Placeholder for OUP logo oup.pdf

Placeholder for journal logo cde.pdf

Journal of Computational Design and Engineering, 20xx, 0(0), 5-0

doi: 10.1093/jcde/
Journal homepage: www.jcde.org
Advance Access Publication Date: 19 August 2020

RESEARCH ARTICLE

HFMOEA: A Hybrid Framework for Multi-objective Feature Selection

Rohit Kundu[1] and Rammohan Mallipeddi[2,*]

[1]Department of Electrical Engineering, Jadavpur University, India and [2]Department of Artificial Intelligence, School of Electronics Engineering, Kyungpook National University, Daegu, South Korea

[*]Corresponding author. E-mail: mallipeddi.ram@gmail.com

Abstract

In this data-driven era, where a large number of attributes are often publicly available, redundancy becomes a major problem which leads to large storage and computational resource requirement. Feature selection is a method for reducing the dimensionality of the data by removing such redundant or misleading attributes. This leads to a selection of optimal feature subsets that can be used for further computation like the classification of data. Learning algorithms, when fit on such optimal subsets of reduced dimensions perform more efficiently and storing data also becomes easier. However, there exists a trade-off between the number of features selected and the accuracy obtained and the requirement for different tasks may vary. Thus, in this paper, a Hybrid Filter Multi-Objective Evolutionary Algorithm (HFMOEA) has been proposed based on the Non-Dominated Sorting Genetic Algorithm (NSGA-II) coupled with filter-based feature ranking methods for population initialization to obtain optimal trade-off solution set to the problem. The two competing objectives for the algorithm are the minimization of the number of selected features and the maximization of the classification accuracy. The filter ranking methods used for population initialization to help in faster convergence of the NSGA-II algorithm to the PF. The proposed HFMOEA method has been evaluated on 18 UCI datasets and 2 deep feature sets (feature extracted from image datasets using deep learning models) and the performance obtained to justify the viability of the approach with respect to the state-of-the-art. The relevant codes of the proposed approach is available in https://github.com/Rohit-Kundu/HFMOEA.

Keywords: Hybrid Optimization; Multi-Objective Optimization Problem (MOOP); Feature Selection; Filter Ranking

1. Introduction

Feature selection is an important aspect in pattern recognition problems especially in this big data era, where a large number of attributes are readily available due to the improved data collection methods. Having a large number of features defining a single sample increases the computational complexity and space requirements manifold, of which all the features might not be as informative, or might even be misleading and degrade the classification performance. Feature selection strategies aim to find the optimal feature subset and discard the redundant information.

However, an exhaustive search for the optimal feature subset selection is an NP-hard problem. Instead, intelligent frameworks are used for the task. Such frameworks are divided into three categories based on their fundamental working principle: (1) filter methods (2) wrapper methods and (3) embedded methods. Filter-based feature selection methods (Bommert et al., 2020) determine feature importance using several scoring metrics in an unsupervised fashion, requiring no classifiers in its core. Wrapper methods (El Aboudi & Benhlima, 2016) on the other hand, use supervised learning for selecting the optimal feature subset. They use classifiers to gauge the classification performance of the se-

Received: a; **Revised:** b; **Accepted:** c

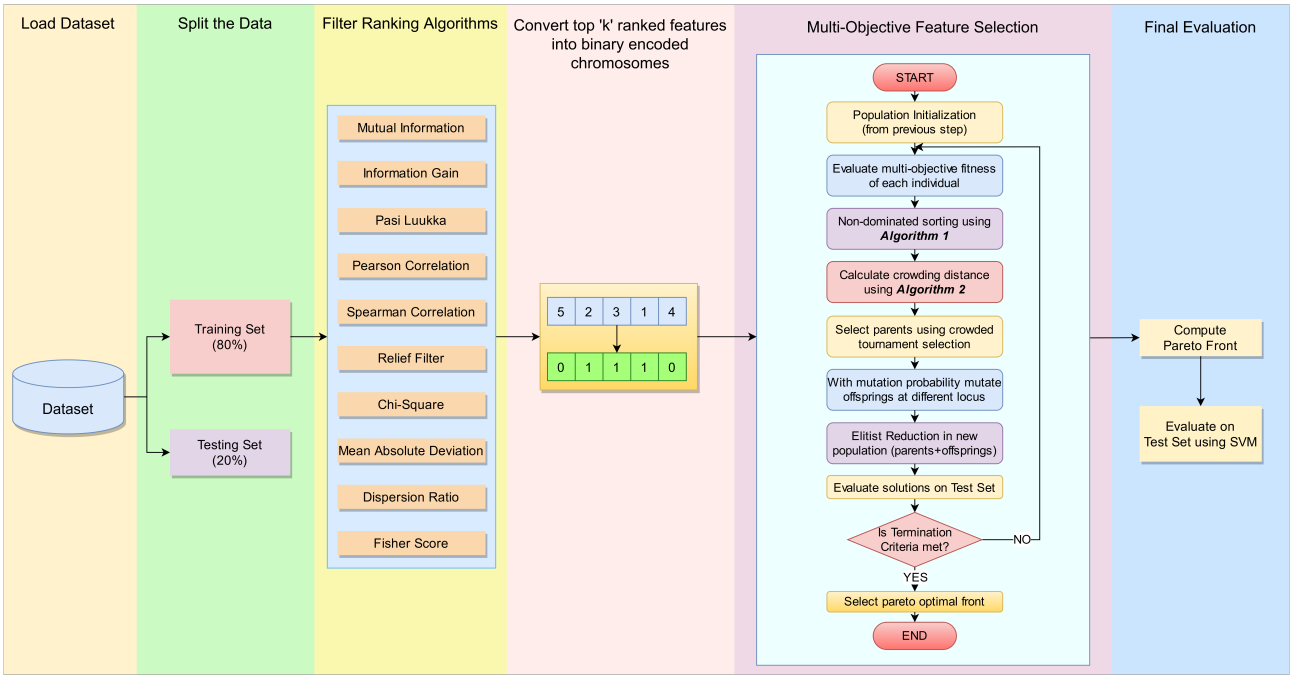


Figure 1: Overall workflow of the proposed HFMOEA framework for multi-objective feature selection.

lected feature subset in every iteration. This leads to an increased computational cost than filter-based methods, but such methods also perform several times better. Embedded methods (Maldonado & López, 2018) are similar to the wrapper-based methods, but they use an intrinsic model building metric during the learning phase. Embedded methods are able to capture more complex patterns than filter methods. According to Lal et al., 2006, embedded methods outperform filter methods when the number of training samples available are large.

Evolutionary meta-heuristics have been used intensively for the feature selection that is formulated as a single-objective optimization problem, i.e., one optimal solution is provided corresponding to best classification performance (El Aboudi & Benhlima, 2016). The fitness functions in these algorithms are usually associated with regularization parameters that control the trade-off between the accuracy and the number of features selected. Generally, a weighted sum of the accuracy of the learning algorithm and the fraction of features not selected. A higher value of this weighted sum indicates a higher fitness of the solution. Thus, to obtain the optimal results, this weight needs to be tuned which itself becomes an optimization problem.

However, in several applications, it is beneficial to have multiple trade-off solutions to the feature selection problem rather than one single optimum. In such cases, according to the need of the application, a solution can be chosen from the trade-off solution set. For example, in some scenarios, there are two solutions to the feature selection problem: accuracy of 95% with 5 features selected, and accuracy of 97% with 10 features selected. In such a case, the increase in the number of features from 5 to 10 might not be worth the computational cost to get only a 2% increase in the accuracy, while in others, accuracy might be of prime importance, regardless of the number of features selected. For scenarios like this, a multi-objective approach is beneficial where the optimal solution can be chosen according to the problem requirement from the set of trade-off solutions, which is referred to as the Pareto Front (PF) in the objective space. Recently, evolutionary algorithms garnered attention in solving multi-objective optimization problems due to their ability to provide the entire PF in a single simulation run as they evolve a population of solutions.

Hence, our research hypotheses is two-fold. First, it is believed that multi-objective optimization is a better feature selection strategy than single-objective optimization which is popular in the literature, since obtaining a set of trade-off solutions is better than one solution so that users can choose the solution they need based on the application domain. Secondly, it is believed that rather than using a randomly initialized population for multi-objective feature selection, some guidance to the initial population can lead to faster and better convergence to the Pareto front. Now, this guidance should not be a very computationally costly operation, because it is an extra step from the traditional multi-objective optimization. Thus, 10 filter-based feature ranking methods are used to initialize part of the initial population of the proposed algorithm, which is computationally cheap, but also leads to fast convergence.

The main highlighting points of our paper are as follows:

(i) A Hybrid Filter Multi-Objective Evolutionary Algorithm (HFMOEA) has been developed, using a hybrid of filter ranking methods and the Non-Dominated Sorting Genetic Algorithm-II (NSGA-II) for feature selection.
(ii) The Multi-Objective nature of the feature selection method gives users a choice to select trade-off solutions, depending on the application. Sometimes, the dimensionality of the feature subset is more important than the accuracy, and single-objective feature selection methods fail to address this.
(iii) Ten filter-based feature ranking methods have been used to generate ten members of the initial population for the multi-objective optimization. This guidance to the population helps in faster convergence of the HFMOEA algorithm to the Pareto front.
(iv) The proposed HFMOEA method has been tested on 18 datasets from the UCI machine learning repository and two deep-

feature sets extracted from biomedical images using CNNs and demonstrate the efficacy of our method.

(v) A statistical analysis of the obtained results has been performed to justify the significance of the proposed HFMOEA method.

The workflow diagram for the proposed HFMOEA framework is shown in Figure 1.

The rest of the paper has been organized as follows: Section 2 outlines the development of state-of-the-art feature selection techniques over the years; Section 3 explains in detail the proposed HFMOEA method for multi-objective feature selection adopted in this research; Section 4 evaluates the proposed HFMOEA algorithm on publicly available datasets to justify the viability of the approach; and finally Section 5 concludes the findings from this research and discusses the possibility for future research in the domain.

2. Related Work

Scalability is an essential characteristic of modern machine learning algorithms that are being developed in this era of big data. Soheili and Haeri, 2021 proposed the Scalable Global Mutual Information framework to deal with large-scale datasets, where they generate histograms of paired columns in a similarity matrix. Then they calculate the dependency criteria based on these histograms and fit an optimization algorithm to generate feature rankings. Other such methods aimed to address the scalability issues include the : Distributed quadratic programming based feature selection (DQPFS) by Soheili and Eftekhari-Moghadam, 2020 and, the BELIEF algorithm by López et al., 2021.

Optimal feature subset selection using exhaustive search is an NP-hard problem, so metaheuristic algorithms are used which aim to find the optimal solutions in the search space using certain strategies (inspired from evolution, hunting strategies, physics, etc.) that explore and exploit the search space efficiently without requiring exhaustive search. Metaheuristic algorithms are popularly used for feature selection as shown in survey papers (Chandrashekar & Sahin, 2014; Venkatesh & Anuradha, 2019; B. Xue et al., 2015). A large number of single-objective optimization algorithms have been proposed over the years. Nature-inspired algorithms like the classical Genetic Algorithm (Holland, 1992) and Harmony Search (Geem, 2009) algorithms are largely popular. Recently proposed algorithms include Whale Optimization Algorithm (Mirjalili & Lewis, 2016) and its upgraded versions like the AltWOA (Kundu et al., 2022) algorithm, Red Deer Algorithm (Fathollahi-Fard et al., 2020) and Aquila Optimizer (Abualigah, Yousri, et al., 2021). Algorithms inspired from other domains like Mathematics and Physics like the Arithmetic Optimization algorithm (Abualigah, Diabat, et al., 2021), Equilibrium Optimizer (Faramarzi et al., 2020) and Archimedes Optimization Algorithm (Hashim et al., 2021) have also been proposed which can be extended to address the single-objective feature selection problem. Kou et al., 2020 evaluated feature selection methods for text classification.

