applied sciences

MDPI

Article

Selecting Features That Influence Vehicle Collisions in the Internet of Vehicles Based on a Multi-Objective Hybrid Bi-Directional NSGA-III

Mubarak S. Almutairi [1], Khalid Almutairi [2] and Haruna Chiroma [3,*]

[1] College of Computer Science and Engineering, University of Hafr Al Batin, Hafr Al Batin 39923, Saudi Arabia
[2] Mechanical Engineering Technology, Applied College, University of Hafr Batin, Hafr Al Batin 39923, Saudi Arabia
[3] Computer Science and Engineering Technology, Applied College, University of Hafr Al Batin, Hafr Al Batin 39923, Saudi Arabia
* Correspondence: charuna@uhb.edu.sa or haruna.chiroma@acm.org

Abstract: The smart platform of generating, collecting, managing and processing dynamic data from different sources in the Internet of Vehicles (IoV) pave the way for a large-scale dataset to be accumulated. The dataset can contain records running into hundreds of thousands and even millions of relevant, irrelevant and redundant features. Therefore, feature selection to select only the significant features for developing vehicle collision detection alarm systems for deployment in the IoV edge is critical. However, previous studies on vehicle collision detection in the IoV have not conducted rigorous feature selection. Limited studies have mainly applied Pearson correlation coefficient (PCC) to select subset features influencing vehicle collision in the domain of IoV. However, PCC can cause relevant features to be discarded if the correlation of the non-linear association is too small, thereby providing incorrect feature ranking, which, in turn, increases the chances of developing a model that will give a poor outcome. To close this gap, this paper proposed a multi-objective, filter-based hybrid non-dominated sorted genetic algorithm III with a gain ratio and bi-directional wrapper for the selection of subset features influencing vehicle collision in the IoV. The proposed approach selected the minimal most significant subset features for developing a vehicle collision detection classifier with maximum accuracy for deployment in the IoV. A comparative study proves that the proposed approach performs better than the compared algorithms across hybrid-, wrapper- and filter-based feature selection methods within the family of the NSGA. Further, a comparative analysis with other evolutionary algorithms proves the superiority of the proposal. This study can help researchers in the future by avoiding the use of large-scale computing resources in acquiring data to develop vehicle collision alert systems in the IoV. This can be achieved since only the subset features discovered in this study are collected, as opposed to collecting large features, thus saving time and resources in the subsequent vehicle collision detection data collection in the IoV.

Keywords: feature selection; filter; gain ratio; Internet of Vehicles; multi-objective NSGA III; vehicle collision detection; sensors; wrapper

Citation: Almutairi, M.S.; Almutairi, K.; Chiroma, H. Selecting Features That Influence Vehicle Collisions in the Internet of Vehicles Based on a Multi-Objective Hybrid Bi-Directional NSGA-III. *Appl. Sci.* **2023**, *13*, 2064. https://doi.org/10.3390/app13042064

Academic Editor: Valentino Santucci

Received: 19 January 2023
Revised: 31 January 2023
Accepted: 2 February 2023
Published: 5 February 2023

1. Introduction

The Internet of Vehicles (IoV) has the potential to generate a large-scale dataset with features running into millions beyond what is currently being witnessed, due to the increase in the number of sensors to be embedded in emerging vehicles for communicating with the IoV environment (Chiroma et al., 2021) [1] . Smart ways of generating, collecting, managing and processing dynamic data from different sources in the IoV pave the way for a large dataset to be accumulated. The dataset can contain records running into hundreds of thousands and even millions of relevant, irrelevant and redundant features. Liu [2] argued that the information expected to be generated by the IoV will be greater than that of the telecommunications industries.

Data generation in the real-world environment is mostly performed in a very large scale which makes the management of the data very complex. Such data typically comprises large numbers of features. In reality, it is not always true that all the features generated from the data are significant for obtaining useful information. Irrelevant and redundant features with the tendency to reduce the algorithm performance can be contained in the features. As such, removing the redundant and irrelevant features while maintaining the performance accuracy of the algorithm is the major aim of conducting feature selection [3].

The availability of large-scale data with hundreds of features has prompted the research community to propose different feature selection methods. Feature selection has the advantage of removing noisy, irrelevant or redundant data; it reduces the convergence time of the algorithm; it improves the quality of the data and the model accuracy can also be improved; and it can save resources in subsequent data collection [4]. Therefore, the selection features should contain only the relevant information required to solve a particular problem [5].

