POSITION UPDATE AND SEARCH STRATEGY BASED BINARY PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION

TIJJANI SANI

UNIVERSITI SAINS MALAYSIA

2025

POSITION UPDATE AND SEARCH STRATEGY BASED BINARY PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION

by

TIJJANI SANI

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2025

ACKNOWLEDGEMENT

I would like to thank my creator, Almighty Allah, the most beneficent, and the most merciful, for all his guidance and giving while I was preparing, doing, and finishing this research work. May peace, mercy, and blessing of Allah be upon his final messenger, Prophet Muhammad (S.A.W), upon his families and all his companions (R.A). I wish to express my deepest gratitude to my supervisor Ass Prof. Dr. Mohd Halim Bin Mohd Noor, for his guidance, advice, correction, encouragement, and insight throughout that research. My sincere gratitude also goes to my co-supervisor Dr. Mohd Nadhir Ab Wahab for his kind assistance toward the success of this research work. I really appreciate their kindness in allowing me to carry out this work under their supervision. My due appreciation to Prof. Dr. Salwani Abdullah, Ass Prof. Dr. Chew Xingying, and Dr. Zarul Fitri Zaaba for their unforgettable contribution toward the success of this research work. Moreover, I want thanks Ass Prof. Dr. Wong Li Pei and Dr. Fadratul Hafinaz Hassan for their contribution toward the success of this research work. Furthermore, I thank my colleague Dr. Abdullahi Uwaisu Muhammad for his assistance that helped the success of this work. I would like to express my sincere gratitude to my beloved parents for their unforgettable responsibility, guidance, and prayers. In addition, I acknowledged the support of my great country Nigeria for sponsoring this program. May Allah protect and help our great nation. I would like to thank my beloved wife, uncles, brothers, sisters, friends, and the entire family for their endless love, encouragement, and support throughout my life. Lastly, I would like to thank all of those who supported me in any respect during the completion of this study.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

C1	Positive acceleration coefficient 1
C2	Positive acceleration coefficient 2
$\mathbf{G_{best}}$	Position of the best particle in a population
$\mathbf{P_{best}}$	Particle's best position
R	A random number between 0 and 1
$\mathbf{V_{(t)}}$	Current particle's velocity
W	Inertia weight
$\mathbf{X_{(t)}}$	Current particle's position

LIST OF APPENDICES

ABC	Artificial Bee Colony
Acc	Accuracy
ACO	Ant Colony Optimization
AI	Artificial Intelligence
Ave	Average
BHOA	Binary Horse Herd Optimization Algorithm
BOA	Butterfly Optimization Algorithm
BOAPSO	Butterfly Optimization Algorithm and Particle Swarm Optimization
CPU	Central Processing Unit
CS-BPSO	Chi-Square and Binary Particle Swarm Optimization
EA	Evolutionary Algorithms
Fs	Feature Selection
GA	Genetic Algorithm
GAFS	Genetic Algorithm for Feature Selection
GOA-EPD	Grasshopper Optimization and Evolutionary Population Dynamics
GSK	Gaining-Sharing Knowledge
HGSA	Hybrid Gravitational Search-based Algorithm
HPSO_SSM	Hybrid Particle Swarm Optimization with a Spiral-Shaped Mechanism
IBPSO	Improved Binary Particle Swarm Optimization
IWSS	Incremental Wrapper Subset Selection
KNN	K – Nearest Kneighbour
ML	Machine Learning
MPPSO	Multi-Population-based Particle Swarm Optimization
NB	Naïve Bayes
NFL	No-Free-Lunch
PSO	Particle Swarm Optimization
PSOFS	Particle Swarm Optimization for Feature Selection
RAM	Random Access Memory
RSO	Rat Warm Optimization
SI	Swarm Intelligence
Std	Standard Deviation

SVM	Support Vector Machine
TLPSO	Three-layer particle swarm optimization
TMGWOA	Grey Wolf Optimizer Algorithm with a Two-phase Mutation
UCI	University of Califonia Irvine

PENGOPTIMUMAN KERUMUNAN ZARAH BINARI BERASASKAN KEMASKINI KEDUDUKAN DAN STRATEGI CARIAN UNTUK PEMILIHAN CIRI

ABSTRAK

Pengembangan data yang pesat disebabkan oleh kemajuan dalam sains dan teknologi menyebabkan peningkatan dalam kedua-dua bilangan kejadian dan ciri (dimensi). Apabila dimensi data meningkat, kos pengiraan juga meningkat, kebanyakannya secara eksponen. Oleh itu, mereka bentuk teknik pemilihan ciri yang boleh memilih ciri yang relevan adalah sangat penting. Algoritma Perduaan Particle Swarm Optimization (BPSO) sering menumpu kepada optimum tempatan dengan cepat (stagnasi optima tempatan), kehilangan peluang lebih baik yang membawa kepada penumpuan pramatang. Selain itu, perumus kemas kini kedudukan asal tidak dapat mengimbangi eksploitasi (carian tempatan) dan penerokaan (carian global) secara berkesan dalam proses carian. Oleh itu, kerja penyelidikan ini mencadangkan kaedah pengoptimuman kawanan zarah binari yang diubah suai untuk pemilihan ciri. Ciri utamanya yang pertama, ialah penerapan pengurangan pengiraan yang mengurangkan bilangan ciri berlebihan dan kos pengiraan. Kedua, menentukan kebarangkalian zarah bertukar kedudukan menggunakan nilai kedudukan dan bukannya nilai halaju yang mempercepatkan penumpuan dan membatalkan penumpuan pramatang secara serentak. Dua kaedah varian untuk menentukan kebarangkalian menukar kedudukan unsur zarah telah diperkenalkan. Ini menghasilkan dua varian pengoptimuman kawanan zarah binari yang dipertingkatkan (EBPSO) untuk pemilihan ciri, dipanggil EBPSO1 dan EBPSO2. Ketiga, mencadangkan strategi carian yang dipertingkatkan yang mengimbangi

