A Meta-Analysis Survey on the Usage of Meta-Heuristic Algorithms for Feature Selection on High-Dimensional Datasets

LI YU YAB, NOORHANIZA WAHID, AND RAHAYU A. HAMID
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor 86400, Malaysia
Corresponding author: Noorhaniza Wahid (nhaniza@uthm.edu.my)
This work was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme under Grant FRGS/1/2018/ICT02/UTHM/02/6.

ABSTRACT Feature selection (FS) using meta-heuristic algorithms on high-dimensional datasets (HDD) is becoming more prevalent due to the continuous advancement in data mining. However, the difficulty in identifying the threshold of features in a dataset to be categorised as HDD remains an issue due to the different schools of thought on this matter. Therefore, this survey intended to determine the threshold for a number of features to be HDD, and subsequently identify the trend or potential FS method for HDD and the most preferred meta-heuristic algorithms and classifiers for both wrapper-based and filter-based FS methods to analyse HDD. This study performed an extensive systematic literature review by implementing the PRISMA guidelines on 62 research articles that were published between 2016 to 2021. This survey proposed a novel grouping technique called literal grouping and data grouping (LGDG) to accurately group the chosen articles based on HDD. The LGDG method serves as a guide for other researchers who intend to perform FS research related to HDD. Literal grouping refers to searching for selected papers using specific keywords, like HDD in this case. While data grouping compares the number of features in datasets towards the threshold, which is set at 2,000 features by the majority. Based on the analyses of all the LGDG groupings, the filter-based FS method gained more attention in recent years with competent results no less than wrapper-based, especially on HDD. Besides that, Moth Flame Optimisation works well in filter-based methods, whereas Cuckoo Optimisation Algorithm works well in wrapper-based, while Whale Optimisation Algorithm works well in both FS methods. As for the classifier’s preferences, SVM, DT, and NB are preferred by the filter-based, while KNN is preferred by the wrapper-based method. It can be recommended that reviewing other aspects such as multi-objective FS on HDD and including more FS methods could be included in future studies as an extension to this survey.

INDEX TERMS Feature selection, filter, wrapper, meta-heuristics, high-dimensional dataset.

I. INTRODUCTION
The expansion of science and technology in the present era has resulted in a tremendous increase in data size and dimensionality. A high-dimensional dataset (HDD) is a collection of data with a large number of features [1]. However, HDD leads to insurmountable memory restrictions, expensive training, and computation costs, resulting in the “curse of dimensionality” [1], [2]. Consequently, it is necessary to execute dimensionality reduction by adopting feature selection (FS) to improve classification performance. FS is a process of identifying the most significant features [3]. It can be classified into wrapper-based, embedded-based, and filter-based methods [3]. Each FS method has its advantages and weaknesses; therefore, researchers often integrate effective meta-heuristic algorithms to improve the performance. According to previous studies, the meta-heuristic optimisation algorithms can simplify optimisation problems [4], [5], classification [6], [7], and FS [8], [9]. Some examples of meta-heuristic algorithms include Ant Lion Optimiser (ALO) [10], Particle
Swarm Optimisation (PSO) [11], and Whale Optimisation Algorithm (WOA) [12]. Due to the presence of many available methodologies, this study conducted a systematic literature review of the most recent works on the subject from 2016 to 2021. This survey intends to determine the threshold of the number of features in a dataset in order to be categorised as a HDD. Besides that, this study also aims to identify the trend or potential FS method for HDDs, along with the most preferred meta-heuristic algorithms and classifiers when dealing with HDDs for both wrapper-based and filter-based FS methods. Several digital libraries and databases were used in this study to gather research articles, namely IEEE Xplore, ScienceDirect, Scopus, Springer, Taylor & Francis, Emerald Insight, and ACM.

The remaining sections of this systematic literature review are organised as follows. Section II provides an overview and the most important definitions used in FS using meta-heuristic algorithms on HDDs. Section III discusses the research questions and selection criteria. Section IV contains information about data extraction and analysis of chosen articles. Section V groups the selected articles into studies employing HDDs and discusses the scale of datasets, meta-heuristic algorithms used, FS methods applied, yearly publication growth, and classifier preferences. Section VI focuses on analysing HDDs by comparing the dataset distribution of each FS method. Section VII presents an overall discussion based on the content covered in the research questions. Finally, section VIII concludes and summarises the entire survey and provides suggestions for future work.

II. BACKGROUND

This section provides a concise summary of FS, meta-heuristic algorithms, and HDDs.

For the past decades, data mining has remained a hot research topic for researchers from various domains. Data mining is a broad field of data science that finds patterns and characteristics in massive amounts of data. It includes regression, clustering, detection of an anomaly, and classification [13]. Data classification entails assigning a class label of an instance based on a previously trained model [14]. In recent years, classification has relied heavily on FS which refers to the selection of the most meaningful inputs [15]. Omitting irrelevant and non-essential features in HDDs can also be defined as FS [16], [17], [18]. FS intends to reduce time complexity and increase predictive precisions [16], [17], [18]. Therefore, this data pre-processing step is very important to generate compact and quality datasets for classification purposes. FS aims to choose suitable features for the classification model to achieve higher accuracy.

Researchers often deal with different kinds of datasets in feature selection. Datasets are interpreted as a matrix, with rows representing the instances and columns representing the features [1]. Datasets with many features are categorised as HDDs. High dimensionality leads to unmanageable memory constraints and high training computing costs, called the “curse of dimensionality” [1], [2]. Thus, feature selection has two key competing goals: (1) optimising classification efficiency and (2) minimising feature numbers to solve the “curse of dimensionality” [19]. Moreover, feature selection is perceived as a multi-objective challenge to balance the trade-off between the two opposing priorities. Hence, dimensionality reduction needs to be performed to reduce the number of features without compromising the retrieval of useful information from HDD to ensure classification performance.

Wrapper-based, embedded-based, and filter-based are the three types of feature selection methods [3]. Wrapper-based feature selection uses the strength of base classifiers to determine the best features in a dataset. Contraction, embedded-based feature selection occurs during model training in the machine learning algorithm [20]. Both wrapper-based and embedded-based methods result in higher time complexity due to the intervention of classifiers in the feature selection process. Meanwhile, the filter-based feature selection method relies on the mutual information in a dataset. It ranks the features by generating a score for each without using the classification model [20]. Moreover, wrapper-based methods are computationally less friendly for HDDs. Embedded-based methods require predictive models, whereas filter methods can be combined with any predictive model to easily integrate HDDs [2]. Among the three methods, filter-based feature selection selects a subset of features without using any learning algorithm, thus, it is relatively faster than the wrapper-based method and is feasible in HDDs. Moreover, filter-based methods possess low complexity among the feature selection methods and are compatible with diverse datasets, including HDDs [20]. Since wrapper-based FS obtains high classification accuracy and filter-based FS maintains lower time complexity, this review only involves wrapper-based and filter-based FS methods. The study also investigated their performance when dealing with HDDs.

For decades, meta-heuristic optimisation algorithms have gained popularity in handling multi-objective FS issues due to the advantages of the two fundamental components, namely exploitation and exploration [21]. In exploration, the optimiser must contain operators to explore the search space globally, whereby motions should be random during this phase. Whereas, the promising section of the search space found during exploration is investigated thoroughly during exploitation [12]. Numerous studies employed meta-heuristic algorithms in FS to deal with HDDs, such as Ant Colony Optimisation (ACO) Algorithm [22], [23], [24], Ant Lion Optimisation Algorithm [25], Butterfly Optimisation Algorithm (BOA) [26], [27], [28], [29], [30], Cat Swarm Optimisation (CSO) Algorithm [31], [32], Chaotic Competitive Swarm Optimisation Algorithm [33], Coral Reefs Optimisation Algorithm [34], Coyote Optimisation Algorithm [35], Crow Search Algorithm (CSA) [36], Cuckoo Optimisation Algorithm (COA) [37], [38], Dragonfly Algorithm [39], [40], Firefly Algorithm [41], Frog Leaping Algorithm [42], Fruit Fly Algorithm [43], Grasshopper Optimisation Algorithm [31], [44], [45], [46], Grey Wolf Optimiser (GWO) [47], [48], [49], [50], [51], [52], [53], [54], [55], Harris Hawk Optimiser
L. Y. Yab et al.: Meta-Analysis Survey on the Usage of Meta-Heuristic Algorithms for FS on HDDs

(HHO) [56], [57], [58], [59], Hyena Optimisation Algorithm [60], Manta Ray Foraging Optimisation [61], Mayfly and Harmony Search Algorithm [62], Moth Flame Optimisation Algorithm (MFOA) [63], [64], Particle Swarm Optimisation Algorithm [65], [66], [67], [68], [69], [70], [71], Salp Swarm Algorithm (SSA) [72], [73], [74], Seagull Optimisation Algorithm (SOA) [75], Social Spider Optimisation (SSO) [76], and WOA [8], [9], [20], [53], [77], [78], [79]. These meta-heuristic algorithms obtain high classification accuracy using several classifiers such as K-Nearest Neighbor (KNN) [80], Decision Tree (DT) [81], Support Vector Machine (SVM) [82], Naïve Bayes (NB) [83], Random Forest (RF) [84], and Multilayer Perceptron (MLP) [85].

III. SYSTEMATIC LITERATURE REVIEW
The formulated research questions of this systematic literature review that needed to be answered are as follow:

1) What is the threshold for the number of features categorised as HDDs?
2) What is the current trend or potential FS method for HDDs?
3) When dealing with HDDs, which meta-heuristic algorithms do the researchers prefer for each FS method?
4) When dealing with HDDs, which classifiers do the researchers prefer for each FS method?

A. SEARCH PROCESS
Related research studies were retrieved from various databases available in the digital libraries of Universiti Tun Hussein Onn Malaysia (UTHM PTTA). The internet search was conducted by exploring advanced-search functions of the digital libraries and databases represented in Fig. 1. The keywords used include “feature selection,” “metaheuristic,” “nature-inspired,” “optimisation,” and “high dimension.” The Boolean operations ‘AND’ and ‘OR’ narrowed the search results with more accurate outcomes. For instance, the searching query used: “feature selection (in the title field) AND metaheuristic (in the abstract field) OR nature-inspired (in the abstract field) AND optimisation (in all metadata fields) AND high dimension (in all metadata fields).”

B. RESEARCH SELECTION CRITERIA
Since various publications on FS using meta-heuristics algorithms on HDDs were retrieved, the following inclusion and exclusion criteria were employed to ensure that the search was oriented and relevant.

1) INCLUSION CRITERIA
a) The research articles must be published between 2016 and 2021.
b) The research articles must be published in peer-reviewed journals.
c) The research articles must be written in English.
d) Only research articles written as technical papers are included.
e) The research articles with more than 1 dataset.

2) EXCLUSION CRITERIA
a) The research articles published before 2016.
b) The research articles unrelated to the search topic.
c) The research articles written in other languages.
d) The research articles written as survey/review papers.
e) The research articles with only 1 dataset.

C. DOCUMENT AND BIBLIOGRAPHY MANAGEMENT
This systematic literature review applies the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart [86] as depicted in Fig. 2. PRISMA is an evidence-based method for reporting systematic reviews and meta-analyses, which includes identifying resources, eligibility-checking resources, and indexing resources. Table 1 summarises the number of research articles discovered.
During the stage of identifying resources, 533 publications were identified in 7 databases. The number of resources obtained with specific keywords used in the searching query of advanced-search function were less but focused. This step ensured that the documents found were relevant to the search criteria, minimizing the number of less relevant or irrelevant records.

As for the stage of resource eligibility-checking, 433 publications were excluded after going through the abstract. Duplicate entries were also eliminated at this stage. Of the 100 eligible publications, 62 were chosen after reading the full-text content. Mendeley Desktop was used to organise and manage all the bibliographic details and references.

IV. EXTRACTION AND ANALYSIS OF DATA

This section summarises the content extraction of the 62 selected publications. The publication title, methodology used, related findings, and FS methods are tabulated in Table 2 and are sorted by most recent year first, followed by the publication title arranged alphabetically. The results from each research work in Table 2 were achieved using their own experimental settings. Each paper was denoted with a key number, as indicated in the rightmost column.

Based on Table 2, eight main issues motivated the studies. Those issues are summarised in Table 3. Although the issues are undertaken by each study varied, the findings were somehow similar, where most of the articles achieved their goals and proved that the implementation of proposed methods outperformed other methods through experimental results, either in terms of classification accuracy, number of selected features, execution time, convergence, or fitness values.

Over the years, the publications related to FS using meta-heuristic algorithms increased as depicted in Fig. 3. Besides that, based on the FS methods used in these 62 articles, the wrapper-based FS method was most preferred by researchers as it appeared in majority of articles, 80.6% (50 publications), while the filter-based FS method was only reported in 9 publications (14.5%), leaving the hybrid FS method the least preferable with 4.8% (3 publications).

