ENHANCED FILTER-WRAPPER FEATURE SELECTION USING ENSEMBLE AND METAHEURISTIC APPROACHES FOR HIGH-DIMENSIONAL MULTI-CLASS TEXT CLASSIFICATION

IGE OLUWASEUN PETER

UNIVERSITI SAINS MALAYSIA

2025

ENHANCED FILTER-WRAPPER FEATURE SELECTION USING ENSEMBLE AND METAHEURISTIC APPROACHES FOR HIGH-DIMENSIONAL MULTI-CLASS TEXT CLASSIFICATION

by

IGE OLUWASEUN PETER

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

May 2025

ACKNOWLEDGEMENT

I acknowledge the almighty God for the privilege, insight, strength and provision to undergo this programme. I am also indebted to my supervisor, in person of Assoc. Prof. Dr. Keng Hoon Gan for her guidance, understanding and words of encouragement throughout the duration of the programme. Ma, I am grateful.

Words of month cannot be sufficient to appreciate the members of the panel, who have been painstakingly guiding my work from the beginning till this moment. Your wealth of experience and guidance have really sharpened my path.

I am also grateful to the Executive Secretary of Universal Basic Education Commission (UBEC), in person of Dr Hamid Bobboyi for approving my study leave. I thank the Director, Planning Research and Statistics, members of staff of the Department and ICT Unit in particular. I will specially appreciate Dr Ibrahim Bala Bakare for his support throughout the programme. I am grateful.

My unreserved appreciation goes to my children - Oluwaseun, Oluwaseyi and Oluwasola for their understanding and support during these periods. I thank my mother Deaconness R. O. Ige for her words of encouragement.

My unconditional gratitude goes to my “Ololufe”. She has always been there for me. I appreciate your ruggedness, courageous words and comfort at all times. I am indeed grateful.

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **ii**

TABLE OF CONTENTS ..... **iii**

LIST OF TABLES ..... **ix**

LIST OF FIGURES ..... **xii**

LIST OF ABBREVIATIONS ..... **xiv**

ABSTRAK ..... **xvi**

ABSTRACT ..... **xvii**

CHAPTER 1 INTRODUCTION **1**

1.1 Research Background 1

1.2 Research Motivation 5

1.3 Problem Statement 7

1.4 Research Objectives 10

1.5 Research Scope 11

1.6 Research Contribution 11

1.7 Thesis Outline 12

CHAPTER 2 LITERATURE REVIEW **13**

2.1 Introduction 13

2.2 Text Classification 13

2.3 High Dimension Multiclass Text 15

2.4 Dimensionality Reduction 16

2.5 Feature Selection 17

2.6 Ensemble Methods 21

2.7 Related works on filter, filter-wrapper and ensemble methods 23

2.7.1 Filter Methods 23

2.7.2 Filter-Wrapper Methods 26

2.7.3 Ensemble Methods 30
2.8 Metaheuristic Algorithms 33
2.8.1 Metaheuristic Algorithm Search (MHAS) 35
2.8.2 Neighborhood Selection Mechanism 37
2.8.3 Similarity Measures between two candidate solutions 38
2.8.4 Overview of Genetic Algorithm 40
2.8.4(a) Selection Mechanism in GA 41
2.8.4(b) Crossover operators in GA 42
2.8.4(c) Mutation operators in GA 43
2.9 The conventional ABC and its variants 43
2.9.1 The conventional ABC 44
2.9.2 The Variants of ABC 46
2.9.3 Existing Hybrid of ABC-GA 49
2.10 Gap Analysis 50
2.10.1 Why Ensemble Feature Selection is needed 50
2.10.2 Why ABC+GA is needed 51
2.10.3 Why are tournament and employed bee index needed 52
2.11 Chapter Summary 52
CHAPTER 3 METHODOLOGY 54
3.1 Introduction 54
3.2 Research Methodology 54
3.3 Phase 1: Data Preprocessing 55
3.3.1 Data Description 56
3.3.2 Data Transformation 60
3.4 Phase 2: Requirement 61
3.4.1 Feature Selection Problem Formulation 61
3.4.2 Fitness Function 62

3.5 Phase 3: Development 64

3.5.1 Multi-Univariate Ensemble Feature Selection Approach (MUNIFES) 64

3.5.2 Hybrid of MUNIFES with Artificial Bee Colony (ABC) and Genetic Algorithm (GA) – MUNIFES-ABC+GA 64

3.5.3 Enhanced MUNIFES-ABC+GA with improved neighborhood selection mechanism (employed bee index and Tournament) 65

3.6 Phase 4: Evaluation 65

3.6.1 Significance Test 66

3.7 Summary 67

CHAPTER 4 MULTI-UNIVARIATE FILTER ENSEMBLE FEATURE SELECTION (MUNIFES) FOR HIGH DIMENSIONAL DATASETS 68

4.1 Introduction 68

4.2 Multi-Univariate Filter Ensemble Feature Selection (MUNIFES) 70

4.2.1 Filter Feature Selection Methods 70

4.2.2 The Proposed Method - MUNIFES 72

4.2.2(a) Preprocessing 73

4.2.2(b) Concatenated Voting Ensemble with Weighting condition 74

4.2.2(c) MUNIFES working with features examples 77

4.2.2(d) Tackling Feature Redundancy in MUNIFES 79

4.2.2(e) Handling Multiclass issues in MUNIFES 82

4.2.2(f) Classification Methods (Classifiers) and its justification 83

4.2.2(g) Experiment Parameters 86

4.3 Experimental Results 86

4.3.1 Performance analysis of MUNIFES 87

4.3.1(a) Performance of MUNIFES with 20-Newsgroup 87

4.3.1(b) Evaluation of MUNIFES effectiveness on 20-Newsgroup 90

4.3.1(c) Performance of MUNIFES with 17-Newsgroup 91

4.3.1(d) Evaluation of MUNIFES effectiveness on 17-Newsgroup 93

4.3.1(e) Performance of MUNIFES with other methods on 20-Newsgroup 94

4.3.1(f) Comparison with RFEFS on 20-Newsgroup 96

4.3.1(g) Statistical Significance 99

4.3.1(h) Discussions 100

4.4 Summary 101

CHAPTER 5 HYBRID OF MULTI-UNIVARIATE FILTER ENSEMBLE FEATURE SELECTION (MUNIFES) WITH ARTIFICIAL BEE COLONY AND GENETIC ALGORITHM (ABC+GA) FOR FEATURE OPTIMIZATION ……..102

5.1 Introduction 102

5.2 Hybrid of Multi-Univariate Filter Ensemble Feature Selection (MUNIFES) with Artificial Bee Colony and Genetic Algorithm (MUNIFES-ABC+GA) 105

