percent higher.

Table 3.14 summarizes the results for CMC data set. CMC is another data set that is widely acknowledged as incredibly difficult to analyze in the literature. The majority of the F-measure values found for this data set are less than 0.4. However, when compared to published findings, ACOVNS outperforms 16 of the 21 algorithms, although it is the slowest for this data set.

The algorithm results for the Zoo data set are presented in Table 3.15. This data set is distinct from the others in that it includes categorical features in addition to numerical features. ACOVNS outperforms 3 of 4 algorithms.

Finally, the results for Wdbc data set are given in Table 3.16. Compared with the six algorithms in the literature, ACOVNS performs better than 5 of them. However, the highest result is almost the same as the result obtained by ACOVNS with a very small difference of 0.001.

Table 3.11: Comparisons on Glass Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.431	1
GA	Singh and Kumar (2020)	0.333	410
PSO	Singh and Kumar (2020)	0.359	400
ACO	Singh and Kumar (2020)	0.364	395
CSS	Y. Kumar and Sahoo (2015)	0.446	418
CCSSA	Y. Kumar and Sahoo (2015)	0.453	402
MCSS	Y. Kumar and Sahoo (2015)	0.449	406
MCSS-PSO	Y. Kumar and Sahoo (2015)	0.454	391
CSO	Singh and Kumar (2020)	0.416	-
P-CSO	Singh and Kumar (2020)	0.424	-
ACO-SA	Niknam et al. (2008)	0.431	49
PSO-SA	Niknam et al. (2008)	0.43	38
SA	Fathian and Amiri (2008)	0.347	410
TS	Fathian and Amiri (2008)	0.351	410
HBMO	Fathian and Amiri (2008)	0.401	390
PSO-ACO	Niknam et al. (2008)	0.434	35
PSO-ACO-k	Niknam and Amiri (2010)	0.435	31
EM	Mehdizadeh et al. (2017)	0.53	-
K-EM	Mehdizadeh et al. (2017)	0.554	-
MEM	Mehdizadeh et al. (2017)	0.554	-
K-MEM	Mehdizadeh et al. (2017)	0.558	-
ACOVNS	Proposed Algorithm	0.389	65

Table 3.12: Comparisons on Thyroid Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.731	0.4
GA	Singh and Kumar (2020)	0.763	153
PSO	Singh and Kumar (2020)	0.778	102
ACO	Singh and Kumar (2020)	0.783	98
CSS	Kumar and Sahoo (2015)	0.789	108
CCSSA	Kumar and Sahoo (2015)	0.792	105
MCSS	Kumar and Sahoo (2015)	0.793	103
MCSS-PSO	Kumar and Sahoo (2015)	0.794	94
CSO	Singh and Kumar (2020)	0.774	-
P-CSO	Singh and Kumar (2020)	0.783	-
ACOVNS	Proposed Algorithm	0.717	54

Table 3.13: Comparisons on Liver Disease Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.467	0.45
GA	Singh and Kumar (2020)	0.482	145
PSO	Singh and Kumar (2020)	0.493	117
ACO	Singh and Kumar (2020)	0.487	122
CSS	Y. Kumar and Sahoo (2015)	0.491	124
CCSSA	Y. Kumar and Sahoo (2015)	0.494	119
MCSS	Y. Kumar and Sahoo (2015)	0.495	123
MCSS-PSO	Y. Kumar and Sahoo (2015)	0.498	104
CSO	Singh and Kumar (2020)	0.485	-
P-CSO	Singh and Kumar (2020)	0.496	-
ACOVNS	Proposed Algorithm	0.549	45

In general, ACOVNS produces promising and superior solutions for the majority of data sets. However, because data sets vary in their characteristics, its performance is inferior on some data sets. Specifically, ACOVNS yields solutions with a lower quality than the k-means method in the Glass and Thyroid data sets. In Chapter 4, we modify the ACOVNS algorithm to carry out feature selection and clustering simultaneously and improve its clustering performance by selecting only the relevant features.

Table 3.14: Comparisons on CMC Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Mehdizadeh et al. (2017)	0.334	0.5
GA	Singh and Kumar (2020)	0.324	160
PSO	Singh and Kumar (2020)	0.331	131
ACO	Singh and Kumar (2020)	0.328	127
CSS	Y. Kumar and Sahoo (2015)	0.359	129
CCSSA	Y. Kumar and Sahoo (2015)	0.367	126
MCSS	Y. Kumar and Sahoo (2015)	0.368	127
MCSS-PSO	Y. Kumar and Sahoo (2015)	0.374	123
CSO	Singh and Kumar (2020)	0.334	-
P-CSO	Singh and Kumar (2020)	0.336	-
ACO-SA	Niknam et al. (2008)	0.333	89
PSO-SA	Niknam et al. (2008)	0.333	73
SA	Fathian and Amiri (2008)	0.325	150
TS	Fathian and Amiri (2008)	0.327	155
HBMO	Fathian and Amiri (2008)	0.33	122
PSO-ACO	Niknam et al. (2008)	0.333	35
PSO-ACO-k	Niknam and Amiri (2010)	0.334	31
EM	Mehdizadeh et al. (2017)	0.4	-
K-EM	Mehdizadeh et al. (2017)	0.356	-
MEM	Mehdizadeh et al. (2017)	0.4	-
K-MEM	Mehdizadeh et al. (2017)	0.405	-
ACOVNS	Proposed Algorithm	0.368	281

Table 3.15: Comparisons on Zoo Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Sun et al. (2019)	0.712	-
DBSCAN	Sun et al. (2019)	0.792	-
GWO	Sun et al. (2019)	0.852	-
GSA	Sun et al. (2019)	0.789	-
ACOVNS	Proposed Algorithm	0.795	19

Table 3.16: Comparisons on Wdbc Data Set

Algorithms	Reference	F-measure	Runtime (seconds)
k-means	Sridevi and Murugan (2014)	0.731	-
Single Linkage	Sridevi and Murugan (2014)	0.729	-
DBSCAN	Sridevi and Murugan (2014)	0.878	-
Spectral Clustering	Sridevi and Murugan (2014)	0.760	-
CSM	Sridevi and Murugan (2014)	0.720	-
CHAMELEON	Sridevi and Murugan (2014)	0.851	-
ACOVNS	Proposed Algorithm	0.877	75

CHAPTER 4

SOLUTION APPROACH FOR SIMULTANEOUS FEATURE SELECTION AND CLUSTERING: F-ACOVNS

We extend the ACOVNS algorithm to perform feature selection and clustering simultaneously. The modified version of ACOVNS (F-ACOVNS) selects a subset of features while performing the clustering. To accomplish this, the algorithm memorizes information on features and processes these information to select the most relevant features. The overall structure of the F-ACOVNS algorithm is presented in Figure 4.1.

The details of F-ACOVNS and its feature selection mechanism are discussed in Section 4.1. The additional notation and formulations used in F-ACOVNS are presented in Sections 4.2 and 4.3, respectively. Following that, the pseudocode of the algorithm is given in Section 4.4. The real-world and synthetic data sets used in experiments are presented in Section 4.5.1. In Section 4.5.2 the parameter tuning results and performance measures are explained. The experimental results for real-world data sets and synthetic data sets are given in Sections 4.5.3 and 4.5.4, respectively.

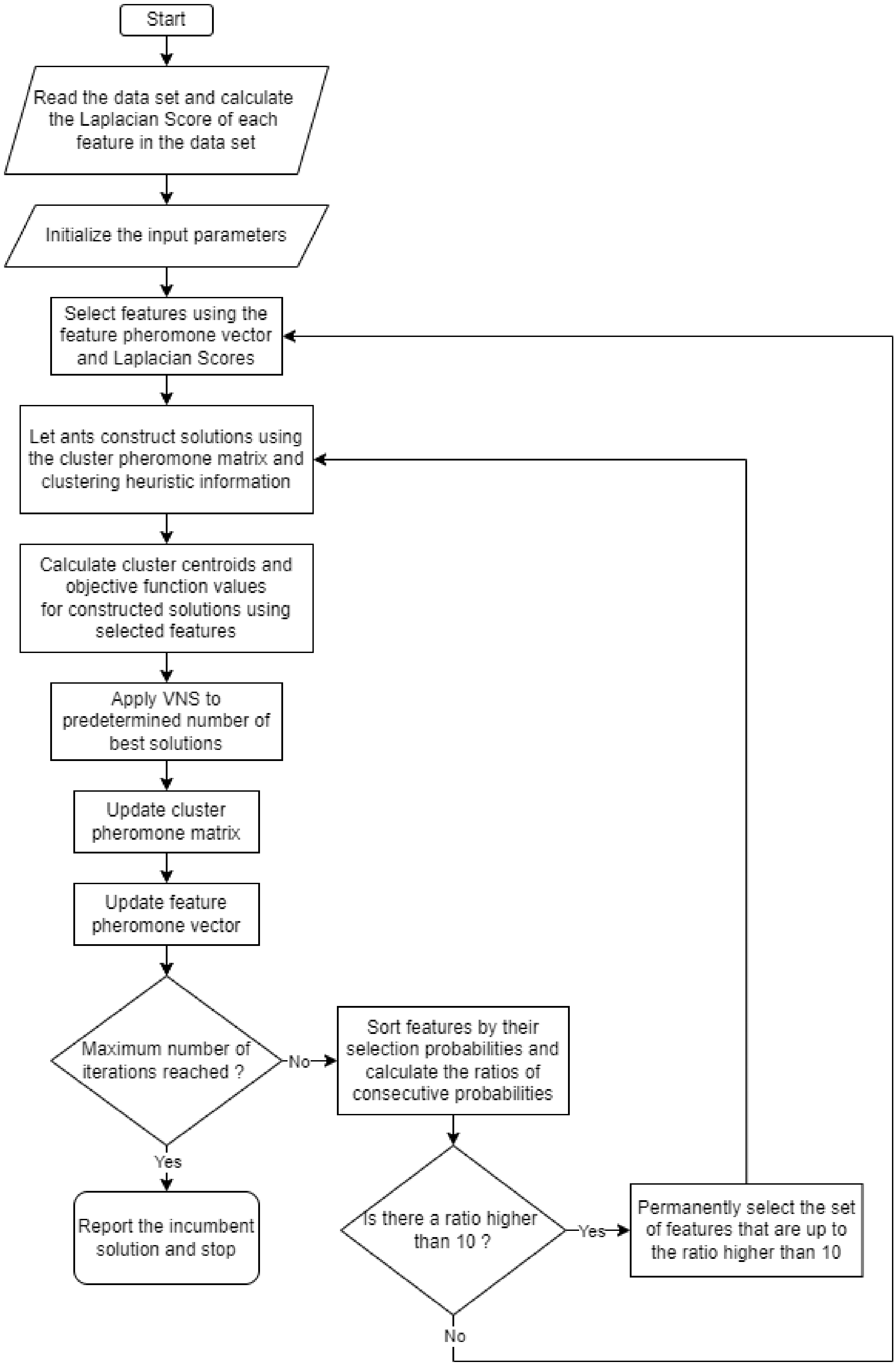

Figure 4.1: Flowchart of the F-ACOVNS algorithm

4.1 F-ACOVNS Algorithm Features

F-ACOVNS algorithm uses ACOVNS as the clustering method, however it also selects a subset of features simultaneously. The algorithm employs two new mechanisms which are the feature pheromone vector and the feature heuristic information. The feature pheromone vector works in the same way as the clustering pheromone matrix. It stores the pheromone value for each feature. However, the feature heuristic information is an entirely novel method. The Laplacian Score (LS) is used for feature heuristic information, therefore it is a static measure. LS is computed for each feature at the beginning of the algorithm, and these scores are used throughout the remainder of the algorithm.

Feature selection in F-ACOVNS is not straightforward. We calculate each feature's selection probability using the feature pheromone vector and heuristic information, similar to cluster selection probabilities. However, there may be situations where none of the features are selected. To deal with such situations, we divide each feature's probability by the maximum of all features. Hence, we guarantee that at least one of the features (the one with the maximum probability) will be selected for sure and probabilities of the others will be in range (0,1). Furthermore, this way we prevent other feature probabilities from decreasing rapidly across the iterations, which may result in premature convergence.

The algorithm selects features using their probabilities at the start of each iteration. Then, it follows the steps of ACOVNS using only the selected features until the end of pheromone update. After updating the clustering pheromone matrix, F-ACOVNS updates the feature pheromone vector and sorts all features in descending order of their probabilities of selection. In the sorted feature list, the algorithm checks the ratio of probabilities for each pair of consecutive features. If there is a ratio higher than 10, F-ACOVNS permanently selects the features in the list up until the one with the ratio of 10. In the rest of the iterations, it performs clustering by using only these selected features. The threshold value of 10 for the ratios is determined by examining progress of selection probabilities in pilot runs and whether or not features selected in this manner lead to good clustering solutions. If reached, the threshold indicates a break point between relevant and redundant features. Otherwise, if the threshold is not exceeded, the algorithm repeats the simultaneous feature selection and clustering procedure until maximum number of iterations is reached.

Additionally, since we concluded that the ACOVNS setting where both feature normalization and heuristic information are used is the best among others, we used this setting in F-ACOVNS.

4.2 Additional Notation for Simultaneous Feature Selection and Clustering

The following notation is used in the forthcoming formulations and pseudocode of the algorithm in addition to the notation given in Section 3.2.

L_f	Laplacian score of feature f
z_{if}	1 if feature f is selected in ant i's solution, 0 otherwise
sT_f	Pheromone value of selecting feature f
sh_f	Heuristic information for feature f
sP_f	Probability of selecting feature f
$s\rho$	Feature evaporation rate
$s\tau$	Initial pheromone vector for feature selection where $sT_f = 1$ for all f
I	Number of selected features in the incumbent solution
$FeatureSelectionOn$	1 if feature selection is continuing, 0 otherwise

4.3 Formulations for Simultaneous Feature Selection and Clustering

Formulations for computing cluster centroids, Euclidean distances, heuristic information for clustering, probabilities of cluster selection, and pheromone update for clustering are the same as in Section 3.3. Additional formulations are given below.

Objective Function

F-ACOVNS employs a single objective function that is based on Euclidean distances but uses only selected features. The objective function value for each solution is computed by adding the Euclidean distances between each sample and its allocated cluster centroid. The lower the objective function value is, the higher the solution quality is. Since we use different numbers of selected features while calculating the objective function values, Equation (4.1) replaces Equation (3.4) in F-ACOVNS.

$$F_i = \frac{1}{|S_i|} \sum_{j=1}^{N} \sum_{k=1}^{N_c} y_{ijk} d_i(x_j, c_{ik}) \quad \forall i \tag{4.1}$$

As the number of features used in each ant's solution may be different, the number of terms summed up to compute the Euclidean distances in Equation (3.3) also differs. Hence, the objective function value is divided by the number of selected features to make different solutions comparable.

Heuristic Information for Feature Selection

The inverse of the LS is used as feature heuristic information in the algorithm as can be seen in Equation (4.2) since a lower LS value means the feature is more relevant. LS for feature f is computed by Equation (2.1) where neighborhood size is taken as $M = 0.2N$.

$$sh_f = \frac{1}{L_f} \quad \forall f \tag{4.2}$$

Probability of Feature Selection

Probability of selecting a feature is calculated by using both the feature pheromone vector and the feature heuristic information. At first individual probability of selection for each feature is calculated by Equation (4.3). Then, the probability of the feature with the highest probability of selection is set to 1 by dividing it to itself. This ensures selection of at least one feature. After that, other feature probabilities are also divided by the highest one as in Equation (4.4) to upscale them proportionally to the highest probability.

$$sP'_f = \frac{sT_f sh_f}{\sum_{m=1}^{N_f} sT_m sh_m} \quad \forall f \tag{4.3}$$

$$sP_f = \frac{sP'_f}{\max_f\{sP'_f\}} \quad \forall f \tag{4.4}$$

Pheromone Update for Feature Selection

Pheromone update for feature selection is similar to that for clustering. The pheromone values of features are increased if they are selected in the L best solutions as can be seen in Equation (4.5). However, using the inverse of objective function value directly occasionally results in very large pheromone values for a few features, which may cause them to dominate the clustering process. To avoid this, we use the inverse of the number of features selected in the incumbent solution $(1/I)$ as a scaling factor, which is a common practice in ACO.

$$sT_f = (1 - s\rho)sT_f + \sum_{i=1}^{L} z_{if} \frac{\frac{1}{I}}{F_i} \quad \forall f \tag{4.5}$$

4.4 Description of the F-ACOVNS Algorithm

Pseudocode of F-ACOVNS is given in Algorithm 2.

Initialization step of the algorithm is given in lines 1-5. In line 2, algorithm parameters are initialized. Problem parameters are initialized and data set is read in line 3. In line 4, the data set is normalized and in line 5, Laplacian Scores are calculated using normalized feature values.

From line 6 to 34, simultaneous feature selection and clustering is executed. Between lines 8 and 18, feature subset selection is made. In lines 19 and 21, the algorithm calls ACOVNS to construct a clustering solution and perform local search, respectively. Between lines 22 and 33, feature pheromone values are updated, feature selection probabilities are recalculated and sorted, and then ratios of consecutive features are checked. If there is a ratio higher than 10, the algorithm exits the loop between lines 6 and 34, and directly moves to line 35.

In line 36, the algorithm checks whether a feature subset is selected permanently. If so and N_{iter} iterations have not been finished yet, in the rest of the iterations the algorithm performs clustering only using the selected features between lines 37-39. Otherwise, the algorithm goes to line 41 and terminates immediately.

Algorithm 2 Pseudocode of the F-ACOVNS Algorithm

1: **Initialization:**
2: Input algorithm parameters N_{ants}, N_{iter}, ρ, $s\rho$, τ, $s\tau$, P_{lowest}, V_{MNS}, L, M, $FeatureSelectionOn$
3: Read problem parameters N, N_c, N_f, and feature values of samples in the data set, x'_{jf} for all j and f
4: Normalize the feature values using Equation (3.1)
5: Calculate Laplacian Scores using Equation (2.1) and heuristic information for features using Equation (4.2)
6: **for** $iter = 1$ to N_{iter} **do**
7:     **for** $i = 1$ to N_{ants} **do**
8:         **Construct a solution:**
9:         Calculate probabilities of feature selection for all f using Equations (4.3) and (4.4).
10:         **for** $f = 1$ to N_f **do**
11:             Generate a uniform random number r between 0 and 1
12:             **if** $r \leq sP_f$ **then**
13:                 Set z_{if} = 1
14:                 Add the selected feature f to set S_i
15:             **else**
16:                 Set z_{if} = 0
17:             **end if**
18:         **end for** ***f***
19:         Invoke lines 12-33 of Algorithm 1 for clustering by using Equation (4.1) instead of Equation (3.4) for calculating the objective function value.
20:     **end for** ***i***
21:     Invoke lines 35-53 of Algorithm 1 for local search and update of clustering pheromone matrix by using Equation (4.1) instead of Equation (3.4) for calculating the objective function values.
22:     **Feature selection pheromone update:**
23:     Update feature selection pheromone vector for all f using Equation (4.5)

24: **Permanent selection of features:**
25: Calculate probability of selection for each feature using Equations (4.3) and (4.4).
26: Sort features in descending order of their selection probabilities and calculate ratio of probabilities for each pair of consecutive features.
27: **for** $f = 2$ to N_f **do**
28: **if** $sP_{f-1}/sP_f \geq 10$ **then**
29: Permanently select all features before f and update S_i
30: Set $FeatureSelectionOn = 0$ and let $currentIter$ = $iter$
31: **exit for** *iter* and go to line 35
32: **end if**
33: **end for** *f*
34: **end for** *iter*
35: **Clustering with permanently selected features:**
36: **if** $FeatureSelectionOn = 0$ **then**
37: **for** $iter = 1$ to $N_{iter} - currentIter$ **do**
38: Invoke lines 10-53 of Algorithm 1 by using feature subset S_i and Equation (4.1) instead of Equation (3.4) for calculating the objective function values.
39: **end for**
40: **end if**
41: **return** N_{ants} Solutions

4.5 Computational Results for F-ACOVNS Settings

The F-ACOVNS algorithm is also implemented in MATLAB and experiments are made by using the same computer as ACOVNS. For calculating the Laplacian Scores, feature selection repository scikit-feature in Python developed by Data Mining and Machine Learning Laboratory at Arizona State University is used (Li et al., 2018). To begin with, the evaporation rate for feature pheromone update in F-ACOVNS is fine-tuned in Section 4.5.2. F-ACOVNS is then tested on both real-world and synthetic data sets. In Section 4.5.3, the performance of F-ACOVNS is first compared with ACOVNS by using the data sets presented in Section 3.6.1. Then, for some additional real-world data sets given in Section 4.5.1, the algorithm is compared with both single and multi-objective algorithms from literature. In Section 4.5.4, the algorithm is compared with MOCNC-F and Δ-MOCK-F (Alakuş, 2018) on synthetic data sets.

4.5.1 Data Sets

In addition to the data sets used in clustering experiments, nine more real-world data sets are used in computational experiments for simultaneous feature selection and clustering. The details of these additional real-world data sets are given in Table 4.1.

Table 4.1: Properties of Real-World Data Sets Used for Feature Selection

Data Set	Sample Size	Number of Features	Number of Clusters	Data Range
Ionosphere	351	34	2	-1 - 1
Sonar	208	60	2	0 - 1
Dermatology	358	34	6	0 - 75
Appendicitis	106	7	2	0 - 1
Breast Tissue	106	9	6	-9.25 - 174480.5
Pima	768	8	2	0 - 846
Ecoli	336	7	5	0 - 1
Parkinsons	195	22	2	-7.96 - 592.03
Spectf	267	44	2	3 - 89

These data sets are frequently used in the feature selection literature due to their characteristics. Some of these data sets are high dimensional and the number of features selected in them varies in the literature. Therefore, we use these data sets to compare performance of F-ACOVNS with some other algorithms.

Additionally, synthetic data sets generated for simultaneous feature selection and clustering by Alakuş (2018) are used to evaluate the performance of F-ACOVNS. For each setting given in the first column of Table 4.2, 10 different generated data sets are used in the experiments. In the names of synthetic data sets, the number before the letter "d" indicates the original number of features, the number before the letter "c" shows the number of clusters, and finally the number after the underscore refers to the number of redundant features added to the data set.

Table 4.2: Properties of Synthetic Data Sets Used for Feature Selection

Data Set	Sample Size	Number of Original Features	Number of Redundant Features Added	Number of Clusters
5d-5c_1	1561.2	5	1	5
5d-5c_3	1561.2	5	3	5
5d-5c_5	1561.2	5	5	5
10d-5c_2	1460.4	10	2	5
10d-5c_5	1460.4	10	5	5
10d-5c_10	1460.4	10	10	5
20d-5c_4	1405.3	20	4	5
20d-5c_10	1405.3	20	10	5
20d-5c_20	1405.3	20	20	5
40d-5c_8	1557	40	8	5
40d-5c_20	1557	40	20	5
40d-5c_40	1557	40	40	5

The sample size column in Table 4.2 is the average number of samples in 10 different data sets generated for the same setting.

4.5.2 Parameter Settings and Performance Measures

F-ACOVNS takes two additional parameters as input, which are the feature evaporation rate and the neighborhood size for LS. Neighborhood size used in LS calculation is taken as one fifth of the sample size as a result of the pilot runs. However, detailed feature evaporation rate runs are made for tuning the parameter. For each of the settings, 5 algorithm replications are made and the average Rand Index values are found. Results of the feature evaporation rate tuning are illustrated in Figures 4.2, 4.3, and 4.4. Detailed results for each replication of each data set are given in Tables B.1 - B.21 (see Appendix B).