Y. Xue et al., 2019 developed a Self-adaptive Particle Swarm Optimization for feature selection and classification. Mahmoud et al., 2021 developed a Multi-Objective PSO (MOPSO) integrated with a fuzzy logic based on the "Technique for Order of Preference by Similarity to Ideal Solution", to address the Labyrinth Weir problem. Amoozegar and Minaei-Bidgoli, 2018 proposed a MOPSO algorithm for the feature selection problem, where the particles are guided by feature ranks computed based on their frequencies in the archive set. Zhang et al., 2015 developed a cost-based MOPSO algorithm for feature selection. The authors integrated a probability-based encoding scheme and an effective hybrid operator to enhance the performance of the MOPSO method.

Raj et al., 2020 proposed an Opposition-based Crow Search Algorithm for feature selection in the biomedical domain. Basak et al., 2021 proposed a two-step feature enhancement approach where they extracted features from the cervical cytology images and reduced them first using Principle Component Analysis preserving 99% variance and then applied the Grey Wolf Optimizer (Mirjalili et al., 2014) for the final feature selection and classification. Chattopadhyay et al., 2021 integrated a local search algorithm (the Adaptive Beta Hill Climbing algorithm) into a global optimization algorithm (the Sine-Cosine Algorithm) for feature selection on deep feature sets for classification of pneumonia disease.

Hybrid optimization algorithms have also been proposed over the years for feature selection to combine the salient properties of its constituent algorithms. Bhattacharyya et al., 2020 proposed a hybrid of Mayfly Algorithm (Zervoudakis & Tsafarakis, 2020) and Harmony Search (Geem, 2009), Sheikh et al., 2020 proposed a hybrid of Harmony Search and Artificial Electric Field optimization. Arivalagan and Venkatachalapathy, 2012 used a hybrid of Genetic Algorithm and Bacterial Foraging Optimization for feature selection in the face recognition domain. (Shunmugapriya & Kanmani, 2017) proposed a hybrid of Ant and Bee Colony Optimizers for feature selection on UCI datasets. Most of the proposed algorithms combine two wrapper-based algorithms which require large computational power for a very nominal increment in performance.

Single-objective optimization algorithms return one single solution to the problem which might not serve the purpose of all applications, thus hurting its generalization capability. In some cases, it might be more prudent to choose a lesser number of features and sacrificing the classification accuracy, rather than having to store a large number of features for obtaining a slight boost in performance. Multi-objective feature selection has thus, gained some attention from researchers over the years as well. Morita et al., 2003 proposed a method for unsupervised multi-objective feature selection and classification using NSGA-II and applied it on a handwritten word recognition problem while Hamdani et al., 2007 applied the NSGA-II algorithm on pattern recognition problems.

Lac and Stacey, 2005 used the NSGA-II algorithm and proposed a least crowded selection algorithm for multi-objective feature selection and classification on ionosphere data. B. Xue et al., 2012 proposed the Particle Swarm Optimization (PSO) algorithm in a multi-objective setting while Zhang et al., 2020 developed a binary Differential Evolution algorithm embedded with self-learning strategies for multi-objective feature selection. Recently, Dhiman et al., 2021 also proposed the Multi-Objective Seagull Optimization Algorithm (MOSOA) inspired from the foraging strategies of seagulls and applied it on 24 benchmark continuous optimization problems achieving performance comparable to state-of-the-art. All these multi-objective approaches described use random population initialization. However, using a guided initialization can greatly improve the convergence to the PF.

In this paper, the problem of multi-objective feature selection has been addressed by using a hybrid mechanism. However, unlike the popular approaches that combine two wrapper-based optimization algorithms incurring a heavy computation cost, filter-based feature ranking methods has been used, that incorporate the intrinsic properties of the dataset for setting ranks to the features and does not use a learning algorithm in its core thus being

computationally efficient. The filter methods has been used for initializing some of the population members in the NSGA-II multi-objective optimization algorithm that leads to faster convergence of the algorithm to the PF. To demonstrate the efficacy, the proposed Hybrid Filter Multi-Objective Evolutionary Algorithm (HFMOEA) has been evaluated on 18 datasets from the UCI machine learning repository and two deep feature sets extracted from a Pneumonia dataset (Kermany et al., 2018) and a Colorectal Cancer dataset (Kather et al., 2016).

3. Proposed Method

In this section, the proposed HFMOEA method has been described, wherein the filter-based ranking methods are used to generate the initial population for the NSGA-II algorithm. The filter methods used for this purpose and the NSGA-II algorithm are described in the following subsections.

3.1 Filter Methods

Filter-based feature selection methods use the intrinsic properties of the data to assert scores and thus rank the importance of the features. In this research, ten filter-based methods has been used for generating a part of the initial population of the NSGA-II algorithm for faster convergence of the solutions to the PF. These 10 filter methods are discussed in brief in this section.

3.1.1 Mutual Information

Mutual Information (MI) aims to measure the correlation between a class and the features. It is an entropy-based filter method, and for a random variable X, it attempts to quantify the degree of uncertainty in the distribution of events of X. The higher the probability of the occurrence of an event, the lower is its entropy, since in such cases the uncertainty is lower. On the other hand, for equally likely events, the entropy is highest since the certainty for any one particular outcome is minimum. Shannon's entropy (Shannon & Weaver, 1962) for a feature set X can be defined by Equation 1, where p_i is the probability mass function for $x_i \in X$.

$$H(X) = -\sum_i p_i \times \log_2 p_i \tag{1}$$

The conditional entropy $H(X|Y)$, that is the amount of uncertainty in determining X when Y is given, is given as in Equation 2, where $p(y, x)$ is the joint probability mass function of Y and X.

$$H(X|Y) = \sum_{y \in Y} p(y) \times H(X|Y = y) = \sum_{x \in X, y \in Y} p(y, x) \times \log_2 \frac{p(y)}{p(y, x)} \tag{2}$$

The minimum value of this conditional entropy is zero, and it occurs when X is dependent on Y statistically, i.e., the knowledge of Y leaves no uncertainty in the determination of X. Similarly, the maximum value of the conditional entropy $H(X|Y)$ occurs when X and Y are statistically independent, i.e., the information provided by Y does not provide any new information for the determination of X. Thus, MI is given by Equation 3 and Equation 4.

$$MI(X, Y) = \begin{cases} H(X) - H(X|Y) \\ H(Y) - H(Y|X) \\ H(X) + H(Y) - H(X, Y) \end{cases} \tag{3}$$

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \times \log_2 \frac{p(x, y)}{p(x) \times p(y)} \tag{4}$$

A higher value of mutual information between the feature X and class label Y signifies the higher importance of the feature for the classification of the sample. The primary advantage of MI is that it can detect general and possibly non-linear dependency relationships. However, the method is also computationally more expensive than other filter-based feature ranking methods.

3.1.2 Information Gain

Information Gain (IG) is an entropy-based measure that uses the Shannon's entropy (Shannon & Weaver, 1962) (Equation 1) to assert feature importance. Information quantifies the unexpectedness of an event. So, a lower probability event has more information than a higher probability event. The information gain for an attribute X having a class label Y is given by Equation 5.

$$IG(X, Y) = H(X) - IV(X) \tag{5}$$

$IV(X)$ represents the "Intrinsic Value" of the feature X and is expressed as Equation 6, where $n(X_i)$ represents the number of values in X having class label i and $values(Y)$ represents the number of unique classes in Y.

$$IV(X) = -\sum_{i \in values(Y)} \frac{n(X_i)}{X} \times \log_2 \left(\frac{n(X_i)}{X} \right) \tag{6}$$

IG is a good measure of feature importance, wherein it analyses comprehensively, the consequences along each branch of the decision tree and identifies the nodes which require further analysis. However, the IG method tends to overfit the training data and a small change in data can create large changes in the calculated IG for the attributes.

3.1.3 Method of Pasi Luukka

This method proposed by Luukka, 2011 uses a fuzzy-entropy measurement to assert feature importance. The fuzzy entropy (H) for attribute $x_i \in X$ is expressed as Equation 7, where μ_X are the fuzzy measures.

$$H(X) = -\sum_{i=1}^{n} \left[\mu_X(x_i) \times \ln(\mu_X(x_i)) + (1 - x_i) \times \ln(1 - \mu_X(x_i)) \right] \tag{7}$$

A lower value of this fuzzy entropy indicates greater feature importance. The primary advantage of this method of using fuzzy entropy is the low computational cost and its capability of robustly handling noise in the data.

3.1.4 Pearson's Correlation Coefficient

The Pearson's Correlation Coefficient (PCC) is a similarity-based measure that works on the following principle: the feature im-

portance is determined by the amount of correlation of the feature with the corresponding label. The higher the correlation, the more important is the feature. The correlation between two variables, say feature 'X' and label 'Y', lies between –1 and +1, where –1 represents total negative correlation (increase in X decreases Y), 0 means no correlation and +1 represents total positive correlation (increase in X increases Y). Thus PCC_X for feature x and label y is given by Equation 8, where $cov(X, Y)$ represents the covariance between X and Y, and the variances of distributions X and Y are given by σ_x and σ_y respectively.

$$PCC_X = \frac{cov(X, Y)}{\sigma_x \sigma_y} \tag{8}$$

The equation for PCC_X can be rewritten as in Equation 9 for $x_i \in X$, $y_i \in Y$, where n is the number of elements in X.

$$PCC_X = \frac{n \times \sum_{i=1}^{n}(x_i y_i) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n \times \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n \times \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} \tag{9}$$

A lower value of PCC_X indicates higher importance of the feature x. PCC can quantify the correlation between attributes in the feature space since it uses the method of covariance. It measures both the magnitude and direction of association between the features.

3.1.5 Spearman's Correlation Coefficient

Spearman's Correlation Coefficient (SCC) is a statistics-based measure and is defined as the PCC between the rank variables. For a sample of size n, the n attributes X_i, and labels Y_i are converted to ranks rg_{X_i}, rg_{Y_i} and the Spearman rank r_S is computed by Equation 10, where $cov(rg_{X_i}, rg_{Y_i})$ represents the covariance of rg_{X_i} and rg_{Y_i}, and the standard deviations of rg_{X_i} and rg_{Y_i} are denoted respectively by σ_{rg_X} and σ_{rg_Y}.

$$r_s = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}} \tag{10}$$

SCC, similar to PCC, helps to measure the strength and direction of the association between two attributes in the feature space when ranked by each of their quantities. SCC is useful to identify the relationship (sensitivity) between the measured results to influencing factors.