eksploitasi dan penerokaan BPSO yang dipertingkatkan dengan mengubah suai perumus kemas kini kedudukan PSO asal menggunakan parameter kawalan pengoptimuman kawanan tikus (RSO) dan feromon pengoptimuman koloni semut (ACO). Beberapa eksperimen telah dilakukan untuk membandingkan prestasi kaedah pemilihan ciri berasaskan BPSO yang diubah suai dengan teknik pemilihan ciri terkini berdasarkan beberapa set data penanda aras yang kerap digunakan. Kaedah yang dicadangkan menduduki tempat pertama dari segi ketepatan, pemilihan ciri dan kadar ralat. Ketepatan yang diperoleh menggunakan 18 data terkenal dari repositori UCI adalah antara 82.41% hingga 100%. Pemilihan ciri adalah antara 4.55 hingga 24.47 dan kadar ralat adalah antara 0.008 hingga 0.1772 bergantung pada set data.

POSITION UPDATE AND SEARCH STRATEGY BASED BINARY PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION

ABSTRACT

The rapid expansion of data due to the advances in science and technology causes an increase in both the number of instances and features (dimensions). When the dimensionality of data increases, the computational cost also increases, mainly exponentially. Consequently, designing a feature selection method that can select relevant features is very important. The binary Particle Swarm Optimization (BPSO) algorithm often converges to a local optimum quickly (local optima stagnation), missing better opportunities that lead to premature convergence. In addition, the original position update formula cannot effectively balance exploitation (local search) and exploration (global search) in the search process. Hence, this research work proposes a modified binary particle swarm optimization method for feature selection. Its main feature first, is the application of a computational reduction which lowers the number of redundant features and computation costs. Secondly, determines the probability of the particles switching positions using position values rather than velocity values which speed up convergence and overcome the premature convergence simultaneously. Two variant methods for determining the probability of changing the position of a particle element were introduced. These result in the two variants of enhanced binary particle swarm optimization (EBPSO) for feature selection, called EBPSO1 and EBPSO2. Third, proposed an improved search strategy that balances exploitation and exploration of the enhanced BPSO by adapting the position update formula using Rat swarm optimization (RSO) control parameters and Ant colony optimization (ACO) pheromone. Several experiments

were performed to compare the performance of the modified BPSO-based feature selection method with the state-of-the-art feature selection techniques based on some frequently used benchmark datasets. The proposed method ranked first in terms of accuracy, features select and error rate. The accuracy obtained using the 18 well known data from UCI repository is range from 82.41% to 100%. The feature selection is range from 4.55 to 24.47 and error rate is range from 0.008 to 0.1772 depending on the dataset.

CHAPTER 1

INTRODUCTION

1.1 Background

Nowadays, the exponential growth of data due to the advances in science and technology (Moradi & Gholampour, 2016), makes data scientists and researchers focus on data analysis and extracting relevant information from data (Brezočnik et al., 2018). The datasets have expanded in terms of both the number of occurrences and features due to the gradual rise of information and the plenty of data (Khurma et al., 2022). The computing cost rises together with the dimensionality of the data, mainly exponentially (Brezočnik et al., 2018). Finding a method to reduce the size of the dataset is required to solve these issues caused by the high dimensionality of the data (Brezočnik et al., 2018).

Generally, researchers used mainly two methods (Jindal, 2017). The first is called feature extraction, and it entails generating a new, low-dimensional feature space (Brezočnik et al., 2018). The second method is feature selection, which identifies the key elements of the original set of features while discarding those that are superfluous or redundant features (Brezočnik et al., 2018), (Subanya & Rajalaxmi, 2014). Unlike feature extraction, feature selection does not transform the features, it only chooses the chiefly advantageous feature subset and discards the redundant features (Vieira et al., 2012). The feature extraction is subdivided into linear and non-linear while the feature selection method is subdivided into wrapper, filter, and hybrid methods (Ahmad & Bou, 2022). The feature selection method is used to identify the key elements of the original set of features while discarding those that are unnecessary or redundant. The feature selection involves the following steps shown in Figure 1.1.