From the 62 selected articles, 10 most active researchers were identified and tabulated in Table 4. They have published remarkable research articles in the related field with at least 3 publications from 2016 to 2021. These researchers contributed to the field of FS with meta-heuristic algorithms by constantly proving the competency of the proposed methods with new improvements each time. The number of publications was calculated by including the names of authors regardless of whether they are the first or co-authors.

V. STUDIES USING HDD

The purpose of this section is to identify true HDD. Not all the 62 selected articles listed in Table 2 dealt with HDD. Thus, a well-planned grouping technique was necessary to accurately perform the HDD analysis with ease. This study aims to introduce a novel grouping technique called literal grouping and data grouping (LGDG). This technique helps to group the selected papers into studies using HDD. The LGDG framework is depicted in Fig. 4.

A. LITERAL GROUPING

Literal grouping represents the searching of selected articles with HDD keywords. All the 62 selected articles were grouped using specific keywords. Those with the keywords were categorised as research articles employing HDD.

1) HDD KEYWORDS

The keywords used in searching for research article content were “HDD,” “high dimensional,” or “high-dimensional.”
| Publication Title                                                                 | Methodology                                                                 | FS Methods                  | Key |
|----------------------------------------------------------------------------------|-----------------------------------------------------------------------------|-----------------------------|-----|
| A hybrid feature selection method based on information theory and binary butterfly optimization algorithm [26] | Combined the MR-MNC1 method with a new version of S-bBOA.                    | Filter                      | K01 |
|                                                                                  | Produced better results in classification accuracy and the number of selected features in most datasets. The proposed method was more suitable for medical datasets with continuous-valued attributes. The computational time of IG-bBOA method was greater than SbBOA. |                             |     |
| A two-stage hybrid ant colony optimization for high-dimensional feature selection [22] | Used the interval strategy, features inherent relevance attributes, and the classification performance to guide optimum feature subset search. TSHFS-ACO is suitable for high-dimensional FS. The optimal FS has state-of-the-art performance on most datasets. TSHFS-ACO has a shorter running time compared with other ACO-based FS methods. | Wrapper                     | K02 |
| An analytical study of modified multi-objective Harris Hawk Optimizer towards medical data feature selection [56] | Multi-Objective Quadratic Binary HHO (MOQBIHHO) was proposed. Crowding Distance (CD) value was used to pick the best non-dominated solutions. 4 quadratic transfer functions mutated the continuous MOHHO to a binary form. Binary conversion used S-shaped transfer functions. The proposed method outperformed other suggested techniques in most datasets in terms of the number of selected features, classification accuracy, and computational time. | Wrapper                     | K03 |
| Feature selection by using chaotic cuckoo optimization algorithm with levy flight, opposition-based learning and disruption operator [37] | Combines Chaos Theory, COA, levy flight, opposition-based learning, and disruption operator. Increased exploration capability. Fast convergence. Obtained high accuracy. | Wrapper                     | K04 |
| Hybrid Binary Grey Wolf with Harris Hawks Optimizer for Feature Selection [57]   | A hybrid BGWO and HHO called HBGWOHKO was proposed. Binary conversion using S-shaped transfer functions. Wrapped with KNN classifier. The proposed hybrid method outperformed the original BGWO algorithm and other benchmark binary algorithms in terms of accuracy, selected feature size, and computational time. | Wrapper                     | K05 |
| Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach [77] | Combined filter and wrapper models using mutual information and KNN. Utilised hyperbolic tangent function to deal with discrete problems. The proposed algorithm outperformed other approaches in terms of the number of features selected and classification performance in datasets with lower class number. | Hybrid                      | K06 |
| Hybridization of Moth Flame Optimization Algorithm and quantum computing for gene selection in microarray data [63] | A Quantum Moth Flame Optimisation Algorithm (QMFoA) for gene selection was proposed. Phase 1: Measuring redundancy and relevance using mRMR. Phase 2: Hybridising MFOA with quantum computing and SVM, where the new update moth operation was integrated using Hamming distance and Archimedes spiral. | Filter                      | K07 |
| Immune Grey Wolf Optimization for Attribute Reduction: Application for Medical Systems [47] | Hybrid IGWO was proposed based on the strengths of GWO and Artificial Immune System (AIS). Three immune operators (selection, cloning, and mutation) were embedded. Alpha, Beta, and Delta wolves were cloned and mutated before selecting the first, second, and third best solutions. The proposed method outperformed other suggested techniques in terms of classification accuracy and computational time in most datasets. | Wrapper                     | K08 |
| Novel self-adapted particle Swarm Optimization Algorithm for feature selection [65] | Used a new learning model of particles to enhance their diversity. Adopted a one-flip neighborhood search strategy to strengthen the local search ability. The population replacement process was performed. The proposed method outperformed the other suggested techniques. Employing the one-flip neighborhood search strategy improved the local search ability of the swarm. Self-adjusted population replacement is beneficial to enhance the diversity of the swarm. | Wrapper                     | K09 |
| Solving feature selection problems by combining mutation and crossover operations with the monarch Butterfly Optimization Algorithm [27] | Monarch Butterfly Optimisation with enhanced crossover operator (MBOICO) and Lévy flight (MBOLF) was proposed, wrapped with KNN classifier. MBOICO possesses a high classification accuracy rate of 93% on average for all datasets and can significantly reduce selection. The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | Wrapper                     | K10 |
| Spatial bound Whale Optimization Algorithm: an efficient high-dimensional feature selection approach [78] | It uses a spatial bounding mechanism to regulate the dimensions of each search agent to improve their search capabilities. Another simplified version, S-SBWOA, was also used for fast computation. Obtained the highest accuracy by selecting the smallest number of features for most datasets. | Wrapper                     | K11 |
| S-shaped versus V-shaped transfer functions for binary Manta Ray Foraging Optimization in feature selection problem [61] | Manta Ray Foraging Optimization (MRFo) was proposed as a binary FS method. The binary conversion used V-shaped or S-shaped transfer functions. The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | Wrapper                     | K12 |
| Two-stage improved Grey Wolf Optimization algorithm for feature selection on high-dimensional classification [48] | MLP was trained using the proposed algorithm to pre-select features and optimise the hidden layer structure. Then, MLP was retrained using compressed datasets. Removed almost more than 95.7% of the features in all datasets to obtain high accuracy. | Wrapper                     | K13 |
| A dynamic locally multi-objective Salp Swarm | Used dynamic time-varying strategy and local fittest solutions to assist the SSA algorithm in balancing exploration and exploitation. | Wrapper                     | K14 |
| Algorithm for feature selection [72] | By avoiding local optima and fast convergence, the proposed method outperformed other multi-objective evolutionary algorithms. | Wrapper | K15 |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------|--------|-----|
| A hybrid grasshopper and new CSA algorithm for feature selection and optimization of multi-layer perceptron [31] | Hybridised SM-BGOA, GOA, and NCSOA with two penalty functions to generate an optimal MLP. SM-QNCSOA yielded better results on most datasets due to its capability to balance exploration, exploitation, and avoid local minima. | Wrapper | K16 |
| A wrapper-filter Feature Selection technique based on Ant Colony Optimization [23] | Filter and wrapper methods were used to measure heuristic desirability in ACO. The study used fitness-based memory and updated pheromone quantity using both accuracy and number of features. | Hybrid | K16 |
| The proposed method outperformed all other suggested techniques. | | | |
| An Asymmetric Chaotic, Competitive Swarm Optimization algorithm for Feature Selection in high dimensional data [33] | An ACCSO was proposed. ACCSO prefers zero to one based on the asymmetrical property of the proposed chaotic map. The fitness function depends on the number of selected features and classification error rate. | Wrapper | K17 |
| | The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | | |
| An Improved Harris Hawk Optimizer Algorithm with Simulated Annealing for Feature Selection in the Medical Field [58] | Used chaotic maps during the initialisation phase to improve the diversity of the solutions. Simulated Annealing (SA) algorithm with the HHH algorithm was used to enhance exploitation and avoid being stuck in local optima. | Wrapper | K18 |
| | The proposed method outperformed the other suggested techniques in most datasets. Chaotic Harris Hawks Optimization (CHHO) with the Sine map can significantly improve the performance of the standard HHO in terms of classification performance, the number of selected features, and convergence rates. Applying SA during the exploitation phase enhanced the local search. The proposed method was well-balanced between exploration and exploitation. | | |
| An intelligent Feature Selection approach based on Moth Flame Optimization for medical diagnosis [64] | Binary conversion by V-shaped or S-shaped transfer functions uses Levy flight operator to increase the diversity of the population. | Wrapper | K19 |
| | The proposed method outperformed the other suggested techniques by demonstrating remarkable effectiveness of Levy flight operator and transfer functions on the performance of MO. | | |
| Binary Coyote Optimization Algorithm for Feature Selection [35] | Developed based on a hyperbolic transfer function. The V-shaped transfer function was used in the binary conversion, wrapped with a Naive Bayes classifier. | Wrapper | K20 |
| | Demonstrated high precision, a smooth convergence curve, and the highest average training accuracy when the predictive accuracy criteria were used. Selected fewer features but with high computational cost. | | |
| Binary Multi-Objective Grey Wolf Optimization for Feature Selection in Classification [49] | A binary version of the Multi-objective GWO based on a sigmoid transfer function called BMOGWO was proposed. The binary version used S-shaped transfer functions, wrapped with the ANN classifier. BMOGWO-S can effectively determine a set of non-dominated solutions. The proposed method outperformed other suggested techniques in terms of feature reduction, classification error rate, and computational time in most datasets. | Wrapper | K21 |
| Bio-Inspired Feature Selection: An Improved Binary Particle Swarm Optimization Approach [66] | Introduces a local search factor based on the Lévy flight, a global search factor relying on weighting inertia coefficient, a population diversity improvement factor based on mutation, and a binary mechanism. | Wrapper | K22 |
| | The proposed method outperformed some of the other suggested techniques in most datasets. | | |
| Dynamic Butterfly Optimization Algorithm for Feature Selection [29] | Used local search algorithm based on mutation operator to avoid local optima and improve solution diversity. | Wrapper | K23 |
| | The proposed method outperformed the other suggested techniques in most datasets as it was able to select the most informative features and discard the irrelevant features. Capable of exploring unseen areas by other baseline algorithms. | | |
| Enhanced Crow Search Algorithm for Feature Selection [36] | An enhanced version of CSA (ECSA) was proposed. The dynamic local neighborhood guided the local search, where a novel global search strategy was implemented to increase the global exploration capability. | Wrapper | K24 |
| | The proposed method outperformed other suggested techniques in most datasets in terms of classification accuracy, fitness values, number of selected features, and computational time. | | |
| Improved Binary Grey Wolf Optimization and its application for Feature Selection [50] | Redefined the formula to increase linearly from 0 to 2. Variant with transfer functions, wrapped with KNN classifier. More accurate at classifying than the original variant. | Wrapper | K25 |
| Improved Harris Hawk Optimizer Using Elite Opposition-Based Learning and Novel Search Mechanism for Feature Selection [59] | Improved Harris Hawk Optimizer (IHHO) was proposed. The Elite Opposition-Based Learning (EOBL) strategy was applied to improve the population diversity and the exploration phase of HHO, to improve the accuracy and accelerate the convergence rate. Three Search Strategies (TSS), including mutation, mutation neighborhood search, and rollback strategies, were proposed to raise the search capabilities and avoid local optima. | Wrapper | K26 |
| | Fast convergence. The method was well-balanced between exploration and exploitation. The proposed method outperformed other suggested techniques in terms of classification accuracy, fitness value, number of selected features, and statistical tests. | | |
| Improved Salp Swarm Algorithm for feature selection [73] | Improved Salp Swarm Algorithm (ISSA) was proposed, where the inertia weight parameter was integrated, wrapped with KNN classifier. The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | Wrapper | K27 |
| | Mayfly-Harmony Search (MA-HS) was proposed based on Mayfly Algorithm and Harmony Search. Binary conversion used the S-shaped transfer functions. | | |
TABLE 2. (Continued.) Methodology, findings, and FS methods of the chosen research studies.