3.1.6 Relief Filter

The Relief algorithm, proposed by Kira and Rendell, 1992 is a similarity-based metric that computes the Euclidean distance to assert feature importance. For this, a feature X is selected randomly, and its two nearest neighbours, based on Euclidean distance are selected as such: one from the same class as X, represented as H, called nearest-hit, and the other neighbour from a different class, represented as M, called nearest-miss. Extension of "Relief" to the multi-class problem by using the Manhattan distance metric by Kononenko et al., 1996 is termed as "ReliefF". For two given instances I_1 and I_2, the difference function for attribute X is defined by Equation 11.

$$diff(X, I_1, I_2) = \frac{|val(X, I_1) - val(X, I_2)|}{max(X) - min(X)} \tag{11}$$

The weight for feature number i is updated as Equation 12, where, $x \in X$ are the features, W_{new} and W_{old} are the new and old weights respectively of the corresponding features, h_i is the nearest hit and m_i is the nearest miss. This expression is iterated over all the samples for the attribute. A higher value of weight indicates higher feature importance.

$$W_{new_i} = W_{old_i} - (x_i - h_i)^2 + (x_i - m_i)^2 \tag{12}$$

The Relief filter methods do not assume conditional independence of the features in the feature space to estimate their importance, unlike many heuristic algorithms. Relief is an efficient algorithm that can capture contextual information and assess the attribute quality in cases where there are strong dependencies between the attributes.

3.1.7 Chi-Square

The Chi-square measure is based on a statistical test that is used to compute the divergence from the expected distribution. The chi-square (χ^2) tests checks if a feature f is independent of the class value. It is given by Equation 13, where d is the number of distinct features, C is the number of distinct classes, n_{is} is the frequency of the i^{th} element having class $c \in C$.

$$\chi_f^2 = \sum_{i=1}^{d}\sum_{c=1}^{C} \frac{(n_{ic} - \mu_{ic})^2}{\mu_{ic}} \tag{13}$$

μ_{js} is given by Equation 14, where, n_{*c} represents the total number of elements in class 'c' and n_{i*} is the frequency of the i^{th} element.

$$\mu_{js} = \frac{n_{*c} \times n_{i*}}{n} \tag{14}$$

The higher the value of the X_f^2 metric, the higher is the importance of the feature f. The Chi-square measure is robust to the distribution of the data, can be computed with ease (i.e., low computational cost) and can capture detailed information from the data. The drawback of the method is that it is unable to interpret data classified into a large number of categories.

3.1.8 Mean Absolute Deviation

It is a statistics-based measure that computes the absolute deviations from a central point (mean) in the data. It is given by Equation 15, where n is the number of samples in the data, X is the feature whose mean absolute deviation (MAD) is being calculated, $x_i \in X$ and $\bar{X}$ is the mean of X.

$$MAD(X) = \frac{1}{n}\left[\sum_{i=1}^{n} |x_i - \bar{X}|\right] \tag{15}$$

A greater $MAD(X)$ indicates greater feature importance since a greater deviation means it carries more information. The advantage of using the MAD method is the simplicity of its formulation, the efficiency of the algorithm and its robustness towards extreme scores since there is no squaring operation as in standard deviation calculation.

3.1.9 Dispersion Ratio

Dispersion Ratio (DR) is a statistics-based measure that is the normalized measure of the dispersion of a distribution. DR quantifies whether a set of attributes are clustered (similar to each other) or dispersed as compared to a standard statistical model. It is given by Equation 16, where σ_X^2 represents the variance of attribute X and μ_X represents the mean of X.

$$DR(X) = \frac{\sigma_X^2}{\mu_X} \tag{16}$$

A higher $DR(X)$ value indicates greater importance of the feature X. The simplicity in the formulation of the metric is its main advantage. However, it is sensitive to outlier data points and does not utilize all the observations in the feature set.

3.1.10 Fisher Score

The Fisher Score is a statistics-based method that aims to find a subset of features in the data where the Euclidean distance between the attributes belonging to the same class are as small as possible and the distance between the attributes belonging to different classes are as high as possible. For a feature $x_i \in X$, the fisher score for the feature ($F(x_i)$) is computed using Equation 17, where C is the number of classes in the data, n is the number of samples, μ_c^i and σ_c^i are the mean and standard deviations of class $c \in C$ corresponding to feature number i and μ^i is the mean of the i^{th} feature.

$$F(x_i) = \frac{\sum_{c=1}^{C} n(\mu_c^i - \mu^i)^2}{\sum_{c=1}^{C} n(\sigma_c^i)^2} \tag{17}$$

The Fisher scoring method is computationally efficient and has transformation invariance properties (Osborne, 1992) making it a good filter ranking method.

3.2 Proposed HFMOEA Method

In general, multi-objective optimization algorithms optimize two or more competing objectives simultaneously. In this study, the two competing objective functions are: (1) maximizing classification accuracy and (2) minimizing the number of selected features. A reduction in the number of features means lesser information available for the classification task which leads to a decrease in the classifier performance, thus making the two objectives competing. The general framework of proposed HFMOEA is based on Non-dominated Sorting Genetic Algorithm-II (NSGA-II) proposed by Deb et al., 2000. In addition, binary chromosomes are used in this study to determine whether a feature is selected or not. The classification accuracy was computed by fitting the training data with the selected features to a Support Vector Machines (SVM) classifier.

HFMOEA, similar to NSGA-II, uses a fast non-dominated sorting scheme along with a crowding distance for ensuring sparsity in solutions in the PF and employs bi-modal crossover and polynomial operators to generate offspring from the population in a generation. Figure 2 shows the flowchart for the proposed HFMOEA algorithm. The best individuals in the PF are selected based on non-dominance and diversity. The method for finding the non-dominated solutions, calculating the crowding distance and the crowding selection is explained in the following subsections.

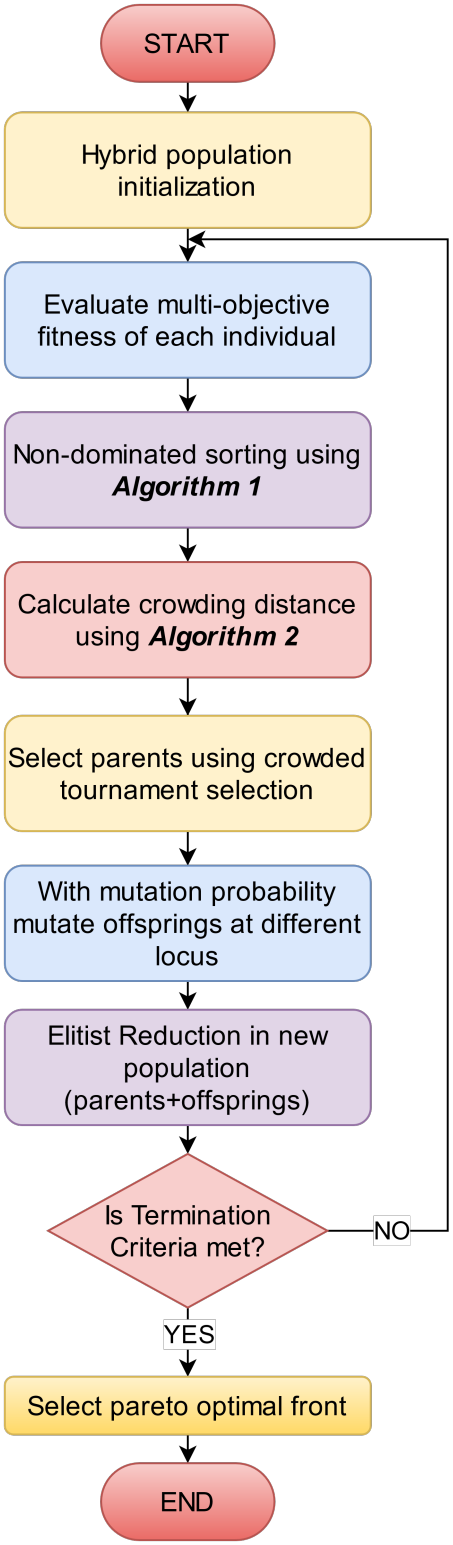


Figure 2: Flow chart diagram for the workflow of HFMOEA algorithm.

3.2.1 Hybrid Initialization

In the HFMOEA, the 10 filter methods described are used to rank the features in the data. Then, the top 'k' ranked features are selected, i.e., set as '1' is the binary encoded chromosome of the population and the rest of the features are not selected, i.e., set as '0'. Now, since a population size of 10 is too small, another 90 individuals are generated through random initialization, thus forming a total population of size 100. The inclusion of the individuals from the filter methods ensures a faster convergence to the PF. For our research, the top 50% of the ranked features are selected from the filter methods (set to 1 in the chromosome) and the rest are set to 0.

Several hybrid methods in the feature selection literature use wrapper methods to initialize the population, which is very computationally expensive for only a small relative boost in the performance. Thus, in the HFMOEA algorithm, only filter methods have been used for guiding the population.

The guided initialization through the filter ranking methods produces superior individuals than the randomly initialized population. These 10 solutions obtained are diverse, which is confirmed by the correlation analysis shown in Section 4 (Table 5). However, apart from these 10 individuals, 90 other individuals are generated randomly, which ensures that the population is diverse. Now, in the NSGA-II algorithm during the offspring generation phase, these superior genes are exchanged between individuals. This spreads the superior genes to more individuals throughout the generations, and the mutation phases ensure diversity of solutions. During the evolution process over the iterations, the unimportant features are discarded due to the guided initialization. This process leads to the faster convergence of the HFMOEA to the PF than randomly initialized NSGA-II. Since such chromosome crossover and mutation operators are characteristic of the NSGA-II algorithm, this algorithm has been chosen for the guided initialization even though other multi-objective algorithms are popularly used in literature. Thus, good results are obtained with fast convergence on the feature selection problem.

At first glance, it seems like the use of filter methods to generate part of the initial population for the HFMOEA algorithm increases the computational complexity of the method. However, due to such a guided initialization scheme, the algorithm converges to the PF much faster than traditional multi-objective algorithms, making the overall computational time required much lower, that is, better convergence behavior is obtained in only a few iterations.

3.2.2 Dominance Ranking

For a problem having M (competing) objectives, a solution s_1 dominates over solution s_2 (represented as $s_1 \succ s_2$) if the following two conditions are satisfied:

(i) s_1 is not worse than s_2 for all the M objectives.
(ii) s_1 is strictly better than s_2 in atleast one of the M objectives.

The algorithm for finding the non-dominated members of the population (P) in a generation is shown in Algorithm 1.

3.2.3 Crowding Distance

Crowding distance measures the density of solutions in the search space, i.e., it provides an estimate of the number of solutions that surround a particular solution under consideration. The algorithm for computing the crowding distance is shown in Algorithm 2, where the objective functions are $\{f_1, f_2, \ldots, f_M\}$, N is the number of solutions in a generation and c_i is the crowding distance of that solution.

Algorithm 1 Determination of Domination Rank

1: Initialize rank counter r to 1 and population P
2: **while** P is not empty **do**
3: N = non-dominated individuals from P based on the definition of domination
4: Assign rank r to population N
5: Remove N from P
6: $r := r + 1$
7: **end while**
8: Return sorted population with their ranks

Algorithm 2 Crowding Distance Computation

1: Initialize crowding distance $c_i = 0$ for $i = 1, 2, \ldots, N$.
2: **for** $o = 1$ **to** M **do**
3: Objective function: f_o
4: Set $c_1 = c_N = \infty$
5: **for** $j = 2$ **to** $N - 1$ **do**
6: $c_j = c_j + (f_{o_{j+1}} - f_{o_{j-1}})$
7: **end for**
8: **end for**
9: Return crowding distance c.