One of the problems of vehicular mobility in the IoV is the collision of vehicles that can lead to injury, permanent disability, vehicle damage, traffic congestion, and the loss of lives and property. Detecting the possibility of vehicle collision in the IoV ahead of time can help the driver of a conventional vehicle or a self-driving vehicle to avoid collision. As the IoV is an emerging research area under development, the subset of features required for developing a collision alarm system for vehicle collision detection in the IoV are not fully understood as research is still active.

Feature selection has the advantage of isolating noisy, irrelevant or redundant data, resulting in less algorithm convergence time and improved data quality. The model accuracy can be improved and resources can be saved in subsequent data collection [4]. Generally, feature selection methods can be classified into one of three categories: filter, wrapper and hybrid/embedded. In the filter method, the evaluation of subset features is independent of the model. The wrapper method evaluates the subset features by computing feature score while maximizing accuracy. On the other hand, the hybrid/embedded method combines both the wrapper and the filter methods [6]. In the embedded feature selection method, the training process typically involves the selection of features to be trained for the modeling algorithm to create a classifier [6].

In machine learning, feature selection is a challenging task [3]. Attempts were made by researchers to develop classifiers for vehicle collision detection in the IoV. The focus of this paper is on the features that influence vehicle collision in the IoV. Rigorous studies on selecting features that influence vehicle collision are limited in the literature, and the few studies that conducted feature selection mainly depended on the correlation coefficient for selecting features influencing vehicle collision in the IoV. Brik and Ksentini [7] proposed deep learning for the development of vehicle collision detection systems for deployment in multi-access edge computing in the IoV. The study used deep recurrent for the modelling to predict vehicle collision in the IoV. The work was limited to the recurrent network prediction of collisions without discussing the features influencing collisions in the IoV and feature selection was absent in the study. Similarly, Chang et al. [8] used YOLOv4 for the development of a vehicle avoidance system in the IoV by predicting future vehicle positions and similar to Brik and Ksentini [7], no feature selection was conducted and a discussion about the features influencing collision in the IoV was out of the scope of the study. Chang et al. [9] developed collision detection alarm system for IoV, features were extracted based on the inbuilt feature extraction mechanism in Sqeeze-and-excitation network. However, feature selection was not covered. Chen et al. [10] developed an IoV-based model for the detection of collisions in the IoV. Like Brik and Ksentini [7]; in Chang et al. [8]; in Chang et al. (2019) [9] study on vehicle collision feature selection was not included in the scope of the studies. On the other hand, Wang et al. [10] proposed an IoV-based vehicle collision detection system. The features influencing vehicle collision were collected and discussed in the study. The correlation coefficient was used to select the subset features used for modeling the algorithm to detect vehicle collision. In another study, the features influencing vehicle collision in the IoV were selected using the correlation coefficient to reduce the number of available features generated from the IoV before developing the classifier for collision detection [11]. Similarly, Almutairi et al. (2023) [12] adopted the features used in the study of Chen et al. [13] to generate the vehicle collision data used for the development of classifiers to detect vehicle collision in the IoV.

From the literature, it is established that the research community in the domain of IoV heavily relied on the correlation coefficient, a traditional technique for the selection of subset features influencing vehicle collision in the IoV environment. However, the correlation coefficient can cause relevant features to be discarded if the correlation of the non-linear association is too small, thereby providing incorrect feature ranking [14] which increases the chances of developing a model that will give a poor outcome. Similarly, the association of two random variables may not be uniformly strong enough which may lead to small values of correlation, resulting in the assumption that association between the variables does not exist [15].

Most of the traditional techniques for the selection of optimal feature subsets mostly experience enormous complexity, premature convergence and expensive computational cost. As a result, nature-inspired metaheuristic algorithms were proposed to mitigate the limitations of the traditional methods which are found to be efficient and effective in minimizing feature subsets and maximizes the model accuracy [3]. For example, the non-dominated sorted genetic algorithm III (NSGA3) has proven to outperform many feature selection algorithms such as the indicator-based evolutionary algorithm, NSGA2, the strength Pareto evolutionary algorithm II and the kernel-based nonlinear extension of Arps decline [16].

To deviate from the limitations of the correlation coefficient feature selection method typically used in the domain of IoV, this paper proposed a multi-objective, filter-based hybrid non-dominated sorted genetic algorithm III (NSGA3) with a gain ratio (GR) and a bi-directional wrapper for the selection of subset features influencing vehicle collision in the IoV to minimize features and maximize accuracy of the model. In a situation where the filter or wrapper cannot effectively produce the best subsets of features, a hybrid approach is applied. Hence, this study uses the benefits of the filter and wrapper to minimize the limitations of the filter and wrapper to improve classification performance with the best subsets of features.