| Methodology, findings, and FS methods of the chosen research studies. | Methodology, findings, and FS methods of the chosen research studies. |
|---|---|
| Mayfly in Harmony: A New Hybrid Meta-Heuristic Feature Selection Algorithm [62] | The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. Fast convergence. |
| Quantum-like mutation-induced dragonfly-inspired optimization approach [39] | A quantum-behaved and Gaussian mutational dragonfly algorithm (QGDA) was proposed. The Quantum rotation gate is used by agents to shift position. The Gaussian mutation is for the strong local search capability. | Wrapper K29 |
| The monarch Butterfly Optimization Algorithm for solving Feature Selection problems [28] | Obtained high accuracy (overall accuracy rate of 88%) and well-performed feature reduction. |
| A Feature Selection method based on hybrid improved Binary Quantum Particle Swarm Optimization[67] | A feature selection method called hybrid improved binary quantum particle swarm optimization (H-II-BQPSSO) was proposed. Crossover and mutation were used to avoid local convergence. The weighted average principle was used in the fitness function. | Filter K31 |
| A filter-based bare-bone Particle Swarm optimization algorithm for unsupervised Feature Selection [68] | Used space reduction strategy based on average mutual information and local filter search strategy based on feature redundancy. | Wrapper K32 |
| A new binary Grasshopper Optimization Algorithm for Feature Selection problem [44] | Initialised grasshopper locations with binary values before updating them with simple operators. Wrapped with 5-Nearest-Neighbor (5-NN). | Wrapper K33 |
| A new hybrid algorithm based on Grey Wolf Optimization and Crow Search Algorithm for Unconstrained Function Optimization and Feature Selection [51] | A hybrid GWO with the CSA called GWOCSA was proposed. CSA and GWO, respectively, handle exploration and exploitation. The alpha updating direction only updates individuals in the population to act as a shrinking strategy to escape from the local optimum. Adaptive balance probability was used to achieve acceleration throughout the early steps of the optimization process. Non-linear control parameter (a) was used to effectively explore the search space. | Wrapper K34 |
| A new hybrid Seagull Optimization Algorithm for Feature Selection [75] | The iteration process uses a roulette wheel to randomly select an algorithm for a location update. TEO algorithm's location update formula was used following the SOA iteration. Used the TEO algorithm's heat exchange formula to improve the predation mode of SOA. SOA-TEO3 outperformed all suggested techniques, with better exploitation ability, reduced execution time, and well-balanced exploitation and exploration. |
| An opposition-based Social Spider Optimization for Feature Selection [76] | Used Opposition-Based Learning (OBL) strategy, in which the OBSSO rules selected the best fitness value solutions. | Wrapper K36 |
| BIFFOA: A Novel Binary Improved Fruit Fly Algorithm for Feature Selection [43] | A binary improved fruit fly optimization algorithm (BIFFOA) was proposed. 4 different strategies were employed based on evolutionary population dynamics (EPD) and new mutation operators to enhance the BIFFOA. | Wrapper K37 |
| Binary Butterfly Optimization approaches for Feature Selection [30] | Binary conversion using V-shaped or S-shaped transfer functions, wrapped with KNN classifier. | Wrapper K38 |
| Binary Grasshopper Optimization Algorithm approaches for Feature Selection problems [45] | Binary conversion using V-shaped or S-shaped transfer functions. A mutation operator was used to improve exploratory behaviour. | Wrapper K39 |
| Binary optimization using hybrid Grey Wolf Optimization for Feature Selection [52] | Binary conversion using V-shaped or S-shaped transfer functions, wrapped with KNN classifier. | Wrapper K40 |
| Efficient Feature Selection method using real-valued Grasshopper Optimization Algorithm [46] | Used statistical measures during iterations to replace the duplicate features with the most promising features. The proposed method was compared with BGA, ACO, SA, and PSO in terms of classification accuracy, feature subset size, fitness value, and convergence rate. | Wrapper K41 |
| Efficient hybrid nature-inspired binary optimizers for Feature Selection [53] | Three hybrids of GWO and WOA (HSGW, RSGW, and ASGWO) were proposed. Hybrid serial GWO-WOA (HSGW) used operators of GWO to reposition the solution; the fittest was used as the leader for WOA. Random Switcher GWO-WOA (RSGW) used a random number [0,1] to employ either GWO or WOA. While the Adaptive Switcher GWO-WOA (ASGWO) used an adaptive mechanism to control the selection process of operators from the GWO or WOA to avoid local optima. | Wrapper K42 |
TABLE 2. (Continued.) Methodology, findings, and FS methods of the chosen research studies.

| Methodology, findings, and FS methods of the chosen research studies. | Filter | Wrapper |
|---|---|---|
| Frequency-based Feature Selection method using Whale Algorithm [20] | Eliminated half of the unimportant attributes by ranking and arranging the important attributes using Mutual Congestion using forward FS and majority voting function. | Filter K43 |
| Hybrid Binary Coral Reefs Optimization algorithm with Simulated Annealing for Feature Selection in high-dimensional biomedical datasets [34] | Obtained higher average accuracy in classification compared to the original WOA and Mutual Congestion alone. | Wrapper K44 |
| Hybrid self-inertia weight adaptive Particle Swarm Optimization with a local search using C4.5 decision tree classifier for Feature Selection problems [69] | The tournament selection strategy increased the diversity of the initial population of individuals. BCROSAT can obtain the highest accuracy and selects the smallest number of features for most datasets. | Wrapper K45 |
| ISSA based on Particle Swarm Optimization for Feature Selection [74] | An improved self-adaptive, inertia weighted PSO with local search combined with C4.5 classifiers for the FS algorithm was proposed. The gradient-based local search handles the searching process. The improved self-adaptive inertia weight PSO handles convergence to a solution. The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | Filter K46 |
| Spotted Hyena Optimization Algorithm with simulated annealing for Feature Selection [60] | SA is embedded in the spotted hyena optimization (SHO) algorithm to enhance the optimal solution found by SHO. SA enhances the final solution obtained by SHO. SHOSA-1 outperformed all other suggested techniques. The average accuracy of the SHOSA-1 algorithm on 20 datasets increased by 9.82%, with an average reduction of 4.64 features and an average fitness improvement of 9.22%. | Wrapper K47 |
| Whale Optimization Algorithm for high-dimensional small-instance Feature Selection [79] | Binary conversion using V-shaped or S-shaped transfer functions, wrapped with KNN classifier. V-shaped functions outperformed the S-shaped functions and other suggested techniques in most datasets in terms of classification accuracy, the number of selected features, and fitness values. | Wrapper K48 |
| A novel hybrid algorithm for Feature Selection [87] | Used maximum Spearman and minimum covariance strategy. It selected the optimal feature subset according to the probability relationship. Fast convergence, better classification accuracy. | Hybrid K49 |
| An improved Feature Selection algorithm based on graph clustering and Ant Colony Optimization [24] | Features were split into multiple clusters in the entire feature space represented as a graph by an efficient community detection algorithm. MGACAO achieved superior performance over the GCACO by reducing the number of features and maintaining the maximum classification accuracy. | Filter K50 |
| Binary Dragonfly Optimization for Feature Selection using time-varying transfer functions [40] | Binary conversion using V-shaped or S-shaped transfer functions, wrapped with KNN classifier. The time-varying S-shaped BDA techniques improved exploration and exploitation while performing FS tasks. Outperformed all other techniques evaluated. | Wrapper K51 |
| Cuckoo inspired algorithms for Feature Selection in heart disease prediction [38] | The cuckoo search algorithm (CSA) and COA were used. Both algorithms were implemented in filter-based FS. Classification performance was compared with other meta-heuristic algorithms. SVM classifier yielded the best results. CSA outperformed other suggested techniques in terms of selected features and classification accuracy in most datasets. | Filter K52 |
| Feature Selection for Optimized High-Dimensional Biomedical Data Using an Improved Shuffled Frog Leaping Algorithm [42] | The improved shuffled frog leaping algorithm (ISFLA) was proposed where the chaos memory weight factor was used to balance exploration and exploitation. The absolute balance group strategy was used to sort individuals based on fitness value. The adaptive transfer factor was used to convert it to a binary form. Binary conversion using S-shaped transfer functions. | Wrapper K53 |
| Hybrid Binary Bat Enhanced Particle Swarm Optimization algorithm for solving Feature Selection problems [70] | The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. Hybrid Binary Bat Enhanced Particle Swarm Optimization (HBBEPSO) algorithm was proposed. BAT echolocation for exploring and enhanced PSO for converging to the best solution. | Wrapper K54 |
| Large-dimensionality small-instance set Feature Selection: A hybrid bio-inspired heuristic approach [54] | A hybrid antlion-grey wolf algorithm (ALO-GWO) was proposed. Roulette wheel selection of agents and adaptive size of the random walks was used, wrapped with KNN classifier. Well-balanced between exploration and exploitation. Avoided premature convergence. The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | Wraper K55 |
| Whale Optimization Algorithm approaches for wrapper Feature Selection [9] | The proposed method used the Tournament and Roulette Wheel selection mechanisms in the search phase. Crossover and mutation operators were utilised to improve the WOA algorithm’s exploitation. Efficient in rapidly searching for optimal/near-optimal subsets. Crossover and mutation variants outperformed other techniques. Fast convergence, well-balanced between exploration and exploitation, avoid local optimal, high accuracy rate. Used the return-cost attractiveness. Selection based on Pareto dominance. Binary movement by using the adaptive jump. | Wrapper K56 |
| Large-dimensionality small-instance set Feature Selection: A hybrid bio-inspired heuristic approach [54] | A hybrid antlion-grey wolf algorithm (ALO-GWO) was proposed. Roulette wheel selection of agents and adaptive size of the random walks was used, wrapped with KNN classifier. Well-balanced between exploration and exploitation. Avoided premature convergence. The proposed method outperformed other suggested techniques in terms of the number of selected features and classification accuracy in most datasets. | Wraper K55 |
| Whale Optimization Algorithm approaches for wrapper Feature Selection [9] | The proposed method used the Tournament and Roulette Wheel selection mechanisms in the search phase. Crossover and mutation operators were utilised to improve the WOA algorithm’s exploitation. Efficient in rapidly searching for optimal/near-optimal subsets. Crossover and mutation variants outperformed other techniques. Fast convergence, well-balanced between exploration and exploitation, avoid local optimal, high accuracy rate. Used the return-cost attractiveness. Selection based on Pareto dominance. Binary movement by using the adaptive jump. | Wrapper K56 |
The Boolean operator ‘OR’ was used to join or exclude terms from a search, resulting in more targeted and productive results. It also reduced the time and effort by lowering the number of irrelevant matches. The keywords searched are case insensitive, hence, the search results were not affected by character case. The 3 fields where the keywords were used include title, issue, and dataset description. For instance, K02’s research title was "A two-stage hybrid ACO for high-dimensional feature selection," and its issue was to select the optimum feature subset in HDDs and avoid local optimum. The datasets used in K02 were described as 11 high-dimensional low-sample datasets. Since the 3 fields of title, issue, and dataset descriptions contained the searched keywords, it was categorised as a study utilising HDD. A complete list of studies utilising HDD based on literal grouping is discussed in the following subsection.

2) RESEARCH STUDIES BASED ON LITERAL GROUPING
Of the 62 selected articles, 19 fell under research topics concerning HDD based on literal grouping. The meta-heuristic algorithm, FS methods, datasets, the average number of features, and classifiers used in these 19 articles are listed in Table 5. The number of datasets used ranged from 6 to 30, while the average number of features ranged between 39 to 10,408 features. As for 10 of the articles, the dataset was retrieved from the UC Irvine Machine Learning Repository (UCI ML) [88], 2 from the Arizona State University repository (ASU) [89], 1 from Kent Ridge Bio-medical Dataset (KRBD), and 7 from unspecified sources.

The top 3 studies with the highest average number of features (Table 5) implemented the Cuckoo Optimisation Algorithm (K04), MFOA (K07), and WOA (K48). These 3 algorithms can perform FS in the HDD based on literal grouping. Besides that, GWO (K13, K40, K55), PSO

---

**TABLE 2.** (Continued.) Methodology, findings, and FS methods of the chosen research studies.

| Methodology, findings, and FS methods of the chosen research studies. | Methodology, findings, and FS methods of the chosen research studies. |
|---|---|
| A return-cost-based binary firefly algorithm for Feature Selection [41]. | A Pareto-dominance-based selection technique may help reduce resource wasting. The proposed method was well-balanced between the return and cost. The adaptive jump technique utilised linear and exponential methods to alter the jump probability. It can solve the FS challenges. |
| Hybrid Whale Optimization Algorithm with simulated annealing for Feature Selection [8]. | Hybridised with SA as a local search operator to amplify nearby areas with the best solution identified in each WOA iteration. Used Tournament Selection mechanism to choose the search agents. |
| Binary Ant Lion approaches for Feature Selection [25] | Binary conversion using V-shaped or S-shaped transfer functions, wrapped with KNN classifier. |
| Binary Grey Wolf Optimization approaches for Feature Selection [55] | Achieved the required population diversity. Well-balanced between exploration and exploitation. Avoided premature convergence. V-shaped functions outperformed the S-shaped functions. |
| Feature Selection is based on an improved Cat Swarm Optimization algorithm for big data classification [32] | A crossover operator was used to generate candidate solutions. Modified the position-changing method by making sure that the size of the change and the current position were negatively correlated. |
| Investigation on Particle Swarm Optimization for Feature Selection on high dimensional data: local search and selection bias [71] | The proposed ICOS yielded fast convergence and outperformed traditional CSO in classification accuracy. The proposed method resulted in a greater number of selected features. Using term frequency-inverse document frequency (TF-IDF) with ICOS for FS yielded better accuracy than TF-IDF alone. |
| The Boolean operator 'OR' was used to join or exclude terms from a search, resulting in more targeted and productive results. It also reduced the time and effort by lowering the number of irrelevant matches. The keywords searched are case insensitive, hence, the search results were not affected by character case. The 3 fields where the keywords were used include title, issue, and dataset description. For instance, K02’s research title was “A two-stage hybrid ACO for high-dimensional feature selection,” and its issue was to select the optimum feature subset in HDDs and avoid local optimum. The datasets used in K02 were described as 11 high-dimensional low-sample datasets. Since the 3 fields of title, issue, and dataset descriptions contained the searched keywords, it was categorised as a study utilising HDD. A complete list of studies utilising HDD based on literal grouping is discussed in the following subsection. |