3.2.4 Crowding Selection

For comparison between two individuals (solutions of NSGA-II) s_1 and s_2, a crowded tournament selection is employed, wherein solution s_1 is selected over s_2 if Equation 18 is satisfied, where r_s is the domination rank of solution s and c_s is the crowding distance of the solution s.

$$s_1 \succ s_2, \text{ iff } \left[r_{s_1} < r_{s_2}\right] \textbf{ or } \left[(r_{s_1} = r_{s_2}) \textbf{ and } (c_{s_1} > c_{s_2})\right] \tag{18}$$

Thus, a solution with lower domination is rank is selected, and if the ranks for the two solutions are the same, the one with a larger crowding distance is selected.

4. Results and Discussion

To evaluate the HFMOEA method of multi-objective feature selection, the algorithm is evaluated on 18 datasets from the UCI machine learning repository (Dua & Graff, 2017) and 2 image datasets from which deep features have been extracted and classified after feature selection. The results for the experiments are reported in the following subsections.

4.1 Results on the UCI Datasets

A brief description of the 18 UCI datasets used in this research is presented in Table 1.

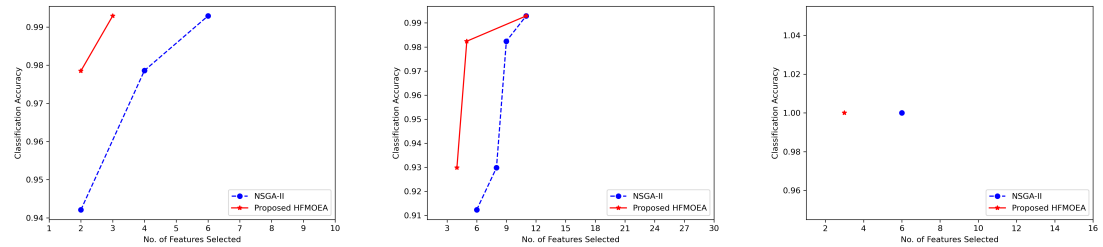
The results (Pareto-fronts) obtained by the proposed HFMOEA framework on the 18 UCI datasets are shown in Figure 3 and Figure 4 and are compared with the PF obtained by the NSGA-II algorithm with random population initialization. In some cases, such as in datasets "CongressEW", "PenglungEW", "Sonar", "Vote", "WaveformEW", "WineEW" and "Zoo", only one point has been obtained as the PF by the hybrid MOEA framework. In most of these cases, the maximum possible accuracy is achieved with only a few features and thus no other non-dominated solution exists. As can be seen from the figures, the hybrid MOEA framework performs significantly better than the NSGA-II algorithm with random initialization (or in some cases, similar performance is obtained). In general, the performance of the proposed algorithm is optimal and superior. The results obtained justify the use of the filter

Table 1: Description of the 18 UCI datasets used in this research to evaluate the performance of the proposed HFMOEA framework.

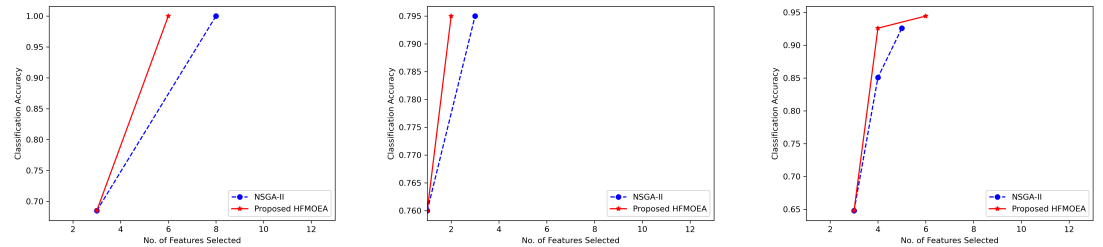
SL No.	Dataset	No. of Attributes	No. of Samples	No. of Classes	Domain
1	BreastCancer	10	699	2	Biology
2	BreastEW	30	568	2	Biology
3	CongressEW	16	434	2	Politics
4	Exactly	13	1000	2	Biology
5	Exactly2	13	1000	2	Biology
6	HeartEW	13	270	2	Biology
7	Ionosphere	34	351	2	Electromagnetism
8	KrVsKpEW	36	3196	2	Game
9	Lymphography	18	148	4	Biology
10	M-of-n	13	1000	2	Biology
11	PenglungEW	325	73	7	Biology
12	Sonar	60	208	2	Biology
13	SpectEW	22	267	2	Biology
14	Tic-tac-toe	9	958	2	Game
15	Vote	16	300	2	Politics
16	WaveformEW	40	5000	3	Physics
17	WineEW	13	178	3	Chemistry
18	Zoo	16	101	7	Artificial

Table 2: Description of the datasets used for deep feature extraction and classification using hybrid MOEA framework.

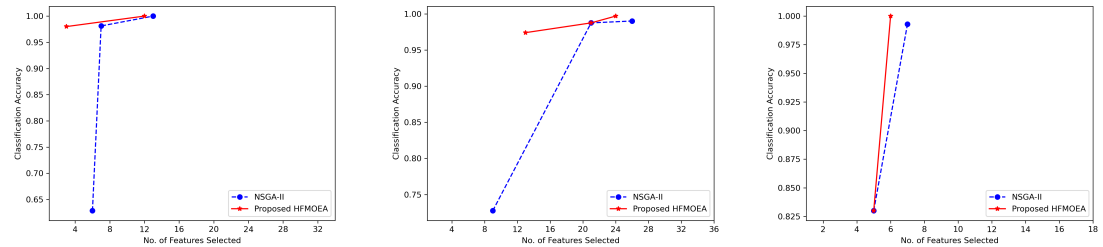
Dataset	Class	Category	No. of Images
Pneumonia (Kermany et al., 2018)	1	Pneumonia	4273
	2	Normal	1583
Colorectal Cancer (Kather et al., 2016)	1	Tumour Epithelium	625
	2	Simple Stroma	625
	3	Complex Stroma	625
	4	Immune Cells	625
	5	Debris	625
	6	Normal Mucosal Glands	625
	7	Adipose Tissue	625
	8	Background (no tissue)	625



(a) BreastCancer **(b)** BreastEW **(c)** CongressEW

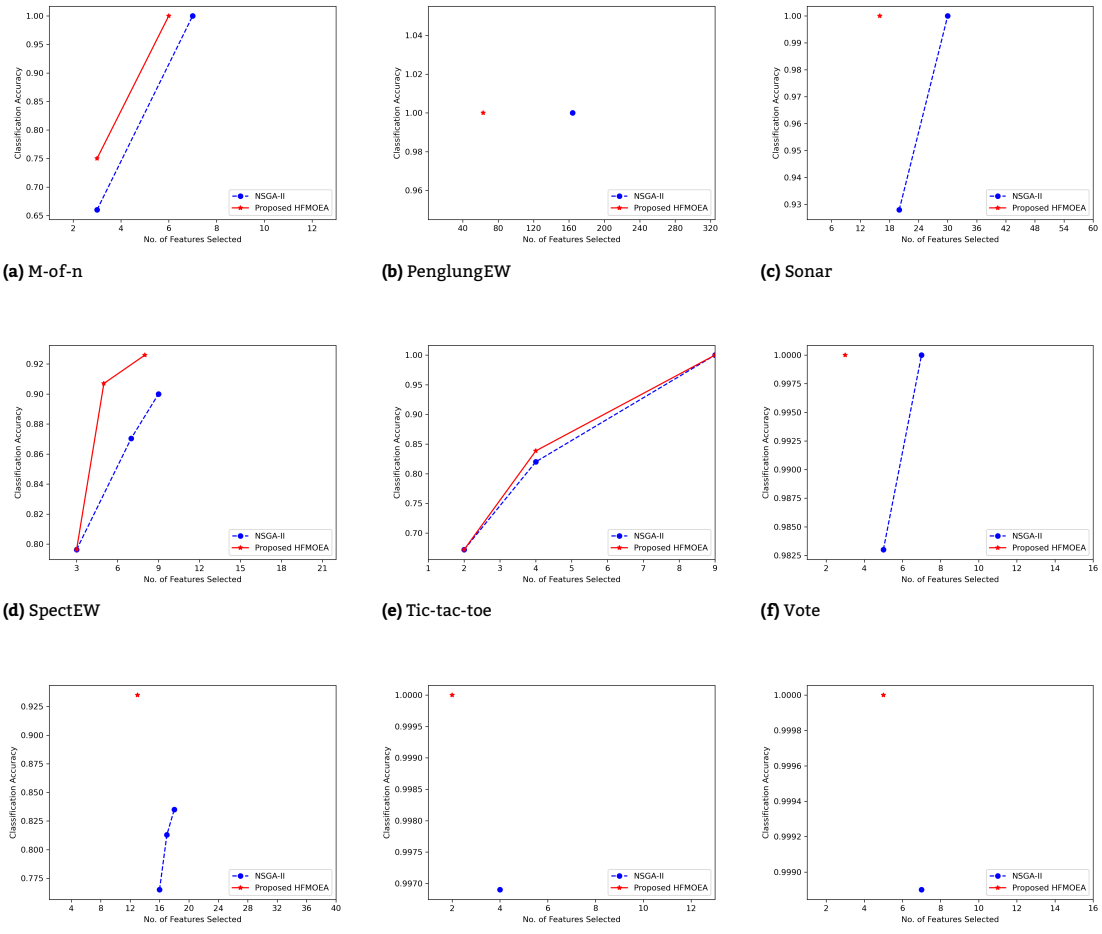


(d) Exactly **(e)** Exactly2 **(f)** HeartEW



(g) Ionosphere **(h)** KrVsKpEW **(i)** Lymphography

Figure 3: Pareto-optimal fronts obtained on the UCI datasets 1-9 by the proposed HFMOEA algorithm (red stars, solid lines) and the NSGA-II algorithm with random initialization (blue circles, dotted lines). The vertical axis represents the accuracy and the horizontal axis represents the number of features selected.



(a) M-of-n **(b)** PenglungEW **(c)** Sonar

(d) SpectEW **(e)** Tic-tac-toe **(f)** Vote

(g) WaveformEW **(h)** WineEW **(i)** Zoo

Figure 4: Pareto-optimal fronts obtained on the UCI datasets 10-18 by the proposed HFMOEA algorithm (red stars, solid lines) and the NSGA-II algorithm with random initialization (blue circles, dotted lines). The vertical axis represents the accuracy and the horizontal axis represents the number of features selected.

Table 3: Pareto optimal solutions obtained using the hybrid MOEA framework on the datasets that have 2 solutions in the PF. *Acc*: Accuracy, *FS*: Number of Features Selected.

Dataset	PF (2 solutions)	
	Acc	FS
BreastCancer	0.9785	2
	0.9929	3
Exactly	0.6850	3
	1.0000	6
Exactly2	0.7600	1
	0.7950	2
Ionosphere	0.9800	3
	1.0000	12
Lymphography	0.8300	5
	1.0000	6
M-of-n	0.7500	3
	1.0000	6

Table 4: Pareto optimal solutions obtained using the hybrid MOEA framework on the datasets that have 3 or more solutions in the PF. *Acc*: Accuracy, *FS*: Number of Features Selected.

Dataset	PF ($\geq$3 solutions)	
	Acc	FS
BreastEW	0.9298	4
	0.9824	5
	0.9929	11
HeartEW	0.6480	3
	0.9260	4
	0.9444	6
KrVsKpEW	0.9740	13
	0.9875	21
	0.9969	24
SpectEW	0.7963	3
	0.9070	5
	0.9259	8
Tic-tac-toe	0.6719	2
	0.8385	4
	1.0000	9

methods for initializing some of the individuals in the population of the NSGA-II algorithm for faster and optimal convergence.