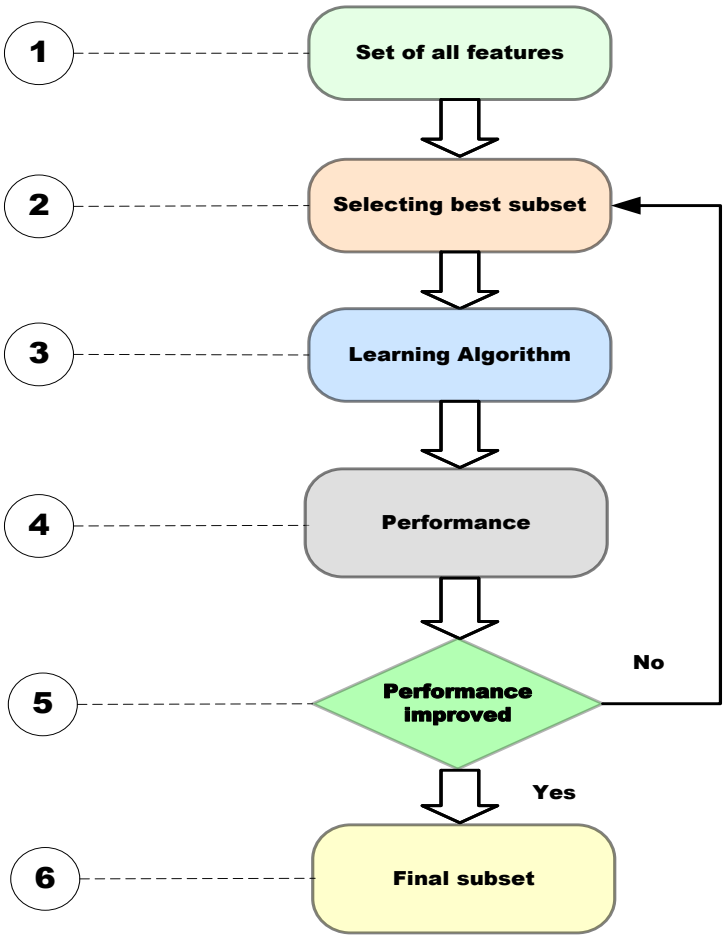

Figure 1.1 The Basic feature selection process

However, a learning algorithm cannot have good performance in a high-dimensional environment (Khurma et al., 2022), (Xue et al., 2014). The reason of this is, high dimensionality results in noisy features, which can deceive the classifier and lead to the overfitting of the data (Khurma et al., 2022). If a classifier has been over-trained using the data and has learned all training instances, comprising redundant and irrelevant features, an overfitting issue arises. Therefore, an overfitting issue refers to the situation where a classifier is over-trained on the data and has learned all the training instances, containing redundant and irrelevant features (Khurma et al., 2022). The previous belief in machine learning (ML), that "the more variables, the greater the performance" has been unacceptable for a while now

(Brezočnik et al., 2018). This is due to the fact that, for logical reasons, the classifier can be more accurate by learning from relevant features (Khurma et al., 2022). Hence, how data mining and machine learning employ feature selection methods has been rapidly increasing (Brezočnik et al., 2018). Hence, this research work proposes a modified particle swarm optimization method for feature selection.

1.2 Motivation

The rapid expansion of data due to the advances in science and technology causes an increase in both the number of instances and features (dimensions). When the dimensionality of data increases, the computational cost also increases, mainly exponentially. Consequently, data scientists and researchers focus on data analysis and extracting relevant information from data. According to International Data Corporation shown in Figure 1.2, the amount of digital data expected to be created by 2025 will be greater than twice the amount of data created since the advent of digital storage (International Data Cooperation, 2023). Therefore, designing a feature selection method that can select relevant features is very important.

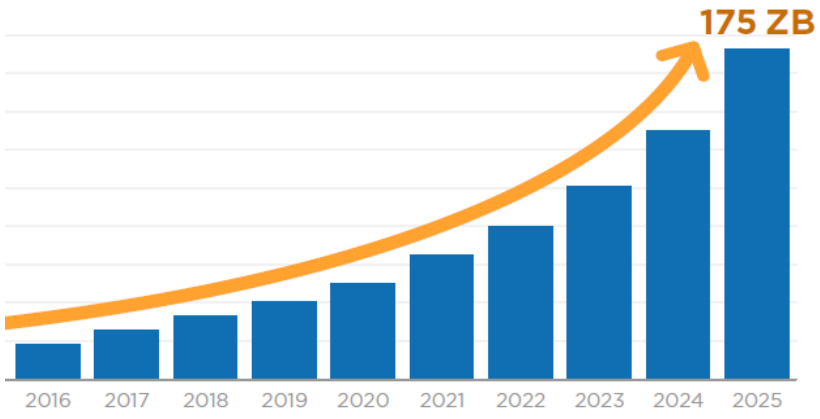

Figure 1.2 Amounts of digital data expected to be created by 2025 (International Data Cooperation, 2023).

The feature selection method is used to identify the key elements of the original set of features while discarding those that are unnecessary or redundant. Although many different search methods have been applied to feature selection, the majority of these algorithms still suffer from the problem of being computationally expensive (Xue et al., 2014). In addition, other feature selection methods face the drawback of selecting redundant features (Rajamohana & Umamaheswari, 2018) which degrade the performance of the model. Moreover, sometimes data obtained in the real world are not completely clean and reliable. Furthermore, not all the features contribute equally to the classification process, and including all features creates a problem known as the curse of dimensionality (Sensor et al., 2020). Consequently, designing an optimization method that can select relevant features and maintain adequate performance is very important. According to the no-free-lunch (NFL) theory, there will be no algorithm that is suitable for solving all the optimization problems (Arora & Anand, 2019),(Aljarah et al., 2020),(Ji et al., 2020). Hence, this observation has motivated us to propose a feature selection method using metaheuristic algorithms for reducing complexity and improving the performance of the learning tool.

1.3 Problem Statement

In recent years, research in feature selection has been rapidly developed in line with the era of huge datasets. This subject has attracted the attention of researchers who have become more interested in developing new feature selection (Fs) methods or improving current technologies. Over the past few years, a variety of novel meta-heuristic algorithms have been proposed (Moazen et al., 2023). These algorithms have demonstrated tremendous promise for solving practical optimization

issues, particularly in the area of feature selection (Mafarja et al., 2018), (Tu et al., 2019). Among those metaheuristic algorithms, Particle swarm optimization (PSO) has gradually gained popularity as a topic for resolving FS in recent years (Song et al., 2021),(Pang et al., 2023), and is widely used because of its quick convergence, stability with particle collaboration (Zhou et al., 2020), a few parameters tuning and low computational cost (Kılıç et al., 2021), (Hu & Li, 2023). PSO has been used in feature selection currently and works very well when dealing with high-dimensional data (Zhou et al., 2020). Many academics have switched to the BPSO to determine the best subset of features which is currently utilized to address the feature selection problem (Taradeh et al., 2019). Despite the high number of PSO variant that have been proposed, they still suffer from the drawbacks of the original PSO (Moazen et al., 2023), (Zhu et al., 2023). Thus, escaping from premature convergence and maintaining good balance between exploitation and exploration are still challenges (Zhang & Kong, 2022). Therefore, the current version of BPSO suffers from the following drawbacks: **First**, the BPSO algorithm often converges to a local optimum quickly (local optima stagnation), missing better opportunities that lead to premature convergence (Kousounadis-Knousen et al., 2023), (Hu & Li, 2023), (Ming & Xie, 2024). **Secondly**, the original position update formula cannot effectively balance exploitation and exploration in the search process (Chen et al., 2019), (Ahmad et al., 2023). Consequently, there is a lack of balancing between local search (exploitation) and global search (exploration) in the BPSO (Moazen et al., 2023). **Third**, using position values rather than velocity values is not enough to speed up convergence and overcome the premature convergence. For this reason, the positions update formula used in the enhanced BPSO cannot control the pressure of local and global searches. Therefore, an exploitation and exploration balancing strategy is needed.