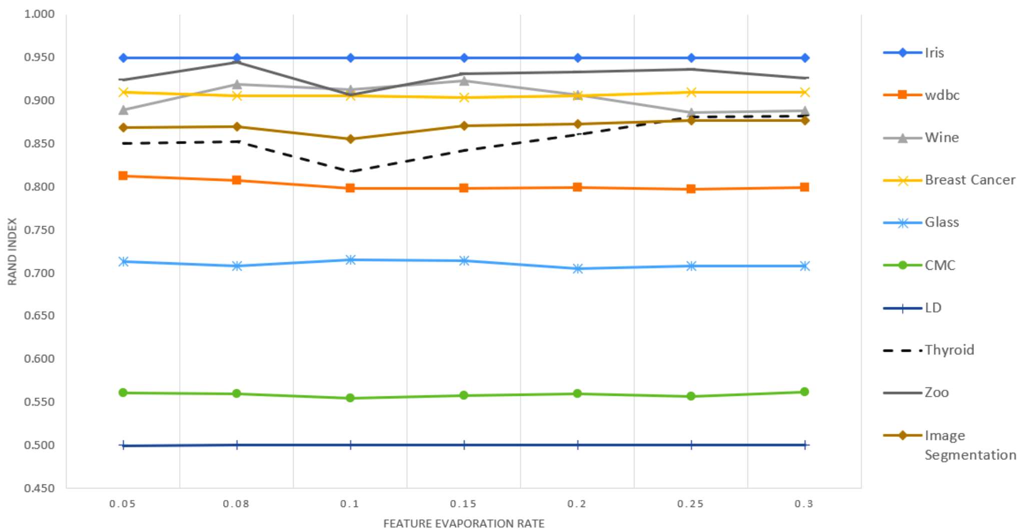

Figure 4.2: Feature Evaporation Rate Tuning on Real-World Data Sets

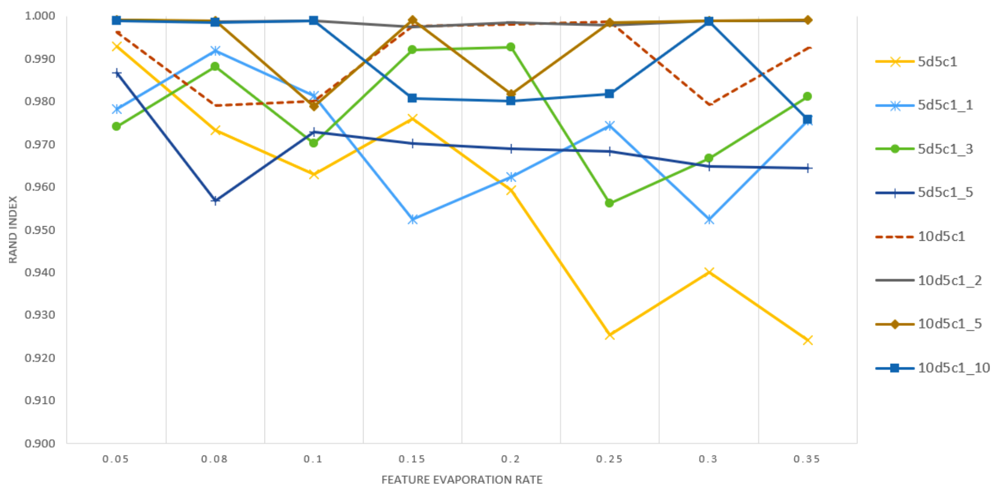

Figure 4.3: Feature Evaporation Rate Tuning on Synthetic Data Sets with at Most 10 Original Features

The figures demonstrate that the best feature evaporation rate may be different for different data sets.. In the majority of real-world data sets, the algorithm's performance

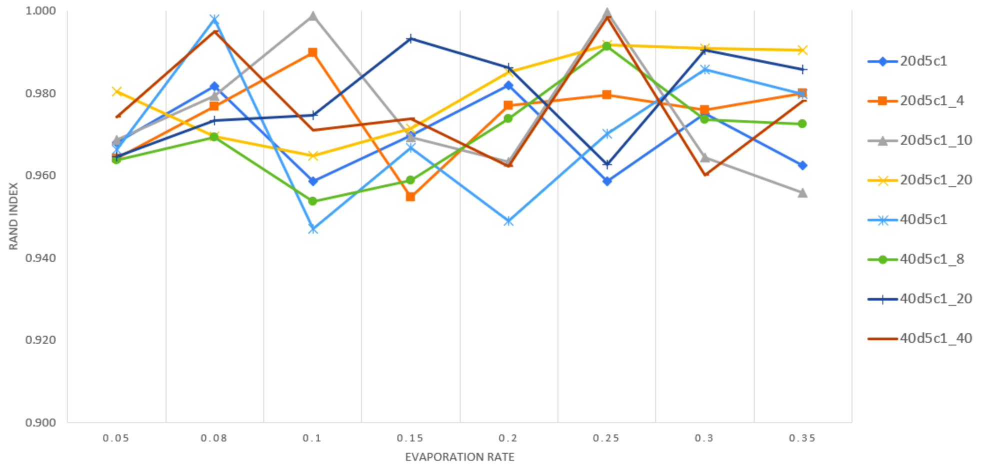

Figure 4.4: Feature Evaporation Rate Tuning on Synthetic Data Sets with More Than 10 Original Features

is unaffected by the feature evaporation rate. In Thyroid, Zoo, and Wine data sets, however, the algorithm's performance changes depending on the feature evaporation rate. Therefore, we take the individual best setting for each real-world data set.

We examined the effect of the feature evaporation rate on synthetic data sets in two cases. The first contains data sets with at most 10 original features, while the second contains the remaining data sets with more features. Since the algorithm's performance on the synthetic data sets varied widely, we chose the parameters that produced better results for each data set group and increased the number of replications to 10 for these settings. For this purpose, feature evaporation rates 0.05, 0.08, 0.1, and 0.15 were tried for data sets with at most 10 original features, and rates 0.25, 0.3, and 0.35 were tried for the rest of the data sets. The average Rand Index results obtained by 10 replications are given in Figures 4.5 and 4.6. Detailed results for each replication of each data set are given in Tables B.22 - B.41 (see Appendix B).

As a result of these experiments, we choose the settings that give the best results in most of the data sets for each case. Therefore, feature evaporation rate is set as 0.15 for the data sets with at most 10 original features and 0.3 for the data sets with more features.

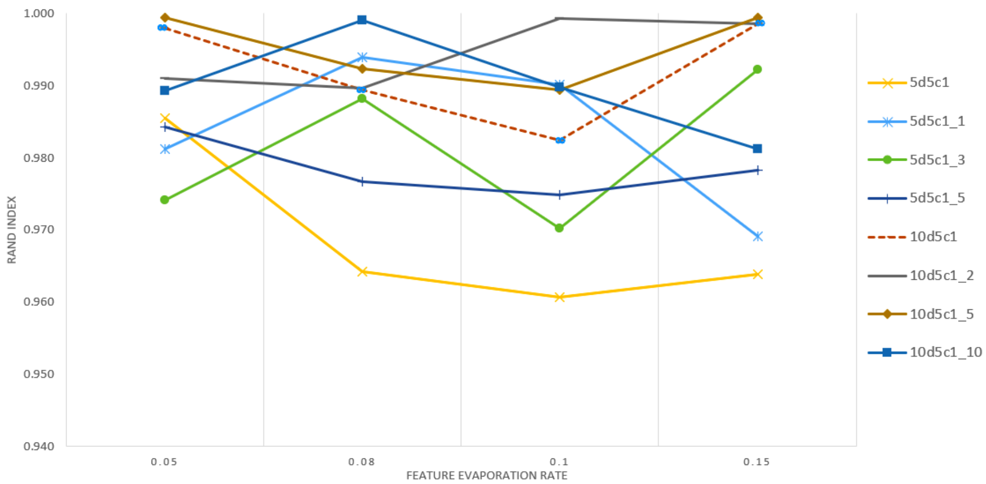

Figure 4.5: Feature Evaporation Rate Tuning on Synthetic Data Sets with at Most 10 Original Features (10 replications)

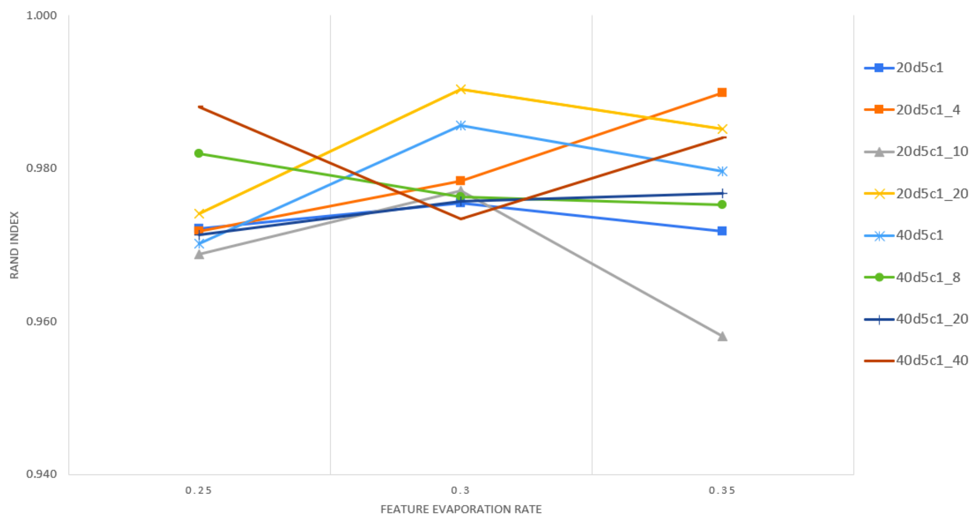

Figure 4.6: Feature Evaporation Rate Tuning on Synthetic Data Sets with More Than 10 Original Features (10 replications)

An alternative objective function for F-ACOVNS is also considered. Unlike Equation (4.1), distances are scaled by the square root of the number of selected features in this objective function. Formulation of the alternative objective function is given in

Equation (4.6).

$$F_i = \sum_{j=1}^{N}\sum_{k=1}^{N_c} y_{ijk}\sqrt{\frac{d_i(x_j, c_{ik})^2}{|S_i|}} = \sum_{j=1}^{N}\sum_{k=1}^{N_c} y_{ijk}\frac{d_i(x_j, c_{ik})}{\sqrt{|S_i|}}$$
$$= \frac{1}{\sqrt{|S_i|}}\sum_{j=1}^{N}\sum_{k=1}^{N_c} y_{ijk} d_i(x_j, c_{ik}) \quad \forall i \tag{4.6}$$

Performance of the two objective functions are compared using selected feature evaporation rates on the real world data sets used for ACOVNS and some of the synthetic data sets. Table 4.3 shows the results of F-ACOVNS with different objective functions.

Table 4.3: F-ACOVNS Performance Comparison with Different Objective Functions (Average of 10 Replications)

Data Set	**Using Objective Function** (4.6)			**Using Objective Function** (4.1)		
	Number of Selected Features	**F-Measure**	**Rand Index**	**Number of Selected Features**	**F-Measure**	**Rand Index**
Iris	2	0.812	0.875	3	**0.925**	**0.950**
Wine	1	0.430	0.605	8.1	**0.885**	**0.922**
Breast Cancer	2	0.803	0.775	5.6	**0.918**	**0.910**
Glass	8	0.426	0.708	8	**0.427**	**0.715**
Thyroid	7.8	0.429	0.705	4	**0.897**	**0.882**
Liver Disease	1	**0.599**	**0.501**	4.5	0.557	0.500
CMC	1	**0.380**	0.554	7.3	0.367	**0.562**
Zoo	1.5	0.704	0.855	7.2	**0.878**	**0.945**
Wdbc	1	0.813	0.784	13.7	**0.834**	**0.812**
5d-5c_1	4	0.893	0.950	5	**0.988**	**0.994**
5d-5c_3	4.3	0.867	0.937	5	**0.986**	**0.993**
20d-5c_20	2.5	0.781	0.890	16.4	**0.981**	**0.990**
40d-5c_40	4.2	0.538	0.782	26.6	**0.982**	**0.991**

Table 4.3 shows that, for almost all data sets, objective function (4.6) results in selection of fewer features and worse solutions compared to objective function (4.1). To investigate this behavior, we have examined the progress of feature selection probabilities over the iterations. We have also tried other evaporation rates, feature selection ratio thresholds, and removal of (1/I) from feature pheromone update, and seen that the results do not change.

We interpret the above results as follows. The marginal contributions of individual features in terms of their discriminative power are naturally different. Even though a second selected feature might have lower contribution than the first one, its addition still improves the clustering accuracy significantly. In objective function (4.1), the marginal cost of adding this second feature is also lower (since the distances in objective function are scaled by the inverse of the number of selected features). This way, the marginal contribution and the marginal cost of an additional feature seem to be balanced well. According to objective function (4.6), however, while the marginal contribution of the second feature remains the same, its marginal cost becomes higher (since the distances in objective function are scaled by the inverse of the square root of the number of selected features). In other words, using additional features seems to be penalized excessively. This results in rapidly decreasing feature selection probabilities in fewer iterations, premature convergence, and lower F-measure and RI values. After comparing the performances of the two objective functions, we have decided to use objective function (4.1) in our comparisons with the literature.

Both Rand Index and F-measure are used as performance measures while comparing F-ACOVNS with ACOVNS. However, only Rand Index is used in making comparisons for synthetic data sets. Rand Index is also used while comparing F-ACOVNS with single-objective algorithms on additional real-world data sets, but F-measure is used in comparisons with multi-objective algorithms. These performance measures differ because studies in the literature use different performance metrics in reporting their results and we want to compare our algorithm with all possible studies.

4.5.3 Computational Results and Comparisons on Real-World Data sets

First of all, we compare F-ACOVNS with ACOVNS in order to observe the change in clustering performance as a result of applying feature selection. Comparisons with ACOVNS are presented in Table 4.4. The F-measure and Rand Index values are the averages of 10 replications for each algorithm, and the superior results are presented in boldface.

It is interesting to see that clustering performance improves in some of these well-known data sets when feature selection is applied to them. Table 4.4 shows that

Table 4.4: Comparison of ACOVNS and F-ACOVNS (Averages for 10 Replications)

Data Set	ACOVNS			F-ACOVNS			
	F-measure	**Rand Index**	**Computing Time (seconds)**	**F-measure**	**Rand Index**	**Number of Selected Features**	**Computing Time (seconds)**
Iris	0.835	0.889	10	**0.925**	**0.950**	3	13
Wine	**0.926**	**0.950**	20	0.885	0.922	8.1	32
Breast Cancer	**0.943**	**0.938**	87	0.918	0.910	5.6	130
Glass	0.389	0.696	65	**0.427**	**0.715**	8	65
Thyroid	0.717	0.708	54	**0.897**	**0.882**	4	32
Liver Disease	0.549	**0.500**	45	**0.557**	**0.500**	4.5	54
CMC	**0.368**	0.556	281	0.367	**0.562**	7.3	451
Zoo	0.795	0.914	19	**0.878**	**0.945**	7.2	21
Wdbc	**0.877**	**0.867**	75	0.834	0.812	13.7	172

F-ACOVNS can find superior or equally good solutions in six of the nine data sets by choosing fewer features. However, in the Wine, Breast Cancer, and Wdbc data sets, ACOVNS produces superior solutions in terms of both F-measure and Rand Index. Although ACOVNS performs better in certain data sets, the differences are at most 4 percent. In contrast, F-ACOVNS achieves considerable enhancements in some data sets, including Iris, Thyroid, and Zoo. In the Iris data set, selecting three of the four features gives an improvement of approximately 9 percent, while selecting four of the five features yields an improvement of approximately 18 percent in the Thyroid data set. Also, in Zoo data set, F-ACOVNS selects 7.2 of 17 features on the average and obtains almost 8 percent better solutions in terms of F-measure. Despite the fact that the outcomes for both performance measures are similar, there are cases where one of the algorithms is better in terms of F-measure and the other one is better in terms of Rand Index. In these circumstances, F-ACOVNS and ACOVNS yield fairly similar results.

As for the computation times, it is evident that the feature selection method of F-ACOVNS increases computation times for the majority of data sets. With the exception of Breast Cancer, Wine, CMC, and Wdbc, the average increase in computing time is below 30 percent. However, in the Thyroid data set, F-ACOVNS outperforms ACOVNS. For the remaining data sets, the computation times are similar. The details of computation times are given in Table 4.5.

Table 4.5 demonstrates that feature selection does not considerably increase compu-

Table 4.5: Computation Times of ACOVNS and F-ACOVNS on Real-World Data Sets (Average of 10 replications)

Data Set	**ACOVNS**		**F-ACOVNS**		
	Number of Iterations	**Computing Time (seconds)**	**Number of Iterations**	**Computing Time for Clustering (seconds)**	**Computing Time for Feature Selection (seconds)**
Iris	150	10	200	8	5
Breast Cancer	400	87	450	114	16
Wine	250	20	300	24	8
Glass	450	65	500	53	12
Thyroid	600	54	630	29	3
Liver Disease	400	45	430	49	5
CMC	500	282	550	420	31
Zoo	250	20	350	8	13
Wdbc	400	75	600	82	90

tation time for the majority of data sets. Except Zoo and Wdbc data sets, clustering is the main part of the time spent by F-ACOVNS. Also, feature selection and clustering times are quite similar in Iris data set. However, for the remaining data sets, feature selection requires substantially less time than clustering.

The comparisons with some single-objective algorithms (Prakash and Singh, 2015) are given in Table 4.6. The best results are reported for the algorithms from literature, therefore we also report our best result rather than the average results for each data set. The superior result for each data set is shown in boldface.

Table 4.6: Comparisons with Single-Objective Algorithms

Data Set	**F-ACOVNS**		**BPSO-X**		**BPSO**		**GA**	
	Rand Index	**Number of Selected Features**	**Rand Index**	**Number of Selected Features**	**Rand Index**	**Number of Selected Features**	**Rand Index**	**Number of Selected Features**
Ionosphere	**0.585**	31	0.571	3	0.554	3	0.564	9
Parkinsons	0.566	19	**0.602**	2	**0.602**	3	**0.602**	3
Sonar	**0.500**	19	0.498	14	0.498	15	0.498	16
Spectf	0.523	43	**0.667**	10	0.619	11	0.615	12
Wdbc	**0.854**	21	0.534	3	0.534	7	0.534	6

Table 4.6 shows that F-ACOVNS finds the best result in 3 of the 5 data sets. In Wdbc, F-ACOVNS significantly outperforms other algorithms in terms of Rand Index

with over 0.30 difference. However, in Parkinsons data set, F-ACOVNS is the worst performing algorithm with 0.036 difference and in Spectf the difference is about 0.088 from the closest one. Also, other algorithms choose fewer features for all of the data sets and this results in better solutions in some cases. In some data sets such as Spectf, F-ACOVNS may be converging to the wrong feature subset.

The comparisons with multi-objective algorithms IMBGSAFS, MBGSAFS, MOPSO, NSGA-II, and FM-CC (Prakash and Singh, 2019) are given in Table 4.7. The results of the multi-objective algorithms are the solutions with the best F-measure values in pareto fronts. Therefore, to be comparable we reported our best result for each data set.

Table 4.7 shows that multi-objective algorithms obtain significantly better results in most of the data sets. F-ACOVNS produced the worse outcome for the majority of data sets. The exception is the Wdbc data set, for which F-ACOVNS found the best solution in terms of F-measure. Additionally, F-ACOVNS also outperforms FM-CC in the Dermatology data set, which is a second exception.

Also, the comparisons with another multi-objective algorithm MODE-cfs (Hancer, 2020) are given in Table 4.8. The results of this algorithm are reported as average F-measure of multiple algorithm replications. Thus, while comparing with MODE-cfs we reported our solutions in the same way.

Table 4.7: Comparisons with Multi-Objective Algorithms

Data Set	F-ACOVNS		IMBGSAFS		MBGSAFS		MOPSO		NSGA-II		FM-CC	
	F-measure	Number of Selected Features	F-measure	Number of Selected Features	F-measure	Number of Selected Features	F-measure	Number of Selected Features	F-measure	Number of Selected Features	F-measure	Number of Selected Features
Ionosphere	0.602	31	0.726	17	0.727	17	0.715	15	0.714	12	0.697	8
Dermatology	0.621	10	0.901	17	0.901	23	0.879	16	0.898	12	0.311	9
Parkinsons	0.617	19	0.735	7	0.729	6	0.735	7	0.729	9	0.733	3
Sonar	0.530	19	0.667	47	0.649	50	0.654	26	0.580	32	0.588	19
Spectf	0.651	43	0.772	23	0.772	16	0.735	7	0.772	19	0.722	5
Wdbc	0.870	21	0.822	28	0.686	6	0.813	28	0.799	28	0.768	3

Table 4.8: Comparisons with Multi-Objective Algorithm MODE-cfs

Data Set	F-ACOVNS		MODE-cfs	
	F-measure	Number of Selected Features	F-measure	Number of Selected Features
Ionosphere	**0.600**	29.8	0.593	21.4
Dermatology	0.463	9	**0.923**	27.6
Sonar	**0.522**	19.7	0.451	29.5
Wdbc	0.834	13.7	**0.870**	16.1
Liver Disease	**0.557**	4.5	0.511	4
Appendicitis	0.816	6.9	**0.820**	5.1
Pima	0.570	1	**0.572**	3.7
Iris	**0.925**	3	0.755	2.9
Thyroid	**0.897**	4	0.584	2.8
Wine	0.885	8.1	**0.913**	10.6
Ecoli	0.630	6	**0.778**	6.7
Breast Tissue	0.438	3	**0.696**	6.7

According to Table 4.8, F-ACOVNS obtains better F-measure values in 5 of the 12 data sets. F-ACOVNS can find competitive outcomes in some data sets despite having a single-objective function. In addition to the data sets in which F-ACOVNS obtains better results, the results in Appendicitis and Pima data sets are almost the same. Also, in Pima data set F-ACOVNS selects only 1 feature and its F-measure is almost the same with MODE-cfs which selects 3.7 features on the average. In the majority of data sets, MODE-cfs provides superior solutions. In the Dermatology data set, the average F-measure of MODE-cfs is 0.46 higher than that of F-ACOVNS. In contrast, F-ACOVNS performs with 0.18 and 0.31 higher F-measures in Iris and Thyroid data sets, respectively.

These comparisons demonstrate that F-ACOVNS can perform significantly better than ACOVNS. Moreover, compared to single-objective algorithms, F-ACOVNS can achieve extremely competitive results. However, based on its performance compared to multi-objective algorithms, F-ACOVNS cannot compete with multi-objective algorithms in the majority of data sets, but it can produce competitive or even superior solutions in certain data sets.

4.5.4 Computational Results and Comparisons on Synthetic Data sets

The performance of F-ACOVNS on synthetic data sets is compared with two multi-objective evolutionary algorithms for simultaneous feature selection and clustering. MOCNC-F is proposed by Alakuş (2018) and Δ-MOCK-F is adapted for feature selection again by Alakuş (2018). As reported by Alakuş (2018), MOCNC-F and Δ-MOCK-F are coded in C and all the experiments were conducted on a PC with a 3.6 GHz 4-Core Intel Core i7-4790 processor and 8GB of RAM. In the second algorithm, given the selected feature subset, clustering is performed using Δ-MOCK proposed by Garza-Fabre et al. (2017). F-ACOVNS is run on the same data sets generated in Alakuş (2018) study, and the comparisons are made in terms of Rand Index. The comparisons on synthetic data sets are given in Table 4.9. There are 10 instances for each pair of problem setting and number of redundant features, and the values in each row are the averages of these 10 instances. The averages of the RI values across 10 instances, the average number of original and redundant features selected are reported for each algorithm.

The results indicate that MOCNC-F yields the best results, although its computation times are much longer than those of the other algorithms. F-ACOVNS finds better solutions than Δ-MOCK-F for data sets with fewer features. However, while the performance of Δ-MOCK-F increases with the number of features, F-ACOVNS is affected negatively by high dimensions and its solution quality decreases. Regarding redundant features, F-ACOVNS is the most precise of the three algorithms. When there are 20 or more redundant features, MOCNC-F and Δ-MOCK-F tend to choose such features. However, F-ACOVNS can almost fully distinguish between original and redundant features.

Table 4.9: Comparisons on Synthetic Data Sets

Data Set	RdF	MOCNC-F				Δ-MOCK-F				F-ACOVNS			
		RI	**OF**	**RF**	**Computing Time (seconds)**	**RI**	**OF**	**RF**	**Computing Time (seconds)**	**RI**	**OF**	**RF**	**Computing Time (seconds)**
5d-5c	1	**0.996**	4.7	0	3454	0.784	3.8	1.7	252	0.994	4.9	0.0	430
	3	**0.997**	4.6	0.1	3008	0.792	4	1.7	245	0.993	5.0	0.0	375
	5	**0.996**	4.6	0.1	2838	0.808	4.7	1.5	243	0.977	5.0	0.2	479
10d-5c	2	**1**	7.6	0	2414	0.717	1	0	228	0.991	9.8	0.0	451
	5	**1**	7.8	0	2575	0.778	2.7	0	227	0.986	9.9	0.1	529
	10	**1**	8.8	0	2716	0.943	6.9	0	232	0.992	10.0	0.0	641
20d-5c	4	**1**	12.9	0	3044	0.993	11.9	0	231	**1**	19.8	0.0	530
	10	**1**	15.8	1	3423	0.997	12.4	0.1	250	0.985	19.2	0.2	687
	20	**1**	17.5	4.6	4707	0.997	16.7	0.9	254	0.987	18.1	0.0	689
40d-5c	8	**1**	30	0.9	19383	0.997	18	0	312	0.978	31.9	0.0	663
	20	**1**	30.3	5.4	29126	0.997	24.4	0.8	332	0.975	26.8	0.0	735
	40	0.972	30	12.6	42988	**0.996**	31.2	5.7	377	0.982	26.6	0.0	742

RdF: Number of redundant features added to the original data set

OF : Average number of original features selected in the solutions

RF : Average number of redundant features selected in the solutions

CHAPTER 5

CONCLUSIONS

In this study, we address the clustering problem where the number of clusters is known and their shapes and densities are variable. Data sets contain only numerical features. There may be a large number features and some of them may be redundant. We propose two new algorithms, namely ACOVNS for clustering and F-ACOVNS for simultaneous feature selection and clustering.

The complexity of the clustering problem is increased by the numerous properties that data sets may have. There are various objective functions for addressing different data set and problem properties. The choice of the objective function has a direct impact on the clustering process. Therefore, it is one of the most crucial selections. In our proposed algorithms ACOVNS and F-ACOVNS, one of the most commonly used objective functions, the sum of Euclidean distances, is used.