Some of the datasets have two solutions in their PF. The Pareto-optimal solutions for these are shown in Table 3. Here, the options are binary: either a solution with a lesser number of features, i.e., requiring lesser storage space or a solution with higher accuracy of classification. In some cases, like in datasets "Exactly" and "M-of-n", the two solutions are markedly different, i.e., the accuracy difference and the difference in the number of features selected between the two solutions are large. For example, in dataset "Exactly", a 68.50% accuracy is obtained with 3 features while a 100% accuracy is obtained with 6 features. In some other cases, like in datasets "BreastCancer" and "Exactly2," the difference between the solutions is not much: the solutions differ by 1 number of features and only about 2-3% accuracy difference. However, in dataset "Lymphography", it can be noted that a difference of 1 feature between the two solutions (5 and 6 features selected by the two solutions) is changing the accuracy by 17%. In such cases, it is more prudent to choose the one with higher accuracy. On the other hand, in dataset "Ionosphere", between the two solutions, there is only an accuracy difference of 2% (98% and 100%) while the number of features differs by 9 (3 and 12 features selected by the two solutions). In such cases, selecting the solution with the lower accuracy is more justifiable.

Table 4 tabulates the Pareto-optimal solutions from the datasets that have three or more solutions in the PF. From the table, it can be seen that for cases like datasets "HeartEW", "SpectEW" and "Tic-tac-toe", the solutions differ a lot both in terms of accuracy and the number of features selected, so the trade-off solutions are well captured by the HFMOEA algorithm and the appropriate solution can be chosen based on the requirement of the particular application. On the other hand, in cases like datasets "BreastEW" and "KrVsKpEW", the difference in accuracies obtained by the different solutions are small, however, the number of features selected differ significantly. However, in all these cases, a significant amount of choices are present to select the solution best suited for the particular application: one with high accuracy, one with a markedly less number of features, and one in the intermediate region.

To justify that the ranks assigned to the features of the same dataset by the 10 different filter methods used in this study are diverse, the Kendall's Rank Correlation test or Kendall's Concordance test (Abdi, 2007) is performed using one of the datasets as an example. It is a non-parametric statistical test used to measure the ordinal association between ranked data. The BreastCancer data has been used to evaluate the filter ranking methods and the results for the same are shown in Table 5. A greater absolute value of the test indicates a stronger correlation between the ranks and a lower correlation is indicative of the diversity between the ranks. From the table, clearly, the correlation values for most of the methods are low (less than 0.5 in maximum cases), which supports the argument that the filter methods produce distinct individuals in the population. Note that, in the table, each of the diagonal entries is 1.00, i.e., each method is perfectly correlated to itself, and the table is symmetric along the diagonal.

The best-accuracy solutions obtained by the proposed HFMOEA methods are compared to the filter-based ranking methods (combined with SVM classifier) and the results are reported in Table 6. As seen from the table, HFMOEA consistently outperforms all the filter methods, with Mutual Information being the second-best method.

Some of the popular single-objective wrapper-based metaheuristics in the literature are as follows:

(i) Mayfly Algorithm (MA) by Zervoudakis and Tsafarakis, 2020
(ii) Harmony Search (HS) by Geem, 2009
(iii) Genetic Algorithm (GA) by Holland, 1992
(iv) Particle Swarm Optimization (PSO) by Kennedy and Eberhart, 1995
(v) Whale Optimization Algorithm (WOA) by Mirjalili and Lewis, 2016
(vi) Gravitational Search Algorithm (GSA) by Rashedi et al., 2009

The solutions with the highest accuracies in the PF are are compared with the best results obtained by different wrapper-based single-objective optimization algorithms (where the algorithms only focus on maximizing accuracy) on the datasets. The parameter settings for these metaheuristics used for comparison are shown in Table 7 and the results obtained on comparison are shown in Table 8. From the table it is evident that the best-accuracy results obtained from the PF generated by the proposed HFMOEA framework outperforms the best results obtained by the single objective metaheuristics which focus on maximizing the accuracy. The Mayfly Algorithm (MA) performs closest to the HFMOEA framework in some cases but is still markedly inferior. This proves the robustness in the performance of the HFMOEA algorithm.

Table 5: Results from Kendall's Concordance test for determining rank correlation on the BreastCancer dataset.

Kendall's Concordance	MI	SCC	Relief	PCC	Chi Square	IG	MAD	DR	Pasi Luukka	Fisher Score
MI	1.00	-0.33	0.06	0.22	-0.67	0.00	-0.33	0.00	0.11	0.44
SCC	-0.33	1.00	0.28	0.00	0.33	0.00	0.44	-0.22	-0.78	-0.33
Relief	0.06	0.28	1.00	0.39	-0.28	-0.06	0.06	-0.72	-0.50	0.06
PCC	0.22	0.00	0.39	1.00	0.00	-0.56	0.33	-0.22	-0.22	-0.33
Chi Square	-0.67	0.33	-0.28	0.00	1.00	-0.11	0.56	0.33	-0.11	-0.67
IG	0.00	0.00	-0.06	-0.56	-0.11	1.00	-0.11	0.11	0.11	0.22
MAD	-0.33	0.44	0.06	0.33	0.56	-0.11	1.00	0.22	-0.33	-0.89
DR	0.00	-0.22	-0.72	-0.22	0.33	0.11	0.22	1.00	0.44	-0.33
Pasi Luukka	0.11	-0.78	-0.50	-0.22	-0.11	0.11	-0.33	0.44	1.00	0.22
Fisher Score	0.44	-0.33	0.06	-0.33	-0.67	0.22	-0.89	-0.33	0.22	1.00

Table 6: Comparison of the best accuracy solutions obtained by the proposed HFMOEA and the accuracies obtained by the filter methods used in this paper.

Dataset	HFMOEA	MI	SCC	Relief	PCC	Chi-Square	IG	MAD	DR	Pasi Luukka	Fisher Score
BreastCancer	**0.9929**	0.6357	0.6357	0.6357	0.6357	0.6357	0.6357	0.6357	0.6357	0.6357	0.6357
BreastEW	**0.9929**	0.9386	0.9386	0.9386	0.9386	0.9386	0.9386	0.9386	0.9386	0.9386	0.9386
CongressEW	**1.0000**	0.5632	0.5632	0.5632	0.5632	0.5632	0.5632	0.5632	0.5632	0.5632	0.5632
Exactly	**1.0000**	0.6550	0.6550	0.6550	0.6550	0.6550	0.6550	0.6550	0.6550	0.6550	0.6550
Exactly2	**0.7950**	0.7550	0.7550	0.7550	0.7550	0.7550	0.7550	0.7550	0.7550	0.7550	0.7550
HeartEW	**0.9444**	0.5741	0.5741	0.5741	0.5741	0.5741	0.5741	0.5741	0.5741	0.5741	0.5741
Ionosphere	**1.0000**	0.6286	0.6286	0.6286	0.6286	0.6286	0.6286	0.6286	0.6286	0.6286	0.6286
KrVsKpEW	**0.9969**	0.9969	0.9969	0.9969	0.9969	0.9969	0.9969	0.9969	0.9969	0.9969	0.9969
Lymphography	**1.0000**	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333
M-of-n	**1.0000**	0.6250	0.6250	0.6250	0.6250	0.6250	0.6250	0.6250	0.6250	0.6250	0.6250
PenglungEW	**1.0000**	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667
Sonar	**1.0000**	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333	0.8333
SpectEW	**0.9259**	0.7963	0.7963	0.7963	0.7963	0.7963	0.7963	0.7963	0.7963	0.7963	0.7963
Tic-tac-toe	**1.0000**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Vote	**1.0000**	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667
WaveformEW	**0.9350**	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740
Wine	**1.0000**	0.9167	0.9167	0.9167	0.9167	0.9167	0.9167	0.9167	0.9167	0.9167	0.9167
Zoo	**1.0000**	0.4500	0.4500	0.4500	0.4500	0.4500	0.4500	0.4500	0.4500	0.4500	0.4500

Table 7: Parameter settings used for the state-of-the-art metaheuristics used for comparison.

OA	Parameter(s)	Value(s)
MA	Positive Attraction Constants (a_1, a_2)	$a_1 = 1$, $a_2 = 1.5$
	Probability of Mutation (M)	$M = 0.3$
	Gravitational Coefficient (g)	$g = 0.8$
	Visibility Coefficient (β)	$\beta = 2$
	Coefficient of Nuptial Dance (d)	$d = 0.1$
HS	Harmony Memory Size (HMS)	HMS is same as population size
	Harmony Memory Considering Rate ($HMCR$)	$HMCR = 0.9$]
	Pitch Adjustment Rate (PAR)	$PAR = 0.3$
GA	Probability of crossover (C)	$C = 90\%$
	Probability of Mutation (M)	$M = 6\%$
PSO	Inertia weight (I)	I lies in [0, 1]
	Constants (C_1, C_2)	$C_1 = 2$, $C_2 = 2$
WOA	Encircling parameter (a)	a lies in [0, 2]
	Shape of spiral (b)	$b = 1$
GSA	Initial Gravitational constant ($G_{initial}$)	$G_{initial} = 6$
	Decay factor (α)	$\alpha = 20$

Table 8: Comparison of the best-accuracy solutions obtained by the proposed HFMOEA and the optimal solutions obtained by single-objective wrapper-based evolutionary metaheuristics. *Acc*: Accuracy, *FS*: Number of Features Selected.

Dataset	HFMOEA		MA		HS		GA		PSO		WOA		GSA	
	Acc	FS	Acc	FS	Acc	FS	Acc	FS	Acc	FS	Acc	FS	Acc	FS
BreastCancer	**0.9929**	3	0.9785	2	0.9420	3	0.9740	4	0.9630	4	0.9571	5	0.9686	4
BreastEW	**0.9929**	11	0.9400	6	0.9560	10	0.9754	8	0.9719	9	0.9553	21	0.9544	10
CongressEW	**1.0000**	3	0.9770	3	0.9540	6	0.9679	2	0.9633	3	0.9290	11	0.9633	4
Exactly	**1.0000**	6	0.9900	7	0.6850	3	1.0000	6	1.0000	6	0.7500	6	0.9940	4
Exactly2	**0.7950**	2	0.7600	1	0.7600	3	0.7700	1	0.7680	1	0.6900	6	0.7700	1
HeartEW	**0.9444**	6	0.8700	4	0.8510	5	0.8741	5	0.8370	3	0.7600	9	0.8296	3
Ionosphere	**1.0000**	12	0.9100	12	0.9000	12	0.9489	7	0.9489	7	0.8900	22	0.9432	9
KrVsKpEW	**0.9969**	24	0.9710	21	0.9380	17	0.9850	11	0.9731	12	0.9151	28	0.9549	14
Lymphography	**1.0000**	6	0.8600	8	0.8330	7	0.8378	5	0.8919	5	0.7850	11	0.8649	6
M-of-n	**1.0000**	6	1.0000	6	0.7500	3	1.0000	6	1.0000	6	0.8500	10	0.9940	5
PenglungEW	**1.0000**	63	1.0000	150	1.0000	165	0.9189	84	0.9189	130	0.7290	145	0.8333	140
Sonar	**1.0000**	16	0.9280	28	0.8570	20	0.9904	19	0.9423	22	0.8543	43	0.9135	24
SpectEW	**0.9259**	8	0.8700	9	0.7960	5	0.8955	5	0.8881	6	0.7870	11	0.8433	5
Tic-tac-toe	**1.0000**	9	0.8200	5	0.7760	4	0.7996	5	0.7996	6	0.7511	11	0.7766	4
Vote	**1.0000**	3	0.9830	5	0.9330	5	0.9733	5	0.9600	3	0.9387	8	0.9600	4
WaveformEW	**0.9350**	13	0.8130	17	0.7880	16	0.7836	15	0.7560	15	0.7120	33	0.7344	14
WineEW	**1.0000**	2	0.9700	4	0.9160	3	0.9888	4	0.9775	5	0.9280	9	0.9775	4
Zoo	**1.0000**	5	1.0000	10	1.0000	10	0.9020	4	0.9608	5	0.9647	10	0.9804	6

Table 9: Solutions on the Pareto-Front (PF) obtained by the HFMOEA algorithm on the two deep feature sets. *Acc*: Accuracy, *FS*: Number of features selected out of 512 features.