However, the research gaps identified are: a) Inefficient Position update strategy in the current BPSO. Since; the BPSO uses velocity value to update particles positions, which may lead to premature convergence. Therefore, there is need of alternate position update strategy that utilizes position value instead of velocity value. b) In addition, the position update of the existing method experiences high exploration when the velocity contribution to the position update is large. Similarly, when the velocity contribution to the position update is too small, this will promote exploitation. For this reason; the influence of velocity is not properly controlled in the original position update equation, leading to this unsteady behaviour. Thus, having more exploitation than exploration, the algorithm will stuck in locally optimal solution. At the same time, the algorithm will lose some best solutions if it is go far in exploration. Therefore, there is need for a modified position update equation in a way that can efficiently balance the local and global searches to improve the feature selection.

1.4 Research Questions

These research questions aim to find specific issues to be addressed to alleviate data pre-processing and feature selection issues which include classifier overfitting, the curse of dimensionality, selecting redundant features, high computational cost, premature convergence. These questions include:

1. How to address the limitations of some existing BPSO-based feature selection methods regarding premature convergence and trapping in local minima?
2. What is the solution to the lack of balancing between global search (exploitation) and global search (exploration) suffered by the current BPSO?

3. How to address the BPSO issues regarding premature convergence and imbalance between exploitation & exploration using a single model?

1.5 Objectives

This research aims to develop a feature selection method that not only reduces the data dimension but also improves the accuracy of the learning tool. Specifically, the goal will be achieved through the following objectives:

1. To improve the particles switching position strategy of BPSO-based feature selection by determining the probability using position values rather than velocity values.
2. To improve the position update strategy of BPSO-based feature selection using EBPSO2 position update strategy and adapting position update formula using formulations from RSO and ACO.
3. To improve the search strategy of BPSO-based feature selections using the EBPSO1 position update strategy and formulations from RSO and ACO.

1.6 Expected Contributions

This research work proposes a modified particle swarm optimization method for feature selection. The main contributions of this research work are summarized as follows:

1. An improved BPSO-based feature selection that determines the probability of the particles switching positions using position values rather than velocity values to overcomes premature convergence which results in the two variants of enhanced binary particle swarm optimization (EBPSO) for feature selection, called EBPSO1 and EBPSO2

2. An enhanced BPSO-based feature selection with an improved search strategy using EBPSO2 position update strategy and modified standard PSO position update formula using formulations from Rat swarm optimization (RSO) control parameters & Ant colony optimization (ACO) pheromone that balances its local search (exploitation) and global search (exploration).
3. An improved BPSO-based feature selection with EBPSO1 position updates strategy and formulations from Rat swarm optimization (RSO) control parameters & Ant colony optimization (ACO) pheromone value that help to better overcome the premature convergence and improve search strategy.

1.7 Scope of Studies

The scope of this research work will be limited to investigating and enhancing BPSO-based feature selection. Several experiments intend to be performed to compare the performance of the modified BPSO-based feature selection method with the state-of-the-art feature selection techniques. The data intended to use, are some well-known benchmark data sets from the UCI repository.

The novelties of the proposed work include the following:

- **Alternative position update method:** Enhanced BPSO-based feature selection by determining the probability of the particles switching positions using position values rather than velocity values to overcome the premature convergence which results in the two variants of enhanced binary particle swarm optimization (EBPSO) for feature selection, called EBPSO1 and EBPSO2
- **Modified standard PSO position update equation:** Enhanced BPSO-based feature selection by improving its search strategy using EBPSO2 position

update strategy and adapting the position update formula using formulations from Rat swarm optimization (RSO) control parameters & Ant colony optimization (ACO) pheromone to balance exploitation and exploration of the enhanced BPSO.

- **More effective PSO-based feature selection method:** Introduced BPSO-based feature selection using the EBPSO1 position update strategy and formulations from Rat swarm optimization (RSO) control parameters & Ant colony optimization (ACO) pheromone value formulation to better overcome the premature convergence and improve search strategy.

1.8 Organization of the Research Work

The remainder of this research work is organized as follows: CHAPTER 2 provides an application of feature selection, an overview of the literature on Feature selection methods, Types of feature selection, Metaheuristics algorithms, and Particle swarm optimization. CHAPTER 3 describes the methodology of building our proposed model. CHAPTER 4 describes enhanced binary particle swarm optimization with position updates for optimal feature selection. CHAPTER 5 describes binary particle swarm optimization with exploitation and exploration balancing methods for optimal feature selection. CHAPTER 6 describes the proposed feature selection method based on an enhanced particle swarm optimization with position update strategy and method for exploitation and exploration balancing. Finally, CHAPTER 7 highlights the research summary, contributions to knowledge, practical applications, and future research directions based on the shortcomings of the proposed methods.

This study proposes an enhanced particle swarm optimization with position updating for feature selection is proposed. The research steps are:

a) Providing the background information and formulating the research questions and objectives
b) Conducting an in-depth review of the existing literature and identifying the research gaps.
c) Describing the research design, methodology, and datasets
d) Describes in details how objective 1 addresses the research question 1. Highlights the results and key findings.
e) Describes in details how objective 2 addresses the research question 2. Highlights the results and key findings.
f) Describes in details how objectives 3 addresses the research question 3. Highlights the results and key findings.
g) Concludes by providing the key findings, the limitations of the study, and recommendations for feature research

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

One of the usual drawbacks encountered when modeling real systems is noise and redundancy in data. Hence, it is very important, when preprocessing data, to select the optimal feature subset (Vieira et al., 2013). Feature selection refers to a technique in which a subset of features is chosen from the original dataset (Ibrahim et al., 2021). The main purpose of feature selection is to reduce the number of features which leads to reducing the processing cost (Ibrahim et al., 2021) and increases the performance of the model and the accuracy of classification (Rajamohana & Umamaheswari, 2018).