ACOVNS is a hybrid swarm intelligence algorithm based on ACO and VNS. The algorithm inherits the exploration capability of ACO and increases ACO's exploitation capability by using VNS as a local search method. In this manner, ACOVNS is the first hybridization of ACO and VNS for clustering. ACOVNS, unlike conventional ACO algorithms, includes an additional exploration mechanism, inspired by the mutation operator of the Genetic Algorithm, that allows the selection of object-cluster assignments with the lowest pheromone values. Moreover, ACOVNS uses heuristic information that enables the algorithm to make selections based not just on the pheromone matrix formed by the entire colony, but also on the individual solutions constructed by the ants. A predetermined number of best solutions constructed by the ants are improved using VNS by assigning varying number of objects to their closest clusters.

Two performance measures F-measure and Rand Index are used to assess the performance of ACOVNS. These two metrics are widely used in the literature. F-measure considers the geometric mean of the distances of object pairs that belong to the same cluster in algorithm's solution and true partition, relative to the proportion of pairs that belong to the same cluster in each partition. On the other hand, the Rand Index calculates the proportion of pairs of objects in both the algorithm's solution and the true partition that agree by belonging to the same or distinct clusters.

The performance of the ACOVNS algorithm is compared to those of various known algorithms, and it is observed to achieve very competitive results. The computational experiments demonstrate that ACOVNS can converge to good solutions in the majority of data sets in relatively low computing times. It should be noted, however, that some of the compared algorithms are quite old, and their computing times would be reduced on more modern computers. In Wine and Liver Disease data sets, the algorithm's solution quality is significantly superior to those of other algorithms in the literature. However, ACOVNS performs poorly in Glass and Thyroid data sets. For the remaining data sets, ACOVNS achieves results that are highly competitive with other well-performing algorithms.

In the second part of the study, we present F-ACOVNS, an extended version of ACOVNS designed for simultaneous feature selection and clustering. Feature selection and clustering are interconnected processes that could lead to information loss if considered separately. Irrelevant and redundant features may degrade the efficiency of the clustering algorithm or even mislead it. Therefore, F-ACOVNS is proposed to address feature selection and clustering problems simultaneously.

F-ACOVNS employs two additional mechanisms, namely the feature pheromone vector and the feature heuristic information. Similar to cluster pheromone matrix, feature pheromone vector contains the pheromone values of individual features. On the other hand, feature heuristic information is the inverse of each feature's Laplacian Score (LS) value. Our algorithm is the first to use LS as heuristic information in the ACO framework. LS provides information regarding the ranks of the features where a smaller LS value indicates that the feature may have higher discriminative power. However, LS only assigns scores to features and ranks them, but it cannot determine the number of features to select. Combining LS with the feature pheromone vector, F-ACOVNS determines the number of features to be selected.

The performance of the F-ACOVNS algorithm is evaluated on both real-world and synthetic data sets. On real-world data sets, F-ACOVNS is compared with ACOVNS, single-objective clustering algorithms, and multi-objective clustering algorithms. The results show that F-ACOVNS increases the performance of ACOVNS by selecting more relevant feature subsets in most of the data sets. In six of the nine data sets, F-ACOVNS finds better or equally good solutions. Additionally, F-ACOVNS detects smaller feature subsets for all of the data sets and this improves the clustering performance for most of the data sets. Especially in some data sets such as Iris and Thyroid, F-ACOVNS achieves significantly higher F-measure and Rand Index values by selecting a very small subset of original features. However, in some other data sets such as Wine, Breast Cancer, and Wdbc, F-ACOVNS obtains lower solution quality than ACOVNS. Therefore, it is possible that certain features may mislead the algorithm in feature selection.

The comparisons on real-world data sets show that F-ACOVNS is a competitive algorithm among single-objective algorithms. In particular for the Wdbc data set, F-ACOVNS can obtain significantly better accuracy results compared to other algorithms. However, when compared to multi-objective algorithms, F-ACOVNS is one of the worst performing algorithms. Multi-objective algorithm results are significantly better than the solutions of F-ACOVNS in terms of F-measure, hence it can be concluded that using multiple objectives is a good idea in solving feature selection and clustering problems simultaneously. Nevertheless, F-ACOVNS with a single objective obtains a superior result for the Wdbc data set compared to multi-objective algorithms.

The comparisons on synthetic data sets, on the other hand, show that F-ACOVNS can separate original and redundant features almost perfectly. In most of the data sets, F-ACOVNS selects only original features while MOCNC-F and Δ-MOCK-F select some redundant features as well. In terms of clustering performance, F-ACOVNS stands in between MOCNC-F and Δ-MOCK-F. Therefore, F-ACOVNS might be the most effective in terms of the computation time and solution quality trade-off.

To further improve the performance of both ACOVNS and F-ACOVNS, a different neighborhood structure for VNS or a different local search method can be implemented. Due to the constructive nature of ACO, it is difficult for either algorithm to reach the optimal solution precisely, hence an efficient and effective improvement step is critical. Also, ACOVNS can be transformed into a multi-objective algorithm and objective functions can be used to evaluate compactness, connectedness, and separation distinctly. An additional objective can also be considered in F-ACOVNS for feature selection.

In this study, the number of clusters is considered as known and taken as an input parameter. Thus, both ACOVNS and F-ACOVNS can be adapted to find the number of clusters in the data sets since the number of clusters in the data set may not be available in some cases. By employing an additional objective function such as Silhouette Index, the number of clusters may be decided.

The focus of this study is the data sets with numerical features. However, categorical features are widely used in real-world applications. Therefore, another future research direction may be extending the solution approach to address the clustering problem where data sets may contain categorical or mixed features. To this end, a method for converting categorical features to numerical features may be employed or different objective functions may be considered to handle these features.

REFERENCES

Abd-Alsabour, N., Hefny, H., & Moneim, A. (2013). Heuristic information for ant colony optimization for the feature selection problem. *IEEE Conference Anthology*, 1–5.

Abualigah, L., Gandomi, A. H., Elaziz, M. A., Hussien, A. G., Khasawneh, A. M., Alshinwan, M., & Houssein, E. H. (2020). Nature-inspired optimization algorithms for text document clustering—a comprehensive analysis. *Algorithms*, *13*(12), 345.

Alakuş, C. (2018). *Neighborhood construction-based multi-objective evolutionary clustering algorithm with feature selection* (Master's thesis). Middle East Technical University.

Alam, A., Muqeem, M., & Ahmad, S. (2021). Comprehensive review on clustering techniques and its application on high dimensional data. *International Journal of Computer Science & Network Security*, *21*(6), 237–244.

Alam, S., Dobbie, G., Koh, Y. S., Riddle, P., & Rehman, S. U. (2014). Research on particle swarm optimization based clustering: A systematic review of literature and techniques. *Swarm and Evolutionary Computation*, *17*, 1–13.

Alam, S., Dobbie, G., & Rehman, S. U. (2015). Analysis of particle swarm optimization based hierarchical data clustering approaches. *Swarm and Evolutionary Computation*, *25*, 36–51.

Almeida, J., Barbosa, L., Pais, A., & Formosinho, S. (2007). Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemometrics and Intelligent Laboratory Systems*, *87*(2), 208–217.

Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2015). Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM transactions on computational biology and bioinformatics*, *13*(5), 971–989.

Arora, S., & Chana, I. (2014). A survey of clustering techniques for big data analysis. *2014 5th International Conference-Confluence The Next Generation Information Technology Summit (Confluence)*, 59–65.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and information systems*, *34*(3), 483–519.

Boryczka, U. (2009). Finding groups in data: Cluster analysis with ants. *Applied Soft Computing*, *9*(1), 61–70.

Brun, M., Sima, C., Hua, J., Lowey, J., Carroll, B., Suh, E., & Dougherty, E. R. (2007). Model-based evaluation of clustering validation measures. *Pattern recognition*, *40*(3), 807–824.

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79.

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

Chuang, L.-Y., Chang, H.-W., Tu, C.-J., & Yang, C.-H. (2008). Improved binary pso for feature selection using gene expression data. *Computational Biology and Chemistry*, *32*(1), 29–38.

Cilibrasi, R., & Vitányi, P. M. (2005). Clustering by compression. *IEEE Transactions on Information theory*, *51*(4), 1523–1545.

Dash, M., & Liu, H. (2000). Feature selection for clustering. *Pacific-Asia Conference on knowledge discovery and data mining*, 110–121.

Dinh, D.-T., Huynh, V.-N., & Sriboonchitta, S. (2021). Clustering mixed numerical and categorical data with missing values. *Information Sciences*, *571*, 418–442.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *26*(1), 29–41.

Dutta, D., Dutta, P., & Sil, J. (2013). Simultaneous continuous feature selection and k clustering by multi objective genetic algorithm. *2013 3rd IEEE International Advance Computing Conference (IACC)*, 937–942.

Dutta, D., Dutta, P., & Sil, J. (2014). Simultaneous feature selection and clustering with mixed features by multi objective genetic algorithm. *International Journal of Hybrid Intelligent Systems*, *11*(1), 41–54.

Dy, J. G., & Brodley, C. E. (2000). Feature subset selection and order identification for unsupervised learning. *ICML*, 247–254.

Dy, J. G., & Brodley, C. E. (2004). Feature selection for unsupervised learning. *Journal of machine learning research*, *5*(Aug), 845–889.

El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. *2016 International Conference on Engineering & MIS (ICEMIS)*, 1–5.

Esmin, A. A., Coelho, R. A., & Matwin, S. (2015). A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data. *Artificial Intelligence Review*, *44*(1), 23–45.

Ezugwu, A. E., Shukla, A. K., Agbaje, M. B., Oyelade, O. N., José-Garcıa, A., & Agushaka, J. O. (2021). Automatic clustering algorithms: A systematic review and bibliometric analysis of relevant literature. *Neural Computing and Applications*, *33*(11), 6247–6306.

Fathian, M., & Amiri, B. (2008). A honeybee-mating approach for cluster analysis. *The International Journal of Advanced Manufacturing Technology*, *38*(7), 809–821.

Ferreira, A. J., & Figueiredo, M. A. (2012). Efficient feature selection filters for high-dimensional data. *Pattern recognition letters*, *33*(13), 1794–1804.

Garza-Fabre, M., Handl, J., & Knowles, J. (2017). An improved and more scalable evolutionary approach to multiobjective clustering. *IEEE Transactions on Evolutionary Computation*, *22*(4), 515–535.

Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.

Guha, S., Rastogi, R., & Shim, K. (1998). Cure: An efficient clustering algorithm for large databases. *ACM Sigmod record*, *27*(2), 73–84.

Hancer, E. (2018). A differential evolution approach for simultaneous clustering and feature selection. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1–7.

Hancer, E. (2020). A new multi-objective differential evolution approach for simultaneous clustering and feature selection. *Engineering Applications of Artificial Intelligence*, *87*, 103307.

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

Hatamlou, A., Abdullah, S., & Nezamabadi-Pour, H. (2012). A combined approach for clustering based on k-means and gravitational search algorithms. *Swarm and Evolutionary Computation*, *6*, 47–52.

Haupt, R. L. (2000). Optimum population size and mutation rate for a simple real genetic algorithm that optimizes array factors. *IEEE Antennas and Propagation Society International Symposium. Transmitting Waves of Progress to the Next Millennium. 2000 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting (C*, *2*, 1034–1037.

He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. *Advances in neural information processing systems*, *18*.

Huang, W., & Ribeiro, A. (2018). Hierarchical clustering given confidence intervals of metric distances. *IEEE Transactions on Signal Processing*, *66*(10), 2600–2615.

Inkaya, T., Kayalıgil, S., & Özdemirel, N. E. (2016). Swarm intelligence-based clustering algorithms: A survey. *Unsupervised learning algorithms* (pp. 303–341). Springer.

Jabbar, A. M., Ku-Mahamud, K. R., & Sagban, R. (2018). Ant-based sorting and aco-based clustering approaches: A review. *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 217–223.

Javani, M., Faez, K., & Aghlmandi, D. (2011). Clustering and feature selection via pso algorithm. *2011 international symposium on artificial intelligence and signal processing (AISP)*, 71–76.

José-Garcia, A., & Gómez-Flores, W. (2016). Automatic clustering using nature-inspired metaheuristics: A survey. *Applied Soft Computing*, *41*, 192–213.

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, 1200–1205.

Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (abc) algorithm. *Journal of global optimization*, *39*(3), 459–471.

Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2014). A comprehensive survey: Artificial bee colony (abc) algorithm and applications. *Artificial Intelligence Review*, *42*(1), 21–57.

Karaboga, D., & Ozturk, C. (2011). A novel clustering approach: Artificial bee colony (abc) algorithm. *Applied soft computing*, *11*(1), 652–657.

Karypis, G., Han, E.-H., & Kumar, V. (1999). Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, *32*(8), 68–75.

Krishnasamy, G., Kulkarni, A. J., & Paramesran, R. (2014). A hybrid approach for data clustering based on modified cohort intelligence and k-means. *Expert Systems with Applications*, *41*(13), 6009–6016.

Kumar, A., Kumar, D., & Jarial, S. (2017). A review on artificial bee colony algorithms and their applications to data clustering. *Cybernetics and Information Technologies*, *17*(3), 3–28.

Kumar, Y., & Sahoo, G. (2015). Hybridization of magnetic charge system search and particle swarm optimization for efficient data clustering using neighborhood search strategy. *Soft Computing*, *19*(12), 3621–3645.

Kumari, B., & Swarnkar, T. (2011). Filter versus wrapper feature subset selection in large dimensionality micro array: A review.

Kuwil, F. H., Atila, Ü., Abu-Issa, R., & Murtagh, F. (2020). A novel data clustering algorithm based on gravity center methodology. *Expert Systems with Applications*, *156*, 113435.

Kuwil, F. H., Shaar, F., Topcu, A. E., & Murtagh, F. (2019). A new data clustering algorithm based on critical distance methodology. *Expert Systems with Applications*, *129*, 296–310.

Lensen, A., Xue, B., & Zhang, M. (2016). Particle swarm optimisation representations for simultaneous clustering and feature selection. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, *50*(6), 94.

Lingras, P., & Huang, X. (2005). Statistical, evolutionary, and neurocomputing clustering techniques: Cluster-based vs object-based approaches. *Artificial Intelligence Review*, *23*(1), 3–29.

Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, *17*(4), 491–502.

Mageshkumar, C., Karthik, S., & Arunachalam, V. (2019). Hybrid metaheuristic algorithm for improving the efficiency of data clustering. *Cluster Computing*, *22*(1), 435–442.

Mane, S. U., & Gaikwad, P. G. (2014). Nature inspired techniques for data clustering. *2014 International conference on circuits, systems, communication and information technology applications (CSCITA)*, 419–424.

Mavrovouniotis, M., Müller, F. M., & Yang, S. (2016). Ant colony optimization with local search for dynamic traveling salesman problems. *IEEE transactions on cybernetics*, *47*(7), 1743–1756.

Mehdizadeh, E., Teimouri, M., Zaretalab, A., & Niaki, S. T. A. (2017). A combined approach based on k-means and modified electromagnetism-like mechanism for data clustering. *International Journal of Information Technology & Decision Making*, *16*(05), 1279–1307.

Mehta, V., Bawa, S., & Singh, J. (2020). Analytical review of clustering techniques and proximity measures. *Artificial Intelligence Review*, *53*(8), 5995–6023.

Mittal, H., Pandey, A. C., Saraswat, M., Kumar, S., Pal, R., & Modwel, G. (2021). A comprehensive survey of image segmentation: Clustering methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*, 1–26.

Mladenović, N., & Hansen, P. (1997). Variable neighborhood search. *Computers & operations research*, *24*(11), 1097–1100.

Nanda, S. J., & Panda, G. (2014). A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm and Evolutionary computation*, *16*, 1–18.

Niknam, T., Olamaei, J., & Amiri, B. (2008). A hybrid evolutionary algorithm based on aco and sa for cluster analysis. *Journal of Applied sciences*, *8*(15), 2695–2702.

Niknam, T., & Amiri, B. (2010). An efficient hybrid approach based on pso, aco and k-means for cluster analysis. *Applied soft computing*, *10*(1), 183–197.

Niknam, T., Fard, E. T., Pourjafarian, N., & Rousta, A. (2011). An efficient hybrid algorithm based on modified imperialist competitive algorithm and k-means for data clustering. *Engineering Applications of Artificial Intelligence*, *24*(2), 306–317.

O'Neill, D., Lensen, A., Xue, B., & Zhang, M. (2018). Particle swarm optimisation for feature selection and weighting in high-dimensional clustering. *2018 IEEE Congress on Evolutionary Computation (CEC)*, 1–8.

Pacheco, T. M., Gonçalves, L. B., Ströele, V., & Soares, S. S. R. (2018). An ant colony optimization for automatic data clustering problem. *2018 IEEE Congress on Evolutionary Computation (CEC)*, 1–8.

Prakash, J., & Singh, P. K. (2013). Partitional algorithms for hard clustering using evolutionary and swarm intelligence methods: A survey. *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, 515–528.

Prakash, J., & Singh, P. K. (2015). Particle swarm optimization with k-means for simultaneous feature selection and data clustering. *2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI)*, 74–78.

Prakash, J., & Singh, P. K. (2019). Gravitational search algorithm and k-means for simultaneous feature selection and data clustering: A multi-objective approach. *Soft Computing*, *23*(6), 2083–2100.

Prakash, J., Singh, P. K., & Kishor, A. (2019). Integrating fitness predator optimizer with multi-objective pso for dynamic partitional clustering. *Progress in Artificial Intelligence*, *8*(1), 83–99.

Ramadas, M., & Abraham, A. (2019). *Metaheuristics for data clustering and image segmentation*. Springer.

Rana, S., Jasola, S., & Kumar, R. (2011). A review on particle swarm optimization algorithms and their applications to data clustering. *Artificial Intelligence Review*, *35*(3), 211–222.

Saha, S., Ekbal, A., Alok, A. K., & Spandana, R. (2014). Feature selection and semi-supervised clustering using multiobjective optimization. *SpringerPlus*, *3*(1), 1–12.

Saha, S., Spandana, R., Ekbal, A., & Bandyopadhyay, S. (2015). Simultaneous feature selection and symmetry based clustering using multiobjective framework. *Applied Soft Computing*, *29*, 479–486.

Sarkar, S., Roy, A., & Purkayastha, B. S. (2013). Application of particle swarm optimization in data clustering: A survey. *International Journal of Computer Applications*, *65*(25).

Sarvari, H., Khairdoost, N., & Fetanat, A. (2010). Harmony search algorithm for simultaneous clustering and feature selection. *2010 International Conference of Soft Computing and Pattern Recognition*, 202–207.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681.

Shelokar, P., Jayaraman, V. K., & Kulkarni, B. D. (2004). An ant colony approach for clustering. *Analytica Chimica Acta*, *509*(2), 187–195.

Sheng, W., Liu, X., & Fairhurst, M. (2008). A niching memetic algorithm for simultaneous clustering and feature selection. *IEEE Transactions on Knowledge and Data Engineering*, *20*(7), 868–879.

Singh, H., & Kumar, Y. (2020). A neighborhood search based cat swarm optimization algorithm for clustering problems. *Evolutionary Intelligence*, *13*(4), 593–609.

Solorio-Fernández, S., Carrasco-Ochoa, J. A., & Martınez-Trinidad, J. F. (2020). A review of unsupervised feature selection methods. *Artificial Intelligence Review*, *53*(2), 907–948.

Son, L. H. (2016). Generalized picture distance measure and applications to picture fuzzy clustering. *Applied Soft Computing*, *46*(100), 284–295.

Sridevi, T., & Murugan, A. (2014). An intelligent classifier for breast cancer diagnosis based on k-means clustering and rough set. *International Journal of Computer Applications*, *85*(11).

Stetco, A., Zeng, X.-J., & Keane, J. (2015). Fuzzy c-means++: Fuzzy c-means with effective seeding initialization. *Expert Systems with Applications*, *42*(21), 7541–7548.

Sun, L., Tao, T., Zheng, X., Bao, S., & Luo, Y. (2019). Combining density peaks clustering and gravitational search method to enhance data clustering. *Engineering Applications of Artificial Intelligence*, *85*, 865–873.

Swetha, K., & Susheela Devi, V. (2012). Simultaneous feature selection and clustering using particle swarm optimization. *International Conference on Neural Information Processing*, 509–515.

Szekely, G. J., & Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of classification*, *22*(2), 151–184.

Taib, H., & Bahreininejad, A. (2021). Data clustering using hybrid water cycle algorithm and a local pattern search method. *Advances in Engineering Software*, *153*, 102961.

Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybernetics and information technologies*, *19*(1), 3–26.

Vipin, K., & Minz, S. (2014). Feature selection: A literature review. *SmartCR*, *4*(3), 211–229.

Visalakshi, S., & Radha, V. (2014). A literature review of feature selection techniques and applications: Review of feature selection in data mining. *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 1–6.

Wang, X., & Xu, Y. (2019). An improved index for clustering validation based on silhouette index and calinski-harabasz index. *IOP Conference Series: Materials Science and Engineering*, *569*(5), 052024.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). Birch: An efficient data clustering method for very large databases. *ACM sigmod record*, *25*(2), 103–114.

Zhao, Y., Karypis, G., & Fayyad, U. (2005). Hierarchical clustering algorithms for document datasets. *Data mining and knowledge discovery*, *10*(2), 141–168.

Zhu, E., Zhang, Y., Wen, P., & Liu, F. (2019). Fast and stable clustering analysis based on grid-mapping k-means algorithm and new clustering validity index. *Neurocomputing*, *363*, 149–170.

Zou, H. (2020). Clustering algorithm and its application in data mining. *Wireless Personal Communications*, *110*(1), 21–30.

Appendix A

EXPERIMENTAL RESULTS OF THE ACOVNS ALGORITHM

In this appendix, we give the detailed experimental results of the ACOVNS algorithm for individual algorithm replications.