Dataset	PF	
	Acc	FS
Pneumonia (Kermany et al., 2018)	0.978	101
	0.980	225
	0.981	240
	0.984	244
Colorectal Cancer (Kather et al., 2016)	0.967	52
	0.968	53
	0.975	91
	0.977	109
	0.979	122

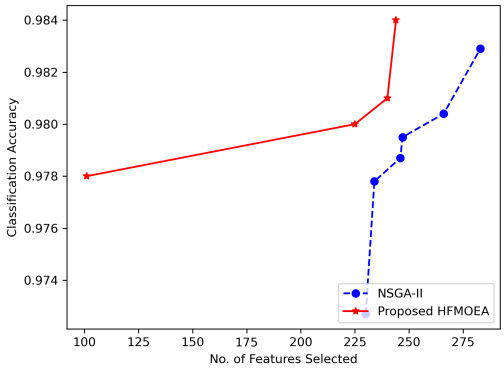
4.2 Results on the Deep Feature Sets

To further justify the robustness in performance of the proposed HFMOEA framework, HFMOEA has been evaluated it on two deep feature sets, i.e, features have been extracted using deep learning models on two image datasets from different domains:

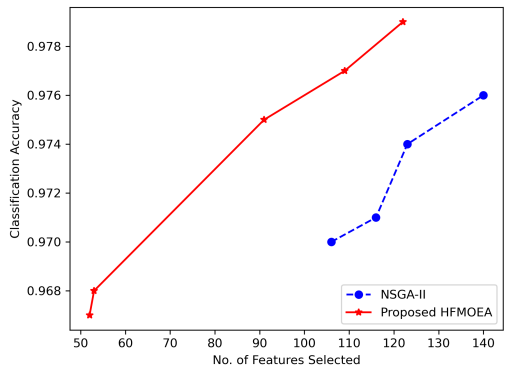
(i) Pneumonia chest X-ray dataset (2-Class) by Kermany et al., 2018
(ii) Colorectal Cancer histopathology dataset (8-class) by Kather et al., 2016

The description of the datasets is shown in Table 2. From the Pneumonia dataset (Kermany et al., 2018) 512 features were extracted from the penultimate layer of the DenseNet-201 model using pre-trained weights of ImageNet (Transfer Learning), and from the Colorectal Cancer dataset (Kather et al., 2016), the MobileNet v2 model is used with transfer learning for feature extraction (256 features). The models have been chosen for optimal performance. The hyperparameters set for the models to extract the features are as follows:

(i) Optimizer: Stochastic Gradient Descent
(ii) Loss Function: Categorical Cross-entropy
(iii) Learning Rate: 1×10^{-5}
(iv) Batch Size: 32



(a) Pneumonia dataset (Kermany et al., 2018)



(b) Colorectal Cancer dataset (Kather et al., 2016)

Figure 5: Pareto fronts obtained by the proposed HFMOEA framework (red stars, solid line) and NSGA-II with random population initialization (blue circles, dotted lines) on the two image datasets from which deep features have been extracted. The vertical axis represents the accuracy and the horizontal axis represents the number of features selected.

Table 10: Comparison of the best-accuracy solutions obtained by the proposed HFMOEA framework with existing methods in the literature on the Pneumonia (Kermany et al., 2018) and Colorectal Cancer (Kather et al., 2016) deep feature sets.

Dataset	Method	Approach	Accuracy (%)
Pneumonia	Mahmud et al., 2020	Novel CovXNet CNN model	98.10
	Zubair, 2020	Transfer Learning with VGG-16	96.60
	Stephen et al., 2019	Novel CNN model	93.73
	Sharma et al., 2020	CNN models	90.68
	Liang and Zheng, 2020	Transfer Learning	90.50
	HFMOEA	**Deep Features+Hybrid MOEA**	**98.40**
Colorectal Cancer	Paladini et al., 2021	Ensemble of CNNs	96.16
	Ohata et al., 2021	Transfer Learning	92.08
	Kather et al., 2016	Texture Analysis	87.40
	Tellez et al., 2019	Unsupervised stain colour normalization	79.66
	HFMOEA	**Deep Features+Hybrid MOEA**	**97.90**

Table 11: Results obtained on performing the McNemar's Statistical Test of the HFMOEA algorithm against the filter methods used for comparison.

Dataset	HFMOEA vs. MI	HFMOEA vs. SCC	HFMOEA vs. Relief	HFMOEA vs. PCC	HFMOEA vs. Chi-Square	HFMOEA vs. IG	HFMOEA vs. MAD	HFMOEA vs. DR	HFMOEA vs. Pasi Luukka	HFMOEA vs. Fisher Score
BreastCancer	2.73E-02	2.85E-02	2.02E-02	2.97E-02	4.85E-02	4.33E-02	4.10E-02	4.00E-02	3.16E-02	1.75E-02
BreastEW	4.63E-02	1.62E-03	2.97E-02	1.18E-02	1.47E-02	3.24E-02	1.31E-02	2.02E-02	1.83E-02	3.01E-02
CongressEW	1.85E-02	4.20E-02	3.08E-02	3.26E-02	4.97E-02	1.80E-02	2.10E-02	8.46E-03	3.05E-02	2.58E-02
Exactly	3.16E-02	1.55E-02	1.32E-02	3.64E-02	4.95E-02	2.53E-02	9.45E-03	4.03E-02	2.97E-02	5.33E-03
Exactly2	2.37E-02	6.12E-04	4.11E-02	4.23E-02	7.01E-03	3.13E-02	4.37E-02	1.84E-03	2.83E-02	4.77E-02
HeartEW	2.37E-03	3.36E-02	1.81E-03	2.00E-02	1.36E-03	4.07E-02	4.22E-02	2.15E-02	2.28E-02	2.47E-02
Ionosphere	1.56E-02	4.18E-02	3.15E-02	4.19E-02	1.14E-03	3.08E-02	2.65E-02	1.30E-02	2.67E-02	3.60E-02
KrVsKpEW	2.07E-02	5.23E-03	2.18E-02	9.13E-03	2.31E-02	3.33E-02	3.96E-02	1.43E-03	4.50E-02	7.43E-04
Lymphography	4.40E-02	2.55E-02	5.74E-03	2.57E-02	1.19E-02	4.15E-03	2.65E-02	4.42E-02	4.59E-02	1.77E-02
M-of-n	3.54E-02	2.70E-02	4.30E-02	2.81E-02	9.97E-03	9.65E-04	4.52E-04	2.68E-02	1.30E-02	2.44E-02
PenglungEW	2.03E-02	3.68E-03	3.51E-03	2.08E-02	2.58E-02	4.35E-02	1.30E-02	4.79E-02	3.80E-02	3.06E-02
Sonar	2.93E-03	1.20E-02	1.91E-02	4.00E-03	3.05E-02	2.40E-03	1.29E-02	2.43E-02	4.82E-02	1.41E-02
SpectEW	3.37E-02	1.56E-02	3.72E-02	5.21E-03	4.25E-02	9.04E-03	1.46E-02	2.44E-02	4.35E-02	1.10E-02
Tic-tac-toe	2.56E-02	3.85E-02	1.55E-02	3.90E-02	8.93E-03	4.16E-02	1.92E-02	3.15E-02	4.51E-03	2.57E-02
Vote	1.54E-02	2.58E-02	1.56E-02	3.31E-02	2.27E-02	3.61E-02	6.76E-05	4.71E-02	4.65E-02	1.94E-02
WaveformEW	2.11E-02	4.63E-02	1.14E-02	1.22E-02	1.14E-02	1.88E-03	4.97E-02	2.46E-02	6.15E-03	2.41E-02
Wine	3.13E-03	3.49E-02	3.02E-02	1.77E-02	3.20E-02	3.11E-02	3.27E-02	7.25E-03	1.63E-03	2.22E-02
Zoo	8.31E-05	8.99E-03	2.59E-02	3.58E-02	3.15E-02	1.12E-02	8.86E-03	4.38E-02	3.69E-02	2.34E-02

(v) Dropout Rate: 20%
(vi) Number of Epochs: 50

The results (PF) obtained by the hybrid MOEA approach and the NSGA-II algorithm (with random population initialization) on the aforementioned two deep feature sets are shown in Figure 5. The solutions of the hybrid MOEA algorithm dominates the solutions of the NSGA-II algorithm. The solutions on the PF of the proposed HFMOEA method are shown in Table 9. Several solutions are present in the PF and the optimal solution can be chosen according to the requirements: higher accuracy or lower number of features or intermediate solutions that presents a satisfactory trade-off.

Table 10 shows the comparison of the best-accuracy solutions obtained by the HFMOEA algorithm with existing methods in literature on these datasets. The results obtained justify the robustness in performance of the HFMOEA approach. On both datasets, the methods that performed closest to our HFMOEA method uses Convolutional Neural Network (CNN) based approaches, while the HFMOEA algorithm of deep feature extraction and classification outperforms these purely deep learning-based methods for end-to-end classification.

4.3 Statistical Analysis

To statistically analyse the significance of the proposed HFMOEA method against the popular wrapper and filter methods used in the feature selection literature, the McNemar's statistical test is performed. The McNemar's test is a non-parametric statistical significance test that assumes the null hypothesis that two models are statistically similar. To reject the hypothesis, the $p - value$ from the test must be less than 5% or 0.05.

Table 11 and Table 12 respectively shows the results from the test obtained from testing our proposed method against filter and wrapper methods. As can be observed from the tables, the p-value is less than 0.05 in every case, justifying that the proposed model is statistically significant, and dissimilar to existing methods.

4.4 Discussion

The time complexity of the proposed HFMOEA algorithm is $O(MN^2)$ where "M" is the number of objectives and "N" is the dimensionality of the dataset. The time complexity is identical to that of the NSGA-II algorithm.

The runtime comparisons of the proposed HFMOEA algorithm and the filter and wrapper methods used for the comparative analysis in this paper is tabulated in Table 13 and Table 14 respectively. When compared to the filter methods, the HFMOEA method inherently takes more time since HFMOEA is essentially a filter-wrapper method and requires validation. But, when compared to the wrapper methods in Table 14, it can be noted that HFMOEA takes roughly the same time as the others but with better convergence due to the guided initialization. Thus, HFMOEA is an efficient algorithm, while also providing multiple trade-off solutions, unlike traditional wrapper methods.

Figure 6 shows the multi-objective solutions that evolved through the generations of the HFMOEA algorithm on the deep feature datasets. The PFs obtained in generations- 1, 2, 3, 5, 10 and 20 (final iteration) are shown along with the PF obtained by

Table 12: Results obtained on performing the McNemar's Statistical Test of the HFMOEA algorithm against the popular wrapper-based metaheuristic algorithms used for comparison.