Nowadays, massive volumes of data are required for the development of technologies comprising artificial intelligence and machine learning. The features that are derived from the data are key factors in how successful these technologies are. Reducing data dimension is one of the crucial pre-processing methods, which is intended to optimize the performance of various data mining jobs by reducing the number of features in a dataset to a minimum (Taradeh et al., 2019). Feature selection plays the main role in extracting useful knowledge from the collected datasets (Mafarja et al., 2018). Feature selection refers to the technique for selecting the smallest possible feature subset from the initial set of features to satisfy a measuring criterion (EL-Hasnony et al., 2022), (Taradeh et al., 2019). In addition, the efficiency of the subsequent data mining operations in terms of computational time and learning performance can be greatly enhanced by removing the redundant and unnecessary features from a dataset (Taradeh et al., 2019). Finding an optimal feature subset from the entire dataset that keeps or improves the learning algorithms' performance is the basic goal of feature selection (EL-Hasnony et al., 2022), (Ji et al., 2020). Therefore,

feature selection is a practical method for improving classification performance in machine learning by reducing the quantity of data features (Ji et al., 2020).

2.2 Approaches of Feature Selection

The researchers are mainly using two techniques to reduce the datasets (Jindal, 2017). The first is called feature reduction, and it entails generating a new, low-dimensional feature space (Brezočnik et al., 2018). The second method is feature selection, which identifies the key elements of the original set of features while discarding those that are superfluous or redundant features (Brezočnik et al., 2018), (Subanya & Rajalaxmi, 2014). Unlike feature reduction, feature selection does not transform the features, it only chooses the best feature subset and discards the redundant features (Vieira et al., 2012).

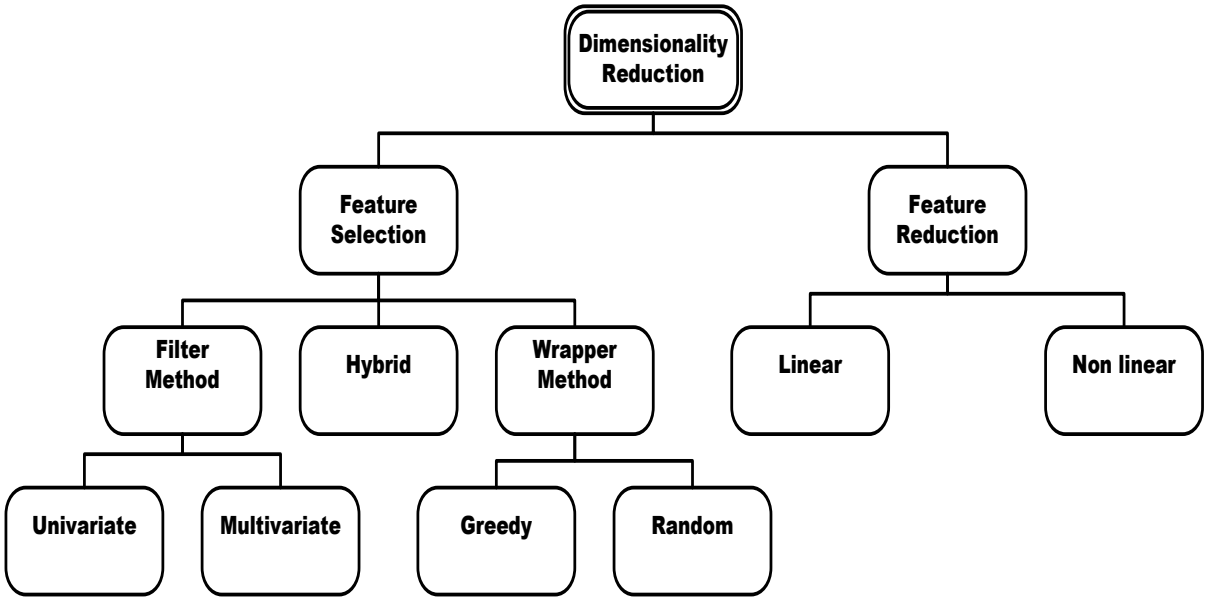

Figure 2.1 The taxonomy of dimensionality reduction methods

The feature reduction is subdivided into linear and non-linear while the feature selection method is subdivided into wrapper, filter, and hybrid methods (Ahmad &

Bou, 2022). Linear feature reduction is used to extract linear features. Example of this method includes principal component analysis (PCA), linear discriminant analysis (LDA), and singular value decomposition (SVD). Non-linear feature reduction is used to extract non-linear features. Example of this method includes kernel PCA, t-distributed Stochastic Neighbor embedding (t-SNE), and multi-dimensional scaling (MDS). The taxonomy of dimensionality reduction methods along with their approaches is shown in Figure 2.1. Feature selection gives the following advantages: 1) Reducing storage requirement. 2) Avoid overfitting. 3) Facilitating data visualization. 4) Speeding up the execution of mining algorithms. 5) Reducing training times (Jindal, 2017). Therefore, a feature selection method decided to be used in this research work to reduce the datasets.

Research in feature selection has followed the general trend of the era of big datasets (Khurma et al., 2022). Thereby, research in this subject has been rapidly developed and has drawn the attention of many scholars, who have grown more interested in producing or improving feature selection methods (Khurma et al., 2022). Feature selection refers to the pre-processing method that allows you to choose a useful subset of features from the initial feature collection (Vieira et al., 2013), (Subanya & Rajalaxmi, 2014). Feature selection aims mainly to eliminate unnecessary and redundant components from the initial feature set and select the best feature subset (Brezočnik et al., 2018).