Table A.1: ACOVNS Results for Each Replication

Data set	Run	Without Normalization and Heuristic Information			With Normalization			With Heuristic Information			With Both Normalization and Heuristic Information		
		Objective Value	F-measure	Rand Index	Objective Value	F-measure	Rand Index	Objective Value	F-measure	Rand Index	Objective Value	F-measure	Rand Index
Iris	1	97.491	0.850	0.899	98.205	0.822	0.881	97.374	0.831	0.887	97.592	0.814	0.875
	2	97.503	0.831	0.887	98.987	0.870	0.913	97.447	0.840	0.893	99.165	**0.880**	**0.920**
	3	97.575	0.849	0.899	99.663	0.850	0.899	97.447	0.840	0.893	97.592	0.814	0.875
	4	97.616	0.869	0.913	98.732	0.830	0.887	97.374	0.831	0.887	98.555	0.821	0.881
	5	97.575	0.849	0.899	97.827	0.841	0.893	97.641	0.859	0.906	99.663	0.850	0.899
	6	97.222	0.833	0.887	99.165	**0.880**	**0.920**	97.322	0.841	0.893	99.165	**0.880**	**0.920**
	7	97.322	0.841	0.893	99.165	**0.880**	**0.920**	97.447	0.840	0.893	98.130	0.813	0.875
	8	97.491	0.850	0.899	99.165	**0.880**	**0.920**	97.374	0.831	0.887	98.544	0.812	0.875
	9	97.844	0.870	0.913	97.824	0.822	0.881	97.616	0.869	0.913	98.982	0.840	0.893
	10	97.590	0.849	0.899	98.498	0.804	0.869	97.246	0.832	0.887	97.824	0.822	0.881
	Averages	97.523	**0.849**	**0.899**	98.723	0.848	0.898	97.429	0.842	0.894	98.521	0.835	0.889
Wine	1	16530.537	0.590	0.721	23369.132	**0.934**	**0.955**	16633.855	0.589	0.719	23178.293	0.904	0.935
	2	16530.537	0.590	0.721	23059.322	0.914	0.942	16579.353	0.593	0.722	23369.132	**0.934**	**0.955**
	3	16530.537	0.590	0.721	24284.338	0.922	0.947	16737.736	0.591	0.719	23757.778	0.933	0.955
	4	16530.537	0.590	0.721	24284.338	0.922	0.947	16657.939	0.589	0.719	23757.778	0.933	0.955
	5	16530.537	0.590	0.721	23369.132	**0.934**	**0.955**	16666.230	0.594	0.722	24676.996	0.901	0.933
	6	16530.537	0.590	0.721	24284.338	0.922	0.947	16894.036	0.595	0.720	23757.778	0.933	0.955
	7	16535.979	0.593	0.723	23369.132	**0.934**	**0.955**	16702.375	0.589	0.719	23757.778	0.933	0.955
	8	16530.537	0.590	0.721	23757.778	0.933	0.955	16794.263	0.604	0.727	23757.778	0.933	0.955
	9	16530.537	0.590	0.721	23369.132	**0.934**	**0.955**	16857.015	0.596	0.721	23587.667	0.923	0.948
	10	16530.537	0.590	0.721	24843.040	0.911	0.940	16538.968	0.588	0.719	23369.132	**0.934**	**0.955**
	Averages	16531.081	0.591	0.721	23798.969	**0.926**	**0.950**	16706.177	0.593	0.721	23697.011	**0.926**	**0.950**

Table A.1: ACOVNS Results for Each Replication

Data set	Run	Without Normalization and Heuristic Information			With Normalization			With Heuristic Information			With Both Normalization and Heuristic Information		
		Objective Value	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**
Breast Cancer	1	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.371	0.941	0.935	2984.068	0.943	0.938
	2	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938
	3	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938
	4	2984.238	**0.945**	**0.940**	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938
	5	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938
	6	2984.238	**0.945**	**0.940**	2984.248	0.940	0.935	2984.068	0.943	0.938	2984.068	0.943	0.938
	7	2984.129	0.940	0.935	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938
	8	2984.289	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938	2984.068	0.943	0.938
	9	2984.068	0.943	0.938	2984.293	0.938	0.932	2984.068	0.943	0.938	2984.068	0.943	0.938
	10	2984.129	0.940	0.935	2984.129	0.940	0.935	2984.371	0.941	0.935	2984.068	0.943	0.938
	Averages	2984.136	**0.943**	**0.938**	2984.115	0.942	0.937	2984.129	0.943	0.937	2984.068	**0.943**	**0.938**
Glass	1	259.460	0.339	0.705	249.439	0.375	0.686	226.352	**0.481**	0.695	239.598	0.383	0.700
	2	257.154	0.349	0.712	283.859	0.362	0.709	222.721	0.408	0.723	249.342	0.377	0.687
	3	258.700	0.336	0.707	279.602	0.349	0.687	225.521	0.436	0.695	238.208	0.408	0.696
	4	254.966	0.354	0.711	274.032	0.360	0.715	219.920	0.432	0.707	240.124	0.384	0.702
	5	257.422	0.346	0.709	280.982	0.337	0.685	219.680	0.429	0.711	237.618	0.406	0.692
	6	259.389	0.343	0.709	281.639	0.346	0.689	222.192	0.394	0.724	249.239	0.374	0.687
	7	257.320	0.356	0.714	245.611	0.407	**0.735**	220.551	0.430	0.698	238.173	0.410	0.696
	8	257.586	0.365	0.716	249.439	0.375	0.686	222.155	0.396	0.724	252.390	0.375	0.684
	9	260.005	0.341	0.706	283.800	0.333	0.701	219.983	0.432	0.705	245.493	0.402	0.730
	10	223.430	0.378	0.726	245.085	0.353	0.702	219.897	0.417	0.709	249.483	0.375	0.686
	Averages	254.543	0.351	**0.711**	267.349	0.360	0.700	221.897	**0.425**	0.709	243.967	0.389	0.696

Table A.1: ACOVNS Results for Each Replication

Data set	Run	Without Normalization and Heuristic Information			With Normalization			With Heuristic Information			With Both Normalization and Heuristic Information		
		Objective Value	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**
Thyroid	1	1984.256	0.700	0.710	2255.863	0.567	0.581	1984.508	0.683	0.696	2086.499	**0.850**	**0.819**
	2	1976.839	0.604	0.626	2026.748	0.586	0.599	1966.687	0.607	0.636	2086.499	**0.850**	**0.819**
	3	1980.288	0.639	0.658	2029.954	0.589	0.602	1975.143	0.667	0.685	2347.968	0.558	0.576
	4	1977.142	0.621	0.640	2013.841	0.591	0.602	1980.231	0.636	0.655	2402.707	0.622	0.629
	5	1981.729	0.657	0.672	2306.194	0.566	0.579	1980.193	0.652	0.668	2306.194	0.566	0.579
	6	1974.831	0.657	0.676	2016.915	0.588	0.601	1982.167	0.651	0.667	2086.499	**0.850**	**0.819**
	7	1974.831	0.657	0.676	2016.407	0.588	0.601	1983.170	0.712	0.720	2086.499	**0.850**	**0.819**
	8	1974.768	0.662	0.681	2027.825	0.586	0.599	1974.889	0.662	0.681	2086.499	**0.850**	**0.819**
	9	1976.983	0.617	0.636	2306.194	0.566	0.579	1968.701	0.615	0.643	2021.734	0.615	0.620
	10	1966.827	0.610	0.638	2301.145	0.568	0.580	1969.135	0.619	0.646	2362.601	0.562	0.578
	Averages	1976.849	0.642	0.661	2130.109	0.580	0.592	1976.482	0.650	0.670	2187.370	**0.717**	**0.708**
Liver Disease	1	9982.215	**0.587**	0.500	11186.651	0.538	0.500	9982.215	**0.587**	0.500	10931.845	0.549	0.500
	2	9982.215	**0.587**	0.500	11186.651	0.538	0.500	9982.215	**0.587**	0.500	11066.875	0.546	**0.501**
	3	9982.215	**0.587**	0.500	11110.541	0.541	0.500	9982.215	**0.587**	0.500	11005.812	0.550	**0.501**
	4	9982.215	**0.587**	0.500	11078.884	0.542	**0.501**	9982.215	**0.587**	0.500	10865.227	0.551	0.500
	5	9982.215	**0.587**	0.500	11186.651	0.538	0.500	9982.215	**0.587**	0.500	10887.125	0.552	0.500
	6	9982.215	**0.587**	0.500	10887.125	0.552	0.500	9982.215	**0.587**	0.500	11003.432	0.549	0.500
	7	9982.215	**0.587**	0.500	10887.125	0.552	0.500	9982.215	**0.587**	0.500	10928.458	0.551	0.500
	8	9982.246	**0.585**	0.500	11110.541	0.541	0.500	9982.215	**0.587**	0.500	10974.104	0.550	**0.501**
	9	9982.215	**0.587**	0.500	11093.762	0.541	0.500	9982.215	**0.587**	0.500	10923.539	0.548	0.500
	10	9982.246	**0.585**	0.500	11093.762	0.541	0.500	9982.246	**0.585**	0.500	11036.770	0.543	0.500
	Averages	9982.221	**0.587**	**0.500**	11082.169	0.543	**0.500**	9982.218	**0.587**	**0.500**	10962.319	0.549	**0.500**

Table A.1: ACOVNS Results for Each Replication

Data set	Run	Without Normalization and Heuristic Information			With Normalization			With Heuristic Information			With Both Normalization and Heuristic Information		
		Objective Value	F-measure	Rand Index	Objective Value	F-measure	Rand Index	Objective Value	F-measure	Rand Index	Objective Value	F-measure	Rand Index
CMC	1	5611.405	0.369	0.558	11091.918	0.364	0.557	5565.597	0.371	0.558	11080.425	0.364	0.556
	2	5633.884	0.362	0.560	11074.585	0.367	0.558	5586.320	0.365	0.561	11072.360	0.366	0.557
	3	5587.095	0.369	0.557	11110.848	0.365	0.557	5562.517	0.367	0.561	10937.858	**0.379**	0.554
	4	5635.214	0.364	0.561	10406.307	0.369	0.560	5575.249	0.365	0.561	11097.961	0.367	0.557
	5	5604.599	0.368	**0.562**	11106.605	0.365	0.557	5557.678	0.369	0.558	11133.168	0.370	0.557
	6	5657.458	0.362	0.561	11118.076	0.366	0.558	5567.934	0.367	0.560	11096.329	0.366	0.558
	7	5621.513	0.365	0.561	11101.985	0.364	0.557	5569.681	0.368	0.561	11094.640	0.368	0.558
	8	5623.960	0.365	0.561	11116.428	0.366	0.557	5570.952	0.364	0.560	11105.316	0.365	0.557
	9	5727.283	0.359	0.559	11092.271	0.365	0.557	5558.484	0.366	0.559	11248.022	0.372	0.550
	10	5597.272	0.367	0.557	11181.647	0.367	0.555	5569.816	0.366	0.561	11100.204	0.367	0.559
	Averages	5629.968	0.365	**0.560**	11040.067	0.366	0.557	5568.423	0.367	**0.560**	11096.628	**0.368**	0.556
Zoo	1	115.400	0.716	0.877	135.359	0.706	0.880	115.693	0.741	0.889	143.564	0.769	0.905
	2	112.928	0.757	0.891	147.208	0.699	0.878	122.108	0.569	0.829	136.330	0.755	0.896
	3	121.708	0.549	0.822	135.359	0.706	0.880	115.400	0.716	0.877	140.928	**0.966**	**0.983**
	4	121.708	0.549	0.822	144.590	0.757	0.901	117.096	0.769	0.898	144.590	0.757	0.901
	5	121.708	0.549	0.822	142.089	0.679	0.869	117.096	0.769	0.898	149.044	0.793	0.911
	6	122.108	0.569	0.829	146.704	0.714	0.883	113.007	0.761	0.893	141.038	0.871	0.942
	7	123.189	0.708	0.872	145.988	0.831	0.927	109.600	0.803	0.915	145.988	0.831	0.927
	8	117.096	0.769	0.898	141.038	0.871	0.942	117.096	0.769	0.898	144.590	0.757	0.901
	9	116.015	0.638	0.856	149.030	0.805	0.914	111.692	0.796	0.911	144.590	0.757	0.901
	10	111.926	0.631	0.851	144.590	0.757	0.901	121.406	0.751	0.892	146.933	0.692	0.877
	Averages	118.379	0.643	0.854	143.196	0.753	0.898	116.019	0.745	0.890	143.760	**0.795**	**0.914**

Table A.1: ACOVNS Results for Each Replication

Data set	Run	Without Normalization and Heuristic Information			With Normalization			With Heuristic Information			With Both Normalization and Heuristic Information		
		Objective Value	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**	**Objective Value**	**F-measure**	**Rand Index**
Wdbc	1	152461.204	0.805	0.774	184935.051	0.876	0.866	151739.908	0.801	0.769	184935.051	0.876	0.866
	2	152495.912	0.809	0.780	184935.051	0.876	0.866	151735.170	0.809	0.779	184935.051	0.876	0.866
	3	152498.372	0.804	0.775	184935.051	0.876	0.866	151738.815	0.803	0.771	184935.051	0.876	0.866
	4	152496.418	0.809	0.780	184935.051	0.876	0.866	151738.815	0.803	0.771	184013.777	**0.879**	**0.869**
	5	152533.685	0.806	0.777	184935.051	0.876	0.866	151738.487	0.807	0.776	184898.486	0.876	0.866
	6	152495.912	0.809	0.780	184935.051	0.876	0.866	151738.815	0.803	0.771	184935.051	0.876	0.866
	7	152495.529	0.807	0.777	184935.051	0.876	0.866	151738.973	0.805	0.774	184935.051	0.876	0.866
	8	152457.023	0.807	0.777	184935.051	0.876	0.866	151738.487	0.807	0.776	184013.777	**0.879**	**0.869**
	9	152506.030	0.802	0.772	184935.051	0.876	0.866	151742.247	0.805	0.774	184013.777	**0.879**	**0.869**
	10	152508.659	0.802	0.772	184935.051	0.876	0.866	151738.815	0.803	0.771	184935.051	0.876	0.866
	Averages	152494.874	0.806	0.776	184935.051	0.876	0.866	151738.853	0.804	0.773	184655.012	**0.877**	**0.867**

Appendix B

EXPERIMENTAL RESULTS OF THE F-ACOVNS ALGORITHM

In this appendix, we give the detailed experimental results of the F-ACOVNS algorithm for individual algorithm replications.

Table B.1: F-ACOVNS Results for Iris Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Iris	0.05	1	54	3	0.925	0.950	11	10	21
		2	42	3	0.925	0.950	5	15	20
		3	71	3	0.925	0.950	8	5	13
		4	62	3	0.925	0.950	6	5	11
		5	63	3	0.925	0.950	7	7	13
		Average	58	3	0.925	0.950	7	8	16
	0.08	1	40	3	0.925	0.950	4	9	13
		2	37	3	0.925	0.950	9	6	15
		3	51	3	0.925	0.950	7	5	12
		4	34	3	0.925	0.950	4	5	9
		5	58	3	0.925	0.950	6	8	14
		Average	44	3	0.925	0.950	6	7	12
	0.1	1	26	3	0.925	0.950	3	5	8
		2	33	3	0.925	0.950	3	6	9
		3	36	3	0.925	0.950	4	5	9
		4	30	3	0.925	0.950	3	5	8
		5	45	3	0.925	0.950	4	9	13
		Average	34	3	0.925	0.950	3	6	10
	0.15	1	21	3	0.925	0.950	2	5	7
		2	29	3	0.925	0.950	3	6	9
		3	22	3	0.925	0.950	2	5	8
		4	28	3	0.925	0.950	3	8	11
		5	25	3	0.925	0.950	2	7	9
		Average	25	3	0.925	0.950	2	6	9
	0.2	1	18	3	0.925	0.950	2	5	7
		2	24	3	0.925	0.950	2	5	8
		3	21	3	0.925	0.950	2	6	8
		4	27	3	0.925	0.950	3	6	8
		5	23	3	0.925	0.950	2	6	8
		Average	23	3	0.925	0.950	2	5	8
	0.25	1	18	3	0.925	0.950	2	5	7
		2	21	3	0.925	0.950	2	6	8
		3	18	3	0.925	0.950	2	7	9
		4	20	3	0.925	0.950	2	5	7
		5	21	3	0.925	0.950	2	5	7
		Average	20	3	0.925	0.950	2	5	7
	0.3	1	18	3	0.925	0.950	2	5	7
		2	21	3	0.925	0.950	2	5	7
		3	17	3	0.925	0.950	2	5	7
		4	20	3	0.925	0.950	2	5	7
		5	20	3	0.925	0.950	2	5	7
		Average	19	3	0.925	0.950	2	5	7
	0.35	1	16	3	0.925	0.950	2	5	7
		2	9	3	0.925	0.950	1	6	7
		3	17	3	0.925	0.950	2	6	8
		4	20	3	0.925	0.950	2	5	7
		5	13	3	0.913	0.942	1	8	9
		Average	15	3	0.923	0.948	1	6	7

Table B.2: F-ACOVNS Results for Wdbc Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Wdbc	0.05	1	234	10	0.833	0.811	92	69	162
		2	316	13	0.836	0.814	121	80	201
		3	260	11	0.829	0.806	94	80	173
		4	222	16	0.834	0.811	81	88	169
		5	211	14	0.829	0.806	83	77	161
		Average	249	13	0.832	0.810	94	79	173
	0.08	1	165	11	0.816	0.790	68	84	152
		2	116	19	0.855	0.837	50	88	138
		3	188	8	0.814	0.787	74	49	123
		4	168	8	0.822	0.798	60	69	129
		5	136	14	0.836	0.814	48	77	126
		Average	155	12	0.829	0.805	60	74	133
	0.1	1	124	12	0.818	0.792	44	80	124
		2	92	17	0.855	0.837	33	83	116
		3	105	12	0.818	0.792	38	76	113
		4	129	9	0.827	0.803	46	78	124
		5	116	13	0.829	0.806	42	72	114
		Average	113	13	0.829	0.806	41	78	118
	0.15	1	83	6	0.814	0.787	30	64	94
		2	93	13	0.829	0.806	33	77	110
		3	70	12	0.831	0.809	25	73	98
		4	89	8	0.822	0.798	32	68	100
		5	92	11	0.827	0.803	33	54	87
		Average	85	10	0.825	0.800	30	67	98
	0.2	1	49	9	0.829	0.806	18	70	87
		2	65	9	0.816	0.790	23	77	100
		3	58	7	0.814	0.787	21	67	88
		4	86	8	0.827	0.803	31	59	90
		5	63	10	0.833	0.811	22	61	84
		Average	64	9	0.824	0.799	23	67	90
	0.25	1	39	6	0.814	0.787	14	74	88
		2	39	6	0.814	0.787	14	66	80
		3	54	7	0.814	0.787	19	70	90
		4	49	11	0.831	0.809	18	71	89
		5	37	19	0.838	0.817	13	81	95
		Average	44	10	0.822	0.797	16	73	88
	0.3	1	45	8	0.820	0.795	16	68	85
		2	31	6	0.814	0.787	12	65	76
		3	34	9	0.829	0.806	12	60	72
		4	35	11	0.820	0.795	13	75	87
		5	38	8	0.822	0.798	14	60	73
		Average	37	8	0.821	0.796	13	66	79
	0.35	1	24	6	0.814	0.787	9	67	76
		2	23	13	0.834	0.811	8	79	87
		3	29	7	0.820	0.795	11	68	79
		4	29	9	0.814	0.787	10	58	68
		5	33	4	0.804	0.776	12	51	63
		Average	28	8	0.817	0.791	10	65	75

Table B.3: F-ACOVNS Results for Breast Cancer Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Breast Cancer	0.05	1	199	5	0.914	0.905	124	125	249
		2	190	8	0.926	0.919	113	84	197
		3	224	4	0.912	0.903	132	65	196
		4	157	8	0.926	0.919	92	109	201
		5	259	7	0.914	0.905	152	89	242
		Average	206	6	0.918	0.910	123	94	217
	0.08	1	138	4	0.909	0.900	81	136	217
		2	106	4	0.912	0.903	72	123	196
		3	105	8	0.926	0.919	63	82	144
		4	145	4	0.912	0.903	85	131	216
		5	103	4	0.909	0.900	60	129	189
		Average	119	5	0.914	0.905	72	120	192
	0.1	1	82	4	0.909	0.900	48	110	159
		2	105	4	0.909	0.900	62	126	188
		3	76	8	0.926	0.919	45	90	135
		4	108	4	0.916	0.908	63	102	166
		5	86	4	0.911	0.903	51	63	114
		Average	91	5	0.914	0.906	54	98	152
	0.15	1	67	6	0.936	0.929	40	117	156
		2	65	4	0.912	0.903	38	95	133
		3	52	8	0.926	0.919	31	86	118
		4	78	3	0.871	0.857	46	96	142
		5	53	7	0.916	0.908	32	127	159
		Average	63	6	0.912	0.903	37	104	142
	0.2	1	37	4	0.912	0.903	23	111	134
		2	63	4	0.909	0.900	37	132	170
		3	43	6	0.921	0.913	25	90	115
		4	36	4	0.914	0.905	22	126	148
		5	48	4	0.914	0.905	28	77	105
		Average	45	4	0.914	0.905	27	107	134
	0.25	1	30	4	0.914	0.905	18	111	129
		2	42	5	0.914	0.905	25	104	129
		3	38	5	0.912	0.903	22	106	129
		4	32	4	0.916	0.908	19	116	135
		5	33	8	0.926	0.919	20	114	133
		Average	35	5	0.916	0.908	21	110	131
	0.3	1	30	4	0.912	0.903	18	130	148
		2	27	8	0.938	0.932	16	134	150
		3	35	5	0.914	0.905	21	122	143
		4	28	4	0.912	0.903	17	125	141
		5	26	6	0.907	0.898	15	123	139
		Average	29	5	0.916	0.908	17	127	144
	0.35	1	20	8	0.926	0.919	11	104	115
		2	22	7	0.928	0.921	13	100	113
		3	35	4	0.898	0.887	21	101	122
		4	24	4	0.914	0.905	14	131	146
		5	21	6	0.904	0.895	12	116	128
		Average	24	6	0.914	0.906	14	110	125

Table B.4: F-ACOVNS Results for Wine Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Wine	0.05	1	123	7	0.913	0.941	22	23	45
		2	100	7	0.633	0.698	19	26	44
		3	166	5	0.757	0.835	24	21	46
		4	165	7	0.882	0.920	24	18	42
		5	163	7	0.855	0.902	20	17	38
		Average	143	7	0.808	0.859	22	21	43
	0.08	1	110	11	0.924	0.948	19	26	45
		2	81	7	0.883	0.921	17	29	46
		3	108	7	0.903	0.934	17	20	37
		4	102	11	0.895	0.928	14	20	34
		5	83	7	0.773	0.845	11	18	30
		Average	97	9	0.876	0.915	16	23	38
	0.1	1	67	7	0.893	0.927	9	18	27
		2	70	7	0.893	0.927	9	18	27
		3	68	6	0.834	0.888	9	19	28
		4	68	7	0.855	0.902	9	19	28
		5	56	6	0.834	0.888	8	19	27
		Average	66	7	0.862	0.907	9	18	27
	0.15	1	49	7	0.893	0.927	6	19	25
		2	46	8	0.934	0.955	6	19	25
		3	33	12	0.923	0.948	4	19	23
		4	48	7	0.893	0.927	7	19	26
		5	40	6	0.834	0.888	5	19	24
		Average	43	8	0.895	0.929	6	19	25
	0.2	1	31	7	0.846	0.896	4	19	23
		2	33	8	0.934	0.955	5	19	24
		3	44	6	0.874	0.915	6	19	25
		4	28	9	0.800	0.864	4	18	22
		5	36	6	0.824	0.881	5	18	23
		Average	34	7	0.855	0.902	5	19	23
	0.25	1	28	7	0.883	0.921	4	18	22
		2	32	8	0.934	0.955	4	18	22
		3	34	5	0.757	0.835	4	17	22
		4	28	9	0.800	0.864	3	17	21
		5	27	7	0.855	0.902	3	17	21
		Average	30	7	0.846	0.895	4	18	21
	0.3	1	20	7	0.893	0.927	2	18	21
		2	18	8	0.778	0.850	2	18	20
		3	31	7	0.837	0.890	4	17	21
		4	20	10	0.886	0.922	2	18	20
		5	23	5	0.757	0.835	3	17	20
		Average	22	7	0.830	0.885	3	18	20
	0.35	1	20	7	0.882	0.920	2	17	20
		2	20	7	0.893	0.927	3	18	20
		3	27	7	0.847	0.896	3	18	22
		4	16	8	0.856	0.903	2	18	20
		5	21	8	0.914	0.942	3	18	21
		Average	21	7	0.878	0.918	3	18	20

Table B.5: F-ACOVNS Results for Glass Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Glass	0.05	1	99	8	0.457	0.735	26	49	75
		2	115	8	0.450	0.719	25	52	77
		3	82	8	0.419	0.711	14	47	61
		4	105	8	0.419	0.711	17	31	48
		5	83	8	0.416	0.706	14	42	56
		Average	97	8	0.432	0.716	19	44	63
	0.08	1	109	8	0.424	0.706	27	48	75
		2	112	8	0.423	0.703	26	48	74
		3	53	8	0.415	0.716	10	47	57
		4	67	8	0.411	0.712	11	48	59
		5	54	8	0.406	0.712	10	48	58
		Average	79	8	0.416	0.710	17	48	65
	0.1	1	39	8	0.411	0.712	7	40	47
		2	78	8	0.428	0.706	13	45	58
		3	48	8	0.432	0.733	9	47	56
		4	62	8	0.414	0.708	10	46	56
		5	42	8	0.429	0.734	7	44	50
		Average	54	8	0.423	0.719	9	44	53
	0.15	1	27	8	0.419	0.711	4	33	37
		2	80	7	0.432	0.704	13	45	58
		3	42	7	0.457	0.735	7	42	49
		4	62	8	0.411	0.712	11	33	44
		5	28	8	0.421	0.740	5	44	49
		Average	48	8	0.428	0.721	8	39	47
	0.2	1	26	8	0.424	0.708	4	38	42
		2	75	7	0.425	0.706	12	45	57
		3	20	8	0.418	0.704	3	34	37
		4	47	7	0.420	0.717	8	34	42
		5	18	8	0.428	0.706	3	46	49
		Average	37	8	0.423	0.708	6	39	46
	0.25	1	21	8	0.417	0.708	4	43	47
		2	24	8	0.416	0.704	4	39	44
		3	16	8	0.430	0.735	3	37	40
		4	25	6	0.447	0.709	4	45	49
		5	16	8	0.427	0.733	3	46	49
		Average	20	8	0.427	0.718	4	42	46
	0.3	1	21	8	0.413	0.706	4	40	44
		2	19	7	0.418	0.701	3	43	46
		3	17	8	0.411	0.712	3	34	37
		4	23	6	0.442	0.703	4	46	50
		5	15	8	0.398	0.707	3	46	49
		Average	19	7	0.416	0.706	3	42	45
	0.35	1	21	8	0.457	0.735	4	38	42
		2	17	7	0.417	0.701	3	46	49
		3	16	8	0.424	0.713	3	45	48
		4	23	6	0.426	0.734	4	46	50
		5	14	8	0.417	0.703	2	35	38
		Average	18	7	0.428	0.717	3	42	45