---

**TABLE 3.** Main issues that motivated the 62 articles.

| Issues | No. of research | List of research |
|---|---|---|
| To tackle the FS search issue in binary space | 18 | K03, K05, K12, K15, K19-K21, K25, K28, K33, K38-K40, K44, K48, K51, K59, and K60. |
| To perform multi-objective FSs (to increase classification accuracy and to decrease the number of selected features) | 18 | K01, K04, K06, K10, K13, K14, K16, K21, K22, K27, K31, K37, K38, K45, K48-K50, and K52. |
| To select the optimum feature subset or reduce unnecessary and redundant features | 16 | K01, K03, K04, K07, K09, K11, K12, K15, K20, K32, K36, K41, K47, K54, K58, and K60. |
| To improve or balance the exploration and exploitation phases of the meta-heuristic algorithms used | 11 | K05, K07, K19, K24, K28, K29, K32, K37, K46, K58, and K60. |
| To select the optimum feature subset in HDDs or simply to tackle the "curse of dimensionality" | 9 | K02, K11, K13, K17, K31, K44, K53, K55, and K62. |
| To avoid local optimum | 9 | K02, K04, K05, K18, K23, K29, K36, and K62. |
| To prevent premature convergence or tackle slow convergence speed | 9 | K08, K10, K24, K27, K29, K32, K34, K42, and K57. |
| To improve the diversity of solutions | 4 | K08, K18, K23, and K26. |
TABLE 4. Active researchers and the number of their publications from the selected 62 articles.

| Authors            | Published between 2016-2021 |
|--------------------|-----------------------------|
| SEYEDALI MIRJALILI | 11                          |
| MAJDI MAFARJA      | 8                           |
| IBRAHIM ALJARAH    | 5                           |
| HOSSAM FARIS       | 4                           |
| MOHAMED ABD ELAZIZ | 4                           |
| ALI ASGHAR HEIDARI | 3                           |
| HOSSAM M. ZAWBAAN  | 3                           |
| MOHAMMAD TUBISHAT  | 3                           |
| QASEM AL-TASHI     | 3                           |
| RAM SARKAR         | 3                           |

FIGURE 5. Year of publication for all 62 selected articles based on literal grouping.

FIGURE 6. Classifier preference for the 19 HDD research articles based on literal grouping.

(K31, K40, K62), HHO (K03, K26), and WOA (K11, K48) were the other most commonly used algorithms (Table 5).

As for the FS methods for the 19 articles based on literal grouping, 84.2% of the articles (16 publications) employed the wrapper-based FS methods while 15.8% (3 publications) used the filter-based FS methods.

From 2016 to 2021, the publications on FS using meta-heuristic algorithms for filter-based and wrapper-based FS methods, specifically those using HDD based on literal grouping manifested an increasing trend as depicted in Fig. 5.

According to Fig. 6, the classifier preference of the 19 HDD research articles based on literal grouping indicated that the filter-based FS methods were more diversified as they can be integrated with SVM (42.9%), RF (14.3%), NB (28.6%), and KNN (14.3%) classifiers, whereas the wrapper-based only used KNN (93.8%) and MLP (6.3%).

In the next section, data grouping is performed to ensure that the journal articles dealt with actual HDD. Although not all the journal articles mentioned HDD research, the size of their dataset was described as HDD. Thus, to categorise these articles as HDD study-related, the threshold of the number of features for a dataset was identified using the minimum HDD feature numbers in all 19 literal-grouped HDD articles.

B. DATA GROUPING

Data grouping identifies articles that are related to HDD even though the articles did not state that the datasets used were HDD, by comparing the number of features in the datasets towards the threshold obtained from literal grouping. For example, if most of the HDD articles from literal grouping have at least \( x \) features in the datasets used, thus, \( x \) is used as the threshold of the minimum feature number. Therefore, the articles that satisfied the threshold were categorised as studies that employed HDD based on data grouping.

The first step in data grouping was determining the thresholds used in all 19 literal-grouped HDD journal articles identified in the previous subsection. In other words, the threshold must be obtained to perform data grouping for all the 62 selected articles.

1) HDD THRESHOLD

The reason to adopt the threshold for the number of features in HDD research is the existence of different schools of thought among the researchers listed in Table 5. Although 19 articles were identified as HDD studies by literal grouping, the dataset dimension varied. For instance, K40 treated the dataset with at least 30 features as HDD, while K62 views the dataset with at least 2308 features as HDD. Therefore, it is advisable to determine the threshold for the number of features adopted in these studies. In this section, the method used to identify the threshold was relatively straightforward, where the 19 literal-grouped HDD papers were read and the minimum number of features used in the datasets were recorded, as depicted in Table 6.

According to Table 6, the minimum features for all datasets used in certain cases were different from the minimum features for HDD, such as K03, K14, K17, K26, K28, K31, K40, K41, and K55. The variation is due to the mixed usage of both HDD and non-HDD in the studies. For instance, K40 had 12 datasets with a minimum of 9 features. However, K03 also specified that only 2 datasets were HDD, where the minimum number of features for the 2 HDDs was 2000. Thus, the common threshold for the number of features to be categorised as HDD is 2000 based on K03. The minimum number of features for HDD are used as thresholds representing each study’s point of view.

According to Table 6, the minimum features for all datasets used in certain cases were different from the minimum features for HDD, such as K03, K14, K17, K26, K28, K31, K40, K41, and K55. The variation is due to the mixed usage of both HDD and non-HDD in the studies. For instance, K40 had 12 datasets with a minimum of 9 features. However, K03 also specified that only 2 datasets were HDD, where the minimum number of features for the 2 HDDs was 2000. Thus, the common threshold for the number of features to be categorised as HDD is 2000 based on K03. The minimum number of features for HDD are used as thresholds representing each study’s point of view.

According to Table 6, there are 8 out of the 19 HDD articles (42.11%), having datasets with at least 2000 features acknowledged as HDD. Another 5 articles (26.32%) mentioned that the datasets with at least 2308 features are categorised as HDD. Meanwhile, only 2 articles (10.53%) required 1024 features like the number of thresholds. The remaining articles were the minority with only 617, 325, 265, and 30 features in their HDD studies. Therefore, the thresholds from the 19 literal-grouped HDD articles...
TABLE 5. General detail of 19 HDD research studies based on literal grouping.

| Key | Meta-heuristic algorithm | FS Methods | Datasets | Average no. of features | Classifiers |
|-----|--------------------------|------------|----------|-------------------------|-------------|
| K01 | IG-bBOA                 | Filter     | 6 HDDs from the UCI ML repository. | 21,847 / 6 3,641 | NB, SVM, RF |
| K02 | TSHFS-ACO               | Wrapper    | 11 high-dimensional low sample datasets. | 91,984 / 11 8,362 | KNN         |
| K03 | MQBHII                  | Wrapper    | 10 benchmark datasets from the UCI ML repository and 2 HDDs. | 5,520 / 12 460 | KNN         |
| K04 | CCOALFDQ                | Wrapper    | 20 HDDs. | 208,160 / 20 10,408 | KNN         |
| K07 | QMFOA                   | Filter     | 13 high-dimensional microarray gene datasets. | 122,986 / 13 9,460 | SVM         |
| K11 | SBWOA                   | Wrapper    | 16 HDDs from Arizona State University. | 119,161 / 16 7,448 | KNN         |
| K13 | IGWO                    | Wrapper    | 10 HDDs. | 81,984 / 10 8,198 | MLP         |
| K14 | MODISSA-lbest           | Wrapper    | 13 benchmark datasets from the UCI ML repository. (3 HDDs). | 23,795 / 13 1,830 | KNN         |
| K17 | ACCSSO                  | Wrapper    | 12 benchmark datasets from UCI and ASU repository (3 HDDs). | 56,702 / 12 4,725 | KNN         |
| K26 | IHHO                    | Wrapper    | 20 benchmark datasets from the UCI ML repository. (8 low, 6 medium, 6 HDDs). | 21,710 / 20 1,086 | KNN         |
| K28 | MA-HS                   | Wrapper    | 18 benchmark datasets from the UCI ML repository and 3 HDDs. | 19,698 / 21 938 | KNN         |
| K31 | HL-BQPSO                | Filter     | 9 gene expression HDDs and 36 UCI datasets | 73,810 / 45 1,640 | NB, SVM, KNN |
| K40 | BGWOPSO                 | Wrapper    | 18 benchmark datasets from the UCI ML repository. (6 HDDs). | 696 / 18 39 | KNN         |
| K41 | GOFS                    | Wrapper    | 10 benchmark datasets from the UCI ML repository. (7 low-dimensional and 3 HDDs) | 11,617 / 10 1,162 | KNN         |
| K44 | BCROSAT                 | Wrapper    | 13 HDDs. | 91,976 / 13 7,075 | KNN         |
| K48 | WOA                     | Wrapper    | 9 high-dimensional low sample datasets. | 78,667 / 9 8,741 | KNN         |
| K53 | ISFLA                   | Wrapper    | 9 HDDs from KRBD. | 62,828 / 9 6,981 | KNN         |
| K55 | ALO-GWO                 | Wrapper    | 18 benchmark datasets from the UCI ML repository and 12 HDDs. | 158,255 / 30 5,275 | KNN         |
| K62 | PSO-LSRG                | Wrapper    | 10 gene expression HDDs | 79,984 / 10 7,998 | KNN         |

* Bold text represents the top 3 datasets with the highest average number of features

TABLE 6. Identifying the threshold for the number of features in 19 HDD research articles based on literal grouping.

| Key | Number of all datasets | Minimum no. of features for all datasets | Number of HDD | Minimum no. of features for HDD | Threshold used |
|-----|------------------------|-----------------------------------------|---------------|---------------------------------|---------------|
| K01 | 6                      | 325                                     | 6             | 325                             | 325           |
| K02 | 11                     | 2,308                                    | 11            | 2,308                           | 2,308         |
| K03 | 12                     | 9                                       | 2             | 2,000                           | 2,000         |
| K04 | 20                     | 265                                     | 20            | 265                             | 265           |
| K07 | 13                     | 2,000                                   | 13            | 2,000                           | 2,000         |
| K11 | 16                     | 1,024                                   | 16            | 1,024                           | 1,024         |
| K13 | 10                     | 2,308                                   | 10            | 2,308                           | 2,308         |
| K14 | 13                     | 3                                       | 13            | 2,000                           | 2,000         |
| K17 | 12                     | 4                                       | 8             | 2,000                           | 2,000         |
| K26 | 20                     | 9                                       | 6             | 617                             | 617           |
| K28 | 21                     | 9                                       | 3             | 2,308                           | 2,308         |
| K31 | 45                     | 3                                       | 9             | 2,000                           | 2,000         |
| K40 | 18                     | 9                                       | 6             | 30                              | 30            |
| K41 | 10                     | 9                                       | 3             | 2,000                           | 2,000         |
| K44 | 13                     | 2,000                                   | 13            | 2,000                           | 2,000         |
| K48 | 9                      | 2,308                                   | 9             | 2,308                           | 2,308         |
| K53 | 9                      | 2,000                                   | 9             | 2,000                           | 2,000         |
| K55 | 30                     | 9                                       | 12            | 1,024                           | 1,024         |
| K62 | 10                     | 2,308                                   | 10            | 2,308                           | 2,308         |

concluded that the threshold for the number of features is at least 2000 features, and this grouping process is called data grouping.

The threshold of 2000 is used across this review article while analysing whether the research studies covered are HDD research based on data grouping. For instance, since
K02 has more than 2000 features in the datasets, it is categorised as a study employing HDD based on data grouping. The 3 situations to apply data grouping on all 62 selected articles are:

a) Data grouping (1-match),
b) Data grouping (mean-match), and
c) Data grouping (all-match).

A complete list of research articles using HDD based on the 3 situations of data grouping is discussed in the following subsections.

2) RESEARCH STUDIES BASED ON DATA GROUPING (1-MATCH)

Data grouping (1-match) was applied to all the 62 selected articles. The 1-match refers to the grouping technique of selecting articles that employed HDD if and only if they have at least 1 dataset that matched the threshold of 2000 features. This technique holds the highest chance of categorising more articles as HDD studies. For instance, if a study consists of 3 datasets (500, 1200, and 2300 features) and obtains a 1-match (the dataset with 2300 features hits the threshold), it can be categorised as a HDD study as it has at least 1 dataset matching the threshold (2300 > 2000).