However, there is a significant increase in the data-producing systems in our daily lives due to the growth of the sources that produce this data (EL-Hasnony et al., 2022). The applications for machine learning frequently depend on datasets with a lot of data (Ji et al., 2020). In addition, the effectiveness of the clustering and classification methods is significantly impacted as the complexity of the datasets rises

(Arora & Anand, 2019). High-dimensional datasets offer several drawbacks, including a lengthy model creation process, redundant data, and decreased performance, which makes data analysis particularly challenging (Arora & Anand, 2019). The main challenge here is how to extract the highly informative data and remove unnecessary and redundant features (EL-Hasnony et al., 2022). Thus, training a classifier with the highest informative data improves performance and reduces the computational complexity of the classifier (EL-Hasnony et al., 2022). Feature selection is a helpful technique because it can remove redundant noise from datasets, allowing machine-learning algorithms to run more quickly and effectively (Ji et al., 2020). There are three classes of feature selection methods: wrapper, filter, and hybrid (Abdel-Basset et al., 2020), (BinSaeedan & Alramlawi, 2021).

2.2.1 Filter-based Feature Selection

In filter-based feature selection methods, each feature is ranked by its probabilistic or statistical characteristics without the use of a learning strategy or classifier model (Moradi & Gholampour, 2016), (Chen et al., 2019). Thus, filter approaches do not rely on any specific learning algorithm(Moradi & Gholampour, 2016), (Xue et al., 2014), (Xue et al., 2012). In this way, statistical metrics are used to evaluate the selected subset (Moradi & Gholampour, 2016), (Taradeh et al., 2019), (Rostami et al., 2021). In other words, filter techniques are independent since they don't use any learning algorithms (Hamid et al., 2021) and select the meaningful feature subset based on statistical data analysis (Moradi & Gholampour, 2016), (Awadallah, Al-Betar, et al., 2022). This can be done by initially, evaluating each feature and assigning a score based on the statistical metrics. Features with scores below a certain level (threshold) are not taken into consideration whereas features

within the predefined level are selected and included in the subset (Awadallah, Al-Betar, et al., 2022). The filter-based approaches are faster than wrapper methods as it does not use a learning tool in its feature selection process (Moradi & Gholampour, 2016), (Awadallah, Al-Betar, et al., 2022). Sometimes the accuracy of the selected feature subset is not adequate (Awadallah, Al-Betar, et al., 2022). The filter-based methods fall into two categories: univariate and multivariate methods. (Moradi & Gholampour, 2016), (Chen et al., 2019).

2.2.1(a) Univariate Approaches

For the univariate approaches, each feature's relevance is independently assessed by a particular statistical criterion (Moradi & Gholampour, 2016), (Chen et al., 2019). This implies that each feature is taken into account independently, neglecting the correlations between features, which could result in less accurate classification (Moradi & Gholampour, 2016), (Chen et al., 2019). The literature has suggested several univariate techniques, such as the chi-square (Bahassine et al., 2020), F-score (Sevani et al., 2019), and information gain (Omuya et al., 2021).

2.2.1(b) Multivariate Approaches

A multivariate approach, on the other hand, has been introduced to solve the drawback of neglecting the correlations among features (Chen et al., 2019). In multivariate approaches, features' relationships are taken into account while evaluating the importance of the features (Moradi & Gholampour, 2016). Due to this, these approaches are slower than univariate methods (Pirgazi et al., 2019). The multivariate method example of filter-based feature selection includes mutual information (Sharmin et al., 2019), Minimal-redundancy- maximal-relevance (Senawi et al., 2017), and Relevance-redundancy feature selection (Tsanas, 2022).

2.2.1(c) Filter-based Method Related Work

The filter-based method was used in many approaches for solving various problems due to its being less computationally complicated. For instance, (Sharmin et al., 2019) first suggest a feature selection method called Joint Bias corrected Mutual Information (JBMI) that takes into account both discretization and feature selection simultaneously. When calculating mutual information for finite samples, this method takes the adjustment into account. Second, a framework called modified discretization and feature selection based on mutual information is suggested. It combines dynamic discretization and JBMI-based feature selection, both of which employ an x^2-based searching strategy.

In another filter-based method proposed by (Hancer, Xue, & Zhang, 2018), the idea of mutual information, ReliefF, and Fisher Score to suggest a new filter-based feature selection method was used. The proposed method attempts to select the highest-rated features as determined by ReliefF and Fisher Score while giving the mutual relevance between features and the class labels instead of employing the mutual redundancy method. Two new differential evolutions (DE) based filter methods are created based on the proposed criterion. The proposed criterion is taken into account in the first method as part of a multi-objective design, as opposed to the second method, which employs it as a single objective problem in a weighted manner. Additionally, single-objective and multi-objective DE algorithms for feature selection were adopted using a well-known mutual information feature selection method (MIFS) based on maximum relevance and minimum redundancy.

In addition, (Senawi et al., 2017) introduce the maximum relevance–minimum multicollinearity (MRmMC) method as a new relevancy-redundancy method for feature selection and ranking that can address some issues with current criteria.

According to the suggested strategy, conditional variance-based correlation characteristics are used to measure relevant variables while redundancy removal is accomplished based on multiple correlation assessments, using an orthogonal projection approach.

Moreover, (Bahassine et al., 2020) developed an effective method for categorizing Arabic text that makes use of the Chi-square feature selection to boost classification effectiveness. This work aims to reduce the amount of data and deliver more accurate classification results. In addition, the improved chi-square compared with three conventional filter-based feature selection methods: mutual information, information gain, and chi-square.

Furthermore, (Omuya et al., 2021) introduce a hybrid approach for feature selection and classifying data based on principal components and Information Gain. This work aims to minimize data dimensions, shorten training time, and improve classification performance by utilizing the chosen features. The hybrid model is a filter-based feature selection method that selects the meaningful features from the original data using Principal Component Analysis (PCA) and information gain and produces good classification accuracy. This approach keeps the computational overheads to a minimum while yet enabling accuracy in the process to be assessed.