5.2.1 Phase I: Multi-Univariate filter ensemble feature selection (MUNIFES) approach 105

5.2.2 Phase II: Hybrid of Artificial Bee Colony and Genetic Algorithm (ABC+GA) 105

5.2.2(a) Feature Subset Generation 107

5.2.2(b) Problem Formulation 108

5.2.2(c) Evaluation Criteria 110

5.2.2(d) Fitness Criteria 111

5.2.2(e) Neighborhood Selection Mechanism 112

5.2.2(f) Experiment Parameter of ABC+GA 116

5.3 Experimental Results and Discussions 117

5.3.1 Performance analysis of MUNIFES-ABC+GA with 20 newsgroup 118

5.3.2 Performance analysis of MUNIFES-ABC+GA with 17 newsgroup 119

5.3.3 Performance based on the number of selected features for 20 newsgroup 120

5.3.4 Performance based on the number of selected features for 17 newsgroup 122

5.3.5 Performance of MUNIFES-ABC+GA with existing methods using 20 newsgroups 123

5.3.6 Performance of MUNIFES-ABC+GA with existing methods using 17 newsgroups 126

5.3.7 Statistically Significant of MUNIFES-ABC+GA for 20 newsgroup 128

5.3.8 Statistically Significant of MUNIFES-ABC+GA for 17 newsgroup 128

5.4 Summary 129

CHAPTER 6 ENHANCED HYBRID OF MULTI-UNIVARIATE FILTER ENSEMBLE FEATURE SELECTION (MUNIFES) WITH ARTIFICIAL BEE COLONY AND GENETIC ALGORITHM (ABC+GA) FOR FEATURE OPTIMIZATION 130

6.1 Introduction 130

6.2 Neighborhood Selection Mechanism 131

6.2.1 Selection Strategy 132

6.2.1(a) Tournament and Employed Bee Index selection in the onlooker bee phase 132

6.2.1(b) Justification of 50-50 probability split for tournament selection and employed bee index 135

6.2.1(c) Balancing Exploration and Exploitation 136

6.3 Experimental Results and Discussions 138

6.3.1 Performance analysis of enhanced MUNIFES-ABC+GA with 20 newsgroup 139

6.3.2 Performance analysis of enhanced MUNIFES-ABC+GA with 17 newsgroup 140

6.3.3 Performance based on the number of selected features for 20 newsgroups 141

6.3.4 Performance based on the number of selected features for 17 newsgroups 143

6.3.5 Performance of enhanced MUNIFES-ABC+GA with existing methods using 20 newsgroup 145

6.3.6 Performance of enhanced MUNIFES-ABC+GA with existing methods using 17 newsgroup 147

6.3.7 Statistically Significant of enhanced MUNIFES-ABC+GA for 20 newsgroup 149

6.3.8 Statistically Significant of enhanced MUNIFES-ABC+GA for 17 newsgroup 149

6.4 Summary 150

CHAPTER 7 CONCLUSION AND FUTURE RECOMMENDATIONS.... 151

7.1 Conclusion..... 151

7.2 Research Summary and Contributions..... 151

7.3 Future Work 155

REFERENCES..... 157

LIST OF PUBLICATIONS

LIST OF TABLES

Page

Table 2.1 Filter-based FS methods summary.....26

Table 2.2 Filter-wrapper methods summary ……..29

Table 2.3 Ensemble FS methods summary ……33

Table 3.1 20Newsgroup dataset 57

Table 3.2 17Newsgroup dataset 58

Table 3.3 Confusion Matrix 66

Table 4.1 10 features summary for the unique features 82

Table 4.2 10 features summary for the final filtered (MUNIFES) ……82

Table 4.3 Experimeter parameter of ANN ……86

Table 4.4 Performance of MUNIFES with 10 classifiers on F1 metric….……94

Table 4.5 Performance of Chi2 with 10 classifiers on F1 metric……...…95

Table 4.6 Performance of ANOVA with 10 classifiers on F1 metric……95

Table 4.7 Performance of Infogain with 10 classifiers on F1 metric…….……95

Table 4.8 Statistical Significant of MUNIFES with Infogain method -20 newsgroup……..99

Table 4.9 Statistical Significant of MUNIFES with Infogain method – 17 newsgroup……..99

Table 5.1 Experiment parameter of ABC+GA 117

Table 5.2 Percentage of Reduction in the number of selected features – 20 newsgroups.....121

Table 5.3 Percentage of Reduction in the number of selected features – 17 newsgroups.....123

Table 5.4a Performance of M-ABC+GA across classes with ABCFS on SVM classifier – 20 newsgroups 125

Table 5.4b Performance of M-ABC+GA across classes with ABCFS on SVM classifier – 20 newsgroups 125

Table 5.5a Performance of M-ABC+GA across classes with ABCFS on SVM classifier – 17 newsgroups 126

Table 5.5b Performance of M-ABC+GA across classes with ABCFS on SVM classifier – 17 newsgroups 126

Table 5.6 T-test comparison between M-ABC+GA and other algorithms – 20 newsgroup 128

Table 5.7 T-test comparison between M-ABC+GA and other algorithms – 17 newsgroup 128

Table 6.1 Percentage of Reduction in the number of selected features – 20 newsgroup (MUNIFES and E(M-ABC+GA)) 142

Table 6.2 Percentage of Reduction in the number of selected features – 20 newsgroup (MUNIFES-ABC+GA and E(M-ABC+GA)) 142

Table 6.3 Percentage of Reduction in the number of selected features – 17 newsgroup (MUNIFES and E(M-ABC+GA)) 142

Table 6.4 Percentage of Reduction in the number of selected features – 17 newsgroup (MUNIFES-ABC+GA and E(M-ABC+GA)) 142

Table 6.5a Performance of the enhanced(M-ABC+GA) across classes with ABCFS on SVM classifiers – 20 newsgroup 142

Table 6.5b Performance of the enhanced(M-ABC+GA) across classes with ABCFS on SVM classifiers – 20 newsgroup 142

Table 6.6a Performance of the enhanced(M-ABC+GA) across classes with ABCFS on SVM classifiers – 17 newsgroup 142

Table 6.6b Performance of the enhanced(M-ABC+GA) across classes with ABCFS on SVM classifiers – 17 newsgroup 142

Table 6.7 T-test Comparison between enhanced(MUNIFES-ABC+GA) and other algorithms – 20 newsgroup.....142

Table 6.8 T-test Comparison between enhanced(MUNIFES-ABC+GA) and other algorithms – 17 newsgroup.....142

LIST OF FIGURES

Page

Figure 2.1 Types of Text Classification 14

Figure 2.2 Dimensionality Reduction 17

Figure 2.3 Approaches to Filter Feature Selection 21

Figure 2.4 Homogenous Ensemble Feature Selection 22

Figure 2.5 Heterogenous Ensemble Feature Selection 23

Figure 2.6 Classification of Metaheuristic Algorithm 34

Figure 2.7 Summary of Genetic Algorithm Operators 41

Figure 2.8 Single Point Crossover Swapping 43

Figure 2.9 Multiple Points Crossover Swapping 43

Figure 3.1 Research Methodology 55

Figure 4.1 The proposed framework – MUNIFES 73

Figure 4.2 The Accuracy comparison between MUNIFES and base FS methods for 20-Newsgroup on 8 classifiers 89

Figure 4.3 The Accuracy comparison between MUNIFES and base FS methods for 20-Newsgroup on ANN and Ensemble classifiers 89

Figure 4.4 The Accuracy comparison between MUNIFES and base FS methods for 17-Newsgroup on 8 classifiers 93

Figure 4.5 The Accuracy comparison between MUNIFES and base FS methods for 17-Newsgroup on ANN and Ensemble classifiers 93

Figure 4.6 RTEFS (Fu et al., 2023) on 20-newsgroup using SVM classifier. 97

Figure 4.7 RTEFS (Fu et al., 2023) on 20-newsgroup using KNN classifier. 97

Figure 4.8 RTEFS (Fu et al., 2023) on 20-newsgroup using Ensemble classifier. 98

Figure 5.1 Schematic diagram of the proposed hybrid (MUNIFES-ABC+GA) for high-dimensional datasets. 107