Of the 62 articles, 26 were identified as journal articles utilising HDD based on data grouping (1-match). The meta-heuristic algorithm, FS methods, datasets, and the average number of features and classifiers used in these 26 articles are summarised in Table 7. The number of datasets used ranged from 4 to 30, while the average number of features was from 384 to 12,852 on average. Whereas, 15 of the 26 articles retrieved the datasets from the UCI ML repository, 3 from the ASU repository, 1 from the KRBD repository, 1 from the Kaggle repository [90], 1 from the Keel repository [91], and 9 from unspecified sources.

The top 3 research articles with the highest average number of features from Table 7 implemented WOA (K43), Cuckoo Optimisation Algorithm (K04), and MFOA (K07). These algorithms can perform well in FS for HDD based on 1-match data grouping. Furthermore, PSO (K31, K32, K45, K62), GWO (K08, K13, K55), and WOA (K11, K43, K48) were the preferable algorithms according to Table 7. Several other meta-heuristic algorithms that appeared more than once were Butterfly Optimisation (K01, K23), HHO (K03, K26), MFOA (K07, K19), and Grasshopper Optimisation Algorithm (K33, K41).

Besides, for the FS methods used in the 26 publications, a majority of articles employed the wrapper-based FS methods (84.6%, 22 publications), while the filter-based FS methods were only employed in 4 publications (15.4%) based on data grouping (1-match) for the 26 HDD studies.

Publications using meta-heuristic algorithms for filter-based and wrapper-based FS methods, specifically those under HDD research studies based on data grouping (1-match) for the years 2016 to 2021 are increasing trends based on Fig. 7.

As for the classifier preferences of the 26 HDD studies based on data grouping (1-match) (Fig. 8), the filter-based FS methods indicated higher adaptability as they can integrate with 5 well-known classifiers namely SVM (40%), RF (10%), NB (30%), KNN (10%), and DT (10%). Meanwhile, the wrapper-based FS methods had only 3 classifier preferences, including KNN (86.95%), DT (8.7%), and MLP (4.35%).

3) STUDIES BASED ON DATA GROUPING (MEAN-MATCH)

Data grouping (mean-match) was applied to all the 62 selected journal articles in this subsection. Mean-match refers to the grouping technique of selecting research articles that used HDD if and only if their average number of features used in datasets match the threshold of 2000 features. For instance, if a research study consists of 3 datasets (with 4000, 2000, and 1500 features) and obtains a mean-match (7500/3 = 2500), it can be categorised as HDD research since its average number of features is greater than the threshold of 2000 features (2500 >= 2000).

Of the 62 publications, 14 were identified as research using HDD based on data grouping (mean-match). The meta-heuristic algorithm, FS methods, datasets, the average number of features, and classifiers used in these 14 publications are summarised in Table 8. Based on Table 8, on average, the number of datasets used ranged from 4 to 30, while the number of features ranged from 3,550 to 12,852. Four of the 14 publications retrieved their datasets from the UCI ML repository, 2 from the ASU repository, 1 from the KRBD repository, and 8 from unspecified sources.

The top 3 studies with the highest average number of features (Table 8) implemented WOA (K43), Cuckoo Optimisation Algorithm (K04), and MFOA (K07). These 3 algorithms...
TABLE 7. General details of the 26 HDD research articles based on data grouping (1-match).

| Key | Meta-heuristic algorithm | FS Methods | Datasets | Average no. of features | Classifiers |
|-----|--------------------------|------------|----------|-------------------------|-------------|
| K01 | IG-bBOA | Filter | 6 HDDs from the UCI ML repository. | = 21,847 / 6 = 3,641 | RF, NB, SVM |
| K02 | TSHFS-ACO | Wrapper | 11 high-dimensional low sample datasets. | = 91,984 / 11 = 8,362 | KNN |
| K03 | MOQBIHIO | Wrapper | 10 benchmark datasets from the UCI ML repository and 2 HDDs. | = 5,520 / 12 = 460 | KNN |
| K04 | CCOALFDO | Wrapper | 20 HDDs. | = 208,160 / 20 = 10,408 | KNN |
| K07 | QMFPA | Filter | 13 high-dimensional microarray gene datasets. | = 122,986 / 13 = 9,460 | SVM |
| K08 | IGWO | Wrapper | 10 benchmark medical datasets. | = 15,200 / 10 = 1,520 | KNN |
| K11 | SBWOA | Wrapper | 16 HDDs from ASU repository. | = 119,161 / 16 = 7,448 | KNN |
| K13 | IGWO | Wrapper | 10 HDDs. | = 81,984 / 10 = 8,198 | MLP |
| K14 | MODSSA-ibest | Wrapper | 13 benchmark datasets from the UCI ML repository. (3 HDDs). | = 23,795 / 13 = 1,830 | KNN |
| K17 | ACCSO | Wrapper | 12 benchmark datasets from UCI and ASU repository (3 HDDs). | = 56,702 / 12 = 4,725 | KNN |
| K19 | LBMFO | Wrapper | 23 benchmark datasets from UCI, Keel, Kaggle data repositories. | = 16,223 / 23 = 705 | KNN |
| K23 | DBOA | Wrapper | 20 benchmark datasets from the UCI ML repository. | = 7,674 / 20 = 384 | KNN |
| K26 | IHHO | Wrapper | 20 benchmark datasets from the UCI ML repository. (8 low, 6 medium, 6 HDDs). | = 21,710 / 20 = 1,086 | KNN |
| K28 | MA-HS | Wrapper | 18 benchmark datasets from the UCI ML repository and 3 high-dimensional microarray datasets. | = 19,698 / 21 = 938 | KNN |
| K31 | HI-BQPSO | Filter | 9 gene expression HDDs and 36 UCI datasets | = 73,810 / 45 = 1,640 | SVM, NB, KNN |
| K32 | FBPSO | Wrapper | 2 benchmark UCI datasets and 4 benchmark ASU datasets. | = 21,300 / 6 = 3,550 | KNN, DT |
| K33 | NBGOA | Wrapper | 20 benchmark datasets from the UCI ML repository. | = 11,077 / 20 = 554 | KNN |
| K37 | BIIFFOA | Wrapper | 25 benchmark datasets from the UCI ML repository. | = 10,365 / 25 = 415 | KNN |
| K41 | GOF | Wrapper | 10 benchmark datasets from the UCI ML repository. (7 low-dimensional and 3 HDDs) | = 11,617 / 10 = 1,162 | KNN |
| K43 | WOA-MC | Filter | 4 benchmark binary medical datasets. | = 51,405 / 4 = 12,852 | DT, SVM, NB |
| K44 | BCROSAT | Wrapper | 13 HDDs. | = 91,976 / 13 = 7,075 | KNN |
| K45 | SIW-APSO-LS | Wrapper | 16 benchmark datasets from the UCI ML repository. | = 6,294 / 16 = 393 | DT |
| K48 | WOA | Wrapper | 9 high-dimensional low sample datasets. | = 78,667 / 9 = 8,741 | KNN |
| K55 | ISFLA | Wrapper | 9 HDDs from KRBD. | = 62,828 / 9 = 6,981 | KNN |
| K55 | ALO-GWO | Wrapper | 18 benchmark datasets from the UCI ML repository and 12 HDDs. | = 158,255 / 30 = 5,275 | KNN |
| K62 | PSO-LSRG | Wrapper | 10 gene expression HDDs | = 79,984 / 10 = 7,998 | KNN |

* Bold text represents the top 3 datasets with the highest average number of features.

can perform FS in the HDD studies based on data grouping (mean-match). Moreover, WOA (K11, K43, K48) and PSO (K32, K62) were recognised as the most preferred algorithms (Table 8).

Besides, the wrapper-based FS methods were employed in 11 publications (78.6%), while the filter-based FS methods were utilised in 3 publications (21.4%) from the 14 HDD studies based on the 14 HDD works in data grouping (mean-match).

The number of publications on FS using meta-heuristic algorithms for filter-based and wrapper-based FS methods, specifically those under HDD studies based on data grouping (mean-match) demonstrated an increasing trend from 2016 to 2021 (Fig. 9).
The trendline gradient for filter-based FS methods was almost as steep as wrapper-based FS methods indicating that this data grouping situation (mean-match) depicts the actual growth of preferences among researchers to adopt filter-based methods in their studies consisting HDD on average.

As for the classifier preference of the 14 HDD studies based on data grouping (mean-match) as depicted in Fig. 10, the filter-based FS methods illustrated more adaptability than that of the wrapper-based as they can integrate with 4 classifiers such as SVM (42.86%), RF (14.3%), NB (28.6%), and DT (14.3%). Whereas, wrapper-based FS methods yielded lesser classifier preference as it only integrated with 3 classifiers, namely 83.33% of KNN and 8.33% for both DT and MLP.

### TABLE 8. General detail of 14 HDD studies based on data grouping (mean-match).

| Key | Meta-heuristic algorithm | FS Methods | Datasets | Average no. of features | Classifiers |
|-----|--------------------------|------------|----------|-------------------------|-------------|
| K01 | IG-bBOA                  | Filter     | 6 HDDs from the UCI ML repository. | 21,847 / 6 | RF, NB, SVM |
| K02 | TSHFS-ACO                | Wrapper    | 11 high-dimensional low sample datasets. | 91,984 / 11 | KNN |
| K04 | CCOAFLDO                | Wrapper    | 20 HDDs. | 208,180 / 20 | KNN |
| K07 | QMFOA                  | Filter     | 13 high-dimensional microarray gene datasets. | 122,986 / 13 | SVM |
| K11 | SBWOA                 | Wrapper    | 16 HDDs from ASU repository. | 119,161 / 16 | KNN |
| K13 | IGWO                  | Wrapper    | 10 HDDs. | 81,984 / 10 | MLP |
| K17 | ACCSO               | Wrapper    | 12 benchmark datasets from UCI and ASU repository (3 HDDs). | 56,702 / 12 | KNN |
| K32 | FBPSO            | Wrapper    | 2 benchmark UCI datasets and 4 benchmark ASU datasets. | 21,300 / 6 | KNN, DT |
| K43 | WOA-MC            | Filter     | 4 benchmark binary medical datasets. | 51,405 / 4 | DT, SVM, NB |
| K44 | BCROSAT         | Wrapper    | 13 HDDs. | 91,976 / 13 | KNN |
| K48 | WOA             | Wrapper    | 9 high-dimensional low sample datasets. | 78,667 / 9 | KNN |
| K53 | ISFLA          | Wrapper    | 9 HDDs from KRBD. | 62,828 / 9 | KNN |
| K55 | ALO-GWO        | Wrapper    | 18 benchmark datasets from the UCI ML repository and 12 HDDs. | 158,255 / 30 | KNN |
| K62 | PSO-LSRG        | Wrapper    | 10 gene expression HDDs | 79,984 / 10 | KNN |

* Bold text represents the top 3 datasets with the highest average number of features

### FIGURE 9. Number of publications per year based on data grouping (mean-match) for 14 HDD studies.

### FIGURE 10. Classifier preferences for 14 HDD studies based on data grouping (mean-match).

4) STUDIES BASED ON DATA GROUPING (ALL-MATCH)

The all-match data grouping was applied to all the 62 selected journal articles in this subsection. All-match refers to the grouping technique that selects articles employing HDD if and only if every dataset used matched the threshold of 2000 features. For instance, if a study consists of 3 datasets (with 2000, 5000, and 7000 features) and obtained an all-match, it can be categorised as HDD research since all the datasets matched the threshold of 2000 features. Therefore, this data grouping technique has the narrowest chance of categorising articles as HDD studies as it requires every dataset to have more than 2000 features.

Eight of the 62 publications were categorised as studies utilising HDD based on the all-match data grouping. Table 9 lists the meta-heuristic algorithm, FS methods,
datasets, the average number of features, and classifiers used in these 8 publications. The number of datasets used ranged from 4 to 13, while the average number of features ranges from 6,981 to 12,852. One of the 8 publications retrieved the datasets from the KRBD repository and 7 from unspecific sources.

The top 3 average number of features were identified from publications K43, K07, and K48 employed WOA (in K43 and K48) and QMFOA (in K07), indicating the superior capability of these two meta-heuristic algorithms to solve FS problems in colossal HDD. Besides that, the authors of the top two publications (K43 and K07) employed the filter-based FS methods over wrapper-based, indicating that the filter-based method was preferred for datasets with extremely high dimensionality. The high preferences for WOA proved its capability to work well in both wrapper-based and filter-based FS methods, especially where the filter-based FS methods from K42, with an average number of features as high as 12,852 being at the top of the list for every data grouping, including 1-match, mean-match, and all-match.

Besides, the wrapper-based FS methods were employed in the majority of publications (6 publications, 75%), while the filter-based FS methods were only utilised in 2 publications (25%) based on the 8 HDD works in data grouping (all-match).

Fig. 11 demonstrated an increase in the number of publications on FS using meta-heuristic algorithms for filter-based and wrapper-based FS methods, specifically those under HDD studies based on data grouping (all-match) from 2016 to 2021. The trendline for filter-based FS methods was as steep as the wrapper-based FS methods. This data grouping technique (all-match) demonstrated the preferences of researchers on adopting filter-based and wrapper-based FS methods are at par.