Table 2.1 Filter-based Feature Selection Methods Analysis

S/N	Paper, Authors & Year	Classification Algorithm	Dataset	Method	Overall Outcomes	Weakness
1.	Simultaneous feature selection and discretization based on mutual information (Sharmin et al., 2019)	SVM	UCI ML Repository & Feature Selection Repository of Arizona State University	Combines dynamic discretization and Joint Bias corrected Mutual Information (JBMI) based feature selection, both of which employ an x^2-based searching strategy.	This work successfully forms a framework that incorporates feature selection and discretization simultaneously.	Need to tune some parameters for each dataset.
2.	Differential evolution for filter feature selection based on information theory and feature ranking (Hancer et al., 2018)	KNN & NB	UCI ML Repository, Biomedical data (DNA) and Text classification data.	A filter-based feature selection method based on information theory, feature ranking, and evolutionary computation (EC) techniques.	Develop a single objective and multi-objective DE based filter feature selection.	Difficulty in identifying redundant features
3.	A new maximum relevance–minimum multicollinearity (MRmMC) method for feature selection and ranking (Senawi et al., 2017).	KNN, NB, SVM & CART	UCI ML Repository	Features are measured by correlation characteristics based on conditional variance while redundancy elimination is achieved according to multiple correlation assessments using an orthogonal projection scheme.	The MRmMC works well with some popular classifiers such as KJNN, NB SVM and CART.	The method may not always find the optimal subset as the search is non-exhaustive
4.	Feature selection using an improved Chi-square for Arabic text classification (Bahassine et al., 2020).	Decision Tree (DT) & SVM	A dataset collected from local and international newspaper websites.	An improved Chi-square was used for feature selection to boost classification effectiveness.	The improved Chi-square achieved better performance than the use of Chi-square, MI and GI in Arabic text classification.	The performance of the algorithm using other datasets is not guaranteed.

Table 2.1 Continued

S/N	Paper, Authors & Year	Classification Algorithm	Dataset	Method	Overall Outcomes	Weakness
5.	Feature Selection for Classification using Principal Component Analysis and Information Gain (Omuya et al., 2021).	DT, NB, SVM and ZeroR		Principal Component Analysis (PCA) and information gain filter-based feature selection method are used to select the meaningful features from the original data.		There is need for a better search mechanism.
6	Adapting Feature Selection Algorithms for the Classification of Chinese Texts (X. Liu et al., 2023)	SVM and NB	Sogou news dataset	An enhanced CHI square with mutual information, a term frequency–CHI square and a term frequency–inverse document frequency was used to adapt feature selection.	Successfully adapt this method for the classification of various categories of Chinese texts	There is need to adjust the classifiers with new methods for better performance of text classification
7	SCADA intrusion detection scheme exploiting the fusion of modified decision tree and Chi-square feature selection (Ahakonye et al., 2023)	XGB, RF, GB, AD, CNN & LSTM	ICS-SCADA & WUSTL-IIoT-2018 datasets.	A Chi-square was used to obtain an optimal subset of features.	A modified decision tree-fused chi-square feature selection successfully used for intrusion detection in supervisory control and data acquisition (SCADA) networks.	Need an improved search strategy.
8.	A New Performance Metric to Evaluate Filter Feature Selection Methods in Text Classification (Çekik & Kaya, 2024).	SVM, KNN and NB	SMS data, Youtube Spam Collection, Sentiment140 dataset, Amazon Reviews and Email Dataset.	Distinguishing feature selector (DFS) was used to obtain an optimal subset of features.	Introduced a performance metric called Selection Error (SE) to determine the actual performance evaluation of filter methods.	The limitation of the SE is the change in the threshold value. Different outcomes can be obtained when the threshold value changes.

Recently, (X. Liu et al., 2023) adapts feature selection algorithms for the classification of Chinese texts using: a) an enhanced CHI square with mutual information (MI) algorithm named (CHMI), b) a term frequency–CHI square (TF–CHI) algorithm and c) a term frequency–inverse document frequency (TF–IDF) algorithm enhanced with the extreme gradient boosting (XGBoost) algorithm named (TF–XGBoost). The CHMI algorithm introduces word frequency and term adjustment; The TF–CHI algorithm enhances weight calculation; and TF–XGBoost which improves the algorithm's ability of word filtering. A modified decision tree-fused chi-square feature selection (Chi+MDT) for intrusion detection in SCADA networks proposed in (Ahakonye et al., 2023). The framework consists of data preprocessing stage in which irrelevant features were eliminated and organized into a coherent dataset. Afterward, a Chi-square was used for obtaining an optimal subset of data features by ranking. Finally, attack detection and classification is done using ML and DL classifiers. In addition, a performance metric called Selection Error (SE) to determine the actual performance evaluation of filter techniques introduce in (Çekik & Kaya, 2024). A distinguishing feature selector (DFS) was used to obtain an optimal subset of features. The limitation of the SE is the change in the threshold value in which different outcomes can be obtained when the threshold value changes. The analysis of related filter-based feature selection algorithms reviewed is tabulated in Table 2.1.

According to analysis made in Table 2.1, many filter-based feature selection approaches were proposed and the main limitations in existing studies are (a) high computational which lead to a decline in real-time detection accuracy (b) lack of efficient search strategy. Therefore, there is need for a feature selection method with a

better search strategy. However, despite the different filter method used in the reviewed papers, the majority of the papers used KNN or SVM as classification algorithm after feature selection.