Figure 5.2 Fitness score comparison between MUNIFES-ABC+GA and other metaheuristic algorithms for 20 newsgroup 118

Figure 5.3 Fitness score comparison between MUNIFES-ABC+GA and other metaheuristic algorithms for 17 newsgroup 119

Figure 5.4 Number of selected features comparison between MUNIFES-ABC+GA and MUNIFES on 20 newsgroup.....120

Figure 5.5 Number of selected features comparison between MUNIFES-ABC+GA and MUNIFES on 17 newsgroup.....122

Figure 6.1 Schematic diagram of the proposed hybrid e(MUNIFES-ABC+GA) for high-dimensional datasets. 131

Figure 6.2 Exploration and Exploitation in the e(MUNIFES-ABC+GA).....131

Figure 6.3 Fitness score comparison between e(MUNIFES-ABC+GA) and other metaheuristic algorithms for 20newsgroup.....131

Figure 6.4 Fitness score comparison between e(MUNIFES-ABC+GA) and other metaheuristic algorithms for 17newsgroup.....131

Figure 6.5 Number of selected features comparison between e(MUNIFES-ABC+GA) and MUNIFES on 20newsgroup. 131

Figure 6.6 Number of selected features comparison between e(MUNIFES-ABC+GA) and MUNIFES on 17newsgroup. 131

LIST OF ABBREVIATIONS

ADA	AdaBoost
ANOVA	Analysis of Variance
ACO	Ant Colony Optimization
ABC	Artificial Bee Colony
ANN	Artificial Neural Network
BA	Bat Algorithm
Chi2	Chi-square
CMFS	Comprehensively Measure Feature Selection
CFS	Correlation-based Feature Selection
DT	Decision Tree
DFSS	Discriminative Feature Selection
DFS	Distinguishing Feature Selector
FA	Firefly Algorithm
GA	Genetic Algorithm
GSA	Gravitational Search Algorithm
IG	Information Gain
KNN	K-Nearest Neighbour
LR	Logistic Regression
MMR	Max-Min Ratio
MHAS	Metaheuristic Algorithm Search
MHAs	Metaheuristic Algorithms
mRMR	Minimum Redundancy Maximum Relevance
MCDM	Multi-Criteria Decision Making
MLP	Multi-Layer Perceptron
NB	Multinomial Naïve Bayes
MUNIFES	Multi-Univariate Ensemble Feature Selection
MI	Mutual Information
NDM	Normalized Difference Measure
PSO	Particle Swarm Optimization
RF	Random Forest

RDC	Relative Discrimination Criterion
RTEFS	Re-ranking and TOPSIS-based ensemble Feature Selection
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
SI	Swarm Intelligence
SU	Symmetric Uncertainty
TV	Term Variance
TC	Text Classification
TCM	Trigonometric Comparison Measure

PEMILIHAN CIRI PENAPIS-PEMBALUT DIPERTINGKAT MENGGUNAKAN PENDEKATAN ENSEMBEL DAN METAHEURISTIK UNTUK PENGELASAN TEKS MULTI KELAS BERDIMENSI TINGGI

ABSTRAK

Kajian ini menangani cabaran dalam klasifikasi teks berbilang kelas berdimensi tinggi, di mana kaedah pemilihan ciri tradisional menghadapi masalah dengan optima setempat dan interaksi ciri. Walaupun kaedah penapis cekap, ia terdedah kepada bias dan ketidakstabilan. Pendekatan ansambel membantu tetapi mempunyai keterbatasan, yang membawa kepada keperluan teknik hibrid. Kajian ini memperkenalkan pendekatan penapis-pembungkus, menggabungkan Multi-Univariate Ensemble Feature Selection (MUNIFES) dengan pengoptimuman hibrid Artificial Bee Colony (ABC) dan Genetic Algorithm (GA). MUNIFES menggabungkan Chi-square, ANOVA, dan Information Gain dengan mekanisme pengundian berwajaran untuk memilih ciri yang diskriminatif dalam 10 pengelas. Kaedah pembungkus mengintegrasikan ABC dan GA dengan MUNIFES untuk meningkatkan pemilihan ciri menggunakan kadar mutasi yang disesuaikan oleh ANN, pemilihan kejohanan, dan pengindeksan lebah pekerja untuk keseimbangan penerokaan-pengeksploitasian. Eksperimen ke atas 20 Newsgroup dan 17 Newsgroup menunjukkan pengurangan lebih 50% ciri sambil mencapai ketepatan klasifikasi 96% dan 100%, mengatasi kaedah terkini. Keputusan ini mengesahkan keberkesanan pendekatan ini dalam mengekalkan kepelbagaian populasi dan meningkatkan prestasi klasifikasi teks.

ENHANCED FILTER-WRAPPER FEATURE SELECTION USING ENSEMBLE AND METAHEURISTIC APPROACHES FOR HIGH-DIMENSIONAL MULTI-CLASS TEXT CLASSIFICATION

ABSTRACT

This study addresses the challenge of high-dimensional multiclass text classification, where traditional feature selection struggles with local optima and feature interaction problems. While filter methods are efficient, they suffer from bias and instability. Ensemble approaches help but have limitations, prompting the need for hybrid techniques. The study introduces a filter-wrapper approach, combining Multi-Univariate Ensemble Feature Selection (MUNIFES) with a hybrid Artificial Bee Colony (ABC) and Genetic Algorithm (GA) optimization. MUNIFES combines Chi-square, ANOVA, and Information Gain multi-univariate features using weighted concatenated voting to select discriminative features across 10 classifiers. The wrapper method integrates ABC and GA with MUNIFES to enhance feature selection using ANN-adjusted mutation rates, tournament selection, and employed bee indexing for exploration-exploitation balance. Experiments on 20 Newsgroup and 17 Newsgroup datasets show over 50% feature reduction while achieving 96% and 100% classification accuracy, outperforming state-of-the-art methods. These results confirm the approach's effectiveness in preserving population diversity and enhancing text classification performance.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Emails, social media sites, and online libraries all produce a tremendous amount of digital data that is added to the Internet every second. The generated data may take the form of figures, text, audio, video, graphs, or other types of information. Such data can be used to extract knowledge of many kinds from a variety of disciplines, including financial, medical, statistical, and logical, among others, to obtain insights and make predictions (Garba Sharifai & Zainol, 2020). Most data that is accessible or available today is text-based (Demoulin & Coussement, 2020). Text-based data, also known as textual data, make up a sizable amount of the generated data. For instance, every month, Facebook and Twitter sites receive close to a billion messages and tweets. Additionally, Wikipedia saw more than a million edits in 2020. Text is unstructured data and makes it difficult to accurately retrieve people's desired informational needs due to the uncertainties in natural languages. Text mining is used to process and analyze such a huge amount of unstructured text data. To categorize texts from the domain that best describes the problem that needs to be solved, the mining process uses a variety of techniques (from the statistical to the artificial intelligence fields) (Mirończuk & Protasiewicz, 2018), (Mujtaba et al., 2019). This classification is known as Text Classification (TC).

The high dimensionality of the feature space brought on by many terms is one of the main issues with text classification. High-dimensional data refers to a dataset with a huge number of attributes in the sample space (Mamdouh Farghaly & Abd El-Hafeez, 2023), which will result in a time-consuming training process because the excessive number of attributes is related to the problem of Curse of Dimensionality.

Due to duplicated or irrelevant terms in the feature space, this issue may increase the computational complexity of machine learning techniques used for text categorization, as well as lead to inefficiency and outcomes with low accuracy (Jia et al., 2022).