The paper is organized as follows. Section 2 briefly discusses the features influencing vehicle collision in the IoV and the basics of the IoV. Section 3 describes the basics operations of feature selection algorithms and the adopted algorithms. Section 4 presents the methodology, Section 5 discusses the results and finally, Section 6 concludes the paper. A summary of the research contributions is as follows:

- A hybrid NSGA III Bi-directional gain ratio is proposed that combines wrapper and filter methods for selecting vehicle collision subset features in the IoV.
- The study reveals the optimal subset features required for vehicle collision detection in the Internet of Vehicles.
- The evaluation study shows that the performance of the hybrid, bi-directional NSGA III is better than the other compared algorithms.
- IoV vehicle collision detection features are reduced to the minimal amount and the accuracy of collision detection in the IoV is maximized.
- It is possible to develop a vehicle collision alert system for collision detection of vehicles in the IoV using only three subset features.
- The proposed algorithm has the potential to be used to develop vehicle collision alarm systems with an improved performance that can assist drivers/self-driving cars to avoid vehicle collision faster and better.

2. Features Influencing Vehicle Collision in Internet of Vehicles

This section is provided for researchers new in the research area to grasp the basic concepts of the IoV for a better understanding of the study. In addition, the features influencing vehicle collision in the IoV are discussed.

The number of vehicles connected to the Internet of things (IoT) is increasing drastically, thus providing the means for vehicle information to be accessed ubiquitously by drivers and passengers. As a result of the increasing number of connected vehicles, new requirements of vehicular networks are emerging, such as a secure and seamless exchange of scalable information between vehicles, humans and roadside infrastructure. This has revolutionized the vehicular ad hoc networks into the IoV [17]. The IoV is transforming as a new emerging research theme and developing from the concept of vehicular ad hoc networks [18]. It is the IoT that drives the revolution of vehicular ad hoc networks into the IoV [19]. In view of the technological advancement in communication systems, vehicles are no longer just a simple means of transportation due to the development of mobile networks and the automotive industry. Vehicles that are connected with sensors, computers, radars, GPS antennas, etc. have the capacity to collect and process a large-scale amount of data as a result of vehicle-to-vehicle communication [20]. Other forms of communication that exist in the IoV environment includes vehicle-to-roadside, vehicle-to-infrastructure, vehicle-to-personal device, vehicle-to-sensor, vehicle-to-pedestrian [17,18,21]; and vehicle-to-home [1], as depicted in Figure 1.

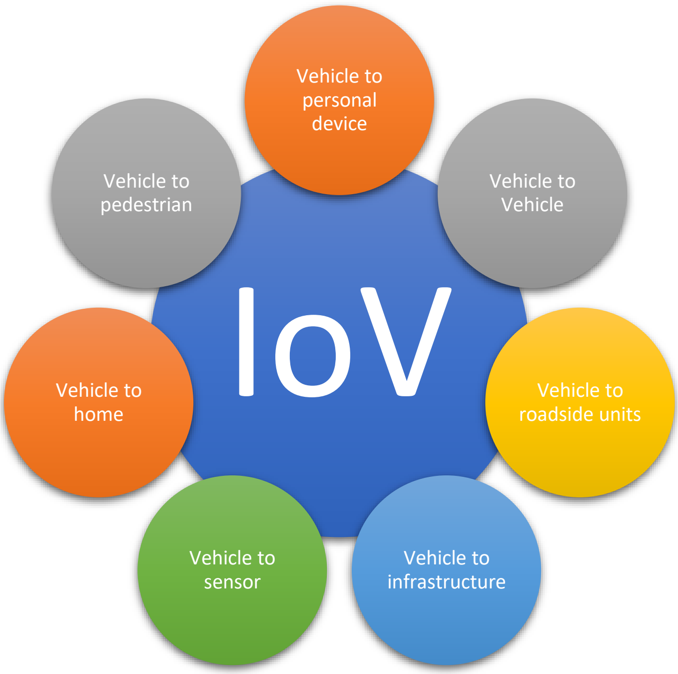

Figure 1. Communications for vehicles in the IoV environment.

The main aim of the IoV is the facilitation of information exchange between the vehicles and other related entities with the objective of reducing vehicle accidents, deviating from traffic congestion and reeling out information services [21]. The active safety system is a technique that provides collision warning to the drivers and the passengers ahead of the coming danger. In the IoV, the active safety systems that integrated wireless communication, GPS and vehicular sensors remain a hot research area [8]. Accidents and traffic congestion, among others already discussed in Section 1, are consequences of vehicle collision.