Based on Fig. 12, filter-based FS methods demonstrated adaptability to integrating 3 classifiers, SVM (50%), NB (25%), and DT (25%) compared to the wrapper-based methods as they can integrate with 3 classifiers such as SVM (50%), NB (25%), and DT (25%) classifiers. Whereas, wrapper-based FS methods were only able to integrate KNN (83.33%) and MLP (16.67%).

VI. HDD ANALYSIS
In this section, HDD analysis was performed based on 4 LGDG grouping techniques introduced in Section V.
TABLE 9. General detail of 8 HDD studies based on data grouping (ALL-match).

| Key | Meta-heuristic algorithm | FS Methods | Datasets | Average no. of features | Classifiers |
|-----|--------------------------|------------|----------|-------------------------|-------------|
| K02 | TSHFS-ACO                | Wrapper    | 11 high-dimensional low sample datasets. | 91,984 / 11 8,362 | KNN         |
| K07 | QMFOA                    | Filter     | 13 high-dimensional microarray gene datasets. | 122,986 / 13 9,460 | SVM         |
| K13 | IGWO                     | Wrapper    | 10 HDDs. | 81,984 / 10 8,198       | MLP         |
| K43 | WOA-MC                   | Filter     | 4 benchmark binary medical datasets. | 51,405 / 4 12,852 | DT, SVM, NB |
| K44 | BCROSAT                  | Wrapper    | 13 HDDs. | 91,976 / 13 7,075       | KNN         |
| K48 | WOA                      | Wrapper    | 9 high-dimensional low sample datasets. | 78,667 / 9 8,741 | KNN         |
| K53 | ISFLA                    | Wrapper    | 9 HDDs from KRBD. | 62,828 / 9 6,981 | KNN         |
| K62 | PSO-LSRG                 | Wrapper    | 10 gene expression HDDs | 79,984 / 10 7,998 | KNN         |

*Bold text represents the top 3 datasets with the highest average number of features*

TABLE 10. Dataset distribution for 19 HDD studies based on literal grouping.

| Key | 1-20 (f) | 21-50 (f) | 51-100 (f) | 101-300 (f) | 301-500 (f) | 501-1999 (f) | => 2000 (f) |
|-----|----------|-----------|------------|-------------|-------------|--------------|-------------|
| K01 | 0        | 0         | 0          | 1           | 0           | 0            | 5           |
| K02 | 0        | 0         | 0          | 0           | 0           | 0            | 11          |
| K03 | 6        | 2         | 0          | 1           | 0           | 1            | 2           |
| K04 | 0        | 0         | 0          | 1           | 1           | 5            | 13          |
| K07 | 0        | 0         | 0          | 0           | 0           | 0            | 13          |
| K11 | 0        | 0         | 0          | 0           | 0           | 1            | 15          |
| K13 | 0        | 0         | 0          | 0           | 0           | 0            | 10          |
| K14 | 5        | 2         | 1          | 0           | 1           | 0            | 4           |
| K17 | 1        | 0         | 1          | 0           | 1           | 1            | 8           |
| K26 | 6        | 6         | 1          | 0           | 1           | 2            | 4           |
| K28 | 11       | 5         | 1          | 0           | 1           | 0            | 3           |
| K31 | 20       | 8         | 7          | 1           | 0           | 0            | 9           |
| K40 | 11       | 5         | 1          | 0           | 1           | 0            | 9           |
| K41 | 4        | 2         | 1          | 0           | 0           | 0            | 3           |
| K44 | 0        | 0         | 0          | 0           | 0           | 0            | 13          |
| K48 | 0        | 0         | 0          | 0           | 0           | 0            | 9           |
| K53 | 0        | 0         | 0          | 0           | 0           | 0            | 9           |
| K55 | 11       | 5         | 1          | 0           | 1           | 1            | 11          |
| K62 | 0        | 0         | 0          | 0           | 0           | 0            | 10          |

C. DATA GROUPING (MEAN-MATCH)

This subsection analyses the datasets used by 14 HDD studies based on data grouping (mean-match). This process determined the number of datasets with specific ranges for the number of features. The details of dataset distribution are tabulated in Table 12. A majority of the datasets, 135 (79.88%) reached the threshold of 2,000 features, while 34 (20.12%) did not. Hence, 79.88% of the datasets qualified to be HDD based on data grouping (mean-match), higher than the literal grouping and data grouping (1-match).

Of the 146 datasets in wrapper-based HDD studies, 113 of them reached the threshold of 2,000 features, while 33 (22.6%) did not. Hence, 77.4% of the datasets in the wrapper-based HDD studies based on data grouping (mean-match) qualified to be HDD (Fig. 17). Only 1 (4.35%) of the filter-based studies did not reach the threshold, whereas 32 (95.65%) reached the threshold of 2,000 features. According to Fig. 18, 95.65% of the datasets...
TABLE 11. Dataset distribution for 26 HDD studies based on data grouping (1-match).

| Key | 1-20 (f) | 21-50 (f) | 51-100 (f) | 101-300 (f) | 301-500 (f) | 501-1999 (f) | >= 2000 (f) |
|-----|-----------|-----------|------------|-------------|-------------|--------------|-------------|
| K01 | 0         | 0         | 0          | 0           | 1           | 0            | 5           |
| K02 | 0         | 0         | 0          | 0           | 0           | 0            | 11          |
| K03 | 6         | 2         | 0          | 1           | 0           | 1            | 2           |
| K04 | 0         | 0         | 0          | 0           | 0           | 1            | 5           |
| K05 | 0         | 0         | 0          | 0           | 0           | 1            | 13          |
| K06 | 2         | 5         | 0          | 0           | 0           | 0            | 3           |
| K11 | 0         | 0         | 0          | 0           | 0           | 1            | 15          |
| K13 | 0         | 0         | 0          | 0           | 0           | 0            | 10          |
| K14 | 5         | 2         | 1          | 0           | 0           | 0            | 4           |
| K17 | 1         | 0         | 0          | 0           | 1           | 1            | 8           |
| K19 | 11        | 8         | 0          | 0           | 0           | 1            | 3           |
| K23 | 7         | 6         | 4          | 0           | 1           | 1            | 1           |
| K26 | 6         | 6         | 1          | 0           | 1           | 2            | 4           |
| K28 | 11        | 5         | 1          | 0           | 0           | 0            | 3           |
| K31 | 20        | 8         | 7          | 1           | 0           | 0            | 9           |
| K32 | 0         | 0         | 0          | 2           | 0           | 0            | 4           |
| K33 | 4         | 7         | 2          | 3           | 2           | 0            | 2           |
| K37 | 12        | 6         | 2          | 2           | 1           | 0            | 2           |
| K41 | 4         | 2         | 1          | 0           | 0           | 0            | 3           |
| K43 | 0         | 0         | 0          | 0           | 0           | 0            | 4           |
| K44 | 0         | 0         | 0          | 0           | 0           | 0            | 13          |
| K45 | 11        | 1         | 2          | 0           | 0           | 0            | 2           |
| K48 | 0         | 0         | 0          | 0           | 0           | 0            | 9           |
| K53 | 0         | 0         | 0          | 0           | 0           | 0            | 9           |
| K55 | 11        | 5         | 1          | 0           | 0           | 1            | 11          |
| K62 | 0         | 0         | 0          | 0           | 0           | 0            | 10          |

TABLE 12. Dataset distribution for 14 HDD studies based on data grouping (mean-match).

| Key | 1-20 (f) | 21-50 (f) | 51-100 (f) | 101-300 (f) | 301-500 (f) | 501-1999 (f) | >= 2000 (f) |
|-----|-----------|-----------|------------|-------------|-------------|--------------|-------------|
| K01 | 0         | 0         | 0          | 0           | 1           | 0            | 5           |
| K02 | 0         | 0         | 0          | 0           | 0           | 0            | 11          |
| K04 | 0         | 0         | 0          | 0           | 0           | 1            | 5           |
| K07 | 0         | 0         | 0          | 0           | 0           | 0            | 13          |
| K11 | 0         | 0         | 0          | 0           | 0           | 0            | 13          |
| K13 | 0         | 0         | 0          | 0           | 0           | 0            | 10          |
| K17 | 1         | 0         | 0          | 0           | 0           | 1            | 1           |
| K19 | 11        | 8         | 0          | 0           | 0           | 0            | 3           |
| K23 | 7         | 6         | 4          | 0           | 1           | 1            | 1           |
| K26 | 6         | 6         | 1          | 0           | 1           | 2            | 4           |
| K28 | 11        | 5         | 1          | 0           | 0           | 0            | 3           |
| K31 | 20        | 8         | 7          | 1           | 0           | 0            | 9           |
| K32 | 0         | 0         | 0          | 2           | 0           | 0            | 4           |
| K33 | 4         | 7         | 2          | 3           | 2           | 0            | 2           |
| K37 | 12        | 6         | 2          | 2           | 1           | 0            | 2           |
| K41 | 4         | 2         | 1          | 0           | 0           | 0            | 3           |
| K43 | 0         | 0         | 0          | 0           | 0           | 0            | 4           |
| K44 | 0         | 0         | 0          | 0           | 0           | 0            | 13          |
| K45 | 11        | 1         | 2          | 0           | 0           | 0            | 2           |
| K48 | 0         | 0         | 0          | 0           | 0           | 0            | 9           |
| K53 | 0         | 0         | 0          | 0           | 0           | 0            | 9           |
| K55 | 11        | 5         | 1          | 0           | 0           | 1            | 11          |
| K62 | 0         | 0         | 0          | 0           | 0           | 0            | 10          |

qualified to be HDD in filter-based HDD research based on data grouping (mean-match), contributing 18.25% higher than the wrapper-based.

D. DATA GROUPING (ALL-MATCH)
This subsection analyses the datasets used by 8 HDD studies based on data grouping (all-match) by determining the number of datasets with specific features range (Table 13).

According to Table 13, all of the 79 datasets (100%) reached the threshold of 2,000 features. It is a predictable outcome since the grouping technique used was all-match, where all datasets were set to reach the threshold. It is also the most convincing way of data grouping, whereby every dataset used in all the studies were HDD. Therefore, the datasets in both the filter-based (17 datasets) and wrapper-based (62 datasets) studies were accepted as HDD, as depicted in Fig. 19 and Fig. 20, respectively.

VII. DISCUSSIONS
This section discusses the trend or potential FS methods in dealing with HDD. The percentage of HDD versus non-HDD in all grouping techniques is listed in Table 14. The values under the four horizontal categories (All FS, wrapper, filter, and hybrid methods) are represented by percentage (%), with the summation of the percentage of non-HDD and HDD should be 100% for each category by each grouping.

Based on Table 14, 82.09% of the non-HDD and 17.91% of the HDD for All FS method represented all the 62 publications. However, not all 62 publications reviewed were
TABLE 13. Dataset distribution for 8 HDD studies based on data grouping (ALL-match).

| Key | 1-20 (f) | 21-50 (f) | 51-100 (f) | 101-300 (f) | 301-500 (f) | 501-1999 (f) | >= 2000 (f) |
|-----|----------|-----------|------------|-------------|-------------|--------------|------------|
| K02 | 0        | 0         | 0          | 0           | 0           | 0            | 11         |
| K07 | 0        | 0         | 0          | 0           | 0           | 0            | 13         |
| K13 | 0        | 0         | 0          | 0           | 0           | 0            | 10         |
| K43 | 0        | 0         | 0          | 0           | 0           | 0            | 4          |
| K44 | 0        | 0         | 0          | 0           | 0           | 0            | 13         |
| K48 | 0        | 0         | 0          | 0           | 0           | 0            | 9          |
| K53 | 0        | 0         | 0          | 0           | 0           | 0            | 9          |
| K62 | 0        | 0         | 0          | 0           | 0           | 0            | 10         |

TABLE 14. Data percentage of HDD versus non-HDD in all grouping techniques.

| Grouping Techniques | Number of research studies | All FS (%) | Wrapper (%) | Filter (%) | Hybrid (%) |
|---------------------|---------------------------|------------|-------------|------------|------------|
|                     |                           | Non-HDD    | HDD         | Non-HDD    | HDD        | Non-HDD | HDD |
| Chosen papers       | 62                        | 82.09      | 17.91       | 82.79      | 17.21      | 72.07   | 27.93 | 100  | 0   |
| Literal grouping    | 19                        | 51.01      | 48.99       | 46.58      | 53.42      | 57.81   | 42.19 | -    | -   |
| Data grouping (1-match) | 26                     | 57.18      | 42.82       | 57.74      | 42.26      | 54.41   | 45.59 | -    | -   |
| Data grouping (mean-match) | 14                  | 20.12      | 79.88       | 22.60      | 77.40      | 4.35    | 95.65 | -    | -   |
| Data grouping (all-match) | 8                      | 0          | 100         | 0          | 100        | 0       | 100   | -    | -   |

* Bold text represents a higher percentage of HDD among FS methods.

FIGURE 13. Dataset distribution for wrapper-based HDD studies based on literal grouping.