Two methods are employed to address this issue: feature extraction and feature selection. Feature extraction is a technique of extracting new features into a separate feature space from the original features (Suhaidi et al., 2021). Several feature extraction techniques, including Principal Component Analysis (PCA) (Lam et al., 1999), (Selamat & Omatu, 2004), latent semantic indexing (Sun et al., 2004), clustering techniques (Slonim & Tishby, 2000), etc., have been employed successfully in text categorization. PCA has garnered a lot of interest among the many feature extraction techniques. It is a statistical method for reducing dimensionality which seeks to minimize variance loss in the original data. It can be thought of as a general-purpose feature extraction tool that is domain independent (Lam et al., 1999).

The feature selection is aimed at selecting only a small portion of the relevant features from the original large dataset to speed up the learning process and improve the performance of the learning model while reducing the classification model complexity and induction time.

In general, there are four different types of feature selection methods: filter, wrapper, embedded and ensemble (Abasabadi et al., 2021). In filter methods, the redundant and unrelated features are removed, and an optimal feature subset is selected. The filter methods do not apply classification algorithms to feature selection processes; therefore, these methods have low computational cost and typically show the best performance in case of high-dimensional datasets. Examples are Information Gain (IG), correlation-based feature selection (CFS), Fisher score, chi-square (chi2), Analysis of Variance (ANOVA) and minimum redundancy maximum relevance

(mRMR). On the contrary, the wrapper methods rely on the predictive performance of a classifier to evaluate the selected feature sets quality. The wrapper methods suffer from high computational cost; however, they finally provide better results compared to filter method. Examples are feature forward selection, backward feature elimination, and recursive feature elimination. On the other hand, the embedded methods admit interactions with the learning algorithm while making use of the internal information of the classification model to perform feature selection. They are more efficient than wrapper methods since they don't need to evaluate feature sets iteratively. Examples are L1 (LASSO) regularization and decision tree (DT). The foundation of ensemble methods is the idea that combining the results of various feature selection techniques is preferable to using the results of a single technique. Different feature selection techniques must be used, each producing an output, before the outputs of the individual models are combined to create a feature selection ensemble. The hybrid methods create a different approach by combining feature selection with other techniques, such as dimensionality reduction, transformation, or wrapper methods. It aims to leverage the strengths of various techniques to achieve better feature selection. While both ensemble and hybrid methods are data-driven, aiming to improve feature selection, they differ in some ways. Firstly, ensemble methods combine multiple feature selection algorithms to make collective decisions while hybrid methods incorporate feature selection with other methods. Secondly, ensemble frequently consider feature selection as a separate step that can be applied to any machine learning process while hybrid integrates feature selection within a specific modeling process. Thirdly, since ensemble methods consider feature selection as a separate step, they tend to be general and can be applied to a wide range of problems while hybrid, because of their integration of feature selection within a model, are more domain-specific and problem

dependent. Ensemble methods aggregate feature selection through voting, bagging, stacking, random forest, boosting (Mohammed & Kora, 2023), while hybrid combine filter and wrapper, wrapper and wrapper, wrapper and embedded methods to form the hybrid (Elemam & Elshrkawey, 2022). Both ensemble and hybrid will be used in the course of this work.

In addition to these feature selection techniques, in recent times, the use of metaheuristics algorithms (MHAs) for feature selection has gained wide popularity. They include Genetic Algorithm (GA) (Katoch et al., 2021), and the Ant Colony Optimization (ACO) algorithm (Mayet et al., 2023), Particle Swarm Optimization (PSO) (V. Kumar et al., 2022), Bat Algorithm (BA) (Al-Betar et al., 2020), Firefly Algorithm (FA) (Bazi et al., 2021), Gray Wolf Optimizer (GWO) (Alyasiri et al., 2021). MHAs are also inspired by physical phenomena, such as Gravitational Search Algorithm (GSA) (Y. Wang et al., 2021). Artificial Bee Colony (ABC) is one of the Swarm Intelligence (SI) algorithms, it was first proposed in 2005 (Karaboga et al., 2014). ABC models the foraging behavior of bee colonies. Since its discovery, it has drawn much attention for its simplicity and ability to solve real world optimization problems, such as quadratic assignment (Akay et al., 2021), sentiment classification (M. Zhang et al., 2021), and automatic programming (Nekoei et al., 2021). However, due to its solution search equation, ABC still suffers from certain drawbacks in solving complex problems. Several studies have shown that the search equation for a solution takes an emphasis on exploration over exploitation which results in slower convergence rates and low convergence accuracy (Bajer & Zorić, 2019; X. Zhou et al., 2021). With a view to addressing these problems, numerous improvements in the ABC were suggested using best individuals for improving exploitation. The ABC variants can produce better performance than the basic ABC, but they also face the difficulty

of how to appropriately utilize candidate solutions without sacrificing population diversity (X. Zhou et al., 2022). There is need to manage the neighborhood selection mechanism to balance between exploration and exploitation.

1.2 Research Motivation

When considering feature subsets, the intricate demands of individual classification algorithms often result in increased runtime. Consequently, filter-based feature selection methods are typically favored over wrappers and embedded methods. Filter-based FS methods are computationally efficient in identifying relevant features, various filtering techniques employ distinct strategies to determine optimal features; hence, the selection process varies in deterministic factors across methods. Relying solely on a single algorithm doesn't guarantee consistent performance across different experiments, a single filter method can achieve favorable results on a certain dataset but perform woefully on another dataset. Therefore, selecting an optimal filter method for a specific dataset will be a difficult task due to the inadequate a priori knowledge about the dataset (Dowlatshahi et al., 2018). Therefore, using a single filter method for feature selection involves conducting numerous attempts to select the appropriate filter method for a specific dataset. This problem leads to high computational complexity because many resources would have been utilized to arrive at a considerable result. Therefore, to ease this irregularity, various studies proposed ensemble methods to aggregate the output of multiple approaches to improve the classification performance of the learning model and decrease the bias of selected features (V. Kumar et al., 2022; Mohammed & Kora, 2023). There are two common approaches to selecting relevant features: Univariate and Multivariate filter feature selection. While the former evaluates the relationship between each individual feature

and the target variable without considering interactions between features, the latter evaluates the relevance of features collectively and capture feature interactions (Labani et al., 2020). Bivariate filter feature selection exists between the two approaches. Bivariate filter assesses the relevance of features by considering pairwise interactions between features and the target variable. Though univariate may not capture feature interaction but performs well on feature relevance to their target classes which has a positive influence on its prediction performance, we are capitalizing on its simplicity, speed, interpretable feature ranking and suitability to work with high-dimensional data while we use ensemble to complement this effort.

Bio-inspired algorithms are appropriate methods for solving complex optimization problems instead of seeking the optimal solution; they tend to search for near optimal solutions that could be found in a reasonable complexity (Jakšić et al., 2023). Two complementary tasks, exploration (diversification) in the search space and exploitation (intensification) of the best solutions found, are essential parts of any metaheuristic algorithm. Artificial Bee Colony (ABC), one of the swarm intelligence (SI) algorithms, is a powerful optimization method, which has strong search abilities to solve many optimization problems. It was first proposed in 2005 (Karaboga et al., 2014). It models the foraging behaviour of bee colonies. The reason for adopting ABC for feature selection with high dimensional problems, are its simplicity, adaptability and robustness (X. Zhou et al., 2022) : i) Simplicity: ABC is comparatively simple to implement, making it available for a wide range of optimization problems. ii) Adaptability: ABC is adaptable to any algorithm and can handle various optimization methods iii) Robustness: ABC can handle noisy and discontinuous objective functions better than some other optimization methods. Besides, ABC has two significant advantages: strong exploration and few control parameters (H. Wang et al., 2020).