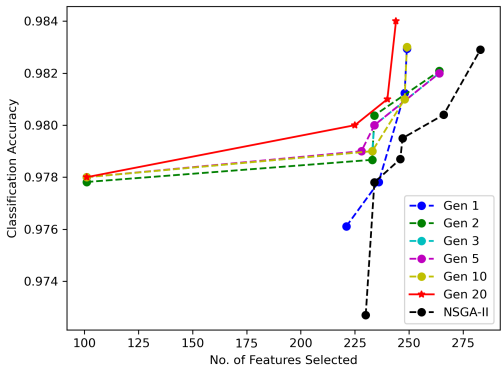
Dataset	HFMOEA vs. MA	HFMOEA vs. HS	HFMOEA vs. GA	HFMOEA vs. PSO	HFMOEA vs. WOA	HFMOEA vs. GSA
BreastCancer	4.23E-02	4.44E-04	9.75E-03	3.43E-02	4.46E-02	4.76E-02
BreastEW	2.41E-02	1.33E-02	4.11E-02	2.79E-02	9.24E-03	4.80E-02
CongressEW	1.53E-02	1.46E-02	1.80E-02	3.31E-02	1.54E-02	3.05E-02
Exactly	1.47E-02	3.82E-03	2.27E-02	2.49E-02	3.30E-02	4.09E-02
Exactly2	1.07E-02	2.48E-02	5.56E-03	3.46E-02	4.64E-02	4.51E-02
HeartEW	1.07E-02	2.97E-02	3.41E-02	1.67E-02	8.42E-03	4.74E-02
Ionosphere	4.26E-02	2.51E-02	1.29E-02	2.17E-02	2.83E-02	3.61E-02
KrVsKpEW	2.88E-02	4.12E-02	3.26E-02	8.28E-03	5.42E-03	3.18E-02
Lymphography	1.41E-02	3.63E-02	8.66E-03	3.52E-02	3.95E-02	1.27E-02
M-of-n	2.64E-02	1.03E-02	3.72E-02	3.73E-02	4.42E-02	2.89E-02
PenglungEW	3.46E-02	3.90E-02	2.09E-02	4.39E-02	3.91E-02	4.56E-02
Sonar	1.55E-02	1.43E-02	4.39E-02	4.63E-02	3.66E-02	4.58E-02
SpectEW	3.13E-02	2.80E-03	1.33E-02	3.70E-02	2.47E-02	2.35E-02
Tic-tac-toe	2.52E-02	3.23E-02	3.66E-02	4.12E-03	6.88E-03	3.02E-02
Vote	4.93E-02	1.82E-02	2.24E-02	4.28E-02	2.30E-02	4.26E-02
WaveformEW	2.50E-02	9.78E-03	1.86E-02	4.76E-02	4.32E-02	3.14E-02
Wine	1.97E-02	2.23E-02	1.88E-02	3.51E-02	1.37E-02	1.70E-02
Zoo	2.31E-02	1.42E-04	4.38E-02	3.83E-02	3.43E-02	1.75E-02

Table 13: Runtime comparison of the proposed HFMOEA algorithm with the filter methods used.

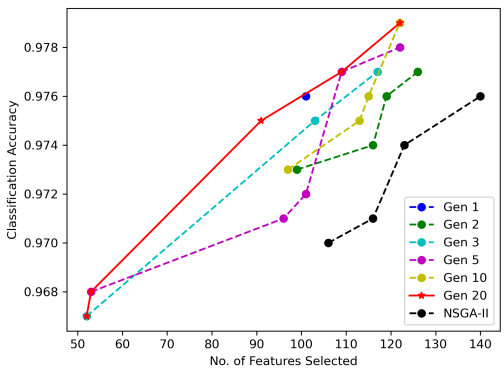
Dataset	Time (s)										
	MI	SCC	Relief	PCC	Chi-Square	IG	MAD	DR	Pasi Luukka	Fisher Score	HFMOEA
BreastCancer	1.86E+00	3.39E-02	8.88E-02	6.98E-03	2.00E-03	8.28E-02	0.00E+00	9.94E-04	4.85E-01	1.00E-03	7.06E+03
BreastEW	4.47E+02	2.04E-01	7.23E-02	5.05E-02	1.56E-03	1.16E-01	1.54E-04	4.17E-04	1.17E+00	4.17E-03	6.52E+02
CongressEW	4.21E-02	3.60E-02	5.99E-02	2.00E-02	3.20E-02	7.59E-02	0.00E+00	0.00E+00	2.32E-01	4.03E-03	5.12E+01
Exactly	4.40E-02	4.39E-02	1.60E-01	1.20E-02	0.00E+00	1.20E-01	0.00E+00	0.00E+00	3.32E-01	4.00E-03	1.95E+02
Exactly2	3.60E-02	3.60E-02	1.40E-01	1.60E-02	0.00E+00	1.32E-01	0.00E+00	0.00E+00	3.28E-01	3.99E-03	1.46E+02
HeartEW	5.27E-01	2.40E-02	2.80E-02	7.99E-03	0.00E+00	4.39E-02	0.00E+00	3.99E-03	8.39E-02	4.00E-03	2.61E+01
Ionosphere	1.29E+02	1.60E-01	5.35E-02	6.19E-02	1.58E-03	9.97E-02	1.40E-04	2.78E-04	8.42E-01	5.97E-03	1.89E+02
KrVsKpEW	4.55E-01	6.46E-01	1.04E+00	1.43E-01	4.00E-03	1.14E+00	0.00E+00	4.02E-03	2.93E+00	1.60E-02	5.79E+03
Lymphography	5.59E-02	2.40E-02	2.00E-02	1.99E-02	0.00E+00	1.00E-01	0.00E+00	0.00E+00	1.32E-01	3.96E-03	5.22E+01
M-of-n	3.20E-02	3.99E-02	1.52E-01	1.60E-02	0.00E+00	1.28E-01	0.00E+00	0.00E+00	3.55E-01	4.03E-03	2.09E+02
PenglungEW	1.48E+01	7.08E+00	1.20E-02	5.63E+00	0.00E+00	1.51E+00	0.00E+00	0.00E+00	1.80E+00	1.12E-01	7.35E+01
Sonar	2.48E+02	3.79E-01	2.56E-02	1.92E-01	1.62E-03	1.47E-01	1.34E-04	2.69E-04	8.71E-01	5.82E-03	3.04E+02
SpectEW	5.09E-02	4.89E-02	3.19E-02	3.09E-02	9.98E-04	7.58E-02	0.00E+00	0.00E+00	1.31E-01	2.99E-03	1.49E+01
Tic-tac-toe	1.60E-02	1.60E-02	9.98E-02	3.99E-03	0.00E+00	6.79E-02	0.00E+00	0.00E+00	1.84E-01	3.99E-03	7.63E+02
Vote	4.00E-02	2.00E-02	2.79E-02	1.20E-02	4.04E-03	5.19E-02	0.00E+00	3.98E-03	1.08E-01	4.00E-03	2.23E+01
WaveformEW	1.19E+03	2.77E+00	2.67E+00	1.27E-01	1.77E-02	1.07E+00	1.86E-03	3.54E-03	2.05E+01	3.86E-02	3.03E+04
Wine	4.01E+00	1.60E-02	1.60E-02	7.99E-03	3.99E-03	4.39E-02	0.00E+00	0.00E+00	8.79E-02	0.00E+00	6.22E+01
Zoo	2.80E-02	1.60E-02	1.20E-02	1.60E-02	0.00E+00	8.40E-02	0.00E+00	0.00E+00	1.32E-01	7.97E-03	1.18E+01

Table 14: Runtime comparison of the proposed HFMOEA algorithm with the popular wrapper-based metaheuristics methods used for comparison.

Dataset	Time(s)						
	MA	HS	GA	PSO	WOA	GSA	HFMOEA
BreastCancer	6.08E+02	5.67E+01	8.08E+00	5.79E+01	5.78E+01	7.80E+01	7.06E+03
BreastEW	1.06E+03	7.44E+01	1.20E+01	9.98E+01	9.91E+01	1.47E+02	6.52E+02
CongressEW	4.82E+02	4.36E+01	6.73E+00	4.57E+01	4.57E+01	7.52E+01	5.12E+01
Exactly	1.04E+03	1.01E+02	1.39E+01	1.08E+02	1.11E+02	1.38E+02	1.95E+02
Exactly2	1.00E+03	8.38E+01	1.23E+01	9.14E+01	9.20E+01	1.16E+02	1.46E+02
HeartEW	3.31E+02	3.10E+01	4.95E+00	3.21E+01	3.20E+01	5.72E+01	2.61E+01
Ionosphere	6.49E+02	6.01E+01	8.59E+00	6.49E+01	6.23E+01	1.29E+02	1.89E+02
KrVsKpEW	7.58E+03	6.13E+02	8.25E+01	6.40E+02	6.66E+02	7.39E+02	5.79E+03
Lymphography	2.59E+02	2.55E+01	4.53E+00	3.36E+01	3.19E+01	5.97E+01	5.22E+01
M-of-n	1.00E+03	8.15E+01	1.18E+01	8.83E+01	9.43E+01	1.14E+02	2.09E+02
PenglungEW	2.71E+02	1.69E+01	3.59E+00	4.27E+01	2.97E+01	4.27E+02	7.35E+01
Sonar	5.09E+02	3.19E+01	6.07E+00	4.45E+01	5.22E+01	1.50E+02	3.04E+02
SpectEW	3.21E+02	3.06E+01	4.75E+00	3.27E+01	3.18E+01	7.14E+01	1.49E+01
Tic-tac-toe	8.51E+02	7.10E+01	1.01E+01	7.53E+01	7.68E+01	9.41E+01	7.63E+02
Vote	3.41E+02	3.22E+01	4.81E+00	3.35E+01	3.34E+01	6.24E+01	2.23E+01
WaveformEW	1.51E+04	1.34E+03	1.67E+02	1.34E+03	1.43E+03	1.49E+03	3.03E+04
Wine	2.68E+02	2.55E+01	4.04E+00	2.66E+01	2.61E+01	5.02E+01	6.22E+01
Zoo	2.10E+02	2.04E+01	3.62E+00	2.13E+01	2.06E+01	4.95E+01	1.18E+01



(a) Pneumonia dataset (Kermany et al., 2018)



(b) Colorectal Cancer dataset (Kather et al., 2016)

Figure 6: Evolution of the pareto-front with generations of the HFMOEA algorithm. The vertical axis represents the accuracy and the horizontal axis represents the number of features selected.

the NSGA-II algorithm without any guided population initialization after 100 generations. It is evident that in both cases, the HFMOEA algorithm converges to a better Pareto Front in only 2 generations, than the NSGA-II algorithm that runs for 100 generations. This justifies the efficacy of the guided population initialization strategy proposed in this research. This is in accordance with our research hypotheses that guiding even a small percentage of the initial population produces faster and better convergence.

The use of the crossover operator in the HFMOEA backbone propagates the superior genes initialized by the filter methods to the other individuals in the population thus leading to faster convergence. And since the HFMOEA method is multi-objective, non-dominated solutions are preserved over the generations, unlike in single-objective optimization algorithms where pre-mature convergence may lead to the algorithm being stuck in a local optima. The comparative results obtained and the statistical analysis further justifies the superiority of the proposed HFMOEA method as compared to existing frameworks.