Avoiding vehicle collision in the IoV requires the development of a classifier that can detect the vehicle collision ahead of time, so that vehicle collision can be avoided [11]. Features are required for developing active safety systems to be deployed at the edge of the IoV to warn the vehicles drivers and passengers. Many features that influence vehicle collisions in the IoV exists. The features are as follows: driver status, gender, age, weather conditions, traffic density, light conditions, road surface conditions, vehicle velocity, inter-vehicle distance, braking intervention, non-skid property, and type of driveway [11]. Lin et al. [22] proposed an alternative scheme for the optimization of single value decomposition as well as the combination of downlink and uplink to optimize the weights of beamforming vectors. Results have found that the proposal was superior when compared to the classical benchmark schemes. Similarly, Lin et al. [23] improved the secrecy energy efficiency by the divide and conquer technique for the improvement of power consumption in beamforming. An et al. [24] employed a base station with multi-antenna as the green interference source to improve transmission that is secured in the satellite network. Research outcome shows that the beamforming scheme proposed in the study outperformed the classical schemes. In another development, An et al. [25] investigated the limit of cognitive-uplink fixed satellite service and terrestrial fixed service performance. The study performed the derivation theoretically before simulations were conducted to verify the effectiveness.

3. Feature Selection Algorithms

The operations of the feature selection algorithms adopted in this study for the IoV data are presented in this section for the readers to understand how the algorithms work to solve the problem of feature selection.

3.1. Basic Operations of the Feature Selection Algorithms

The basic principle of the feature selection algorithm is presented as follows:

Let us assume that the original set of features is Y with the cardinalities n and d representing the required features in the subset X, $X \subseteq Y$; Assuming $J(X)$ is the feature selection criterion function for the features set X, then the higher the value of the J the better the feature subset without any loss of generality. The feature selection problem aims to find the subset $X \subseteq Y$ such that $|X| = d$ and $J(X) = \frac{\max}{Z \subseteq Y, |Z|=d} J(Z)$ [26].

3.2. Multi-Objective NSGA3 and NSGA2

Deb et al. [27] proposed that the NSGA2 was an improvement over the NSGA because it lacks elitism, sharing parameters and speed. The NSGA II has been the most frequently used multi-objective algorithm since its appearance in the literature. However, since the inception of NSGA3, it has been playing the same role as the NSGA2 for performance evaluation [28]. Subsequently, the NSGA3 begin operations using a set of reference points with a description. The offspring of the population is created using the current population of the parent at a particular generation using genetic operators. The combination of the population is sorted into non-domination at different level. The population members that cannot be accommodated in the last front are saved and the remaining members are rejected. The next set of members that are selected for subsequent generation and the other members are selected for the last front. The selection is conducted by performing systematic analysis of the set members with respect to the reference points as supplied, unlike the case in NSGA2 where the front members are chosen based on the large crowd distance value. To obtain an identical range for objective values and supplied reference points, they are first normalized. As such, zero vector becomes the normal point of the set. The members of the set are associated with the reference point based on proximity to the reference line attained by linking the normal point with the point of reference. It is found that this procedure typically assists in determining the number and the population indices for the members associated with the reference point for each supply. As the result, the niching procedure is applied for the selection of the population members not represented in the set using the result of the association procedure as described in the preceding section. The reference point with the least number of associations in the population searches for an associated point. The members are continuously added one-by-one until the population is full [28]. The performance of the NSGA2 and NSGA3 is inconsistent because the performance depends on the datasets or the target problem to be solved. A study indicates that NSGA2 outperform NSGA3 on the knapsack problems, whereas NSGA3 outperforms NSGA2 for distance minimization problems [28].

3.3. Gain Ratio, Filter and Wrapper

Typically, the GR gives priority to the features that have large values, making the GR biased towards the large number of values. The combination of feature 1 and feature 2 has better potential than feature 1 or feature 2 alone, meaning that the combination is better than the constituent features. Split information is mostly used to deviate from the bias [29]. Feature selection can be filter, wrapper or hybrid. Filters select the best subsets of features by ranking all the features, and features with high classification performance are chosen at the detriment of others. Hence, there is no dependency among the subsets of features that are chosen. Filters are computationally fast since there is no interaction among the selected subsets of features. On the other hand, wrappers are computationally expensive since, for each subset of features that is selected, the learning algorithm must be included to determine the performance of the chosen features. In this study, NSGA2 developed by Deb et al. [27] and NSGA3 introduced by Mkaouer et al. [30], along with the GR as the filter evaluation criterion, are used to solve the issue of dependency among selected subsets of features. GR-based filters are now becoming popular in solving various multi-objective optimization problems in feature selection [31].