However, ABC is not good at exploitation (intensification) and cannot obtain high accuracy solutions. This results in a slow convergence speed and low convergence accuracy (Ustun et al., 2022). Given this, a balance between exploration and exploitation is crucial for optimal performance. GA can enhance the exploitation (searching around good or promising solutions to ensure the best possible solution is found in that region) ability of ABC. This means GA helps refine promising solutions in more focused and detailed ways (Too & Abdullah, 2021). This is seen through GA's crossover and mutation strengths. The exploitation ability of ABC is enhanced by replacing the onlooker bee operator with the crossover phase of GA while we replaced the selection method with tournament and employed bee selections to balance exploration and exploitation. The proposed method uses the weights from a pretrained model of dataset as functional weights to determine the mutation rate used for global exploration and modified solution strategy for crossover used to boost local search for exploitation.

1.3 Problem Statement

High-dimensional data refers to datasets with a large number of features relative to the number of samples, making analysis and visualization challenging (Jia et al., 2022). In the 20 Newsgroups dataset, high dimensionality arises because each document is represented as a sparse vector of term frequencies (TF-IDF values) across thousands of unique words (features). With a vocabulary size of approximately 20,000, this creates a sparse feature space, necessitating effective feature selection techniques to extract relevant and non-redundant heterogeneous features for improved classification. Filter-based feature selection is commonly used to reduce dimensionality by evaluating each feature independently without considering feature

interactions. However, different filter methods apply varying criteria and heuristics, leading to inconsistencies in selected features. A single filter method often requires multiple attempts to determine the best approach for a given dataset, increasing computational complexity. To address this, ensemble filter methods aggregate multiple approaches, enhancing discriminative ability while reducing bias and instability (V. Kumar et al., 2022; Mohammed & Kora, 2023). Despite advancements, the use of a multi-univariate ensemble filter selection (MUNIFES) for high-dimensional multiclass datasets remains unexplored. Leveraging multiple univariate filter methods—such as Chi-square, ANOVA, and Information Gain—within an ensemble framework can enhance feature selection by integrating diverse ranking criteria. This approach ensures the selection of highly relevant, non-redundant heterogeneous features, aligning to improve classification performance while addressing the limitations of single-filter methods (Fu et al., 2023; Hashemi et al., 2021).

Although filter algorithms are computationally efficient in selecting a feature subset, they often suffer from the feature interaction problem, where the relevance of a feature depends on its relationship with other features. This limitation can lead to local optimum selection, affecting classification performance (Sharifai & Zainol, 2020). To address this, metaheuristic algorithms provide an effective solution by optimizing feature selection based on both relevance and interaction. Artificial Bee Colony (ABC) is well-suited for high-dimensional feature selection due to its simplicity, adaptability, and robustness (X. Zhou et al., 2022). As a Swarm Intelligence (SI) algorithm, ABC offers strong exploration capabilities with minimal control parameters, making it efficient for searching diverse feature subsets (H. Wang et al., 2020). Meanwhile, the Genetic Algorithm (GA) is effective in exploitation, leveraging

evolutionary operators to refine feature selection and enhance classification accuracy (Too & Abdullah, 2021). By combining ABC and GA within a hybrid filter-wrapper approach (MUNIFES-ABC+GA), this study aims to maximize classification accuracy while minimizing the number of selected features. The proposed method integrates MUNIFES, which selects relevant and non-redundant features, with the optimization power of ABC and GA, using a fitness function to ensure an optimal balance between exploration (diversity preservation) and exploitation (refinement of selected features). This approach directly addresses the feature interaction problem, ensuring a more robust and efficient feature selection process.

Another critical challenge is how to effectively utilize candidate solutions without compromising population diversity. While numerous Metaheuristic Search Algorithms have been developed to improve exploitation and exploration for more efficient search performance (Jain et al., 2019), many of these studies primarily focus on designing search operators or creating algorithm variants (Cheraghalipour et al., 2018; Han et al., 2018). However, research indicates that the success of a search process is not solely dependent on search operators (Cui et al., 2018; Piotrowski & Napiorkowski, 2018). Selection methods and search operators must complement each other (Meidani et al., 2022; Mirjalili et al., 2017). The effectiveness of a search process is determined by how well the search operators synchronize with selection methods in guiding candidate solutions through the search space (Ali et al., 2018). Since no single metaheuristic algorithm can universally solve all feature selection challenges (Piri et al., 2023), improving local and global search strategies is necessary to balance exploration and exploitation for better text classification (Meidani et al., 2022). To address this, the study proposes an enhanced hybrid algorithm (e-MUNIFES-ABC+GA), which integrates an improved neighborhood selection mechanism. This

approach employs an employed bee index to reinforce local search (exploitation) and tournament selection to maintain population diversity (exploration). By refining the search process, this method ensures that candidate solutions are effectively utilized, striking a balance between exploration (diverse solutions) and exploitation (converging on optimal features), ultimately enhancing the accuracy and efficiency of text classification.

1.4 Research Objectives

This research aims to develop filter-wrapper approaches for feature selection with high dimensional multi-class text classification to select the most relevant features that maximize the predictive ability of each class. The main objective is to develop an enhanced method to classify high-dimensional data and show that the proposed methods can perform better than the state of-the-art methods. The research objectives of this thesis are as follows:

(i) To enhance the filter-based method, discriminative ability through an ensemble of multiple univariate filter approaches (MUNIFES) to select relevant and non-redundant heterogenous features with the target classes.

(ii) To propose a hybrid filter-wrapper method (MUNIFES-ABC+GA) using MUNIFES and bioinspired algorithms (ABC+GA) with fitness function to maximize the accuracy while minimizing the number of selected features.

(iii) To further enhance the proposed algorithm (e-MUNIFES-ABC+GA) with an improved neighbourhood selection (employed bee index and Tournament) mechanism to boost local search to balance exploration and exploitation.

1.5 Research Scope

This research focuses on enhancing the filter and wrapper metaheuristic-based approaches for feature selection of High Dimensional Multiclass datasets. The proposed approaches have been assessed with high dimensional roughly balanced datasets in English Language. The effectiveness of the proposed approaches is evaluated in terms of accuracy, precision, recall, F1-score, fitness score and the average number of selected features. The results are compared with other known state-of-the-art methods.

1.6 Research Contribution

The following contributions are extracted from the research objectives:

(i) Enhancing the discriminative abilities of multiple univariate filter approaches through ensemble of heterogenous multi univariate filter feature selection to identify the relevant and non-redundant features with the target classes through concatenated weighted voting. The resulting method is known as Multi Univariate Filter Feature Selection (MUNIFES).

(ii) A proposed hybrid of filter-wrapper approach to address the partial feature interaction problem in filter methods, and poor exploitation problem in ABC to maximize accuracy and minimize the number of selected features. The resulting method is hybrid of MUNIFES and ABC+GA (MUNIFES-ABC+GA).

(iii) The MUNIFES-ABC+GA is enhanced with an improved neighbourhood selection (employed bee index and Tournament) mechanism to further deal with the premature convergence during local search and enhance balance between exploration and exploitation. The resulting method is enhanced MUNIFES-ABC+GA (e-MUNIFES-ABC+GA) to deal with multiclass high dimensional datasets.