Table B.6: F-ACOVNS Results for CMC Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
CMC	0.05	1	104	7	0.365	0.562	152	408	560
		2	119	7	0.366	0.561	173	402	575
		3	108	7	0.364	0.561	150	394	544
		4	105	8	0.369	0.562	145	395	540
		5	116	8	0.363	0.558	160	396	556
		Average	110	7	0.365	0.561	156	399	555
	0.08	1	68	7	0.365	0.560	100	399	498
		2	75	7	0.366	0.561	111	400	511
		3	84	7	0.366	0.559	116	390	506
		4	72	8	0.395	0.551	99	392	491
		5	73	7	0.365	0.562	100	388	489
		Average	74	7	0.371	0.559	105	394	499
	0.1	1	65	7	0.364	0.556	89	394	483
		2	70	8	0.393	0.551	96	393	490
		3	103	2	0.420	0.500	140	194	334
		4	47	8	0.367	0.564	64	390	454
		5	55	7	0.371	0.567	76	390	466
		Average	68	6	0.383	0.548	93	352	445
	0.15	1	37	7	0.365	0.562	51	392	443
		2	47	7	0.368	0.561	65	390	454
		3	46	8	0.367	0.561	63	393	456
		4	32	8	0.396	0.553	44	395	439
		5	34	7	0.365	0.561	47	393	440
		Average	39	7	0.372	0.560	54	392	446
	0.2	1	28	7	0.367	0.564	38	392	431
		2	27	7	0.366	0.561	37	393	430
		3	35	8	0.365	0.560	48	393	441
		4	28	8	0.367	0.548	38	392	430
		5	34	7	0.368	0.565	47	393	440
		Average	30	7	0.366	0.559	42	393	434
	0.25	1	42	2	0.410	0.517	58	235	292
		2	23	7	0.364	0.560	32	393	425
		3	32	6	0.369	0.565	44	389	433
		4	24	8	0.367	0.560	33	393	426
		5	27	7	0.370	0.567	37	392	429
		Average	30	6	0.376	0.554	41	360	401
	0.3	1	23	6	0.365	0.560	31	390	422
		2	22	7	0.366	0.559	30	391	421
		3	29	6	0.368	0.564	40	393	432
		4	24	8	0.365	0.561	33	392	424
		5	19	8	0.371	0.566	26	393	419
		Average	23	7	0.367	0.562	32	392	423
	0.35	1	14	8	0.368	0.562	19	392	411
		2	22	5	0.369	0.558	30	391	420
		3	16	7	0.366	0.563	22	393	415
		4	15	8	0.368	0.549	20	390	410
		5	17	8	0.368	0.564	23	391	414
		Average	17	7	0.368	0.559	23	391	414

Table B.7: F-ACOVNS Results for Liver Disease Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Liver Disease	0.05	1	65	5	0.530	0.500	22	45	67
		2	58	5	0.530	0.500	18	50	68
		3	82	5	0.529	0.500	20	46	66
		4	81	5	0.530	0.500	18	49	66
		5	102	4	0.586	0.501	22	37	59
		Average	78	5	0.541	0.500	20	45	65
	0.08	1	46	4	0.586	0.501	15	40	55
		2	42	5	0.529	0.500	15	43	58
		3	64	4	0.586	0.501	16	40	56
		4	40	5	0.529	0.500	9	46	56
		5	56	5	0.530	0.500	12	38	50
		Average	50	5	0.552	0.500	13	41	55
	0.1	1	41	4	0.586	0.501	9	39	47
		2	21	5	0.530	0.500	5	47	52
		3	41	5	0.530	0.500	9	42	51
		4	36	5	0.529	0.500	8	47	54
		5	54	5	0.530	0.500	11	38	49
		Average	39	5	0.541	0.500	8	43	51
	0.15	1	27	4	0.586	0.501	6	36	42
		2	21	5	0.530	0.500	4	42	47
		3	28	5	0.530	0.500	6	45	51
		4	33	5	0.530	0.500	7	46	53
		5	41	5	0.530	0.500	9	35	43
		Average	30	5	0.541	0.500	6	41	47
	0.2	1	22	4	0.586	0.501	5	35	40
		2	19	5	0.530	0.500	4	41	45
		3	28	5	0.530	0.500	6	45	51
		4	26	4	0.586	0.501	5	37	43
		5	23	5	0.530	0.500	5	38	43
		Average	24	5	0.553	0.500	5	39	44
	0.25	1	19	4	0.586	0.501	4	38	42
		2	14	5	0.517	0.501	3	46	49
		3	16	5	0.530	0.500	3	43	46
		4	22	4	0.586	0.501	5	35	40
		5	17	5	0.530	0.500	4	39	43
		Average	18	5	0.550	0.500	4	40	44
	0.3	1	17	4	0.586	0.501	4	42	45
		2	13	5	0.538	0.500	3	46	48
		3	16	5	0.530	0.500	3	42	45
		4	20	4	0.586	0.501	4	41	45
		5	16	5	0.530	0.500	3	40	43
		Average	16	5	0.554	0.500	3	42	45
	0.35	1	14	5	0.530	0.500	3	35	38
		2	11	5	0.538	0.500	2	42	45
		3	16	5	0.530	0.500	3	43	46
		4	14	5	0.529	0.500	3	45	48
		5	15	5	0.530	0.500	3	34	38
		Average	14	5	0.532	0.500	3	40	43

Table B.8: F-ACOVNS Results for Thyroid Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Thyroid	0.05	1	81	4	0.898	0.884	19	31	50
		2	165	4	0.898	0.884	31	24	55
		3	127	4	0.898	0.884	19	22	41
		4	118	4	0.708	0.704	17	51	69
		5	92	4	0.723	0.719	15	52	66
		Average	117	4	0.825	0.815	20	36	56
	0.08	1	65	4	0.898	0.884	15	24	39
		2	77	4	0.898	0.884	18	27	45
		3	65	4	0.898	0.884	10	26	36
		4	79	4	0.898	0.884	12	40	53
		5	83	4	0.580	0.601	12	22	34
		Average	74	4	0.835	0.828	14	28	41
	0.1	1	45	4	0.591	0.610	7	35	41
		2	69	4	0.898	0.884	11	30	41
		3	57	4	0.898	0.884	9	22	31
		4	74	4	0.899	0.884	11	40	51
		5	54	4	0.589	0.609	8	25	33
		Average	60	4	0.775	0.774	9	30	39
	0.15	1	25	4	0.898	0.884	4	28	31
		2	65	4	0.905	0.891	10	33	43
		3	41	4	0.830	0.816	6	51	57
		4	33	4	0.850	0.835	5	50	55
		5	50	4	0.591	0.610	7	23	30
		Average	43	4	0.815	0.808	6	37	43
	0.2	1	24	4	0.898	0.884	3	23	26
		2	28	4	0.898	0.884	4	23	28
		3	37	4	0.781	0.769	5	51	57
		4	28	4	0.898	0.884	4	27	31
		5	46	4	0.762	0.754	7	50	57
		Average	33	4	0.848	0.835	5	35	40
	0.25	1	23	4	0.898	0.884	4	27	31
		2	21	4	0.898	0.884	3	23	26
		3	15	4	0.898	0.884	2	23	25
		4	26	4	0.898	0.884	4	22	26
		5	24	4	0.898	0.884	3	25	29
		Average	22	4	0.898	0.884	3	24	27
	0.3	1	22	4	0.898	0.884	3	25	28
		2	21	4	0.898	0.884	3	23	26
		3	14	4	0.898	0.884	2	23	25
		4	25	4	0.899	0.884	4	40	44
		5	16	4	0.898	0.884	2	22	24
		Average	20	4	0.899	0.884	3	27	29
	0.35	1	22	4	0.898	0.884	3	28	31
		2	21	4	0.898	0.884	3	23	26
		3	10	4	0.898	0.884	1	24	25
		4	24	4	0.898	0.884	3	23	26
		5	16	4	0.898	0.884	2	26	28
		Average	19	4	0.898	0.884	3	25	27

Table B.9: F-ACOVNS Results for Zoo Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
Zoo	0.05	1	200	5	0.797	0.913	25	9	34
		2	140	4	0.803	0.915	18	5	22
		3	189	5	0.797	0.913	17	6	23
		4	146	13	0.952	0.977	13	8	21
		5	191	5	0.797	0.913	17	7	24
		Average	173	6	0.829	0.926	18	7	25
	0.08	1	96	14	0.981	0.991	12	14	26
		2	114	4	0.799	0.912	14	4	18
		3	60	14	0.939	0.970	5	9	14
		4	108	13	0.900	0.953	9	8	17
		5	92	4	0.803	0.915	8	4	12
		Average	94	10	0.884	0.948	10	8	17
	0.1	1	117	5	0.780	0.904	11	5	16
		2	75	4	0.799	0.912	7	8	14
		3	59	14	0.779	0.906	5	10	15
		4	107	5	0.780	0.904	9	5	14
		5	84	4	0.781	0.904	7	5	13
		Average	88	6	0.784	0.906	8	7	14
	0.15	1	54	11	0.920	0.963	5	11	16
		2	52	4	0.799	0.912	5	4	8
		3	74	5	0.921	0.960	6	5	11
		4	62	5	0.797	0.913	6	10	16
		5	70	4	0.803	0.915	6	4	10
		Average	62	6	0.848	0.933	5	7	12
	0.2	1	46	11	0.892	0.949	4	12	16
		2	52	4	0.804	0.915	4	4	8
		3	48	5	0.797	0.913	4	9	14
		4	33	14	0.935	0.969	3	10	13
		5	32	13	0.940	0.972	3	12	15
		Average	42	9	0.874	0.944	4	9	13
	0.25	1	44	4	0.804	0.915	4	5	9
		2	32	5	0.780	0.902	3	4	7
		3	35	6	0.945	0.974	3	5	8
		4	32	4	0.803	0.915	3	9	12
		5	28	13	0.979	0.990	2	8	11
		Average	34	6	0.862	0.939	3	6	9
	0.3	1	39	5	0.921	0.960	3	6	9
		2	20	12	0.850	0.932	2	11	13
		3	31	6	0.927	0.964	3	5	7
		4	32	5	0.770	0.900	3	7	10
		5	45	7	0.702	0.872	4	5	10
		Average	33	7	0.834	0.926	3	7	10
	0.35	1	33	6	0.876	0.941	3	13	16
		2	19	12	0.751	0.896	2	12	14
		3	24	5	0.921	0.960	2	5	7
		4	25	4	0.803	0.915	2	5	7
		5	26	8	0.821	0.923	2	8	10
		Average	25	7	0.834	0.927	2	8	11

Table B.10: F-ACOVNS Results for 5d5c1_1 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
5d5c1_1	0.05	1	55	5	0.989	0.995	83	336	419
		2	62	5	0.952	0.977	93	335	428
		3	50	5	0.864	0.935	69	327	396
		4	56	5	0.994	0.997	76	316	392
		5	59	5	0.975	0.988	81	325	406
		Average	56	5	0.955	0.978	81	328	408
	0.08	1	38	5	0.989	0.994	57	323	380
		2	47	5	0.952	0.977	71	333	404
		3	39	5	0.991	0.996	56	321	378
		4	43	5	0.990	0.995	59	318	377
		5	41	5	0.994	0.997	56	314	370
		Average	42	5	0.983	0.992	60	322	382
	0.1	1	34	5	0.995	0.998	47	309	355
		2	42	5	0.831	0.920	58	327	386
		3	32	5	0.988	0.994	44	307	351
		4	39	5	0.996	0.998	54	325	379
		5	37	5	0.993	0.997	51	296	347
		Average	37	5	0.961	0.981	51	313	364
	0.15	1	25	5	0.886	0.946	34	327	362
		2	29	5	0.838	0.922	39	308	347
		3	23	5	0.834	0.921	31	312	343
		4	24	5	0.954	0.978	33	329	362
		5	26	5	0.991	0.996	35	315	350
		Average	25	5	0.901	0.952	35	318	353
	0.2	1	20	5	0.977	0.989	27	315	342
		2	20	5	0.991	0.996	27	314	341
		3	19	5	0.873	0.940	26	325	351
		4	18	5	0.783	0.892	25	327	352
		5	22	5	0.991	0.996	30	327	357
		Average	20	5	0.923	0.962	27	322	349
	0.25	1	18	5	0.989	0.994	25	318	343
		2	15	5	0.992	0.996	21	330	351
		3	16	5	0.989	0.995	22	320	341
		4	17	5	0.780	0.891	23	318	341
		5	19	5	0.993	0.996	26	330	356
		Average	17	5	0.949	0.974	23	323	346
	0.3	1	16	5	0.841	0.925	21	318	340
		2	14	5	0.873	0.940	19	326	345
		3	14	5	0.807	0.902	19	318	337
		4	14	5	0.996	0.998	19	325	344
		5	14	5	0.995	0.997	19	307	325
		Average	14	5	0.902	0.953	19	319	338
	0.35	1	13	5	0.882	0.942	17	321	338
		2	14	5	0.901	0.953	19	329	348
		3	13	5	0.992	0.996	17	323	340
		4	12	5	0.988	0.994	16	326	342
		5	14	5	0.986	0.993	19	306	324
		Average	13	5	0.950	0.976	18	321	339

Table B.11: F-ACOVNS Results for 5d5c1_3 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
5d5c1_3	0.05	1	51	5	0.990	0.995	67	279	346
		2	58	5	0.992	0.996	72	277	349
		3	57	5	0.772	0.887	64	267	331
		4	55	5	0.993	0.996	61	270	330
		5	52	5	0.991	0.996	58	299	357
		Average	55	5	0.948	0.974	64	278	343
	0.08	1	40	5	0.976	0.989	56	293	350
		2	43	5	0.982	0.991	57	298	355
		3	39	5	0.961	0.981	47	248	295
		4	40	5	0.972	0.987	40	233	273
		5	41	5	0.988	0.994	43	245	288
		Average	41	5	0.976	0.988	49	263	312
	0.1	1	38	5	0.989	0.995	41	234	274
		2	36	5	0.834	0.921	35	232	268
		3	35	5	0.990	0.995	34	233	267
		4	35	5	0.993	0.997	34	232	267
		5	31	5	0.881	0.944	30	231	261
		Average	35	5	0.938	0.970	35	232	267
	0.15	1	26	5	0.990	0.995	25	231	257
		2	27	5	0.994	0.997	26	232	258
		3	32	5	0.991	0.995	31	232	263
		4	27	5	0.983	0.992	27	231	258
		5	27	5	0.962	0.982	26	232	259
		Average	28	5	0.984	0.992	27	232	259
	0.2	1	24	5	0.990	0.995	25	233	258
		2	23	5	0.982	0.991	23	232	254
		3	25	5	0.980	0.990	24	233	257
		4	24	5	0.986	0.993	23	231	254
		5	19	5	0.988	0.994	18	264	282
		Average	23	5	0.985	0.993	23	238	261
	0.25	1	15	5	0.902	0.954	16	274	291
		2	16	5	0.905	0.955	18	259	276
		3	19	5	0.993	0.996	18	291	310
		4	20	5	0.798	0.898	27	313	339
		5	16	5	0.953	0.977	20	267	287
		Average	17	5	0.910	0.956	20	281	301
	0.3	1	15	5	0.834	0.921	15	230	245
		2	13	5	0.984	0.992	12	233	245
		3	14	5	0.995	0.998	13	232	245
		4	15	5	0.873	0.932	14	231	245
		5	15	5	0.981	0.991	14	231	245
		Average	14	5	0.933	0.967	14	231	245
	0.35	1	13	5	0.962	0.981	12	231	243
		2	12	5	0.988	0.994	11	232	243
		3	13	5	0.887	0.947	12	231	243
		4	12	5	0.990	0.995	12	231	243
		5	12	5	0.977	0.989	11	231	242
		Average	12	5	0.961	0.981	12	231	243

Table B.12: F-ACOVNS Results for 5d5c1_5 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
5d5c1_5	0.05	1	49	5	0.989	0.994	55	220	275
		2	52	5	0.899	0.952	58	218	277
		3	55	5	0.990	0.995	57	216	273
		4	51	5	0.993	0.996	52	221	273
		5	51	5	0.990	0.995	56	222	278
		Average	52	5	0.972	0.987	56	219	275
	0.08	1	39	5	0.958	0.980	43	223	266
		2	39	5	0.894	0.950	42	222	264
		3	36	5	0.855	0.932	37	203	240
		4	42	5	0.988	0.994	41	205	246
		5	37	5	0.847	0.928	37	202	240
		Average	39	5	0.909	0.957	40	211	251
	0.1	1	32	5	0.982	0.991	31	209	240
		2	35	5	0.987	0.994	35	212	247
		3	31	5	0.890	0.948	31	217	248
		4	34	5	0.875	0.941	36	209	245
		5	32	5	0.980	0.990	31	203	234
		Average	33	5	0.943	0.973	33	210	243
	0.15	1	23	5	0.976	0.988	22	202	225
		2	26	5	0.964	0.983	25	207	232
		3	24	5	0.991	0.996	27	206	233
		4	26	5	0.987	0.994	25	205	230
		5	24	5	0.778	0.891	23	202	226
		Average	25	5	0.939	0.970	25	204	229
	0.2	1	19	5	0.950	0.976	18	202	220
		2	21	5	0.875	0.941	20	201	222
		3	20	5	0.899	0.952	20	203	223
		4	19	5	0.968	0.984	18	203	221
		5	18	5	0.984	0.992	17	209	226
		Average	19	5	0.935	0.969	19	204	222
	0.25	1	15	5	0.851	0.925	16	214	229
		2	16	5	0.987	0.993	17	217	234
		3	17	5	0.980	0.990	17	224	241
		4	16	5	0.870	0.939	17	217	234
		5	15	5	0.987	0.994	16	215	232
		Average	16	5	0.935	0.968	17	217	234
	0.3	1	13	5	0.972	0.986	13	209	221
		2	15	5	0.924	0.964	15	212	227
		3	16	5	0.986	0.993	16	216	232
		4	15	5	0.780	0.892	15	214	229
		5	14	5	0.979	0.990	14	213	226
		Average	15	5	0.928	0.965	14	213	227
	0.35	1	12	5	0.978	0.989	12	205	217
		2	13	5	0.846	0.928	12	202	214
		3	11	5	0.835	0.922	11	201	212
		4	12	5	0.979	0.989	12	208	220
		5	12	5	0.987	0.994	12	205	217
		Average	12	5	0.925	0.964	12	204	216

Table B.13: F-ACOVNS Results for 10d5c1_2 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
10d5c1_2	0.05	1	76	10	0.999	0.999	124	402	526
		2	67	10	0.996	0.998	110	448	558
		3	65	10	0.998	0.999	101	435	536
		4	72	10	0.997	0.999	111	412	523
		5	64	10	1.000	1.000	98	404	502
		Average	69	10	0.998	0.999	109	420	529
	0.08	1	49	10	0.999	0.999	82	425	507
		2	54	10	0.995	0.998	90	361	451
		3	46	10	0.999	1.000	71	381	452
		4	55	10	0.996	0.998	84	425	510
		5	51	10	0.998	0.999	79	445	524
		Average	51	10	0.997	0.999	81	408	489
	0.1	1	45	10	0.999	0.999	69	421	490
		2	39	10	0.995	0.998	61	390	451
		3	37	10	0.998	0.999	57	402	459
		4	42	10	0.996	0.998	64	371	436
		5	46	10	0.999	1.000	71	449	519
		Average	42	10	0.997	0.999	64	407	471
	0.15	1	30	10	0.997	0.999	46	392	437
		2	27	10	0.998	0.999	41	363	404
		3	27	10	0.998	0.999	41	396	438
		4	29	10	0.995	0.998	44	431	476
		5	28	10	0.984	0.993	42	442	484
		Average	28	10	0.994	0.998	43	405	448
	0.2	1	29	10	0.997	0.999	44	434	478
		2	22	10	0.998	0.999	34	381	414
		3	22	10	0.994	0.997	34	437	470
		4	23	10	0.995	0.998	35	438	473
		5	23	10	0.999	0.999	35	440	475
		Average	24	10	0.996	0.998	36	426	462
	0.25	1	17	10	0.999	0.999	26	444	470
		2	18	10	0.985	0.993	28	442	469
		3	18	10	0.997	0.999	27	433	460
		4	17	10	0.997	0.999	26	447	473
		5	18	10	0.998	0.999	27	445	472
		Average	18	10	0.995	0.998	27	442	469
	0.3	1	19	10	0.999	0.999	29	443	472
		2	18	10	0.997	0.999	28	408	435
		3	17	10	0.996	0.998	26	365	391
		4	14	10	0.997	0.999	21	395	416
		5	17	10	0.998	0.999	25	444	470
		Average	17	10	0.997	0.999	26	411	437
	0.35	1	13	10	0.996	0.998	19	444	464
		2	17	10	0.997	0.999	26	444	470
		3	16	10	0.998	0.999	24	407	431
		4	13	10	0.997	0.999	20	446	466
		5	17	10	0.999	1.000	25	445	471
		Average	15	10	0.997	0.999	23	437	460

Table B.14: F-ACOVNS Results for 10d5c1_5 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
10d5c1_5	0.05	1	71	10	0.997	0.999	117	426	543
		2	63	10	0.998	0.999	105	440	545
		3	65	10	0.997	0.999	102	337	439
		4	60	10	0.999	1.000	93	420	514
		5	82	10	0.999	0.999	128	456	584
		Average	68	10	0.998	0.999	109	416	525
	0.08	1	52	10	1.000	1.000	88	440	527
		2	50	10	0.998	0.999	84	478	562
		3	45	10	0.998	0.999	70	417	487
		4	45	10	0.996	0.998	70	438	508
		5	49	10	0.996	0.998	76	468	544
		Average	48	10	0.997	0.999	78	448	526
	0.1	1	51	10	0.999	0.999	80	434	514
		2	42	10	0.996	0.998	65	384	449
		3	42	10	0.997	0.999	65	357	423
		4	44	10	0.998	0.999	69	459	527
		5	46	10	0.766	0.899	73	458	531
		Average	45	10	0.951	0.979	70	418	489
	0.15	1	27	10	0.999	1.000	42	456	498
		2	29	10	0.998	0.999	45	428	473
		3	27	10	0.997	0.999	42	416	457
		4	38	10	0.998	0.999	59	469	528
		5	32	10	0.999	1.000	49	398	447
		Average	31	10	0.998	0.999	47	433	481
	0.2	1	21	10	0.978	0.991	32	464	496
		2	23	10	0.994	0.997	35	469	504
		3	24	10	0.832	0.926	37	464	501
		4	28	10	0.994	0.997	43	453	497
		5	23	10	0.997	0.999	35	419	455
		Average	24	10	0.959	0.982	37	454	491
	0.25	1	17	10	0.994	0.997	26	466	493
		2	19	10	0.997	0.999	29	402	431
		3	18	10	0.998	0.999	28	460	488
		4	20	10	0.998	0.999	31	413	444
		5	17	10	0.997	0.999	26	422	447
		Average	18	10	0.996	0.998	28	433	461
	0.3	1	14	10	0.999	0.999	21	468	490
		2	18	10	0.997	0.999	27	387	415
		3	17	10	0.997	0.999	26	456	482
		4	16	10	0.999	0.999	24	432	457
		5	20	10	0.997	0.999	30	463	494
		Average	17	10	0.998	0.999	26	442	467
	0.35	1	13	10	0.999	0.999	20	470	490
		2	15	10	0.999	1.000	23	391	414
		3	15	10	0.997	0.999	23	385	408
		4	13	10	0.997	0.999	20	390	410
		5	20	10	0.998	0.999	31	464	494
		Average	15	10	0.998	0.999	23	420	443

Table B.15: F-ACOVNS Results for 10d5c1_10 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
10d5c1_10	0.05	1	77	10	0.997	0.999	126	420	546
		2	75	10	0.999	0.999	124	435	559
		3	73	10	0.999	1.000	115	441	556
		4	82	10	0.996	0.998	128	417	545
		5	69	10	0.997	0.999	109	370	478
		Average	75	10	0.997	0.999	120	417	537
	0.08	1	61	10	0.998	0.999	100	470	571
		2	54	10	0.989	0.995	90	459	549
		3	48	10	1.000	1.000	75	395	470
		4	52	10	0.998	0.999	81	425	506
		5	50	10	0.997	0.999	79	390	469
		Average	53	10	0.997	0.998	85	428	513
	0.1	1	53	10	0.997	0.998	82	446	529
		2	42	10	0.998	0.999	65	464	529
		3	37	10	0.999	0.999	58	435	493
		4	46	10	0.997	0.999	71	459	531
		5	40	10	0.998	0.999	62	457	519
		Average	44	10	0.998	0.999	68	453	520
	0.15	1	29	10	0.984	0.993	45	461	506
		2	32	10	0.998	0.999	50	393	442
		3	30	10	0.997	0.999	47	368	415
		4	35	10	0.809	0.914	55	461	516
		5	31	10	0.996	0.998	49	452	501
		Average	31	10	0.957	0.981	49	427	476
	0.2	1	33	10	0.997	0.999	52	452	503
		2	23	10	0.781	0.904	36	465	500
		3	30	10	0.998	0.999	46	412	458
		4	27	10	0.999	1.000	42	404	446
		5	23	10	0.998	0.999	36	458	493
		Average	27	10	0.955	0.980	42	438	480
	0.25	1	18	10	0.996	0.998	28	384	411
		2	19	10	0.887	0.951	30	467	497
		3	21	10	0.997	0.999	32	418	450
		4	25	9	0.998	0.999	38	400	439
		5	21	10	0.913	0.962	32	463	495
		Average	21	10	0.958	0.982	32	426	458
	0.3	1	18	10	0.998	0.999	28	461	489
		2	17	10	0.997	0.998	26	404	430
		3	16	10	0.996	0.998	24	452	476
		4	24	9	0.999	1.000	37	427	463
		5	20	10	0.997	0.999	31	420	451
		Average	19	10	0.997	0.999	29	433	462
	0.35	1	18	10	0.998	0.999	27	462	490
		2	16	10	0.951	0.979	25	460	485
		3	17	10	0.781	0.904	26	463	489
		4	15	10	0.997	0.999	23	456	478
		5	17	10	0.997	0.999	26	464	490
		Average	17	10	0.945	0.976	25	461	486