4. Methodology

The description of the IoV environment used for data collection and the stages involved in applying the proposed algorithm for vehicle collision subset feature selection are presented in this section.

4.1. Dataset Description

To collect data for the vehicle collision in the IoV, the environment was created in the VISSIM before generating and collecting the data for vehicle collision. The IoV environment was created for different scenarios. The configurations of the IoV environment in the VISSIM are described as follows: The IoV architecture covers 1500 $\times$ 1500 m for a single vehicle lane road comprising bends, multiple lanes, speed distribution, an area with speed restriction, signal controllers, the view of the IoV in both 2D and 3D with a minimal speed, average speed and maximum speed of 50 km/h, 95 km/h and 140 km/h, respectively. Approximated vehicular mobility traffic was between 30 and 180 vehicles. The vehicles in mobility for the traffic signal heads to the controller showing the signal light displaying green, red and amber. The IoV environment was reconfigured for different scenarios to capture the low-speed distribution required on the roads and the number of vehicles on the road moving to different zones. The low-speed zone on the road varied between 5 km/h and 50 km/h, with a varied number of vehicles moving into different zones [12]. The features for the vehicle collision in the IoV were as follows: Driving state, clearance, spacing, safety distance (gross), safety distance (net), following distance (gross), following distance (net), interaction state, delay time, lane and speed, as indicated in Figure 2. Finally, the features used in the study were as listed as number of lanes, status of the driver, nature of the environment, velocity of the vehicle, distance between vehicles and breaking capability. Figure 2 shows the configuration of the features influencing vehicle collision in the IoV simulated in the VISSIM environment.

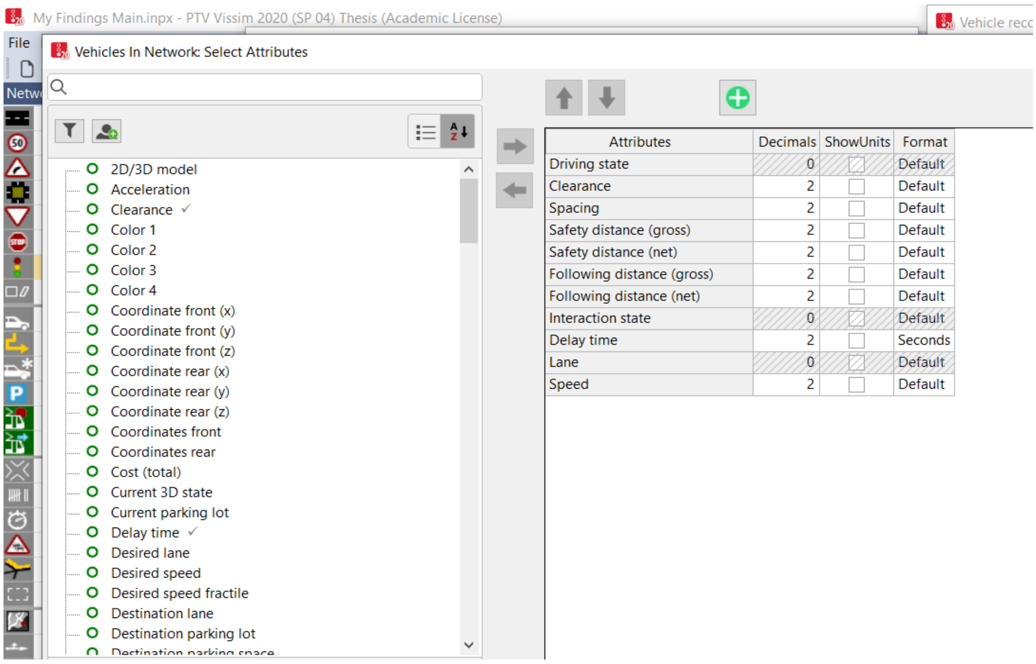

Figure 2. Some of the features influencing vehicle collision in the IoV environment.

4.2. Proposed Multi-Objective Hybrid NSGA3

The experiment platforms were as follows. Hardware platform: Mac M2 chip with an 8-core CPU having core performance and four efficiency cores, 10 cores, a GPU 16-core Neural engine and 100 GB/s bandwidth of memory. Software platform: The experiment was conducted in Anaconda Python platform. Parameter settings in Table 1 were adopted from Kumar and Yadav [32,33]. The parameter settings were used for executing the NSGA3 and NSGA2.

Table 1. NSGA family parameter setting.