2.2.1(d) Limitations of Filter-based

Filter-based are generally have lower performance compared to the wrapper-based feature selection method (Moradi & Gholampour, 2016), (Awadallah, Al-Betar, et al., 2022), (Ganesh et al., 2023), (Houssein et al., 2023). The filter approaches heavily rely on the relationship between the features and interactions between learning models and features have not been completely utilized (Houssein et al., 2023). Consequently, this restricts their classification while trying to find the best feature subset (Arora & Anand, 2019), (Chen et al., 2019), (Gangavarapu & Patil, 2019), (Wei et al., 2021), (Ganesh et al., 2023). The filter approaches evaluate features without considering how they relate to one another. This causes difficulty in identifying redundant features leading to low stability and poor performance (Rajamohana & Umamaheswari, 2018), (Arora & Anand, 2019), (Chen et al., 2019), (Çekik & Kaya, 2024), (Yin et al., 2023). When determining the subset of the features, these methods use statistical information only (Çekik & Kaya, 2024). Therefore, it fails to take the relevance of the various dimensions into account (Arora & Anand, 2019), (Chen et al., 2019).

2.2.2 Wrapper-based Feature Selection

A classifier is used to evaluate the feature subset performance in the wrapper method (Taradeh et al., 2019). Wrapper approaches include a learning algorithm as part of the evaluation function (Xue et al., 2012), (Xue et al., 2014), (Moradi & Gholampour, 2016). The learning tool included in the wrapper feature selection approaches is used to assess the selected subset of the available features during the search stage. Therefore, the wrapper-based approaches yield more promising

prediction results than the filter-based approaches (Xue et al., 2014), (Chen et al., 2019), (Xue et al., 2012), (Rostami et al., 2021). In other words, the wrapper-based feature selection technique utilizes a learning algorithm to evaluate the feature subset performance through an iterative search process such that the search process is guided by the outcomes of the learning algorithm at each iteration (Bahassine et al., 2020), (Wei et al., 2021). Therefore the inclusion of a learning algorithm in the wrapper-based method improves their prediction accuracy (Chen et al., 2019), (Xue et al., 2012), (Bahassine et al., 2020). However, by continuously applying the learning algorithm during the search process, these methods become computationally more expensive, especially for high-dimensional datasets (Bahassine et al., 2020). Generally, the wrapper-based methods can be classified into Greedy (sequential) and Random (global) search approaches (Chen et al., 2019), (Bahassine et al., 2020), (Wei et al., 2021).

2.2.2(a) Greedy (Sequential) Search

In the greedy search approach, the hill-climbing algorithm serves as the foundation for the search method, which adds or removes a single feature iteratively in a greedy manner (Chen et al., 2019),(Bahassine et al., 2020),(Wei et al., 2021). The two well-known greedy search methods are sequential forward selection (SFS) and sequential backward selection (SBS) (Moradi & Gholampour, 2016), (Chen et al., 2019), (Bahassine et al., 2020). The sequential forward search approach starts with an empty feature set and adds a feature to it at each step to improve the classifier's performance (Moradi & Gholampour, 2016), (Xue et al., 2014); until the performance of the classification cannot be further improved by the inclusion of a feature (Xue et al., 2014). However, the sequential backward search strategy starts with the entire collection of features and greedily removes one feature at a time based on the

classifier's performance (Moradi & Gholampour, 2016), (Xue et al., 2014); until the performance of the classification cannot be further improved by the removal of a feature (Xue et al., 2014). The local search is used in sequential search strategies rather than the global search, therefore these algorithms look for solutions that fall between the sub-optimal and near-optimal regions (Moradi & Gholampour, 2016), (Wei et al., 2021). Consequently, the sequential-based feature selection approaches still have several issues, including local optima stagnation and high computational costs (Moradi & Gholampour, 2016), (Chen et al., 2019), (Xue et al., 2012), (Wei et al., 2021). Therefore, creating a successful feature selection algorithm requires an effective global search method (Xue et al., 2012).

2.2.2(b) Global (Random) Search

Global search approaches, on the other hand, incorporate randomization into their search strategy to investigate a huge percentage of the solution space and better address feature selection problems (Moradi & Gholampour, 2016), (Chen et al., 2019). This implies that this search method uses global search in its search strategy to find the optimal feature subset within the range of available features (Chen et al., 2019). Meta-heuristic optimization algorithms have been used to solve feature subset selection problems due to their random search advantages and global search capability (Xue et al., 2012),(Wei et al., 2021). These optimization methods include the genetic algorithm (GA) (Rostami et al., 2021), particle swarm optimization (PSO) (Zhou et al., 2021), ant colony optimization (ACO) (Ghosh et al., 2020), artificial bee colony (ABC) (Subanya & Rajalaxmi, 2014), butterfly optimization algorithm (BOA) (Arora & Anand, 2019), etc. Due to their success in tackling the feature selection problem, metaheuristic algorithms have received a lot of attention recently (Wei et al., 2021).

However, wrapper-based approaches can usually achieve better results than filter approaches (Xue et al., 2012), (Xue et al., 2014). A wrapper-based approaches always employ the specified learning method during the search stage; thus, they are more expensive computationally, especially for complex classification problems (Chen et al., 2019). In addition, there is another class of feature selection method called embedded feature selection method in which the classification method as well as the learning process, is incorporated into the feature search process. (Kabir et al., 2011), (Chen et al., 2019), (Omuya et al., 2021). Hence, it is impossible to separate the processes of learning and feature selection (Omuya et al., 2021). However, since the interaction between feature selection and classification is the primary focus, this strategy performs similarly to the wrapper approach (Kabir et al., 2011).

2.2.2(c) Wrapper-based Method Related Work

However, metaheuristic algorithms in wrapper mode have received a lot of attention recently due to their success in tackling the feature selection problem. For instance, (García-Dominguez et al., 2020) utilized a genetic algorithm for a feature selection to reduce the original size of the dataset used in their proposed recognition model as an important aspect when working with the limited resources of a mobile device. The proposed model is designed for the recognition and classification of children's activities through automatic learning methods, optimized for application on mobile devices.

An improved discretization-based particle swarm optimization (PSO) for FS proposed in (Zhou et al., 2021). The authors applied a moderate pre-screening process to obtain a reduced size of features at first. Then, a ranking-based cut-point table that stores multiple cut-points sorted by an entropy-based cut-point priority for each feature was obtained. To find the optimal combination of the cut points that could best