Table B.16: F-ACOVNS Results for 20d5c1_4 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
20d5c1_4	0.05	1	107	20	0.909	0.954	148	439	587
		2	96	20	0.863	0.934	133	446	579
		3	128	20	0.999	1.000	169	435	604
		4	169	19	0.999	1.000	221	439	660
		5	96	20	0.861	0.933	127	428	555
		Average	119	20	0.926	0.964	160	438	597
	0.08	1	77	20	1.000	1.000	108	395	502
		2	79	20	1.000	1.000	111	398	509
		3	74	20	0.998	0.999	98	428	527
		4	80	20	0.898	0.949	107	415	522
		5	97	20	0.867	0.936	128	434	562
		Average	81	20	0.953	0.977	110	414	524
	0.1	1	65	20	1.000	1.000	86	370	456
		2	73	20	0.999	0.999	96	400	496
		3	64	19	1.000	1.000	84	380	464
		4	55	20	1.000	1.000	73	441	513
		5	85	17	0.899	0.949	111	429	541
		Average	68	19	0.980	0.990	90	404	494
	0.15	1	46	19	1.000	1.000	60	376	436
		2	45	18	0.860	0.933	59	418	477
		3	35	20	0.920	0.959	46	420	466
		4	49	20	0.904	0.952	66	408	474
		5	57	18	0.852	0.929	76	426	502
		Average	46	19	0.907	0.955	61	410	471
	0.2	1	58	16	1.000	1.000	76	365	441
		2	36	18	0.864	0.934	48	426	473
		3	32	20	1.000	1.000	43	441	484
		4	42	17	0.903	0.951	55	428	483
		5	47	16	1.000	1.000	62	384	445
		Average	43	17	0.953	0.977	57	409	465
	0.25	1	30	16	0.991	0.995	39	441	480
		2	46	15	0.896	0.949	60	414	475
		3	52	16	1.000	1.000	68	429	498
		4	40	16	1.000	1.000	52	407	460
		5	25	19	0.908	0.954	33	427	459
		Average	39	16	0.959	0.980	51	424	474
	0.3	1	46	12	0.904	0.952	60	423	483
		2	32	17	1.000	1.000	42	407	448
		3	32	18	1.000	1.000	42	382	424
		4	21	20	0.849	0.928	27	417	444
		5	22	19	1.000	1.000	29	377	406
		Average	31	17	0.951	0.976	40	401	441
	0.35	1	23	19	1.000	1.000	30	403	433
		2	20	20	1.000	1.000	26	396	422
		3	32	16	0.901	0.951	42	399	441
		4	18	20	1.000	1.000	23	434	458
		5	27	17	0.899	0.950	35	422	456
		Average	24	18	0.960	0.980	31	411	442

Table B.17: F-ACOVNS Results for 20d5c1_10 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
20d5c1_10	0.05	1	131	20	1.000	1.000	138	292	431
		2	139	20	0.999	0.999	146	318	464
		3	91	20	0.905	0.953	94	303	396
		4	103	20	0.865	0.934	106	297	402
		5	94	20	0.915	0.957	97	305	403
		Average	112	20	0.937	0.969	116	303	419
	0.08	1	74	20	0.997	0.999	82	310	391
		2	70	20	0.922	0.960	77	302	379
		3	69	20	0.894	0.947	72	300	371
		4	118	19	1.000	1.000	120	287	407
		5	66	20	0.983	0.991	66	304	370
		Average	79	20	0.959	0.979	83	300	384
	0.1	1	74	20	1.000	1.000	73	260	334
		2	62	20	0.992	0.996	62	296	358
		3	52	20	0.998	0.999	52	280	332
		4	89	18	0.999	1.000	90	218	308
		5	59	20	1.000	1.000	59	303	362
		Average	67	20	0.998	0.999	67	272	339
	0.15	1	62	20	0.897	0.949	64	293	357
		2	66	17	0.904	0.951	66	297	362
		3	47	20	1.000	1.000	49	301	350
		4	55	19	1.000	1.000	59	305	364
		5	40	20	0.892	0.946	45	316	361
		Average	54	19	0.939	0.969	57	302	359
	0.2	1	35	20	0.998	0.999	40	331	371
		2	31	20	0.868	0.936	36	307	343
		3	33	20	0.849	0.928	34	303	337
		4	46	18	1.000	1.000	48	284	332
		5	35	20	0.908	0.954	36	302	338
		Average	36	20	0.925	0.963	39	305	344
	0.25	1	33	19	0.998	0.999	33	279	312
		2	34	20	1.000	1.000	34	292	326
		3	39	19	1.000	1.000	39	269	307
		4	35	18	1.000	1.000	34	296	331
		5	36	18	1.000	1.000	35	280	316
		Average	35	19	1.000	1.000	35	283	318
	0.3	1	24	20	0.890	0.945	24	291	315
		2	34	19	0.999	0.999	37	298	335
		3	31	20	0.900	0.950	31	284	315
		4	25	17	0.849	0.928	25	283	308
		5	20	20	0.999	0.999	20	273	293
		Average	27	19	0.927	0.964	27	286	313
	0.35	1	28	17	0.895	0.948	28	283	310
		2	38	14	0.898	0.948	37	278	315
		3	37	15	0.998	0.999	36	264	301
		4	29	17	0.891	0.946	29	283	312
		5	30	17	0.874	0.939	30	283	313
		Average	32	16	0.911	0.956	32	278	310

Table B.18: F-ACOVNS Results for 20d5c1_20 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
20d5c1_20	0.05	1	155	20	0.906	0.953	215	397	612
		2	169	19	0.996	0.998	233	394	627
		3	172	20	0.998	0.999	229	363	593
		4	143	20	1.000	1.000	192	351	543
		5	146	20	0.905	0.952	197	389	586
		Average	157	20	0.961	0.980	213	379	592
	0.08	1	106	19	1.000	1.000	148	337	485
		2	82	20	0.901	0.950	116	404	520
		3	122	20	0.899	0.950	164	392	556
		4	115	16	0.999	1.000	153	388	541
		5	99	20	0.895	0.948	133	396	529
		Average	105	19	0.939	0.969	143	383	526
	0.1	1	90	17	0.999	1.000	120	394	514
		2	70	20	0.998	0.999	95	370	465
		3	76	20	0.894	0.947	101	389	490
		4	99	17	0.854	0.930	131	378	510
		5	76	20	0.897	0.949	102	362	463
		Average	82	19	0.928	0.965	110	379	488
	0.15	1	73	16	0.955	0.976	98	354	452
		2	44	20	0.999	0.999	59	400	459
		3	90	14	0.891	0.945	120	392	511
		4	83	17	0.989	0.994	111	396	507
		5	40	19	0.881	0.942	53	395	448
		Average	66	17	0.943	0.971	88	387	475
	0.2	1	40	20	1.000	1.000	54	384	437
		2	58	16	0.922	0.960	78	374	451
		3	60	14	0.935	0.967	80	375	455
		4	41	18	0.999	0.999	54	375	430
		5	40	19	0.998	0.999	53	392	444
		Average	48	17	0.971	0.985	64	380	444
	0.25	1	52	15	1.000	1.000	69	367	437
		2	25	20	0.922	0.960	34	395	429
		3	62	14	1.000	1.000	83	328	410
		4	27	16	0.999	0.999	36	381	417
		5	37	18	1.000	1.000	49	399	449
		Average	41	17	0.984	0.992	54	374	428
	0.3	1	38	14	0.911	0.954	51	372	423
		2	31	16	1.000	1.000	41	384	425
		3	35	14	1.000	1.000	47	362	409
		4	38	15	1.000	1.000	50	316	366
		5	31	15	1.000	1.000	41	395	436
		Average	35	15	0.982	0.991	46	366	412
	0.35	1	21	18	1.000	1.000	28	396	424
		2	27	15	0.999	0.999	36	387	423
		3	34	15	0.910	0.953	45	383	428
		4	30	14	1.000	1.000	40	322	361
		5	24	17	1.000	1.000	32	359	391
		Average	27	16	0.982	0.990	36	369	405

Table B.19: F-ACOVNS Results for 40d5c1_8 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
40d5c1_8	0.05	1	106	40	0.899	0.951	212	590	802
		2	166	40	0.861	0.935	323	688	1011
		3	121	40	0.998	0.999	234	680	914
		4	156	39	0.860	0.934	299	623	923
		5	163	40	0.998	0.999	314	617	931
		Average	142	40	0.923	0.964	276	640	916
	0.08	1	135	39	0.998	0.999	262	627	889
		2	96	40	0.876	0.941	192	706	898
		3	135	40	0.886	0.945	257	682	939
		4	159	36	0.999	0.999	299	632	931
		5	83	40	0.922	0.961	159	675	834
		Average	122	39	0.936	0.969	234	664	898
	0.1	1	72	40	0.918	0.960	138	559	697
		2	73	40	0.870	0.938	139	661	800
		3	154	31	0.896	0.950	288	670	958
		4	86	40	0.893	0.948	163	603	767
		5	61	40	0.945	0.972	117	585	702
		Average	89	38	0.905	0.954	169	616	785
	0.15	1	143	26	0.882	0.943	267	650	917
		2	104	29	0.928	0.964	195	608	803
		3	42	40	0.903	0.953	81	680	760
		4	68	40	0.860	0.934	129	666	795
		5	137	23	0.998	0.999	256	633	889
		Average	99	32	0.914	0.959	185	648	833
	0.2	1	89	23	0.994	0.997	167	610	777
		2	62	30	0.998	0.999	118	508	626
		3	35	40	0.998	0.999	65	652	718
		4	41	40	0.856	0.933	76	633	709
		5	67	28	0.877	0.941	122	619	741
		Average	59	32	0.945	0.974	110	605	714
	0.25	1	77	25	0.998	0.999	140	619	759
		2	64	25	0.921	0.961	116	606	722
		3	29	39	0.996	0.998	53	632	685
		4	65	27	0.999	0.999	118	595	713
		5	45	34	0.998	0.999	83	655	737
		Average	56	30	0.982	0.991	102	621	723
	0.3	1	56	25	0.998	0.999	101	604	705
		2	28	40	0.938	0.969	51	649	700
		3	55	22	0.855	0.932	98	588	686
		4	23	40	0.939	0.969	43	653	696
		5	49	24	0.998	0.999	89	574	663
		Average	42	30	0.945	0.974	76	614	690
	0.35	1	58	24	0.900	0.951	105	588	692
		2	36	33	0.998	0.999	66	527	593
		3	48	20	0.997	0.998	86	545	632
		4	53	23	0.887	0.945	95	585	680
		5	48	26	0.937	0.969	87	620	707
		Average	49	25	0.944	0.972	88	573	661

Table B.20: F-ACOVNS Results for 40d5c1_20 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
40d5c1_20	0.05	1	126	40	0.896	0.950	252	697	948
		2	155	40	0.999	0.999	307	710	1016
		3	202	39	1.000	1.000	388	704	1092
		4	186	40	0.861	0.935	360	690	1050
		5	202	38	0.870	0.938	390	682	1072
		Average	174	39	0.925	0.965	339	697	1036
	0.08	1	202	30	0.918	0.960	392	676	1069
		2	232	25	0.915	0.958	449	681	1130
		3	119	39	0.896	0.950	233	696	929
		4	239	27	0.999	0.999	456	689	1145
		5	109	39	1.000	1.000	212	713	925
		Average	180	32	0.946	0.973	348	691	1040
	0.1	1	230	24	0.872	0.939	438	664	1102
		2	217	27	0.999	0.999	414	681	1095
		3	126	36	0.865	0.937	243	687	930
		4	188	28	0.999	0.999	359	695	1054
		5	123	35	0.999	0.999	237	696	933
		Average	177	30	0.947	0.975	338	685	1023
	0.15	1	111	24	0.999	1.000	213	676	889
		2	130	26	1.000	1.000	248	679	927
		3	109	30	0.998	0.999	208	686	894
		4	51	40	0.936	0.968	102	699	801
		5	107	29	0.999	0.999	206	688	893
		Average	102	30	0.986	0.993	195	685	881
	0.2	1	91	25	1.000	1.000	174	703	877
		2	46	37	0.998	0.999	89	704	792
		3	42	39	0.857	0.933	82	695	777
		4	86	25	0.999	0.999	168	676	844
		5	86	25	0.999	0.999	165	682	847
		Average	70	30	0.971	0.986	136	692	828
	0.25	1	49	29	0.889	0.947	96	680	776
		2	59	27	0.999	0.999	116	687	802
		3	77	22	0.860	0.934	149	662	811
		4	46	33	0.999	1.000	90	702	793
		5	49	25	0.859	0.934	96	667	763
		Average	56	27	0.921	0.963	109	680	789
	0.3	1	34	32	1.000	1.000	73	704	777
		2	27	37	0.999	0.999	59	704	763
		3	68	23	0.999	1.000	131	670	801
		4	48	29	0.904	0.953	93	675	767
		5	65	24	0.999	0.999	126	683	809
		Average	48	29	0.980	0.990	96	687	783
	0.35	1	45	23	0.997	0.999	87	663	751
		2	53	23	0.999	1.000	100	663	763
		3	40	32	0.999	0.999	76	685	761
		4	34	28	0.999	0.999	65	687	753
		5	49	21	0.855	0.932	94	650	744
		Average	44	25	0.970	0.986	85	670	754

Table B.21: F-ACOVNS Results for 40d5c1_40 Data Set

Data set	Evaporation Rate	Replication	Number of iterations for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
40d5c1_40	0.05	1	201	39	0.859	0.934	285	490	776
		2	202	39	0.998	0.999	282	489	771
		3	169	40	0.998	0.999	238	497	735
		4	190	40	0.872	0.940	260	468	728
		5	297	32	0.999	0.999	398	490	888
		Average	212	38	0.945	0.974	293	487	779
	0.08	1	192	32	0.998	0.999	269	479	748
		2	137	40	0.997	0.998	189	479	668
		3	110	40	0.997	0.998	151	475	625
		4	202	30	0.961	0.980	271	454	726
		5	230	27	0.998	0.999	306	453	759
		Average	174	34	0.990	0.995	237	468	705
	0.1	1	208	24	0.918	0.959	295	453	747
		2	94	37	0.997	0.998	132	477	609
		3	126	29	0.998	0.999	176	468	644
		4	191	29	0.859	0.934	260	460	720
		5	111	35	0.929	0.965	157	479	636
		Average	146	31	0.940	0.971	204	467	671
	0.15	1	128	25	0.870	0.938	178	464	642
		2	142	26	0.856	0.932	204	462	666
		3	132	25	0.999	0.999	184	479	663
		4	135	23	1.000	1.000	188	463	651
		5	143	22	0.998	0.999	191	452	642
		Average	136	24	0.945	0.974	189	464	653
	0.2	1	80	27	0.857	0.933	117	452	569
		2	73	30	0.925	0.963	99	451	550
		3	71	26	0.934	0.967	96	450	546
		4	60	29	0.896	0.950	81	451	532
		5	73	30	0.998	0.999	101	475	576
		Average	71	28	0.922	0.962	99	456	554
	0.25	1	80	19	0.999	1.000	110	457	567
		2	85	20	0.994	0.997	118	460	577
		3	61	25	0.994	0.997	86	455	540
		4	49	29	0.998	0.999	66	469	535
		5	57	21	0.998	0.999	78	450	528
		Average	66	23	0.997	0.998	92	458	550
	0.3	1	52	22	0.907	0.954	70	443	513
		2	64	19	0.920	0.960	91	441	532
		3	58	18	0.923	0.962	78	444	522
		4	35	30	0.923	0.962	50	478	527
		5	55	20	0.925	0.963	77	453	530
		Average	53	22	0.919	0.960	73	452	525
	0.35	1	44	24	0.999	1.000	61	458	519
		2	46	23	0.997	0.999	62	445	507
		3	41	23	0.860	0.934	55	437	493
		4	56	18	0.917	0.958	75	431	506
		5	54	17	0.999	0.999	72	460	532
		Average	48	21	0.954	0.978	65	446	511

Table B.22: F-ACOVNS Results for Iris Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	54	3	0.925	0.950	11	10	21
	2	42	3	0.925	0.950	5	15	20
	3	71	3	0.925	0.950	8	5	13
	4	62	3	0.925	0.950	6	5	11
	5	63	3	0.925	0.950	7	7	13
	6	69	3	0.925	0.950	5	12	17
	7	94	3	0.925	0.950	6	9	15
	8	54	3	0.925	0.950	7	10	16
	9	38	3	0.925	0.950	4	6	10
	10	44	3	0.925	0.950	5	9	13
	Average	59.1	3	0.925	0.950	6	9	15
0.08	1	40	3	0.925	0.950	4	9	13
	2	37	3	0.925	0.950	9	6	15
	3	51	3	0.925	0.950	7	5	12
	4	34	3	0.925	0.950	4	5	9
	5	58	3	0.925	0.950	6	8	14
	6	36	3	0.925	0.950	8	10	18
	7	56	3	0.925	0.950	7	8	15
	8	43	3	0.925	0.950	4	6	10
	9	28	3	0.925	0.950	2	8	10
	10	43	3	0.925	0.950	4	7	11
	Average	42.6	3	0.925	0.950	5	7	13
0.1	1	26	3	0.925	0.950	3	5	8
	2	33	3	0.925	0.950	3	6	9
	3	36	3	0.925	0.950	4	5	9
	4	30	3	0.925	0.950	3	5	8
	5	45	3	0.925	0.950	4	9	13
	6	34	3	0.925	0.950	4	6	10
	7	55	3	0.925	0.950	6	5	11
	8	41	3	0.925	0.950	4	5	9
	9	27	3	0.925	0.950	3	6	8
	10	40	3	0.925	0.950	4	6	10
	Average	36.7	3	0.925	0.950	4	6	10
0.15	1	21	3	0.925	0.950	2	5	7
	2	29	3	0.925	0.950	3	6	9
	3	22	3	0.925	0.950	2	5	8
	4	28	3	0.925	0.950	3	8	11
	5	25	3	0.925	0.950	2	7	9
	6	29	3	0.925	0.950	3	7	9
	7	46	3	0.925	0.950	5	6	10
	8	38	3	0.925	0.950	5	6	11
	9	26	3	0.925	0.950	3	8	11
	10	19	3	0.925	0.950	2	6	8
	Average	28.3	3	0.925	0.950	3	6	9
0.2	1	18	3	0.925	0.950	2	5	7
	2	24	3	0.925	0.950	2	5	8
	3	21	3	0.925	0.950	2	6	8
	4	27	3	0.925	0.950	3	6	8
	5	23	3	0.925	0.950	2	6	8
	6	28	3	0.925	0.950	3	7	10
	7	46	3	0.925	0.950	5	6	12
	8	36	3	0.925	0.950	5	9	14
	9	14	3	0.925	0.950	2	7	9
	10	13	3	0.925	0.950	2	8	10
	Average	25	3	0.925	0.950	3	6	9
0.25	1	18	3	0.925	0.950	2	5	7
	2	21	3	0.925	0.950	2	6	8
	3	18	3	0.925	0.950	2	7	9
	4	20	3	0.925	0.950	2	5	7
	5	21	3	0.925	0.950	2	5	7
	6	28	3	0.925	0.950	4	8	12
	7	37	3	0.925	0.950	5	9	14
	8	31	3	0.925	0.950	4	7	11

Table B.22: F-ACOVNS Results for Iris Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	13	3	0.925	0.950	2	7	9
	10	12	3	0.925	0.950	2	9	11
	Average	21.9	3	0.925	0.950	3	7	9
0.3	1	18	3	0.925	0.950	2	5	7
	2	21	3	0.925	0.950	2	5	7
	3	17	3	0.925	0.950	2	5	7
	4	20	3	0.925	0.950	2	5	7
	5	20	3	0.925	0.950	2	5	7
	6	20	3	0.925	0.950	3	8	10
	7	36	3	0.925	0.950	5	7	12
	8	10	3	0.925	0.950	1	8	10
	9	12	3	0.925	0.950	2	7	9
	10	7	3	0.925	0.950	1	6	7
	Average	18.1	3	0.925	0.950	2	6	8

Table B.23: F-ACOVNS Results for Wdbc Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	234	10	0.833	0.811	92	69	162
	2	316	13	0.836	0.814	121	80	201
	3	260	11	0.829	0.806	94	80	173
	4	222	16	0.834	0.811	81	88	169
	5	211	14	0.829	0.806	83	77	161
	6	204	13	0.834	0.811	83	71	154
	7	287	10	0.814	0.787	114	79	193
	8	195	16	0.834	0.811	73	76	150
	9	241	13	0.832	0.809	90	75	164
	10	180	21	0.870	0.854	67	121	188
	Average	235	13.7	0.834	0.812	90	82	171
0.08	1	165	11	0.816	0.790	68	84	152
	2	116	19	0.855	0.837	50	88	138
	3	188	8	0.814	0.787	74	49	123
	4	168	8	0.822	0.798	60	69	129
	5	136	14	0.836	0.814	48	77	126
	6	155	11	0.832	0.809	63	86	149
	7	209	9	0.814	0.787	84	65	149
	8	115	19	0.850	0.831	43	112	155
	9	187	14	0.832	0.809	69	82	151
	10	200	12	0.836	0.814	74	77	151
	Average	163.9	12.5	0.831	0.807	63	79	142
0.1	1	124	12	0.818	0.792	44	80	124
	2	92	17	0.855	0.837	33	83	116
	3	105	12	0.818	0.792	38	76	113
	4	129	9	0.827	0.803	46	78	124
	5	116	13	0.829	0.806	42	72	114
	6	134	8	0.825	0.800	49	72	121
	7	111	4	0.804	0.776	41	52	93
	8	158	10	0.827	0.803	58	55	113
	9	142	6	0.814	0.787	52	60	113
	10	107	6	0.814	0.787	40	69	108
	Average	121.8	9.7	0.823	0.798	44	70	114
0.15	1	83	6	0.814	0.787	30	64	94
	2	93	13	0.829	0.806	33	77	110
	3	70	12	0.831	0.809	25	73	98
	4	89	8	0.822	0.798	32	68	100
	5	92	11	0.827	0.803	33	54	87
	6	77	11	0.818	0.792	28	97	126
	7	76	10	0.816	0.790	28	74	102
	8	94	12	0.818	0.792	35	80	115
	9	76	10	0.820	0.795	28	92	120
	10	105	11	0.829	0.806	39	80	119
	Average	85.5	10.4	0.823	0.798	31	76	107
0.2	1	49	9	0.829	0.806	18	70	87
	2	65	9	0.816	0.790	23	77	100
	3	58	7	0.814	0.787	21	67	88
	4	86	8	0.827	0.803	31	59	90
	5	63	10	0.833	0.811	22	61	84
	6	67	8	0.818	0.792	24	57	81
	7	66	6	0.816	0.790	24	62	87
	8	51	15	0.831	0.809	19	96	115
	9	51	12	0.818	0.792	21	93	114
	10	62	13	0.831	0.809	28	79	106
	Average	61.8	9.7	0.823	0.799	23	72	95
0.25	1	39	6	0.814	0.787	14	74	88
	2	39	6	0.814	0.787	14	66	80
	3	54	7	0.814	0.787	19	70	90
	4	49	11	0.831	0.809	18	71	89
	5	37	19	0.838	0.817	13	81	95
	6	53	5	0.812	0.787	22	71	94
	7	42	8	0.822	0.798	16	63	79
	8	48	10	0.818	0.792	17	75	92

Table B.23: F-ACOVNS Results for Wdbc Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	33	11	0.825	0.800	13	89	101
	10	42	12	0.831	0.809	16	89	104
	Average	43.6	9.5	0.822	0.797	16	75	91
0.3	1	45	8	0.820	0.795	16	68	85
	2	31	6	0.814	0.787	12	65	76
	3	34	9	0.829	0.806	12	60	72
	4	35	11	0.820	0.795	13	75	87
	5	38	8	0.822	0.798	14	60	73
	6	46	9	0.836	0.814	19	73	91
	7	38	8	0.816	0.790	14	60	75
	8	45	8	0.822	0.798	17	69	86
	9	33	11	0.820	0.795	12	101	113
	10	34	10	0.834	0.811	13	75	88
	Average	37.9	8.8	0.823	0.799	14	71	85