1.7 Thesis Outline

This thesis is divided into seven chapters. It begins with an introduction Chapter that represents an overview of the research. This contains the background, motivation, problem statement, objectives, methodology, scope, and contribution.

Chapter 2 reviews the relevant literature. It started with text classification, the curse of dimensionality and how it could be resolved. This leads to feature selection and the need to combine methods to improve classification accuracy. Types of ensembles. The use of metaheuristic algorithm and its challenges and the need for hybrid algorithm. It finally summarises the entire literature and state the research gap that calls for attention.

Chapter 3 presents the research framework with explanation on each of the phases. Data Preprocessing entails data description and data transformation. Requirement entails feature selection problem formulation and the fitness function. It also discusses the Development and the Evaluation phases.

Chapters 4, 5 and 6 entail the development of MUNIFES, MUNIFES-ABC+GA and enhanced MUNIFES-ABC+GA and their evaluation together with the discussion of the findings.

Chapter 7 covers the summary of findings, conclusion, contribution to knowledge, and future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of the related important issues to this research. The literature review discussed the variety of research conducted in this area to back up the study and find the gaps in feature selection with a specific interest in text classification. This chapter introduces text classification, high dimension multiclass text, dimensionality reduction, feature selection and extraction, ensemble methods and metaheuristic algorithms.

2.2 Text Classification

The process of classifying a document into a predetermined category is known as text classification. Formally, text classification assigns one category, c_j, to a document d_i if d_i is a document from the whole set of documents D and $\{c_1, c_2, c_3,...,c_n\}$ is the set of all the categories (Rosquist, 2021). Documents may be assigned to one class or many classes depending on their attributes. A document is referred to as "single-label" if it is only assigned to one class, and it is referred to as "multi-label" if it is allocated to more than one class (Sajid et al., 2023). This is summarised in Figure 2.1. Most of the multi-label classification methods are of low accuracy and classify research articles into a limited number of categories. The classification of research articles into multiple categories with high accuracy is a challenging task (Bogatinovski et al., 2022). Usually, the Multi-label classification (MLC) task is confused with multi-class classification (MCC). In MCC, there are also multiple classes (labels) that a given document can belong to, but the document can belong to only one of these multiple classes. In that perspective, the MCC task can be seen as a special case of the MLC

task, where exactly one class (label) is relevant for each document (Bogatinovski et al., 2022). If only one of the two classes is assigned to the document, a "single-label" text classification problem can be further subdivided into a "binary class" problem, and if only N mutually exclusive classes are assigned to the document, it becomes a "multi-class" problem. The goal of text classification is to automatically categorize text documents into relevant classes or topics, making it easier to organize, search, and retrieve textual information. In single-label, each document is assigned to the most relevant category as in news articles, sentiment analysis, spam email detection while multi-label assigns each document to multiple categories as seen in assigning keywords to a research paper that can cover multiple subjects.

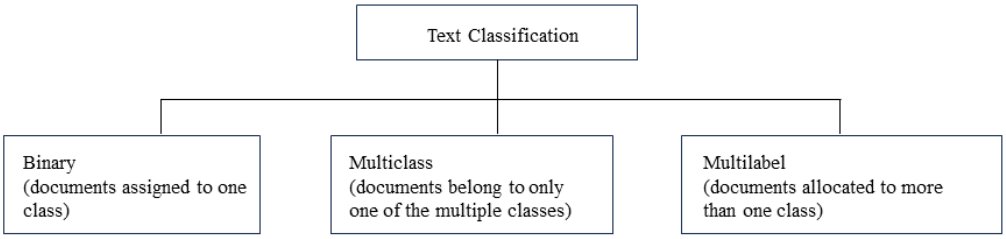

Figure 2.1 Types of Text Classification

The amount of knowledge available on the internet nowadays is enormous and growing quickly. Since the advent of digital documents, automatic text classification has been a significant application and research area for controlling the vast volume of material accessible on the web (Gasparetto et al., 2022). It is based on machine learning techniques, which create a classifier automatically by learning the traits of the categories from a group of articles that have already been classified. It is crucial for text retrieval, question-answering, and information extraction and summarization. Most of the time, heterogeneous data for classification is gathered from the web, newsgroups, message boards, broadcast or printed news, scientific articles, news

reports, movie reviews, and ads. Due to their multi-source nature, they frequently have drastically varying writing styles, preferred vocabulary, and formats, even for documents belonging to the same category. Automatic text classification is therefore crucial because of this heterogeneity and multiplicity. Another concept is the high dimension multiclass text which brings about complexity in text classification.

2.3 High Dimension Multiclass Text

High-dimensional multiclass text refers to text datasets with a large number of features, often represented by words, phrases, or n-grams, and involving classification into multiple categories. In natural language processing (NLP), each word or token becomes a feature, leading to thousands or millions of dimensions, while multiclass classification requires assigning each document to one of several possible categories, beyond simple binary classification. Multiclass classification is inherently more complex than binary classification (Akinola et al., 2022). Multiclass text classification poses several challenges due to the inherent complexity of textual data and the nature of multiclass problems. Some of these challenges are: high dimensionality, class imbalance, ambiguity and overlap of texts across classes, feature sparsity, semantic understanding, scalability and evaluation complexity. The effects of these challenges are increased computational complexity, risk of overfitting, algorithms biases towards dominant classes, higher misclassification rates, and decreased model accuracy. In multiclass problems, the algorithm needs to learn the relationships between many different classes, which increases the likelihood of classification errors. The presence of multiple classes requires sophisticated approaches to handle overlapping features across different classes, which can lead to class confusion (Santos et al., 2023). Addressing multiclass issues need different approaches and multiple considerations.

The next section discusses dimensionality reduction, as one of the ways of addressing the multiclass issues.

2.4 Dimensionality Reduction

Text data are easily available on the internet, in news articles, academic publications, messages on social media due to the advancement in digital technology. Most of these text data are unstructured, difficult to process and even require more storage to analyse (Adikari et al., 2021). Unstructured data remains a challenge in almost all data-intensive application fields such as businesses, universities, research institutions, government funding agencies, and technology-intensive companies (Baharudin et al., 2010). They are in the form of reports, email, views, news, etc. In the text classification task, several features represent a document, and a large collection of documents involves millions of features for each training instance, that is, the number of features defines the dimensionality for the dataset. The number of features is far larger than the number of samples. These huge input features cause the training to be extremely slow, difficult and unnecessarily occupy a large space. This problem is referred to as the curse of dimensionality. A general field of study called dimensionality reduction is concerned with solving the challenges posed by the high-dimensional nature of textual data. It transforms the set of features into a more compact form that will improve the learning algorithm's efficiency which can both work on supervised and unsupervised analytical methods. In addressing the curse of dimensionality, feature selection and feature extraction are often used. While feature extraction focuses on transforming the original document to a new document with reduced dimension, feature selection identifies a subset of the most informative

and discriminative features from the original high dimensional document. The focus of this work is on feature selection. Figure 2.2 shows the details.