The multi-objective nature of the HFMOEA algorithm allows to choose the solution they need based on their application, unlike single-objective algorithms where only one optimal solution is provided. For example, in bio-medical applications accuracy of the method is of utmost importance, and thus in such scenarios, a high-accuracy solution can be chosen from the obtained PF. In other applications, the storage requirements may be of more importance than the accuracy of the solution. Thus, HFMOEA has significant real-world implications, being a multi-objective feature selection approach.

5. Conclusion and Future Research

With the development of data acquisition techniques and the availability of publicly accessible data, the need for feature selection to eliminate redundant data for efficient storage is ever-increasing. However, the feature selection process indulges a trade-off between the performance of a learning model and the number of features selected, and the relative importance between the two varies upon the application. Some applications demand high performance regardless of the storage requirement while for some the storage requirements holds priority of model performance. Thus an automatic system for generalizing this is required.

Single-objective optimization algorithms fail to do this since they return only one solution which is generally the one with better performance. In this paper, a Hybrid Filter Multi-Objective Evolutionary Algorithm (HFMOEA) is proposed, that uses a hybrid of filter-based feature ranking methods and the NSGA-II algorithm to identify the PF in the problem and present a range of solutions from which the application-specific solution can be chosen. The use of the filter ranking methods enhances the performance of the traditional NSGA-II algorithm by ensuring faster convergence, and at the same time, it is more computationally efficient than using two wrapper algorithms for a hybrid, as has been explored in most of the literature. The filter methods used in this study produce diverse feature rankings as justified by the Kendall's concordance test performed. Along with the 10 filter methods, the 90 other randomly initialized population ensure the diversity in the population. Through the evolution over the generations, the superior genes from the filter methods flow into the other individuals via the cross-over operation, and the redundant features are discarded leading to faster convergence. The evaluation of the HFMOEA method on 18 UCI datasets and 2 high dimensional deep feature sets justify that the proposed approach is

superior to existing state-of-the-art methods.

In the future, the authors aim to extend the method to more domains of applications, for example, to DNA and RNA sequence data. The number of filter methods available in the literature in the literature. Thus, 90% of the population had to be initialized randomly and the entire initial population could not be guided. More filter methods may be tested in the future as they become available to initialize an even more prominent part of the population for the HFMOEA algorithm to analyze whether the convergence to the Pareto front can be faster or more computationally efficient.

Conflict of interest statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abdi, H. (2007). The kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, 508–510.

Abualigah, L., Diabat, A., Mirjalili, S., Abd Elaziz, M., & Gandomi, A. H. (2021). The arithmetic optimization algorithm. *Computer methods in applied mechanics and engineering*, *376*, 113609.

Abualigah, L., Yousri, D., Abd Elaziz, M., Ewees, A. A., Al-qaness, M. A., & Gandomi, A. H. (2021). Aquila optimizer: A novel meta-heuristic optimization algorithm. *Computers & Industrial Engineering*, *157*, 107250.

Amoozegar, M., & Minaei-Bidgoli, B. (2018). Optimizing multi-objective pso based feature selection method using a feature elitism mechanism. *Expert Systems with Applications*, *113*, 499–514.

Arivalagan, S., & Venkatachalapathy, K. (2012). Face recognition based on a hybrid meta-heuristic feature selection algorithm. *International Journal of Computer Applications*, *55*(17).

Basak, H., Kundu, R., Chakraborty, S., & Das, N. (2021). Cervical cytology classification using pca and gwo enhanced deep features selection. *SN Computer Science*, *2*(5), 369. https://doi.org/10.1007/s42979-021-00741-2

Bhattacharyya, T., Chatterjee, B., Singh, P. K., Yoon, J. H., Geem, Z. W., & Sarkar, R. (2020). Mayfly in harmony: A new hybrid meta-heuristic feature selection algorithm. *IEEE Access*, *8*, 195929–195945.

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis*, *143*, 106839.

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

Chattopadhyay, S., Kundu, R., Singh, P. K., Mirjalili, S., & Sarkar, R. (2021). Pneumonia detection from lung x-ray images using local search aided sine cosine algorithm based deep feature selection method. *International Journal of Intelligent Systems*.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. *International conference on parallel problem solving from nature*, 849–858.

Dhiman, G., Singh, K. K., Soni, M., Nagar, A., Dehghani, M., Slowik, A., Kaur, A., Sharma, A., Houssein, E. H., & Cengiz, K. (2021). Mosoa: A new multi-objective seagull optimization algorithm. *Expert Systems with Applications*, *167*, 114150.

Dua, D., & Graff, C. (2017). UCI machine learning repository. http://archive.ics.uci.edu/ml

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

Faramarzi, A., Heidarinejad, M., Stephens, B., & Mirjalili, S. (2020). Equilibrium optimizer: A novel optimization algorithm. *Knowledge-Based Systems*, *191*, 105190.

Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., & Tavakkoli-Moghaddam, R. (2020). Red deer algorithm (rda): A new nature-inspired meta-heuristic. *Soft Computing*, 1–29.

Geem, Z. W. (2009). *Music-inspired harmony search algorithm: Theory and applications* (Vol. 191). Springer.

Hamdani, T. M., Won, J.-M., Alimi, A. M., & Karray, F. (2007). Multi-objective feature selection with nsga ii. *International conference on adaptive and natural computing algorithms*, 240–247.

Hashim, F. A., Hussain, K., Houssein, E. H., Mabrouk, M. S., & Al-Atabany, W. (2021). Archimedes optimization algorithm: A new metaheuristic algorithm for solving optimization problems. *Applied Intelligence*, *51*(3), 1531–1551.

Holland, J. H. (1992). Genetic algorithms. *Scientific american*, *267*(1), 66–73.

Kather, J. N., Weis, C.-A., Bianconi, F., Melchers, S. M., Schad, L. R., Gaiser, T., Marx, A., & Zöllner, F. G. (2016). Multi-class texture analysis in colorectal cancer histology. *Scientific reports*, *6*(1), 1–11.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95-international conference on neural networks*, *4*, 1942–1948.

Kermany, D., Zhang, K., & Goldbaum, M. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. https://doi.org/10.17632/rscbjbr9sj.2

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. *Machine learning proceedings 1992* (pp. 249–256). Elsevier.

Kononenko, I., Robnik-Sikonja, M., & Pompe, U. (1996). Relieff for estimation and discretization of attributes in classification, regression, and ilp problems. *Artificial intelligence: methodology, systems, applications*, 31–40.

Kou, G., Yang, P., Peng, Y., Xiao, F., Chen, Y., & Alsaadi, F. E. (2020). Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*, *86*, 105836.

Kundu, R., Chattopadhyay, S., Cuevas, E., & Sarkar, R. (2022). Altwoa: Altruistic whale optimization algorithm for feature selection on microarray datasets. *Computers in Biology and Medicine*, 105349.

Lac, H. C., & Stacey, D. A. (2005). Feature subset selection via multi-objective genetic algorithm. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, *3*, 1349–1354.

Lal, T. N., Chapelle, O., Weston, J., & Elisseeff, A. (2006). Embedded methods. *Feature extraction* (pp. 137–165). Springer.

Liang, G., & Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer methods and programs in biomedicine*, *187*, 104964.

López, D., Ramıórez-Gallego, S., Garcıóa, S., Xiong, N., & Herrera, F. (2021). Belief: A distance-based redundancy-proof feature selection method for big data. *Information Sciences*, *558*, 124–139.

Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, *38*(4), 4600–4607.

Mahmoud, A., Yuan, X., Kheimi, M., Almadani, M. A., Hajilounezhad, T., & Yuan, Y. (2021). An improved multi-objective particle swarm optimization with topsis and fuzzy logic for optimizing trapezoidal labyrinth weir. *IEEE Access*, *9*, 25458–25472.

Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization. *Computers in biology and medicine*, *122*, 103869.

Maldonado, S., & López, J. (2018). Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for svm classification. *Applied Soft Computing*, *67*, 94–105.

Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, *95*, 51–67.

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, *69*, 46–61.

Morita, M., Sabourin, R., Bortolozzi, F., & Suen, C. Y. (2003). Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition. *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, *3*, 666–666.

Ohata, E. F., das Chagas, J. V. S., Bezerra, G. M., Hassan, M. M., de Albuquerque, V. H. C., & Reboucas Filho, P. P. (2021). A novel transfer learning approach for the classification of histological images of colorectal cancer. *The Journal of Supercomputing*, 1–26.

Osborne, M. R. (1992). Fisher's method of scoring. *International Statistical Review/Revue Internationale de Statistique*, 99–117.

Paladini, E., Vantaggiato, E., Bougourzi, F., Distante, C., Hadid, A., & Taleb-Ahmed, A. (2021). Two ensemble-cnn approaches for colorectal cancer tissue type classification. *Journal of Imaging*, *7*(3). https://doi.org/10.3390/jimaging7030051

Raj, R. J. S., Shobana, S. J., Pustokhina, I. V., Pustokhin, D. A., Gupta, D., & Shankar, K. (2020). Optimal feature selection-based medical image classification using deep learning model in internet of medical things. *IEEE Access*, *8*, 58006–58017.

Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2009). Gsa: A gravitational search algorithm. *Information sciences*, *179*(13), 2232–2248.

Shannon, C. E., & Weaver, W. (1962). The mathemtical theory of communication. *Paperback edition, University of Illinois Press, Urbana*.

Sharma, H., Jain, J. S., Bansal, P., & Gupta, S. (2020). Feature extraction and classification of chest x-ray images using cnn to detect pneumonia. *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 227–231.

Sheikh, K. H., Ahmed, S., Mukhopadhyay, K., Singh, P. K., Yoon, J. H., Geem, Z. W., & Sarkar, R. (2020). Ehhm: Electrical harmony based hybrid meta-heuristic for feature selection. *IEEE Access*, *8*, 158125–158141.

Shunmugapriya, P., & Kanmani, S. (2017). A hybrid algorithm using ant and bee colony optimization for feature selection and classification (ac-abc hybrid). *Swarm and Evolutionary Computation*, *36*, 27–36.

Soheili, M., & Eftekhari-Moghadam, A. M. (2020). Dqpfs: Distributed quadratic programming based feature selection for big data. *Journal of Parallel and Distributed Computing*, *138*, 1–14.

Soheili, M., & Haeri, M. A. (2021). Scalable global mutual information based feature selection framework for large scale datasets. *2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC)*, 41–50.

Stephen, O., Sain, M., Maduh, U. J., & Jeong, D.-U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, *2019*.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., & van der Laak, J. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical image analysis*, *58*, 101544.

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

Xue, B., Zhang, M., & Browne, W. N. (2012). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE transactions on cybernetics*, *43*(6), 1656–1671.

Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, *20*(4), 606–626.

Xue, Y., Xue, B., & Zhang, M. (2019). Self-adaptive particle swarm optimization for large-scale feature selection in classification. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *13*(5), 1–27.

Zervoudakis, K., & Tsafarakis, S. (2020). A mayfly optimization algorithm. *Computers & Industrial Engineering*, *145*, 106559.

Zhang, Y., Gong, D.-w., & Cheng, J. (2015). Multi-objective particle swarm optimization approach for cost-based feature selection in classification. *IEEE/ACM transactions on computational biology and bioinformatics*, *14*(1), 64–75.

Zhang, Y., Gong, D.-w., Gao, X.-z., Tian, T., & Sun, X.-y. (2020). Binary differential evolution with self-learning for multi-objective feature selection. *Information Sciences*, *507*, 67–85.

Zubair, S. (2020). An efficient method to predict pneumonia from chest x-rays using deep learning approach. *The Importance of Health Informatics in Public Health during a Pandemic*, *272*, 457.