HDD studies, only 17.91% were HDD studies. Studies that employed the hybrid FS methods did not deal with HDD. The wrapper-based and filter-based FS methods measured 17.21% and 27.93% in HDD studies, indicating a difference of 10.72% between both methods. This percentage suggested that studies with filter-based FS methods focused more on HDD compared to wrapper-based FS methods (Fig. 21).

For literal grouping, wrapper-based FS methods performed better with 53.42% of the studies involving HDD, while 46.58% of studies involved non-HDD. Of which, 42.19% of the HDDs used filter-based FS methods. Both the wrapper-and filter-based methods indicated an increment compared to the previous grouping of all selected publications. This observation implied that the literal grouping successfully differentiated studies employing HDD, as presented in Fig. 22.

As for 1-match data grouping, similar readings were determined compared to literal grouping. The filter-based FS methods outperformed the wrapper-based (42.26%) method with more HDD (45.59%). Despite the slight difference between the two methods, they presented better numbers than those without LGDG grouping. It also indicated that this data grouping method effectively differentiated HDD studies (Fig. 22 (b)).

Meanwhile, the mean-match data grouping adopted the average number of features in the datasets used. It obtained the most accurate numbers compared to other grouping techniques, with an average of 79.88% HDD studies. Of the 14 studies which employed HDD, 79.88% of the datasets had
greater than or equal to 2,000 features per dataset. Studies employing the wrapper-based FS methods counted 77.40% HDD, while filter-based FS methods indicated that 95.65% of the studies utilized HDD (Fig. 22 (c)).

Finally, the all-match data grouping, being the narrowest grouping technique, revealed that 100% of the studies which employed wrapper-based and filter-based FS methods reached the threshold. The datasets used in the 8 studies were extremely high in dimension, and therefore, the group measured 100% HDD regardless of the FS methods (Fig. 22 (d)).

Table 15 presents the percentage of each FS method preference in all grouping techniques. It can be concluded that wrapper-based FS methods were the most preferred compared to the filter-based in the studies reviewed in this survey.

Specifically, K48 pointed out that wrapper-based FS methods are increasingly being employed in place of filter-based FS methods with the expansion of data mining techniques in many sectors [79]. This idea was supported by 50 of the
62 studies reviewed in this survey. HDD studies based on data groupings such as K02 and K44 stated that filter-based FS methods might overlook feature dependencies and preserve redundant or irrelevant features due to the absence of machine learning algorithms in FS [22], [34]. Furthermore, K02, K13, K44, and K62 also suggested that filter-based FS methods obtained low classification accuracy and consume lesser computational cost, unlike wrapper-based methods which need to build learning models to evaluate each selected feature [22], [34], [48], [71]. Hence, meta-heuristic algorithms are widely used in wrapper-based FS to tackle the computational cost issue when dealing with HDD.

Undeniably, all 62 publications proved that meta-heuristics algorithms were helpful in the FS of HDD, as they can help alleviate the computational load of the wrapper-based FS methods. But at the same time, many studies overlooked an essential factor: filter-based FS methods can also provide promising outcomes in the FS of HDD. The studies reviewed in this survey revealed that filter-based FS methods can also achieve high accuracy through experimental results, no less than wrapper-based FS methods. Besides having filter-based FS methods integrated with powerful meta-heuristic algorithms, lesser computational costs, excellent FS, and potentially good classification accuracy were also desirable. Studies utilizing HDD based on LGDG grouping techniques such as K01, K07, K31, and K43 accomplished excellent outcomes from experiments [20], [26], [63], [67].

The graphical presentation of Table 15 is shown in Fig. 23. According to Fig. 23, the ratio of filter-based to wrapper-based methods exhibited a significant increase based on the LGDG grouping techniques.

As illustrated in Fig. 23, the ratio of filter-based to wrapper-based increased when grouping techniques were applied. The tendency of using filter-based FS methods increased especially when the grouping techniques used were of data grouping (mean-match) and data grouping (all-match). Both these grouping techniques are comparatively more accurate because their HDD versus non-HDD was genuinely higher (HDD in mean-match and all-match was 79.88% and 100%, respectively, refer to Table 14 and Fig. 22). Therefore, based on Table 14 and Table 15, the filter-based FS methods demonstrated an invisible up-growing trendline. In short, the number of studies that considered using filter-based methods to deal with HDD has increased.

Without applying LGDG grouping, only 14.5% of the studies from all 62 papers utilised filter-based methods. However, based on the values in mean-match and all-match data grouping in Table 15, filter-based FS methods acquired 21.4% and 25% preferences in HDD studies. Although the filter-based FS method is not commonly used as the wrapper-based FS method, the acceptance among researchers to adopt the filter-based FS methods in HDD has increased.

In short, both wrapper- and filter-based FS methods have their advantages. Both the methods acquired outstanding results with HDD, as discussed in this survey. The key is to choose the right method based on the nature of the datasets used. For instance, wrapper-based FS methods are often used in studies on HDD with low sample numbers (K02 and K48), whereas filter-based FS methods are often used on medical HDD or microarray gene expression HDD (K07 and K43).

Furthermore, integrating suitable meta-heuristic algorithms could also boost the performance of both wrapper- and filter-based FS methods. Table 16 summarises the top 3 meta-heuristic algorithms with the highest average number of features based on the LGDG groupings (refer to Table 5 (literal grouping), Table 7 (1-match), Table 8 (mean-match), and Table 9 (all-match)).
According to Table 16, the top 3 meta-heuristic algorithms have the highest average number of features based on LGDG groupings consisting of Cuckoo Optimisation Algorithm (K04), MFOA (K07), and WOA (K43, K48). These 3 meta-heuristic algorithms appeared in all 4 LGDG groupings of the HDD studies except for the Cuckoo Optimisation Algorithm which was not included in all-match data grouping. Instead, WOA which appeared twice in the all-match data grouping for K43 and K48 indicated its superior competency in datasets with extremely high dimensions. Hence, it can be concluded that MFOA works well with filter-based FS methods, while Cuckoo Optimisation Algorithm works well with wrapper-based. Also, WOA works well in both FS methods.
TABLE 19. Top classifiers based on LGDG groupings in all 62 chosen papers.

| Top classifiers | Description | Research |
|-----------------|-------------|----------|
| SVM             | SVM is a commonly used supervised learning technique for classification tasks. It is particularly well-known for its structural risk minimization capabilities, which enable it to discover solutions with the fewest associated risks. It is relatively easy to train this classifier, and additionally, it performs admirably with HDD. | K01, K07, K31, K43, K49, K50, K52, K61 |
| KNN             | KNN is a supervised machine learning algorithm. It is a straightforward classifier that seeks the K-nearest labelled data to classify unlabeled data using a certain distance. It is prevalent due to its simplicity in implementation and efficient performance in obtaining high classification accuracy. | K01-K06, K08-K12, K14, K16-K19, K22-K42, K44, K46-K48, K51, K53-K60, K62 |
| DT              | DT is a frequently used and well-known classifier. It attempts to construct a pattern for classification purposes using training data and information entropy. It is an accurate way of determining the optimal features for decision-making (the best split). The branches are translated into rules to generate a decision-making pattern. | K32, K43, K45 |
| NB              | The NB classifier is a probabilistic classifier that operates on the Bayes theorem, assuming no dependency between features. It is often used as a baseline approach for many classification problems because of its speed and scalability. The classification with NB converges fast and is resistant to data noise. | K01, K20, K31, K43, K52 |

In short, different FS methods work well with different classifiers, and knowing the best combination increases the classification performance. Table 18 indicated the top classifiers used in the research studies with the highest average number of features based on LGDG groupings (refer to Table 5 (literal grouping), Table 7 (1-match), Table 8 (mean-match), and Table 9 (all-match)).

Based on Table 18, the classifiers were limited to SVM, KNN, DT, and NB. The SVM, DT, and NB were the preferred classifiers for datasets with extremely high dimensions using the filter-based FS methods, whereas KNN was preferred for the wrapper-based FS methods. These classifiers were also widely used in all the 62 selected publications as depicted in Table 19. Each classifier has its strengths and hence is preferred based on the field of study.

However, categorising HDD research using the proposed LGDG technique is done manually, and therefore, time consuming. This is because the searching process of the HDD keywords in the fields of title, issue, and dataset descriptions by literal grouping requires manual effort. Similarly, for data grouping, each research work’s datasets were manually compared to the threshold of 2,000 features to correctly group the dataset as HDD for analysis.

VIII. CONCLUSION AND FUTURE WORKS

Over the years, meta-heuristic algorithms have demonstrated their capability in various domains, including FS. Due to technological advancement, data expansion is unavoidable, where an enormous amount of data are being generated every second in different fields. The application of FS is no longer a task of simply selecting the relevant features that contribute to classification accuracy with a minimum number of selected features. Instead, FS keeps up with the pace of data growth as it also has to tackle the “curse of dimensionality” on HDD. Therefore, it is crucial to integrate meta-heuristic algorithms to aid FS.

With that being said, there are different FS methods available to accomplish the tasks mentioned earlier. For instance, filter- and wrapper-based FS methods are most frequently used when dealing with HDD using meta-heuristic algorithms. Since there are different schools of thought on determining a dataset as HDD, this study surveyed the threshold of the number of features in a dataset used to categorise HDD. This study also aimed to identify the trend or potential FS methods for HDD, together with the most preferred meta-heuristic algorithms and classifiers when employing HDD for both wrapper- and filter-based FS methods. Therefore, an extensive systematic literature review was conducted by implementing the PRISMA guidelines. The 62 journal articles selected for this survey were published between 2016 to 2021 and were retrieved from 7 databases accessed through the digital libraries in Universiti Tun Hussein Onn Malaysia.

To accurately group the 62 publications into HDD studies, this survey proposed a novel grouping technique called the LGDG. Literal grouping represents the searching of selected articles using HDD keywords. A total of 19 literal-grouped articles were categorised as HDD research. The data grouping consisted of 3 subgroups, namely 1-match, mean-match, and all-match. The threshold for the number of features in the 19 articles from literal grouping was set at 2,000 features by the majority. Meanwhile, the 26 publications categorised as HDD research based on data grouping (1-match), require at least 1 dataset with 2,000 features and above. While mean-match data grouping estimated 14 publications as HDD research in which the average number of features had to be greater or equal to the threshold. Finally, 8 articles were categorised as HDD research based on all-match data grouping, where every dataset requires 2,000 features and above. The publications by all LGDG groupings were analysed and discussed on different aspects such as the overall preference by researchers, the increasing trend for different FS methods in the yearly publication charts, suitable classifiers for each FS method, and the meta-heuristic algorithms that are excellent in both FS methods.

Based on the findings, different meta-heuristic algorithms and classifiers were preferred for different FS methods. Moreover, studies with wrapper-based FS methods indicated a remarkable ability in obtaining high classification performance. However, the filter-based FS methods also...
gained more attention in recent years with competent results on HDD.

In conclusion, as suggested by the No Free Lunch theorem, there is no absolute answer to the question of the best FS method of all. Once again, the key is to find a suitable method depending on the dataset’s nature.

In the future, other aspects such as multi-objective FS on extreme HDD could be reviewed. Including other FS methods could be considered as an extension of the current survey. Lastly, the proposed LGDG technique could be improved to cover more extensive and effective searching methods for HDD research.