Parameter Description	Settings
Reference points number	0.9
Crossover probability	0.9
Crossover distribution index	0.25
Mutation probability	0.2
Hypervolume	[0, 0]–[1, 1]
Generations	500
Population size	200
Population initialization	Random
Objectives	2

This study used multi-objective feature selection that evaluates a group of features concurrently. The NSGA3, along with the GR as the filter evaluation measure and bi-directional, is used for the selection of IoV collision features with the most significant informative subset features. For the purpose of comparison within the family of NSGA, NSGA2 with the GR and bi-directional is used across filter-, wrapper- and hybrid-based feature selection methods for the selection of features influencing vehicle collision in the IoV.

In the procedure for executing the multi-objective NSGA3 for the feature selection, binary chromosomes were adopted for the selection of the IoV features influencing vehicle collision within the IoV environment. Because the proposed study is multi-objective, the study aims to minimize the IoV features influencing vehicle collision and maximize the accuracy of vehicle collision detection in the IoV. Each group of the subset features is selected by the NSGA2 and NSGA3. The filter-based method with GR is used to rank all the subsets of features along with the classification performance for both NSGA3 and NSGA2. The features are stored in order of their classification performance. Classifiers are used for the evaluation of the discriminant value of the IoV feature subset selected by the NSGA2 and NSGA3. The study used support vector machine (SVM), random forest (RFC), Gaussian naïve Bayes (GNB), decision tree (DTC), and K-nearest neighbor (KNN) as the learning algorithms for developing the classifier. The 5 algorithms were chosen because the algorithms continue to be relevant in the literature, especially in the situation where data are not in large-scale volume. The IoV feature subset selected was used to develop 5 different classifiers for the purpose of evaluating the performance of the feature subset the chromosomes were able to identify. The 5 classifiers were subjected to detect vehicle collision in the IoV based on the selected optimum feature subset. The vehicle collision detection accuracy to be maximized is returned by each of the classifiers with a different combination of the feature subset.

The experiment is repeated with the wrapper method that uses the bi-directional concept. The bi-directional operation is comparable to the forward way of selection with a distinction of evaluating the significance of any existing features before adding another new subset feature. If vehicle collision features in the data are deemed to be insignificant, the feature is quickly eliminated by backward selection. As a result, it combines backward elimination with forward selection. The stages involved in the bi-directional removal of features includes entering and leaving the model, selection of a significance level, carrying out subsequent forward selection, then complete removal of the feature in backward operation. The optimal subset of features is obtained by repeating the procedure as necessary. The population diversity of the NSGA is twisted and first uses the forward selection then the best subsets of features are used as backward elimination within the non-dominated concept; specifically, NSGA2 uses its adaptive ability via a distance crowding operator. On the other hand, the NSGA3 maintains solution diversity. Lastly, they are fed into the bi-directional wrapper-based approach for both NSGA3 and NSGA2. The complete stages in the framework for the proposed multi-objective hybrid NSGA3 are depicted in Figure 3. To further understand the performance of the proposal, the study compared it to the performance of other feature selection algorithms outside the NSGA family to verify the advantages and novelty of the proposed feature selection algorithm. To be fair in the comparison, we used multi-objective optimization algorithms on the same dataset for the selection of the features influencing vehicle collision in the IoV. We selected the multi-objective algorithms that are well established in the literature for the performance comparison. The algorithms selected and used for the evaluation of the performance are as follows: Pareto envelope-based selection algorithm II [34], multi-objective evolutionary algorithm based on decomposition [35], strength Pareto evolutionary algorithm 2 [36], niched pareto genetic algorithm [37] and multi-objective genetic algorithm [38]. The algorithms were used on the same dataset to select subset features influencing collision in the IoV and detect vehicle collision. The results were recorded and included computational time.