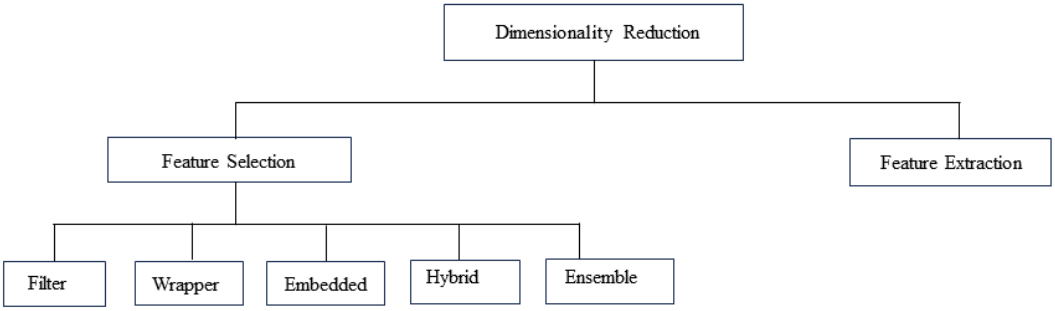

Figure 2.2 Dimensionality Reduction

2.5 Feature Selection

Feature Selection serves as a preprocessing technique for classification problems. This, in turn, performs decent text representation and minimizes computational time by reducing the overfitting and the error rate of the classifier to achieve a precise classification (Iqbal et al., 2020). This field has acquired the attention of many researchers for many years. In literature many techniques have been proposed for feature selection. There are three primary Feature Selection methods for Text Classification: the filter-based approach, the wrapper-based approach, and the embedded based approach (Alyasiri et al., 2022) while the other two (hybrid and ensemble) are derived from the primary methods.

The filter-based approach - This is based on ranking strategies independent of classifiers. It improves generality, speed, scalability and prevents overfitting, but neglects feature interdependency. It performs a statistical analysis over the feature space by sorting the dataset's features according to some univariate [e.g., Information Gain (IG)] or multivariate (e.g., Correlation-based Feature Selection (CFS)) approaches, then choosing the top-N features with the highest-ranking features. The variables that have been given scores according to the proper ranking criteria and the

variables with scores below the cutoff are eliminated. Due to their independence from the supervised learning method, filter-based approaches provide greater generality.

Either a univariate approach or a multivariate approach can be used to evaluate the features in a feature selection scheme. A univariate technique evaluates each feature separately, offers the feature's unique discriminatory qualities, and is quick, but it ignores any potential feature correlation. Contrarily, a multivariate technique evaluates the features while taking feature dependencies into account, however it is quite slow. Due to the large number of features that need to be processed, univariate filter algorithms are frequently utilized in studies on text classification (Abellana & Lao, 2023). The finest N characteristics are chosen while the others are discarded after the individual discriminatory abilities of the features are known. As a result, even though feature dependencies are disregarded, a compact subset of features is obtained through univariate technique. Chi square (chi2), Document Frequency (DF), information gain (infogain), and correlation coefficient scores are a few examples of univariate techniques while Inconsistency criterion, minimum Redundancy maximum Relevance (mRmR), Feature selection for sparse clustering are examples of multivariate techniques.

Wrapper method - Wrapper approaches use learning algorithm and training set heuristics that are more effective at defining optimal features than just relevant features. It views choosing a set of features as a search issue in which many combinations are created, assessed, and contrasted with one another. The technique employs a backward elimination approach to eliminate the subset's unimportant traits. To find the pertinent characteristic, some predetermined learning algorithm is required. The cross-validation method prevents feature overfitting. Due to the repeated learning processes and cross-validation, wrapper approaches are computationally

expensive and take more time than the filter method (Alyasiri et al., 2022), but they produce results that are more accurate than the filter model. As a result, the classification algorithm and the search feature subset interact, improving the feature subset selection process. Wrapper models allow for the acquisition of optimal characteristics as opposed to merely pertinent features. It also retains interdependence between features and feature subsets, which is an advantage. However, a wrapper-based strategy might produce superior outcomes than a filter-based approach if it uses an optimization algorithm as FS (Das et al., 2022). This is because the optimization process considers the classification accuracy and the number of selected features when evaluating the consistent features using an objective function. Recursive feature elimination, sequential (forward/backward) feature selection (SFS) algorithms, and genetic algorithms (GA) are examples of the wrapper-based method.

Embedded method - It uses algorithms with their own built-in feature selection methods; unlike wrapper, it has less computational cost, but it has an issue of difficulties in constructing a suitable optimization function. The method uses the supervised learning algorithm's learning process to try and choose features. The embedded approach can be divided into three groups: regularization models, built-in mechanisms, and pruning methods. In the pruning approach, the support vector machine (SVM) is used to recursively remove the features with lower correlation coefficient values after processing all of the data in the training phase to create the classification model [r]. The C4.5 and ID3 supervised learning algorithms' training phases are used in part to pick features in the built-in mechanism-based feature selection technique. The fitting error is minimized using objective functions in the regularization approach, and features with close to zero regression coefficients are removed (Alyasiri et al., 2022). This includes RIDGE, LASSO, and decision tree.

Ensemble method - The main idea is to blend the predictions of multiple algorithms to improve the overall feature selection process. It combines multiple feature selection models to make collective decisions about feature relevance (Khoder & Dornaika, 2021). Examples include voting, bagging, boosting, random forest.

Hybrid method - Hybrid combines feature selection with other techniques, such as dimensionality reduction or wrapper to leverage on the potencies of different techniques to achieve better feature selection (Abiodun et al., 2021). Examples include Filter-Wrapper hybrid or Wrapper-Embedded Hybrid.

While both ensemble and hybrid methods aim to improve feature selection by selecting most relevant features for a particular problem, they differ in mode of implementation. The former often handle feature selection as a separate step to be applied to any machine learning model hence general while the latter integrate feature selection within a specific modelling process hence domain specific and problem dependent.

There are various approaches to selecting relevant features from a dataset, including filter methods. Filter methods can also be categorized based on the dimensionality of the feature evaluation. Univariate, bivariate and multivariate (Ren et al., 2020). Univariate evaluates the relationship between each individual feature and target class without considering feature interaction, bivariate evaluates the relevance of features by considering pairwise interactions between features and the target variable while multivariate considers interactions and dependencies among multiple features concurrently. Univariates are faster, work well with high dimensional data, simpler to implement and interpret than both bivariate and multivariate that require more computational resources. Examples of the univariates are chi2, ANOVA, Infogain. Examples of bivariate are mutual information, t-test while the multivariate has Recursive Feature Elimination (RFE), L1 Regularization (Lasso), Principal

Component Analysis (PCA), Canonical Correlation Analysis (CCA) as examples. While parametric univariate assumes a specific distribution, non parametric do not make clearcut assumptions about data distribution. Greedy bivariate selects the most informative pair from each step based on a predefined criterion during evaluation while all paired captures all possible combinations of feature pairing and evaluate their relevance concurrently. Figure 2.3 has a breakdown of filter methods.

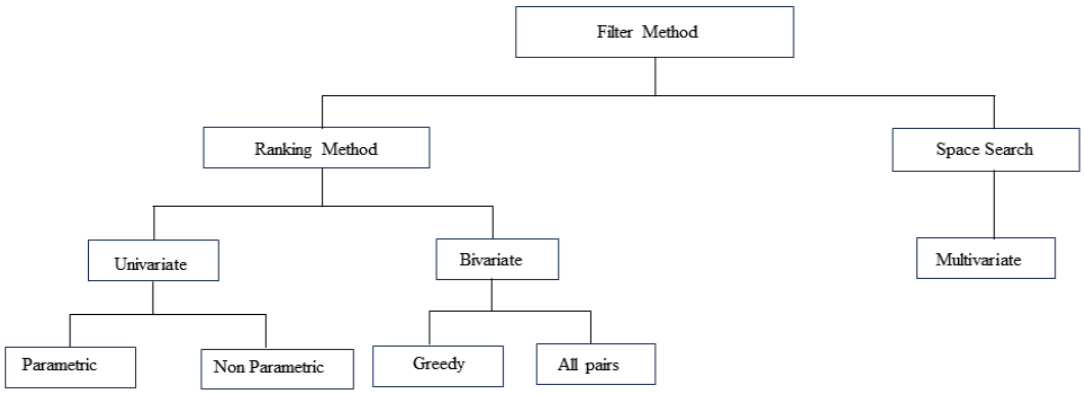

Figure 2.3 Approaches to Filter Feature Selection

What comes to mind is the type of ensemble to consider for harnessing the strength of various feature selection techniques. The subsequent section discusses the ensemble types and feature selection techniques and related works used to tackle feature interaction problem.