Table B.24: F-ACOVNS Results for Breast Cancer Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	199	5	0.914	0.905	124	125	249
	2	190	8	0.926	0.919	113	84	197
	3	224	4	0.912	0.903	132	65	196
	4	157	8	0.926	0.919	92	109	201
	5	259	7	0.914	0.905	152	89	242
	6	234	4	0.912	0.903	113	102	215
	7	173	8	0.926	0.919	109	84	193
	8	203	8	0.926	0.919	120	125	245
	9	153	4	0.914	0.905	85	144	229
	10	212	5	0.912	0.903	102	148	249
	Average	200.4	6.1	0.918	0.910	114	107	222
0.08	1	138	4	0.909	0.900	81	136	217
	2	106	4	0.912	0.903	72	123	196
	3	105	8	0.926	0.919	63	82	144
	4	145	4	0.912	0.903	85	131	216
	5	103	4	0.909	0.900	60	129	189
	6	159	7	0.923	0.916	100	103	203
	7	117	4	0.909	0.900	79	100	178
	8	128	5	0.912	0.903	70	89	158
	9	123	5	0.914	0.905	76	111	187
	10	149	5	0.914	0.905	88	99	187
	Average	127.3	5	0.914	0.905	77	110	188
0.1	1	82	4	0.909	0.900	48	110	159
	2	105	4	0.909	0.900	62	126	188
	3	76	8	0.926	0.919	45	90	135
	4	108	4	0.916	0.908	63	102	166
	5	86	4	0.911	0.903	51	63	114
	6	147	4	0.909	0.900	88	123	211
	7	96	4	0.911	0.903	49	61	111
	8	118	5	0.912	0.903	51	69	120
	9	88	4	0.911	0.903	40	75	115
	10	100	8	0.926	0.919	43	90	133
	Average	100.6	4.9	0.914	0.906	54	91	145
0.15	1	67	6	0.936	0.929	40	117	156
	2	65	4	0.912	0.903	38	95	133
	3	52	8	0.926	0.919	31	86	118
	4	78	3	0.871	0.857	46	96	142
	5	53	7	0.916	0.908	32	127	159
	6	51	4	0.912	0.903	23	97	120
	7	63	4	0.909	0.900	28	98	126
	8	70	4	0.912	0.903	32	114	145
	9	48	4	0.914	0.905	22	106	127
	10	58	4	0.916	0.908	26	108	134
	Average	60.5	4.8	0.912	0.903	32	104	136
0.2	1	37	4	0.912	0.903	23	111	134
	2	63	4	0.909	0.900	37	132	170
	3	43	6	0.921	0.913	25	90	115
	4	36	4	0.914	0.905	22	126	148
	5	48	4	0.914	0.905	28	77	105
	6	47	4	0.914	0.905	20	85	105
	7	50	7	0.917	0.908	21	108	129
	8	50	5	0.912	0.903	22	83	105
	9	44	4	0.916	0.908	19	86	105
	10	42	4	0.909	0.900	18	99	117
	Average	46	4.6	0.914	0.905	24	100	123
0.25	1	30	4	0.914	0.905	18	111	129
	2	42	5	0.914	0.905	25	104	129
	3	38	5	0.912	0.903	22	106	129
	4	32	4	0.916	0.908	19	116	135
	5	33	8	0.926	0.919	20	114	133
	6	45	8	0.938	0.932	20	48	67
	7	33	8	0.926	0.919	14	72	86
	8	36	4	0.914	0.905	15	65	81

Table B.24: F-ACOVNS Results for Breast Cancer Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	37	5	0.905	0.895	16	108	124
	10	36	4	0.914	0.905	16	66	81
	Average	36.2	5.5	0.918	0.910	19	91	109
0.3	1	30	4	0.912	0.903	18	130	148
	2	27	8	0.938	0.932	16	134	150
	3	35	5	0.914	0.905	21	122	143
	4	28	4	0.912	0.903	17	125	141
	5	26	6	0.907	0.898	15	123	139
	6	30	8	0.938	0.932	13	73	86
	7	27	8	0.926	0.919	12	78	90
	8	24	4	0.914	0.905	10	103	113
	9	35	5	0.909	0.900	15	109	124
	10	32	4	0.912	0.903	21	146	167
	Average	29.4	5.6	0.918	0.910	16	114	130

Table B.25: F-ACOVNS Results for Wine Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	123	7	0.913	0.941	22	23	45
	2	100	7	0.633	0.698	19	26	44
	3	166	5	0.757	0.835	24	21	46
	4	165	7	0.882	0.920	24	18	42
	5	163	7	0.855	0.902	20	17	38
	6	130	7	0.855	0.902	24	32	56
	7	168	9	0.866	0.909	45	26	71
	8	122	7	0.893	0.927	22	35	57
	9	133	7	0.893	0.927	25	35	60
	10	137	7	0.893	0.927	25	25	50
	Average	140.7	7	0.844	0.889	25	26	51
0.08	1	110	11	0.924	0.948	19	26	45
	2	81	7	0.883	0.921	17	29	46
	3	108	7	0.903	0.934	17	20	37
	4	102	11	0.895	0.928	14	20	34
	5	83	7	0.773	0.845	11	18	30
	6	73	7	0.893	0.927	26	41	66
	7	86	9	0.866	0.909	25	26	52
	8	76	7	0.855	0.902	16	29	45
	9	91	10	0.914	0.942	16	26	42
	10	87	7	0.893	0.927	17	19	36
	Average	89.7	8.3	0.880	0.918	18	25	43
0.1	1	67	7	0.893	0.927	9	18	27
	2	70	7	0.893	0.927	9	18	27
	3	68	6	0.834	0.888	9	19	28
	4	68	7	0.855	0.902	9	19	28
	5	56	6	0.834	0.888	8	19	27
	6	63	7	0.893	0.927	11	27	39
	7	77	6	0.834	0.888	15	20	34
	8	64	7	0.893	0.927	12	31	43
	9	72	7	0.893	0.927	13	35	48
	10	60	7	0.893	0.927	11	21	32
	Average	66.5	6.7	0.871	0.913	11	23	33
0.15	1	49	7	0.893	0.927	6	19	25
	2	46	8	0.934	0.955	6	19	25
	3	33	12	0.923	0.948	4	19	23
	4	48	7	0.893	0.927	7	19	26
	5	40	6	0.834	0.888	5	19	24
	6	43	11	0.884	0.921	6	30	36
	7	49	9	0.866	0.909	9	20	29
	8	53	7	0.843	0.894	10	34	44
	9	59	7	0.893	0.927	11	32	43
	10	62	7	0.893	0.927	11	25	36
	Average	48.2	8.1	0.885	0.922	8	24	31
0.2	1	31	7	0.846	0.896	4	19	23
	2	33	8	0.934	0.955	5	19	24
	3	44	6	0.874	0.915	6	19	25
	4	28	9	0.800	0.864	4	18	22
	5	36	6	0.824	0.881	5	18	23
	6	37	10	0.850	0.898	7	25	32
	7	45	7	0.893	0.927	9	18	26
	8	36	7	0.873	0.914	6	29	35
	9	38	6	0.834	0.888	8	23	30
	10	39	8	0.895	0.929	7	22	29
	Average	36.7	7.4	0.862	0.907	6	21	27
0.25	1	28	7	0.883	0.921	4	18	22
	2	32	8	0.934	0.955	4	18	22
	3	34	5	0.757	0.835	4	17	22
	4	28	9	0.800	0.864	3	17	21
	5	27	7	0.855	0.902	3	17	21
	6	31	8	0.796	0.862	6	24	30
	7	26	10	0.884	0.922	5	21	26
	8	24	5	0.757	0.835	4	21	25

Table B.25: F-ACOVNS Results for Wine Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	33	5	0.757	0.835	6	18	24
	10	34	8	0.895	0.929	6	27	33
	Average	29.7	7.2	0.832	0.886	5	20	25
0.3	1	20	7	0.893	0.927	2	18	21
	2	18	8	0.778	0.850	2	18	20
	3	31	7	0.837	0.890	4	17	21
	4	20	10	0.886	0.922	2	18	20
	5	23	5	0.757	0.835	3	17	20
	6	25	6	0.834	0.888	4	20	24
	7	27	6	0.834	0.888	5	22	27
	8	23	5	0.757	0.835	4	17	21
	9	23	7	0.893	0.927	3	20	23
	10	27	7	0.873	0.914	3	25	28
	Average	23.7	6.8	0.834	0.888	3	19	23

Table B.26: F-ACOVNS Results for Glass Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	99	8	0.457	0.735	26	49	75
	2	115	8	0.450	0.719	25	52	77
	3	82	8	0.419	0.711	14	47	61
	4	105	8	0.419	0.711	17	31	48
	5	83	8	0.416	0.706	14	42	56
	6	98	8	0.420	0.702	26	61	87
	7	124	8	0.417	0.707	39	84	122
	8	110	8	0.433	0.703	27	72	98
	9	83	8	0.429	0.734	19	56	76
	10	99	8	0.416	0.706	24	72	96
	Average	99.8	8	0.428	0.713	23	56	80
0.08	1	109	8	0.424	0.706	27	48	75
	2	112	8	0.423	0.703	26	48	74
	3	53	8	0.415	0.716	10	47	57
	4	67	8	0.411	0.712	11	48	59
	5	54	8	0.406	0.712	10	48	58
	6	84	8	0.419	0.711	32	76	108
	7	83	8	0.423	0.703	34	84	118
	8	53	8	0.429	0.705	13	81	94
	9	70	8	0.423	0.703	17	40	57
	10	54	8	0.424	0.707	13	62	74
	Average	73.9	8	0.420	0.708	19	58	77
0.1	1	39	8	0.411	0.712	7	40	47
	2	78	8	0.428	0.706	13	45	58
	3	48	8	0.432	0.733	9	47	56
	4	62	8	0.414	0.708	10	46	56
	5	42	8	0.429	0.734	7	44	50
	6	48	8	0.418	0.729	11	57	68
	7	112	8	0.432	0.702	25	70	96
	8	41	8	0.425	0.733	9	74	83
	9	71	8	0.468	0.681	16	56	72
	10	57	8	0.411	0.712	13	52	65
	Average	59.8	8	0.427	0.715	12	53	65
0.15	1	27	8	0.419	0.711	4	33	37
	2	80	7	0.432	0.704	13	45	58
	3	42	7	0.457	0.735	7	42	49
	4	62	8	0.411	0.712	11	33	44
	5	28	8	0.421	0.740	5	44	49
	6	45	8	0.412	0.713	11	65	76
	7	44	8	0.419	0.703	10	69	79
	8	30	8	0.417	0.708	6	43	49
	9	65	7	0.423	0.717	12	43	55
	10	55	8	0.420	0.705	10	56	66
	Average	47.8	7.7	0.423	0.715	9	47	56
0.2	1	26	8	0.424	0.708	4	38	42
	2	75	7	0.425	0.706	12	45	57
	3	20	8	0.418	0.704	3	34	37
	4	47	7	0.420	0.717	8	34	42
	5	18	8	0.428	0.706	3	46	49
	6	27	8	0.419	0.711	5	31	35
	7	28	8	0.423	0.703	5	50	55
	8	26	8	0.424	0.708	5	57	61
	9	37	8	0.420	0.702	6	38	44
	10	54	8	0.506	0.682	10	43	53
	Average	35.8	7.8	0.431	0.705	6	42	48
0.25	1	21	8	0.417	0.708	4	43	47
	2	24	8	0.416	0.704	4	39	44
	3	16	8	0.430	0.735	3	37	40
	4	25	6	0.447	0.709	4	45	49
	5	16	8	0.427	0.733	3	46	49
	6	23	8	0.411	0.712	4	44	48
	7	23	8	0.419	0.711	5	46	51
	8	18	8	0.411	0.712	3	35	38

Table B.26: F-ACOVNS Results for Glass Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	33	6	0.460	0.685	6	25	31
	10	21	8	0.504	0.676	4	31	34
	Average	22	7.6	0.434	0.708	4	39	43
0.3	1	21	8	0.413	0.706	4	40	44
	2	19	7	0.418	0.701	3	43	46
	3	17	8	0.411	0.712	3	34	37
	4	23	6	0.442	0.703	4	46	50
	5	15	8	0.398	0.707	3	46	49
	6	20	8	0.426	0.716	4	62	65
	7	27	6	0.428	0.712	5	44	49
	8	16	8	0.444	0.729	3	53	56
	9	32	6	0.448	0.687	7	29	36
	10	19	8	0.420	0.705	4	49	52
	Average	20.9	7.3	0.425	0.708	4	45	48

Table B.27: F-ACOVNS Results for CMC Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	1	104	7	0.365	0.562	152	408	560
	2	119	7	0.366	0.561	173	402	575
	3	108	7	0.364	0.561	150	394	544
	4	105	8	0.369	0.562	145	395	540
	5	116	8	0.363	0.558	160	396	556
0.05	6	108	8	0.365	0.558	148	458	605
	7	116	7	0.369	0.559	160	442	602
	8	109	7	0.368	0.564	145	401	545
	9	113	7	0.368	0.566	149	412	561
	10	136	7	0.366	0.563	179	444	623
	Average	113.4	7.3	0.366	0.561	156	415	571
	1	68	7	0.365	0.560	100	399	498
	2	75	7	0.366	0.561	111	400	511
	3	84	7	0.366	0.559	116	390	506
	4	72	8	0.395	0.551	99	392	491
	5	73	7	0.365	0.562	100	388	489
0.08	6	70	7	0.365	0.559	98	455	553
	7	74	7	0.365	0.562	105	446	551
	8	70	8	0.366	0.560	93	442	535
	9	71	8	0.368	0.564	93	430	523
	10	78	7	0.368	0.558	103	445	548
	Average	73.5	7.3	0.369	0.560	102	419	521
	1	65	7	0.364	0.556	89	394	483
	2	70	8	0.393	0.551	96	393	490
	3	103	2	0.420	0.500	140	194	334
	4	47	8	0.367	0.564	64	390	454
	5	55	7	0.371	0.567	76	390	466
0.1	6	73	7	0.367	0.564	96	371	467
	7	66	7	0.366	0.561	87	443	530
	8	53	8	0.366	0.561	70	443	513
	9	58	8	0.366	0.562	77	449	526
	10	65	7	0.367	0.557	86	441	527
	Average	65.5	6.9	0.375	0.554	88	391	479
	1	37	7	0.365	0.562	51	392	443
	2	47	7	0.368	0.561	65	390	454
	3	46	8	0.367	0.561	63	393	456
	4	32	8	0.396	0.553	44	395	439
	5	34	7	0.365	0.561	47	393	440
0.15	6	35	8	0.368	0.562	46	437	483
	7	41	7	0.368	0.554	54	435	489
	8	44	8	0.366	0.560	58	431	489
	9	41	7	0.371	0.544	54	443	497
	10	42	8	0.365	0.559	55	427	482
	Average	39.9	7.5	0.370	0.558	54	413	467
	1	28	7	0.367	0.564	38	392	431
	2	27	7	0.366	0.561	37	393	430
	3	35	8	0.365	0.560	48	393	441
	4	28	8	0.367	0.548	38	392	430
	5	34	7	0.368	0.565	47	393	440
0.2	6	33	8	0.368	0.561	43	438	481
	7	34	7	0.369	0.561	45	416	461
	8	34	7	0.367	0.562	45	427	471
	9	31	7	0.368	0.564	41	439	479
	10	27	8	0.396	0.553	35	426	462
	Average	31.1	7.4	0.370	0.560	42	411	453
	1	42	2	0.410	0.517	58	235	292
	2	23	7	0.364	0.560	32	393	425
	3	32	6	0.369	0.565	44	389	433
	4	24	8	0.367	0.560	33	393	426
	5	27	7	0.370	0.567	37	392	429
0.25	6	27	7	0.368	0.565	36	442	477
	7	34	7	0.370	0.554	46	415	461
	8	27	8	0.367	0.564	36	444	480

Table B.27: F-ACOVNS Results for CMC Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	24	7	0.370	0.551	32	468	499
	10	23	8	0.369	0.565	43	444	487
	Average	28.3	6.7	0.372	0.557	40	401	441
0.3	1	23	6	0.365	0.560	31	390	422
	2	22	7	0.366	0.559	30	391	421
	3	29	6	0.368	0.564	40	393	432
	4	24	8	0.365	0.561	33	392	424
	5	19	8	0.371	0.566	26	393	419
	6	20	7	0.366	0.562	32	457	489
	7	19	8	0.365	0.559	26	438	464
	8	18	8	0.369	0.563	30	451	481
	9	21	7	0.366	0.563	30	438	468
	10	19	8	0.369	0.565	28	459	487
	Average	21.4	7.3	0.367	0.562	31	420	451

Table B.28: F-ACOVNS Results for Liver Disease Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	65	5	0.530	0.500	22	45	67
	2	58	5	0.530	0.500	18	50	68
	3	82	5	0.529	0.500	20	46	66
	4	81	5	0.530	0.500	18	49	66
	5	102	4	0.586	0.501	22	37	59
	6	69	5	0.530	0.500	22	56	78
	7	62	4	0.586	0.501	20	45	65
	8	67	4	0.586	0.501	17	40	57
	9	54	5	0.530	0.500	12	47	59
	10	56	5	0.530	0.500	12	49	61
	Average	69.6	4.7	0.547	0.500	18	46	65
0.08	1	46	4	0.586	0.501	15	40	55
	2	42	5	0.529	0.500	15	43	58
	3	64	4	0.586	0.501	16	40	56
	4	40	5	0.529	0.500	9	46	56
	5	56	5	0.530	0.500	12	38	50
	6	54	4	0.586	0.501	19	63	83
	7	47	4	0.586	0.501	22	57	79
	8	47	4	0.586	0.501	17	57	74
	9	45	5	0.530	0.500	12	75	87
	10	46	5	0.530	0.500	13	58	71
	Average	48.7	4.5	0.558	0.500	15	52	67
0.1	1	41	4	0.586	0.501	9	39	47
	2	21	5	0.530	0.500	5	47	52
	3	41	5	0.530	0.500	9	42	51
	4	36	5	0.529	0.500	8	47	54
	5	54	5	0.530	0.500	11	38	49
	6	42	4	0.586	0.501	13	57	70
	7	39	4	0.586	0.501	12	57	68
	8	43	4	0.586	0.501	12	58	71
	9	30	5	0.530	0.500	9	57	66
	10	46	5	0.530	0.500	13	56	69
	Average	39.3	4.6	0.553	0.500	10	50	60
0.15	1	27	4	0.586	0.501	6	36	42
	2	21	5	0.530	0.500	4	42	47
	3	28	5	0.530	0.500	6	45	51
	4	33	5	0.530	0.500	7	46	53
	5	41	5	0.530	0.500	9	35	43
	6	27	4	0.586	0.501	8	53	61
	7	31	4	0.586	0.501	8	51	59
	8	31	4	0.586	0.501	9	56	65
	9	23	5	0.530	0.500	7	47	54
	10	24	5	0.530	0.500	7	68	75
	Average	28.6	4.6	0.553	0.500	7	48	55
0.2	1	22	4	0.586	0.501	5	35	40
	2	19	5	0.530	0.500	4	41	45
	3	28	5	0.530	0.500	6	45	51
	4	26	4	0.586	0.501	5	37	43
	5	23	5	0.530	0.500	5	38	43
	6	23	4	0.586	0.501	7	61	68
	7	24	4	0.586	0.501	7	50	57
	8	24	4	0.586	0.501	6	50	56
	9	15	5	0.530	0.500	4	46	50
	10	24	5	0.530	0.500	7	76	83
	Average	22.8	4.5	0.558	0.500	6	48	54
0.25	1	19	4	0.586	0.501	4	38	42
	2	14	5	0.517	0.501	3	46	49
	3	16	5	0.530	0.500	3	43	46
	4	22	4	0.586	0.501	5	35	40
	5	17	5	0.530	0.500	4	39	43
	6	19	4	0.586	0.501	6	56	62
	7	18	5	0.530	0.500	6	52	58
	8	18	5	0.530	0.500	5	54	60

Table B.28: F-ACOVNS Results for Liver Disease Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	24	4	0.586	0.501	7	61	68
	10	18	4	0.586	0.501	5	64	69
	Average	18.5	4.5	0.557	0.500	5	49	54
0.3	1	17	4	0.586	0.501	4	42	45
	2	13	5	0.538	0.500	3	46	48
	3	16	5	0.530	0.500	3	42	45
	4	20	4	0.586	0.501	4	41	45
	5	16	5	0.530	0.500	3	40	43
	6	16	4	0.586	0.501	5	61	66
	7	18	5	0.530	0.500	5	49	54
	8	14	4	0.586	0.501	3	48	51
	9	19	4	0.586	0.501	5	57	63
	10	17	4	0.586	0.501	5	60	65
	Average	16.6	4.4	0.565	0.500	4	49	53

Table B.29: F-ACOVNS Results for Thyroid Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	81	4	0.898	0.884	19	31	50
	2	165	4	0.898	0.884	31	24	55
	3	127	4	0.898	0.884	19	22	41
	4	118	4	0.708	0.704	17	51	69
	5	92	4	0.723	0.719	15	52	66
	6	102	4	0.898	0.884	27	39	66
	7	84	4	0.898	0.884	20	40	60
	8	94	4	0.898	0.884	17	41	58
	9	94	4	0.898	0.884	16	28	45
	10	110	4	0.898	0.884	17	20	37
	Average	106.7	4	0.862	0.850	20	35	55
0.08	1	65	4	0.898	0.884	15	24	39
	2	77	4	0.898	0.884	18	27	45
	3	65	4	0.898	0.884	10	26	36
	4	79	4	0.898	0.884	12	40	53
	5	83	4	0.580	0.601	12	22	34
	6	86	4	0.898	0.884	20	31	51
	7	66	4	0.898	0.884	18	35	53
	8	56	4	0.871	0.848	11	65	77
	9	68	4	0.898	0.884	10	23	33
	10	47	4	0.896	0.880	8	64	72
	Average	69.2	4	0.864	0.852	14	36	49
0.1	1	45	4	0.591	0.610	7	35	41
	2	69	4	0.898	0.884	11	30	41
	3	57	4	0.898	0.884	9	22	31
	4	74	4	0.899	0.884	11	40	51
	5	54	4	0.589	0.609	8	25	33
	6	44	4	0.905	0.891	7	34	41
	7	64	4	0.898	0.884	10	29	38
	8	51	4	0.896	0.880	8	66	74
	9	67	4	0.898	0.884	10	23	33
	10	41	4	0.781	0.769	6	65	71
	Average	56.6	4	0.825	0.818	9	37	45
0.15	1	25	4	0.898	0.884	4	28	31
	2	65	4	0.905	0.891	10	33	43
	3	41	4	0.830	0.816	6	51	57
	4	33	4	0.850	0.835	5	50	55
	5	50	4	0.591	0.610	7	23	30
	6	37	4	0.898	0.884	6	23	29
	7	55	4	0.898	0.884	8	27	35
	8	29	4	0.871	0.848	5	54	59
	9	37	4	0.898	0.884	6	22	28
	10	34	4	0.898	0.884	5	21	26
	Average	40.6	4	0.854	0.842	6	33	39
0.2	1	24	4	0.898	0.884	3	23	26
	2	28	4	0.898	0.884	4	23	28
	3	37	4	0.781	0.769	5	51	57
	4	28	4	0.898	0.884	4	27	31
	5	46	4	0.762	0.754	7	50	57
	6	30	4	0.898	0.884	5	25	29
	7	45	4	0.898	0.884	7	28	34
	8	23	4	0.898	0.884	3	31	35
	9	36	4	0.898	0.884	6	21	27
	10	30	4	0.905	0.891	5	41	46
	Average	32.7	4	0.874	0.861	5	32	37
0.25	1	23	4	0.898	0.884	4	27	31
	2	21	4	0.898	0.884	3	23	26
	3	15	4	0.898	0.884	2	23	25
	4	26	4	0.898	0.884	4	22	26
	5	24	4	0.898	0.884	3	25	29
	6	25	4	0.898	0.884	4	26	30
	7	36	4	0.898	0.884	6	23	28
	8	21	4	0.899	0.884	3	38	41

Table B.29: F-ACOVNS Results for Thyroid Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	31	4	0.898	0.884	5	22	27
	10	25	4	0.871	0.848	4	53	57
	Average	24.7	4	0.896	0.881	4	28	32
0.3	1	22	4	0.898	0.884	3	25	28
	2	21	4	0.898	0.884	3	23	26
	3	14	4	0.898	0.884	2	23	25
	4	25	4	0.899	0.884	4	40	44
	5	16	4	0.898	0.884	2	22	24
	6	23	4	0.898	0.884	3	25	28
	7	21	4	0.898	0.884	4	24	27
	8	14	4	0.899	0.884	2	41	43
	9	25	4	0.898	0.884	4	24	28
	10	24	4	0.882	0.862	4	45	49
	Average	20.5	4	0.897	0.882	3	29	32

Table B.30: F-ACOVNS Results for Zoo Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	200	5	0.797	0.913	25	9	34
	2	140	4	0.803	0.915	18	5	22
	3	189	5	0.797	0.913	17	6	23
	4	146	13	0.952	0.977	13	8	21
	5	191	5	0.797	0.913	17	7	24
	6	224	5	0.801	0.914	26	13	39
	7	175	11	0.771	0.903	21	14	36
	8	134	11	0.854	0.935	13	14	27
	9	111	14	0.758	0.899	11	14	24
	10	188	5	0.921	0.960	17	5	22
	Average	169.8	7.8	0.825	0.924	18	9	27
0.08	1	96	14	0.981	0.991	12	14	26
	2	114	4	0.799	0.912	14	4	18
	3	60	14	0.939	0.970	5	9	14
	4	108	13	0.900	0.953	9	8	17
	5	92	4	0.803	0.915	8	4	12
	6	133	5	0.921	0.960	19	8	26
	7	120	5	0.921	0.960	17	14	31
	8	120	4	0.799	0.912	16	5	21
	9	102	4	0.799	0.912	13	7	20
	10	139	5	0.921	0.960	17	7	24
	Average	108.4	7.2	0.878	0.945	13	8	21
0.1	1	117	5	0.780	0.904	11	5	16
	2	75	4	0.799	0.912	7	8	14
	3	59	14	0.779	0.906	5	10	15
	4	107	5	0.780	0.904	9	5	14
	5	84	4	0.781	0.904	7	5	13
	6	73	4	0.799	0.912	9	6	16
	7	88	4	0.799	0.912	11	8	19
	8	111	4	0.793	0.909	14	4	18
	9	85	4	0.793	0.909	11	6	17
	10	119	8	0.741	0.891	14	15	28
	Average	91.8	5.6	0.784	0.906	10	7	17
0.15	1	54	11	0.920	0.963	5	11	16
	2	52	4	0.799	0.912	5	4	8
	3	74	5	0.921	0.960	6	5	11
	4	62	5	0.797	0.913	6	10	16
	5	70	4	0.803	0.915	6	4	10
	6	66	4	0.855	0.934	8	8	16
	7	41	14	0.885	0.949	6	16	22
	8	46	16	0.854	0.934	6	25	31
	9	48	4	0.799	0.912	5	7	12
	10	66	5	0.842	0.923	6	5	12
	Average	57.9	7.2	0.847	0.931	6	10	15
0.2	1	46	11	0.892	0.949	4	12	16
	2	52	4	0.804	0.915	4	4	8
	3	48	5	0.797	0.913	4	9	14
	4	33	14	0.935	0.969	3	10	13
	5	32	13	0.940	0.972	3	12	15
	6	46	4	0.803	0.915	4	4	9
	7	33	14	0.941	0.972	3	14	17
	8	33	13	0.778	0.906	4	9	13
	9	39	4	0.793	0.909	4	4	8
	10	43	4	0.803	0.915	4	4	8
	Average	40.5	8.6	0.849	0.933	4	8	12
0.25	1	44	4	0.804	0.915	4	5	9
	2	32	5	0.780	0.902	3	4	7
	3	35	6	0.945	0.974	3	5	8
	4	32	4	0.803	0.915	3	9	12
	5	28	13	0.979	0.990	2	8	11
	6	36	4	0.803	0.915	4	5	8
	7	35	4	0.803	0.915	3	5	8
	8	35	11	0.909	0.958	3	7	11