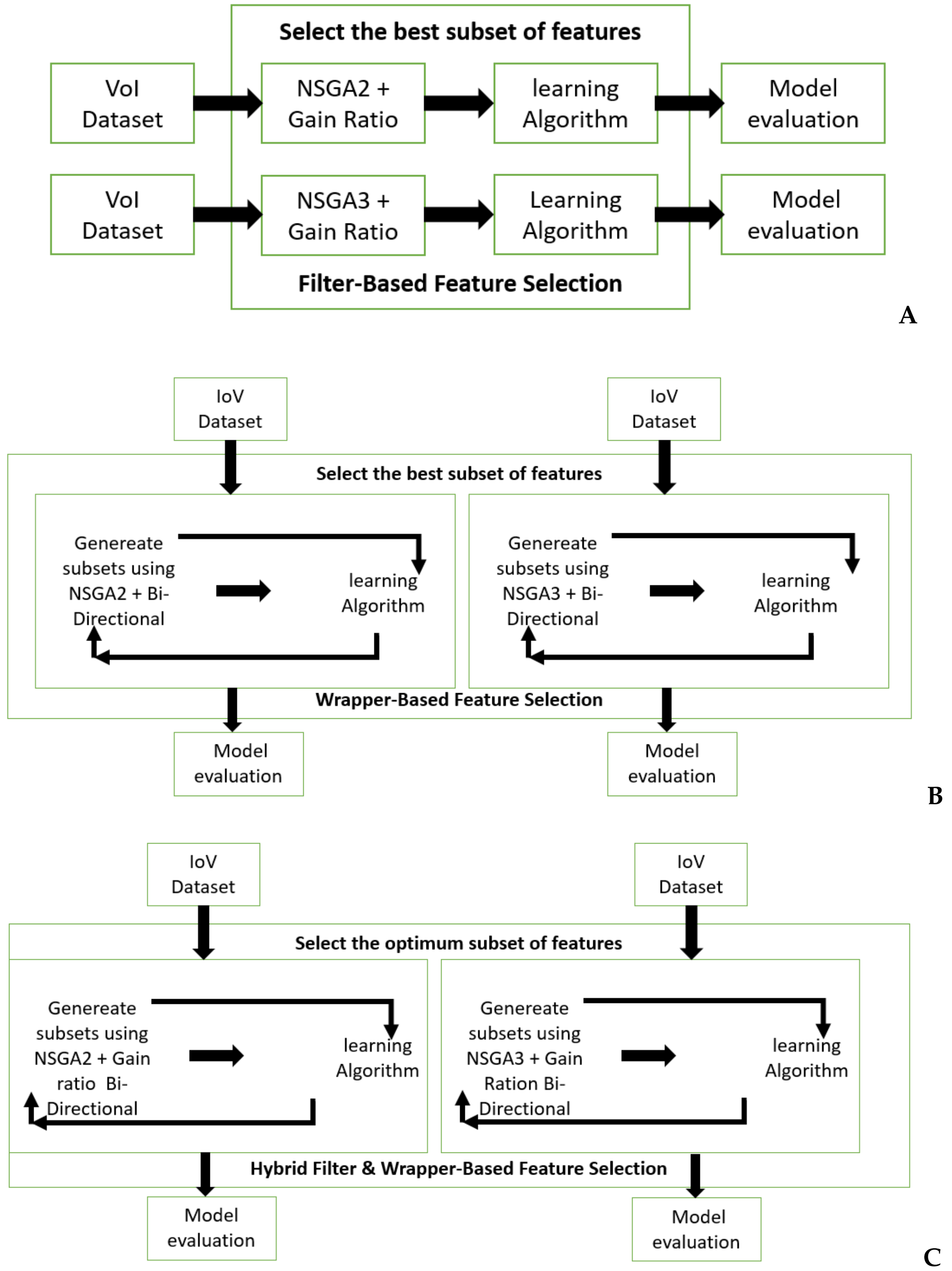

Figure 3. Propose framework for subset selection: (**A**) NSGA3 + GR, (**B**) NSGA3 + Bi-directional and (**C**) the NSGA3 + Bi-directional + GR.

5. Results and Discussion

The default settings of two multi-objective feature selections (NSGA3 and NSGA2) along with filter-based GR were run on the datasets. The performance of the five different classifiers are recorded as shown in Table 2. Table 2 can infer the most relevant number of features with better accuracy. From the graphs plotted in Figures 4–10 , the *x*-axis represents the number of IoV vehicle collision detection selected features, whereas the *y*-axis represents the classification accuracy recorded by each classifier on the forty independent runs for each of the algorithms.

Table 2. Results of the independent runs of the algorithms for IoV vehicle collision feature selection and accuracy.