REFERENCES

[1] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, Feature Selection for High-Dimensional Data, Cham, Switzerland: Springer, 2015.
[2] A. Bommer, X. Sun, B. Bischi, J. Rahnenführer, and M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data,” Comput. Statist. Data Anal., vol. 143, Mar. 2020, Art. no. 106839, doi: 10.1016/j.csda.2019.106839.
[3] L. Hu, W. Gao, K. Zhao, P. Zhang, and F. Wang, “Feature selection considering two types of feature relevancy and feature interdependency,” Expert Syst. Appl., vol. 93, pp. 423–434, Mar. 2018, doi: 10.1016/j.eswa.2017.10.016.
[4] G. Chen, L. Liu, and S. Huang, “Enhanced GSA-based optimization for minimization of power losses in power system,” Math. Problems Eng., vol. 2015, pp. 1–13, Dec. 2015, doi: 10.1155/2015/527128.
[5] X. Li and K. M. Luk, “The grey wolf optimizer and its applications in electromagnetics,” IEEE Trans. Antennas Propag., vol. 68, no. 3, pp. 2186–2197, Mar. 2020, doi: 10.1109/TAP.2019.2938703.
[6] W. Gao, “Improved ant colony clustering algorithm and its performance study,” Comput. Intell. Neurosci., vol. 2016, pp. 1–14, Jan. 2016, doi: 10.1155/2016/483592.
[7] A. S. Assiri, A. G. Hussien, and M. Amin, “Ant lion optimization: Variants, hybrids, and applications,” IEEE Access, vol. 8, pp. 7746–77764, 2020, doi: 10.1109/ACCESS.2020.2909338.
[8] M. M. Mafarja and S. Mirjalili, “Hybrid whale optimization algorithm with simulated annealing for feature selection,” Neuro Comput., vol. 260, pp. 302–312, Oct. 2017, doi: 10.1016/j.neucom.2017.04.053.
[9] M. Mafarja and S. Mirjalili, “Whale optimization approach for wrapper feature selection,” Appl. Soft Comput., vol. 62, pp. 441–453, Jan. 2018, doi: 10.1016/j.asoc.2017.11.006.
[10] S. Mirjalili, “The ant lion optimizer,” Adv. Eng. Softw., vol. 83, pp. 80–98, May 2015, doi: 10.1016/j.advengsoft.2015.01.010.
[11] J. Kennedy and R. Eberhart, “Particle swarm optimization,” Port. Swarm Optim., vol. 1, pp. 1942–1948, Jan. 2010, doi: 10.1007/978-0-470-61216-3.
[12] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” Adv. Eng. Softw., vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.
[13] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, “Feature Selection for High-Dimensional Data,” Adv. Eng. Softw., vol. 159, pp. 270–285, Nov. 2018, doi: 10.1016/j.advengsoft.2018.08.051.
[14] P. Bansal, S. Kumar, S. Parasra, and S. Singh, “A hybrid grasshopper and new cat swarm optimization algorithm for feature selection and evaluation of multi-layer perceptron,” Soft Comput., vol. 24, no. 20, pp. 15463–15489, Oct. 2020, doi: 10.1007/s11227-020-04877-w.
[15] K. C. Lin, K. Y. Zhang, Y. H. Huang, and J. C. Hung, “Feature selection based on an improved cat swarm optimization algorithm for big data classification,” J. Supercomput., vol. 72, no. 8, pp. 3210–3221, 2016, doi: 10.1007/s11227-016-1631-0.
[16] S. Pichai, K. Sunat, and S. Chiewchanwattana, “An asymmetric chaotic feature selection algorithm based on feature selection using redundant removal,” Neurocomputing, vol. 343, no. 14, pp. 11301–11308, 2019, doi: 10.1016/j.neucom.2019.01.006.
[17] L. Y. Yab et al.: Meta-Analysis Survey on the Usage of Meta-Heuristic Algorithms for FS on HDDs
H. Hichem, M. Elkamel, M. Rafik, M. T. Mesaaoud, and C. Ouahiba, "An analytical study of modified multi-objective Harris hawks optimizer for feature selection problem," Int. J. Adv. Inform. Sci., vol. 4, no. 2, pp. 95–106, 2018, doi: 10.26555/ijain.v4i2.245.

C. Yu, Z. Cai, X. Ye, M. Wang, X. Zhao, G. Liang, H. Chen, and C. Li, "Quantum-like mutation-induced dragonfly-inspired optimization approach," Math. Comput. Simul., vol. 178, pp. 259–289, Dec. 2020, doi: 10.1016/j.matcom.2020.06.012.

M. Mafarja, I. Aljarah, M. A. Heidari, H. Faris, P. Fournier-Viger, X. Li, and S. Mirjalili, "Binary dragonfly optimization for feature selection using time-varying transfer functions," Knowl.-Based Syst., vol. 161, pp. 185–204, Aug. 2018, doi: 10.1016/j.knosys.2018.08.003.

Y. Zhang, X.-F. Song, and D.-W. Gong, "A return-cost-based binary firefly algorithm for feature selection," Inf. Sci., vol. 419, pp. 561–574, Dec. 2017, doi: 10.1016/j.ins.2017.08.047.

B. Hu, Y. Dai, Y. Su, P. Moore, X. Zhang, C. Mao, J. Chen, and L. Xu, "Feature selection for optimized high-dimensional biomedical data using an improved shuffled frog leaping algorithm," IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 15, no. 6, pp. 1765–1773, 2018, doi: 10.1109/TCBB.2018.2802263.

Y.-J. Hu, J. Li, H. Faris, and Z. Li, "BBFIOA: A novel binary improved fruit fly algorithm for feature selection," IEEE Access, vol. 7, pp. 81177–81194, 2019, doi: 10.1109/access.2019.2917502.

H. Hichem, M. Elkamel, M. Rafik, M. T. Mesaoud, and C. Ouahiba, "A new binary grasshopper optimization algorithm for feature selection problem," J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 2, pp. 316–328, Feb. 2022, doi: 10.1016/j.jksuci.2019.11.007.

M. Mafarja, I. Aljarah, H. Faris, A. I. Hammoury, A. M. Al-Zoubi, and S. Mirjalili, "Binary grasshopper optimisation algorithm approaches for feature selection problems," Expert Syst. Appl., vol. 117, pp. 267–286, Mar. 2019, doi: 10.1016/j.eswa.2018.09.015.

A. Zakeri and A. Hokmabadi, "Efficient feature selection method using real-valued grasshopper optimization algorithm," Expert Syst. Appl., vol. 119, pp. 61–72, Apr. 2019, doi: 10.1016/j.eswa.2018.10.021.

B. K. Kihel, S. K. Kihel, and S. Chouraqui, "Immune grey wolf optimizer for attribute reduction: Application for medical systems," Int. J. Inform. Eng. Syst., vol. 14, no. 4, pp. 35–43, Aug. 2021, doi: 10.22266/ijies.2021.0831.04.

C. Shen and K. Zhang, "Two-stage improved grey wolf optimization algorithm for feature selection on high-dimensional classification," Complex Intell. Syst., vol. 8, no. 4, pp. 2769–2789, Aug. 2022, doi: 10.1007/s40747-021-00452-4.

Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, H. Alhussean, M. G. Ragab, and A. A. Alquashabi, "Binary multi-objective grey wolf optimizer for feature selection in classification," IEEE Access, vol. 8, pp. 106247–106262, Dec. 2020, doi: 10.1109/access.2020.3029725.

P. Hu, J.-S. Pan, and S.-C. Chu, "Improved binary grey wolf optimizer and its application for feature selection," Knowl.-Based Syst., vol. 195, May 2020, Art. no. 105746, doi: 10.1016/j.knosys.2020.105746.

S. Arora, H. Singh, M. Sharma, S. Sharma, and P. Anand, "A new hybrid algorithm based on Grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection," IEEE Access, vol. 7, pp. 26343–26361, 2019, doi: 10.1109/access.2019.2897325.

Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussean, "Binary optimization using hybrid grey wolf optimization for feature selection," IEEE Access, vol. 7, pp. 39496–39508, 2019, doi: 10.1109/access.2019.2907675.

M. Mafarja, A. Qasem, A. A. Heidari, I. Aljarah, H. Faris, and S. Mirjalili, "Efficient hybrid nature-inspired binary optimizers for feature selection," Cognit. Comput., vol. 12, no. 1, pp. 150–175, Jan. 2020, doi: 10.1007/s12559-019-09668-6.

H. M. Zawbaa, E. Emary, C. Grosan, and V. Snasel, "Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach," Swarm Intell. Comput., vol. 42, pp. 29–42, Oct. 2018, doi: 10.1016/j.eweco.2018.02.021.

E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," Neurocomputing, vol. 172, pp. 371–381, Jan. 2016, doi: 10.1016/j.neucom.2015.06.083.

J. Piri and P. Mohapatra, "An analytical study of modified multi-objective Harris hawks optimizer towards medical data feature selection," Comput. Biol. Med., vol. 135, Aug. 2021, Art. no. 104558, doi: 10.1016/j.compbiomed.2021.104558.
[75] H. Jia, Z. Xing, and W. Song, “A new hybrid seagull optimization algorithm for feature selection,” IEEE Access, vol. 7, pp. 49614–49631, 2019, doi: 10.1109/ACCESS.2019.2909945.

[76] R. A. Ibrahim, M. A. Elaziz, D. Oliva, E. Cuevas, and S. Lu, “An opposition-based social spider optimization for feature selection,” Soft Comput., vol. 23, no. 24, pp. 13547–13567, Dec. 2019, doi: 10.1007/s00500-019-03891-x.

[77] A. Got, A. Moussaoui, and D. Zouache, “Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach,” Expert Syst. Appl., vol. 183, Nov. 2021, Art. no. 115312, doi: 10.1016/j.eswa.2021.115312.

[78] J. Too, M. Mafarja, and S. Mitjalili, “Spatial bound whale optimization algorithm: An efficient high-dimensional feature selection approach,” Neural Comput. Appl., vol. 33, no. 23, pp. 16229–16250, Dec. 2021, doi: 10.1007/s00521-021-06224-y.

[79] M. Mafarja, I. Jaber, S. Ahmed, and T. Thaher, “Whale optimisation algorithm for high-dimensional small-instance feature selection,” Int. J. Parallel Emergent Distrib. Syst., vol. 36, no. 2, pp. 80–96, Mar. 2021, doi: 10.1080/17445760.2019.1617866.

[80] K Nearest Neighbor—An Overview | ScienceDirect Topics. Accessed: Dec. 14, 2021. [Online]. Available: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/k-nearest-neighbor

[81] Decision Tree—An Overview | ScienceDirect Topics. Accessed: Dec. 14, 2021. [Online]. Available: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/decision-tree

[82] Support Vector Machine—An Overview | ScienceDirect Topics. Accessed: Dec. 14, 2021. [Online]. Available: https://www.sciencedirect.com/topics/neuroscience/support-vector-machine

[83] Naive Bayes Classifier—An Overview | ScienceDirect Topics. Accessed: Dec. 14, 2021. [Online]. Available: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/naive-bayes-classifier

[84] Random Forest—An Overview | ScienceDirect Topics. Accessed: Dec. 14, 2021. [Online]. Available: https://www.sciencedirect.com/topics/engineering/random-forest

[85] Multilayer Perceptron—An Overview | ScienceDirect Topics. Accessed: Dec. 14, 2021. [Online]. Available: https://www.sciencedirect.com/topics/engineering/multilayer-perceptron

[86] M. J. Page et al., “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” BMJ, vol. 372, Mar. 2021, doi: 10.1136/BMJ.N71.

[87] Y. Zheng, Y. Li, G. Wang, Y. Chen, Q. Xu, J. Fan, and X. Cui, “A novel hybrid algorithm for feature selection,” Pers. Ubiquitous Comput., vol. 22, nos. 5–6, pp. 971–985, 2018, doi: 10.1007/s00779-018-1156-z.

[88] UCI Machine Learning Repository. Accessed: Nov. 28, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/index.php

[89] Arizona State University. FEATURE SELECTION DATASETS. Accessed: Jun. 17, 2021. [Online]. Available: https://jundongl.github.io/scikit-feature/datasets.html

[90] Find Open Datasets and Machine Learning Projects | Kaggle. Accessed: Nov. 28, 2021. [Online]. Available: https://www.kaggle.com/datasets

[91] KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems (Regression, Classification, Clustering, Pattern Mining and so on). Accessed: Nov. 28, 2021. [Online]. Available: https://sci2s.ugr.es/keel/datasets.php

LI YU YAB was born in Johor, Malaysia, in 1996. She received the Bachelor of Computer Science degree (Hons.) in multimedia computing from Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia, in 2019, and the Diploma degree in robotic process automation developing from UiPath, in 2020. She is currently pursuing the master’s degree in information technology with UTHM.

From 2019 to 2021, she was a System Developer at Ramatex, Malaysia. She has been a Research Assistant with the Faculty of Computer Science and Information Technology, UTHM, since 2021. Her research interests include meta-heuristic algorithms in soft computing and data mining.

Ms. Yab was a recipient of the Best Final Year Project with a Speech Recognition English Learning Application for School Remedial Program, in 2019. She has won a silver medal in Virtual International Research and Innovation Symposium and Exposition (RISE), in 2021.

NOORHANIZA WAHID was born in Johor, Malaysia. She received the Bachelor of Computer Science (Hons.) and M.S. degrees in multimedia system from Universiti Putra Malaysia, in 1997 and 2000, respectively, and the Ph.D. degree in information technology from Sydney University, Australia, in 2011.

From 2019 to 2020, she was a Senior Researcher with the Soft Computing & Data Mining Centre (SMC), Universiti Tun Hussein Onn Malaysia (UTHM). From 2017 to 2022, she was appointed as a Senior Researcher with the Multimedia Research Group (MeRG) and the Center of Excellence of Technical and Vocational Education Training (TVET), Institute for Integrated Engineering. From 2021 to 2022, she was a Principal Researcher with the Media & Signal Processing Focus Group (MSP), UTHM. Her research interests include meta-heuristic algorithms in soft computing, data mining, optimization, deep learning, and multimedia in IT applications, mainly virtual reality and augmented reality.

RAHAYU A. HAMID was born in Johor, Malaysia, in 1978. She received the Bachelor of Computer Science and M.Sc. degrees in computer science from Universiti Teknologi Malaysia (UTM), in 2000 and 2004, respectively, and the Ph.D. degree in computer science from RMIT University, Australia, in 2017. She is currently a Senior Lecturer and the Head of the Multimedia Department, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia. Her research interests include image retrieval, image search behavior, meta-heuristic algorithms, multimedia system and application, and HCI. Her recent research interest includes improving feature selection algorithm through optimization.