Table B.30: F-ACOVNS Results for Zoo Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	9	35	4	0.803	0.915	3	5	8
	10	40	5	0.921	0.960	4	11	15
	Average	35.2	6	0.855	0.936	3	6	10
0.3	1	39	5	0.921	0.960	3	6	9
	2	20	12	0.850	0.932	2	11	13
	3	31	6	0.927	0.964	3	5	7
	4	32	5	0.770	0.900	3	7	10
	5	45	7	0.702	0.872	4	5	10
	6	32	4	0.803	0.915	3	5	8
	7	37	4	0.799	0.912	3	4	7
	8	30	9	0.947	0.975	3	8	11
	9	24	4	0.803	0.915	2	11	14
	10	27	4	0.803	0.915	3	6	8
	Average	31.7	6	0.833	0.926	3	7	10

Table B.31: F-ACOVNS Results for 5d5c1_1 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	55	5	0.989	0.995	83	336	419
	2	62	5	0.952	0.977	93	335	428
	3	50	5	0.864	0.935	69	327	396
	4	56	5	0.994	0.997	76	316	392
	5	59	5	0.975	0.988	81	325	406
	6	53	5	0.997	0.999	79	362	441
	7	57	5	0.997	0.999	86	344	430
	8	55	5	0.985	0.993	75	367	442
	9	60	5	0.997	0.999	82	250	332
	10	58	5	0.856	0.932	79	373	453
	Average	56.5	5	0.961	0.981	81	333	414
0.08	1	38	5	0.989	0.994	57	323	380
	2	47	5	0.952	0.977	71	333	404
	3	39	5	0.991	0.996	56	321	378
	4	43	5	0.990	0.995	59	318	377
	5	41	5	0.994	0.997	56	314	370
	6	42	5	0.997	0.999	63	281	344
	7	47	5	0.997	0.999	71	279	349
	8	40	5	0.997	0.999	57	295	352
	9	43	5	0.999	0.999	59	367	426
	10	37	5	0.970	0.985	50	367	417
	Average	41.7	5	0.988	0.994	60	320	380
0.1	1	34	5	0.995	0.998	47	309	355
	2	42	5	0.831	0.920	58	327	386
	3	32	5	0.988	0.994	44	307	351
	4	39	5	0.996	0.998	54	325	379
	5	37	5	0.993	0.997	51	296	347
	6	35	5	0.999	0.999	48	370	418
	7	39	5	0.997	0.999	54	314	367
	8	32	5	0.998	0.999	44	348	392
	9	33	5	0.997	0.999	45	297	342
	10	34	5	0.997	0.999	46	357	403
	Average	35.7	5	0.979	0.990	49	325	374
0.15	1	25	5	0.886	0.946	34	327	362
	2	29	5	0.838	0.922	39	308	347
	3	23	5	0.834	0.921	31	312	343
	4	24	5	0.954	0.978	33	329	362
	5	26	5	0.991	0.996	35	315	350
	6	26	5	0.996	0.998	35	262	297
	7	25	5	0.991	0.995	34	357	391
	8	28	5	0.999	0.999	38	350	388
	9	28	5	0.867	0.937	38	364	401
	10	25	5	0.997	0.999	34	283	317
	Average	25.9	5	0.935	0.969	35	321	356

Table B.32: F-ACOVNS Results for 5d5c1_5 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	49	5	0.989	0.994	55	220	275
	2	52	5	0.899	0.952	58	218	277
	3	55	5	0.990	0.995	57	216	273
	4	51	5	0.993	0.996	52	221	273
	5	51	5	0.990	0.995	56	222	278
	6	51	5	0.995	0.997	58	244	302
	7	49	5	0.993	0.996	54	239	293
	8	47	5	0.993	0.996	45	294	339
	9	44	5	0.997	0.999	41	297	338
	10	50	5	0.835	0.921	50	297	347
	Average	49.9	5	0.967	0.984	53	247	299
0.08	1	39	5	0.958	0.980	43	223	266
	2	39	5	0.894	0.950	42	222	264
	3	36	5	0.855	0.932	37	203	240
	4	42	5	0.988	0.994	41	205	246
	5	37	5	0.847	0.928	37	202	240
	6	39	5	0.986	0.993	49	285	334
	7	36	5	0.996	0.998	46	282	328
	8	36	5	0.997	0.999	40	265	305
	9	35	5	0.992	0.996	39	277	316
	10	40	5	0.994	0.997	46	200	246
	Average	37.9	5	0.951	0.977	42	236	278
0.1	1	32	5	0.982	0.991	31	209	240
	2	35	5	0.987	0.994	35	212	247
	3	31	5	0.890	0.948	31	217	248
	4	34	5	0.875	0.941	36	209	245
	5	32	5	0.980	0.990	31	203	234
	6	35	5	0.999	0.999	38	285	323
	7	32	5	0.997	0.999	34	277	311
	8	33	5	0.993	0.997	35	266	301
	9	31	5	0.999	0.999	35	289	324
	10	33	5	0.778	0.890	31	279	310
	Average	32.8	5	0.948	0.975	34	245	278
0.15	1	23	5	0.976	0.988	22	202	225
	2	26	5	0.964	0.983	25	207	232
	3	24	5	0.991	0.996	27	206	233
	4	26	5	0.987	0.994	25	205	230
	5	24	5	0.778	0.891	23	202	226
	6	23	5	0.998	0.999	22	266	288
	7	24	5	0.961	0.981	23	292	315
	8	25	5	0.932	0.968	24	294	318
	9	21	5	0.995	0.998	29	260	289
	10	29	5	0.973	0.987	30	285	315
	Average	24.5	5	0.955	0.978	25	242	267

Table B.33: F-ACOVNS Results for 10d5c1_2 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	76	10	0.999	0.999	124	402	526
	2	67	10	0.996	0.998	110	448	558
	3	65	10	0.998	0.999	101	435	536
	4	72	10	0.997	0.999	111	412	523
	5	64	10	1.000	1.000	98	404	502
	6	76	10	0.999	0.999	125	424	549
	7	61	10	0.999	1.000	103	472	575
	8	67	10	1.000	1.000	106	546	652
	9	69	10	0.810	0.916	109	565	674
	10	72	10	0.999	0.999	114	435	549
	Average	68.9	10	0.980	0.991	110	454	564
0.08	1	49	10	0.999	0.999	82	425	507
	2	54	10	0.995	0.998	90	361	451
	3	46	10	0.999	1.000	71	381	452
	4	55	10	0.996	0.998	84	425	510
	5	51	10	0.998	0.999	79	445	524
	6	50	10	1.000	1.000	84	556	640
	7	42	10	0.999	1.000	71	485	557
	8	48	10	0.999	0.999	77	466	542
	9	42	10	1.000	1.000	66	412	478
	10	46	10	0.781	0.904	73	528	601
	Average	48.3	10	0.977	0.990	78	449	526
0.1	1	45	10	0.999	0.999	69	421	490
	2	39	10	0.995	0.998	61	390	451
	3	37	10	0.998	0.999	57	402	459
	4	42	10	0.996	0.998	64	371	436
	5	46	10	0.999	1.000	71	449	519
	6	43	10	0.999	1.000	68	350	418
	7	35	10	1.000	1.000	55	481	535
	8	40	10	0.999	1.000	63	348	411
	9	39	10	0.999	1.000	61	553	614
	10	42	10	0.999	0.999	66	533	598
	Average	40.8	10	0.998	0.999	63	430	493
0.15	1	30	10	0.997	0.999	46	392	437
	2	27	10	0.998	0.999	41	363	404
	3	27	10	0.998	0.999	41	396	438
	4	29	10	0.995	0.998	44	431	476
	5	28	10	0.984	0.993	42	442	484
	6	26	10	0.999	1.000	42	364	406
	7	30	10	1.000	1.000	47	491	538
	8	28	10	0.999	1.000	44	422	465
	9	29	10	0.999	0.999	45	527	572
	10	32	10	0.999	1.000	50	453	503
	Average	28.6	10	0.997	0.999	44	428	472

Table B.34: F-ACOVNS Results for 10d5c1_5 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	71	10	0.997	0.999	117	426	543
	2	63	10	0.998	0.999	105	440	545
	3	65	10	0.997	0.999	102	337	439
	4	60	10	0.999	1.000	93	420	514
	5	82	10	0.999	0.999	128	456	584
	6	68	10	1.000	1.000	113	547	660
	7	68	10	0.999	1.000	115	441	556
	8	70	10	0.999	0.999	109	420	529
	9	73	10	0.999	1.000	114	404	518
	10	72	10	1.000	1.000	112	426	538
	Average	69.2	10	0.999	0.999	111	432	543
0.08	1	52	10	1.000	1.000	88	440	527
	2	50	10	0.998	0.999	84	478	562
	3	45	10	0.998	0.999	70	417	487
	4	45	10	0.996	0.998	70	438	508
	5	49	10	0.996	0.998	76	468	544
	6	47	10	0.841	0.929	79	551	630
	7	50	10	1.000	1.000	85	435	521
	8	54	10	1.000	1.000	86	335	421
	9	55	10	1.000	1.000	87	408	495
	10	55	10	1.000	1.000	86	341	427
	Average	50.2	10	0.983	0.992	81	431	512
0.1	1	51	10	0.999	0.999	80	434	514
	2	42	10	0.996	0.998	65	384	449
	3	42	10	0.997	0.999	65	357	423
	4	44	10	0.998	0.999	69	459	527
	5	46	10	0.766	0.899	73	458	531
	6	44	10	0.999	1.000	69	554	623
	7	48	10	0.999	1.000	76	320	395
	8	42	10	1.000	1.000	66	556	622
	9	44	10	1.000	1.000	69	330	399
	10	44	10	1.000	1.000	70	444	514
	Average	44.7	10	0.975	0.989	70	430	500
0.15	1	27	10	0.999	1.000	42	456	498
	2	29	10	0.998	0.999	45	428	473
	3	27	10	0.997	0.999	42	416	457
	4	38	10	0.998	0.999	59	469	528
	5	32	10	0.999	1.000	49	398	447
	6	41	10	1.000	1.000	64	555	619
	7	34	10	0.999	0.999	53	506	559
	8	33	10	0.998	0.999	51	404	455
	9	30	10	1.000	1.000	46	339	385
	10	28	10	1.000	1.000	44	450	494
	Average	31.9	10	0.999	0.999	49	442	492

Table B.35: F-ACOVNS Results for 10d5c1_10 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.05	1	77	10	0.997	0.999	126	420	546
	2	75	10	0.999	0.999	124	435	559
	3	73	10	0.999	1.000	115	441	556
	4	82	10	0.996	0.998	128	417	545
	5	69	10	0.997	0.999	109	370	478
	6	74	10	0.999	0.999	126	429	555
	7	67	10	0.999	1.000	115	601	716
	8	69	10	0.769	0.900	113	717	830
	9	67	10	1.000	1.000	109	526	635
	10	101	10	1.000	1.000	164	537	701
	Average	75.4	10	0.975	0.989	123	489	612
0.08	1	61	10	0.998	0.999	100	470	571
	2	54	10	0.989	0.995	90	459	549
	3	48	10	1.000	1.000	75	395	470
	4	52	10	0.998	0.999	81	425	506
	5	50	10	0.997	0.999	79	390	469
	6	56	10	0.999	0.999	96	438	534
	7	48	10	1.000	1.000	84	603	687
	8	51	10	1.000	1.000	84	563	646
	9	50	10	0.998	0.999	81	412	493
	10	56	10	1.000	1.000	91	552	644
	Average	52.6	10	0.998	0.999	86	471	557
0.1	1	53	10	0.997	0.998	82	446	529
	2	42	10	0.998	0.999	65	464	529
	3	37	10	0.999	0.999	58	435	493
	4	46	10	0.997	0.999	71	459	531
	5	40	10	0.998	0.999	62	457	519
	6	43	10	0.999	1.000	69	446	516
	7	45	10	0.999	1.000	73	608	681
	8	42	10	1.000	1.000	68	521	589
	9	41	10	1.000	1.000	66	538	604
	10	51	10	0.781	0.904	84	694	778
	Average	44	10	0.977	0.990	70	507	577
0.15	1	29	10	0.984	0.993	45	461	506
	2	32	10	0.998	0.999	50	393	442
	3	30	10	0.997	0.999	47	368	415
	4	35	10	0.809	0.914	55	461	516
	5	31	10	0.996	0.998	49	452	501
	6	34	10	1.000	1.000	55	456	510
	7	39	10	0.795	0.909	63	677	740
	8	29	10	1.000	1.000	47	580	627
	9	33	10	0.999	1.000	54	711	765
	10	32	10	0.999	1.000	52	669	721
	Average	32.4	10	0.958	0.981	52	523	574

Table B.36: F-ACOVNS Results for 20d5c1_4 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.25	1	30	16	0.991	0.995	39	441	480
	2	46	15	0.896	0.949	60	414	475
	3	52	16	1.000	1.000	68	429	498
	4	40	16	1.000	1.000	52	407	460
	5	25	19	0.908	0.954	33	427	459
	6	42	15	0.905	0.952	61	442	474
	7	29	20	0.937	0.968	44	482	526
	8	44	16	0.899	0.949	59	446	505
	9	25	19	1.000	1.000	33	347	380
	10	42	18	0.902	0.951	56	471	526
	Average	37.5	17	0.944	0.972	50	431	478
0.3	1	46	12	0.904	0.952	60	423	483
	2	32	17	1.000	1.000	42	407	448
	3	32	18	1.000	1.000	42	382	424
	4	21	20	0.849	0.928	27	417	444
	5	22	19	1.000	1.000	29	377	406
	6	32	15	1.000	1.000	42	307	441
	7	19	20	1.000	1.000	25	344	369
	8	30	17	1.000	1.000	40	356	396
	9	28	13	0.905	0.953	36	436	472
	10	27	21	0.902	0.951	36	481	516
	Average	28.9	17.2	0.956	0.978	38	393	440
0.35	1	23	19	1.000	1.000	30	403	433
	2	20	20	1.000	1.000	26	396	422
	3	32	16	0.901	0.951	42	399	441
	4	18	20	1.000	1.000	23	434	458
	5	27	17	0.899	0.950	35	422	456
	6	31	15	1.000	1.000	40	413	442
	7	40	14	1.000	1.000	52	344	396
	8	21	19	1.000	1.000	27	374	401
	9	27	15	1.000	1.000	35	316	351
	10	37	10	0.997	0.998	48	451	500
	Average	27.6	16.5	0.980	0.990	36	395	430

Table B.37: F-ACOVNS Results for 20d5c1_10 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
	1	33	19	0.998	0.999	33	279	312
	2	34	20	1.000	1.000	34	292	326
	3	39	19	1.000	1.000	39	269	307
	4	35	18	1.000	1.000	34	296	331
	5	36	18	1.000	1.000	35	280	316
0.25	6	38	18	0.867	0.935	42	454	318
	7	44	20	0.849	0.928	51	491	542
	8	42	19	0.898	0.949	45	468	513
	9	32	19	0.897	0.949	33	478	511
	10	32	21	0.848	0.928	32	483	515
	Average	36.5	19.1	0.936	0.969	38	379	399
	1	24	20	0.890	0.945	24	291	315
	2	34	19	0.999	0.999	37	298	335
	3	31	20	0.900	0.950	31	284	315
	4	25	17	0.849	0.928	25	283	308
	5	20	20	0.999	0.999	20	273	293
0.3	6	36	16	1.000	1.000	36	324	313
	7	26	20	1.000	1.000	25	373	399
	8	19	20	0.899	0.949	19	461	479
	9	27	20	1.000	1.000	26	335	361
	10	20	20	1.000	1.000	19	345	365
	Average	26.2	19.2	0.953	0.977	26	327	348
	1	28	17	0.895	0.948	28	283	310
	2	38	14	0.898	0.948	37	278	315
	3	37	15	0.998	0.999	36	264	301
	4	29	17	0.891	0.946	29	283	312
	5	30	17	0.874	0.939	30	283	313
0.35	6	26	16	1.000	1.000	25	321	310
	7	26	18	0.829	0.916	25	450	475
	8	36	15	0.879	0.937	35	374	409
	9	21	21	1.000	1.000	20	374	395
	10	25	20	0.896	0.948	24	447	471
	Average	29.6	17	0.916	0.958	29	336	361

Table B.38: F-ACOVNS Results for 20d5c1_20 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.25	1	52	15	1.000	1.000	69	367	437
	2	25	20	0.922	0.960	34	395	429
	3	62	14	1.000	1.000	83	328	410
	4	27	16	0.999	0.999	36	381	417
	5	37	18	1.000	1.000	49	399	449
	6	45	16	0.899	0.950	51	448	428
	7	44	14	0.898	0.947	48	443	491
	8	29	20	0.851	0.929	28	423	452
	9	47	14	1.000	1.000	45	297	342
	10	37	18	0.912	0.956	36	431	467
	Average	40.5	16.5	0.948	0.974	48	391	432
0.3	1	38	14	0.911	0.954	51	372	423
	2	31	16	1.000	1.000	41	384	425
	3	35	14	1.000	1.000	47	362	409
	4	38	15	1.000	1.000	50	316	366
	5	31	15	1.000	1.000	41	395	436
	6	26	18	1.000	1.000	25	278	412
	7	22	19	1.000	1.000	22	307	328
	8	26	19	1.000	1.000	25	239	264
	9	29	16	1.000	1.000	28	343	371
	10	29	18	0.898	0.949	29	453	482
	Average	30.5	16.4	0.981	0.990	36	345	392
0.35	1	21	18	1.000	1.000	28	396	424
	2	27	15	0.999	0.999	36	387	423
	3	34	15	0.910	0.953	45	383	428
	4	30	14	1.000	1.000	40	322	361
	5	24	17	1.000	1.000	32	359	391
	6	31	17	1.000	1.000	30	316	405
	7	20	19	0.896	0.948	20	446	465
	8	20	17	0.903	0.952	20	451	470
	9	26	15	1.000	1.000	25	326	352
	10	21	19	1.000	1.000	20	330	350
	Average	25.4	16.6	0.971	0.985	30	371	407

Table B.39: F-ACOVNS Results for 40d5c1_8 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.25	1	77	25	0.998	0.999	140	619	759
	2	64	25	0.921	0.961	116	606	722
	3	29	39	0.996	0.998	53	632	685
	4	65	27	0.999	0.999	118	595	713
	5	45	34	0.998	0.999	83	655	737
	6	59	32	0.922	0.962	112	677	723
	7	25	40	0.879	0.942	51	731	782
	8	69	22	0.999	0.999	124	564	688
	9	67	25	0.999	0.999	121	695	817
	10	75	25	0.919	0.960	136	674	810
	Average	57.5	29.4	0.963	0.982	105	645	744
0.3	1	56	25	0.998	0.999	101	604	705
	2	28	40	0.938	0.969	51	649	700
	3	55	22	0.855	0.932	98	588	686
	4	23	40	0.939	0.969	43	653	696
	5	49	24	0.998	0.999	89	574	663
	6	62	20	0.999	1.000	112	638	690
	7	65	22	0.923	0.962	118	600	719
	8	57	21	0.999	1.000	103	581	684
	9	45	31	0.862	0.935	82	724	805
	10	53	26	0.999	1.000	97	593	690
	Average	49.3	27.1	0.951	0.976	89	620	704
0.35	1	58	24	0.900	0.951	105	588	692
	2	36	33	0.998	0.999	66	527	593
	3	48	20	0.997	0.998	86	545	632
	4	53	23	0.887	0.945	95	585	680
	5	48	26	0.937	0.969	87	620	707
	6	21	37	1.000	1.000	39	588	661
	7	46	22	0.917	0.959	83	694	776
	8	24	39	0.857	0.933	45	660	705
	9	30	35	0.999	0.999	55	632	686
	10	47	20	0.997	0.998	84	653	737
	Average	41.1	27.9	0.949	0.975	74	609	687

Table B.40: F-ACOVNS Results for 40d5c1_20 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.25	1	49	29	0.889	0.947	96	680	776
	2	59	27	0.999	0.999	116	687	802
	3	77	22	0.860	0.934	149	662	811
	4	46	33	0.999	1.000	90	702	793
	5	49	25	0.859	0.934	96	667	763
	6	50	32	0.871	0.939	97	723	789
	7	67	27	1.000	1.000	129	463	593
	8	76	23	0.999	0.999	141	689	830
	9	47	29	0.999	0.999	87	507	593
	10	66	25	0.923	0.962	121	685	807
	Average	58.6	27.2	0.940	0.971	112	647	756
0.3	1	34	32	1.000	1.000	73	704	777
	2	27	37	0.999	0.999	59	704	763
	3	68	23	0.999	1.000	131	670	801
	4	48	29	0.904	0.953	93	675	767
	5	65	24	0.999	0.999	126	683	809
	6	69	21	0.919	0.960	126	693	783
	7	55	23	1.000	1.000	100	463	563
	8	63	23	0.923	0.962	115	706	821
	9	52	26	0.859	0.934	95	694	789
	10	63	23	0.897	0.950	114	684	798
	Average	54.4	26.1	0.950	0.976	103	668	767
0.35	1	45	23	0.997	0.999	87	663	751
	2	53	23	0.999	1.000	100	663	763
	3	40	32	0.999	0.999	76	685	761
	4	34	28	0.999	0.999	65	687	753
	5	49	21	0.855	0.932	94	650	744
	6	45	22	1.000	1.000	82	537	754
	7	42	21	1.000	1.000	77	536	613
	8	38	23	0.892	0.948	70	676	747
	9	43	22	0.897	0.950	78	671	749
	10	36	31	0.875	0.941	66	653	718
	Average	42.5	24.6	0.951	0.977	80	642	735

Table B.41: F-ACOVNS Results for 40d5c1_40 Data Set (10 Replications)

Evaporation Rate	Replication	Number of iters for feature convergence	Number of selected features	F-Measure	Rand Index	Feature Selection Duration (sec)	Clustering Duration (sec)	Total Duration (sec)
0.25	1	80	19	0.999	1.000	110	457	567
	2	85	20	0.994	0.997	118	460	577
	3	61	25	0.994	0.997	86	455	540
	4	49	29	0.998	0.999	66	469	535
	5	57	21	0.998	0.999	78	450	528
	6	49	32	0.897	0.950	78	623	550
	7	53	25	1.000	1.000	78	471	550
	8	77	23	0.999	0.999	106	616	722
	9	70	23	1.000	1.000	97	451	548
	10	59	29	0.872	0.939	83	608	690
	Average	64	24.6	0.975	0.988	90	506	581
0.3	1	52	22	0.907	0.954	70	443	513
	2	64	19	0.920	0.960	91	441	532
	3	58	18	0.923	0.962	78	444	522
	4	35	30	0.923	0.962	50	478	527
	5	55	20	0.925	0.963	77	453	530
	6	60	19	0.999	1.000	84	455	525
	7	49	21	1.000	1.000	68	406	475
	8	59	24	0.999	1.000	81	541	622
	9	57	23	0.999	0.999	83	623	706
	10	52	23	0.861	0.935	73	607	681
	Average	54.1	21.9	0.945	0.973	76	489	563
0.35	1	44	24	0.999	1.000	61	458	519
	2	46	23	0.997	0.999	62	445	507
	3	41	23	0.860	0.934	55	437	493
	4	56	18	0.917	0.958	75	431	506
	5	54	17	0.999	0.999	72	460	532
	6	48	23	1.000	1.000	66	534	511
	7	37	26	1.000	1.000	53	616	670
	8	35	25	0.901	0.952	64	788	853
	9	51	19	0.998	0.999	93	609	702
	10	48	23	0.999	1.000	65	549	614
	Average	46	22.1	0.967	0.984	67	533	591