2.6 Ensemble Methods

Ensemble feature selection (FS) involves the integration of multiple models rather than relying on a single model. This approach can be broadly classified into two main groups based on the type of feature selectors employed: Homogenous and Heterogenous ensembles. In homogeneous ensembles, multiple subsets of the same type of feature selection technique or model are used. It leverages data diversity by

partitioning the dataset into multiple divisions. Within each division, a feature selection method is applied, which leads to the final feature subset (Figure 2.4). They are frequently employed to mitigate the risk of overfitting and strengthen the stability of feature selection processes. By applying the same feature selection method multiple times with different subsets of data, more robust selection of features can be obtained. Common methods are bagging and boosting.

The goal of heterogenous ensemble is to combine the strengths of different types of feature selection techniques or models. The function diversity in the ensemble arises from the use of different algorithms, or strategies for feature selection. They are designed to capture the complementary strengths of different methods and potentially achieve better feature selection results. Heterogeneous ensembles leverage diverse functions, employing multiple feature selection methods on the same dataset. The results from these feature selection methods are combined to identify the optimal subset of features (Figure 2.5). (Hashemi et al., 2022).

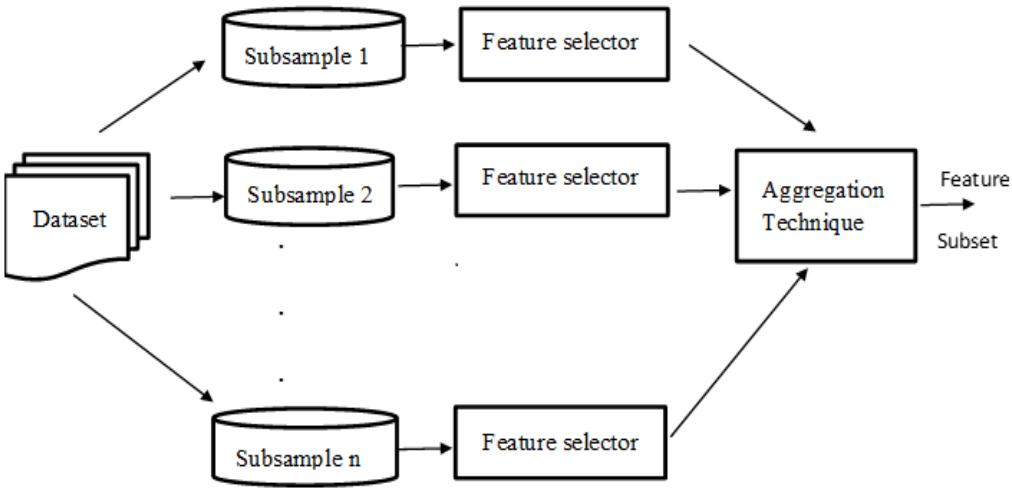

Figure 2.4 Homogenous Ensemble Feature Selection

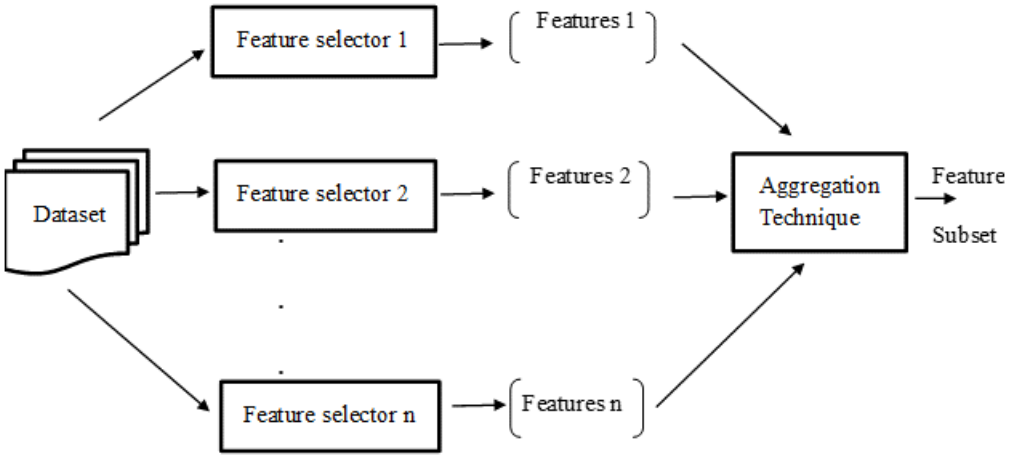

Figure 2.5 Heterogenous Ensemble Feature Selection

2.7 Related works on filter, filter-wrapper and ensemble methods

This section introduces related works on filter, filter-wrapper and ensemble methods that constitute the basis for our proposed work.

2.7.1 Filter Methods

Within the literature, numerous studies focus on filter-based feature selection methods. Notable among these are:

Correlation-based feature selection (CFS)

CFS (Hall, 1999) is an algorithm that couples evaluation formula with a suitable correlation measure and a heuristic search strategy. Experiments on artificial datasets showed that CFS quickly identifies and removes irrelevant, redundant, and noisy features, and classifies relevant features if their relevance does not strongly depend on other features.

Distinguishing Feature Selector (DFS)

DFS (Uysal & Gunal, 2012) selects distinctive features while removing uninformative ones bearing in mind certain requirements on term characteristics.

Experimental results shown that DFS offers a competitive performance with classification accuracy, dimension reduction rate, and processing time.

Comprehensively Measure Feature Selection (CMFS)

CMFS (J. Yang et al., 2012) comprehensively measures the significance of a term both in inter-category and intra-category as against other methods who only consider one category. The experimental results, comparing CMFS with six well-known feature selection algorithms, show that the proposed method CMFS is significantly superior to other methods of comparison.

Discriminative Feature Selection (DFSS)

DFSS (Zong et al., 2015), a novel feature selection method picks features having discriminative power in document first and later calculates the semantic similarity between features and documents. The results show a significant improvement over the state-of-the-art methods.

Relative Discrimination Criterion (RDC)

RDC (Rehman et al., 2015) introduces a new feature ranking metric which observes document frequencies for each term count while estimating the usefulness of a term. The result shows a significant contribution of RDC over other methods.

Normalized Difference Measure (NDM)

NDM (Javed & Babri, 2017) considers the relative document frequencies to prevent assigning equal ranks to terms having different difference. The performance of NDM is investigated against seven feature ranking metrics which recorded 66% cases in terms of macro-F1 measure and in 51% cases in terms of micro F1 measure.

Max–Min Ratio (MMR)

MMR (Rehman et al., 2018) selects smaller subsets of more relevant terms even in the presence of highly skewed classes, a product of max-min ratio of the true