Method	SVM		KNN		GNB		RFC		DTC	
	Selected Features	**Accuracy**	**Selected Features**	**Accuracy**	**Selected Features**	**Accuracy**	**Selected Features**	**Accuracy**	**Selected Features**	**Accuracy**
NSGA2 + GR	All features (6)	75.11	All features	81.24	All features	74.55	All features	80.5	All features	79.5
	F1 (3)	88.22	F3 (4)	88.00	F1 (3)	88.51	F4 (4)	81.00	F3 (4)	81.5
NSGA3 + GR	All features (6)	89.64	All features (6)	84.92	All features (6)	85.00	All features	90.25	All features	89.0
	F1 (3)	**98.22**	F3 (4)	90.34	F4 (4)	91.28	F3 (4)	92.75	F5 (5)	90.5
NSGA2 + Bi-directional	All features (6)	90.2	All features	89.75	All features	89.55	All features	85.65	All features	82.50
	F1 (3)	**91.44**	F2: (3)	90.00	F1 (3)	90.00	F1 (3)	89.00	F6 (2)	89.50
NSGA3 + Bi-directional	All features (6)	93.25	All features (6)	91.05	All features (6)	91.75	All features	92.25	All features	90.65
	F1 (3)	**98.97**	F3 (4)	91.59	F1 (3)	92.25	F2 (3)	93.55	F2 (3)	91.30
NSGA2 + Bi-directional + GR	All features (6)	92.55	All features	90.75	All features	91.05	All features	89.12	All features	89.77
	F1 (3)	**93.00**	F3 (4)	91.75	F1 (3)	91.75	F1 (3)	90.05	F3 (4)	90.50
NSGA3 + Bi-directional + GR	All features (6)	99.25	All features (6)	92.05	All features (6)	93.15	All features	94.25	All features	90.65
	F1 (3)	**99.97**	F2: (3)	92.24	F1 (3)	93.30	F2 (3)	94.58	F2 (3)	92.32

F1: Distance between vehicles, velocity of the vehicle and breaking capability. F2: Breaking capability, velocity of vehicle and number of Lanes. F3: Number of lanes, velocity of vehicle, distance between vehicles and breaking capability. F4: Distance between vehicles, velocity of vehicle, breaking capability and nature of environment. F5: Nature of environment, number of lanes, velocity of vehicle, distance between vehicles and breaking capability. F6: Breaking capability and distance between vehicles.

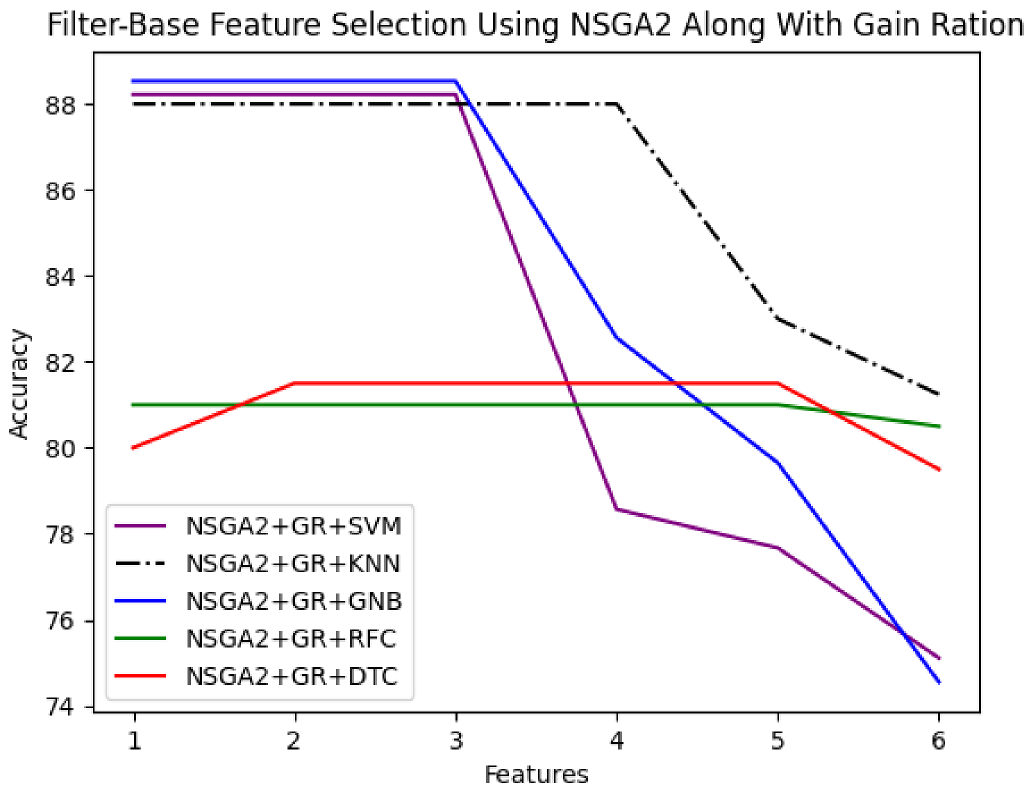

Figure 4. NSGA2 + GR with different classifiers builds using subsets of the IoV vehicle collision detection features.

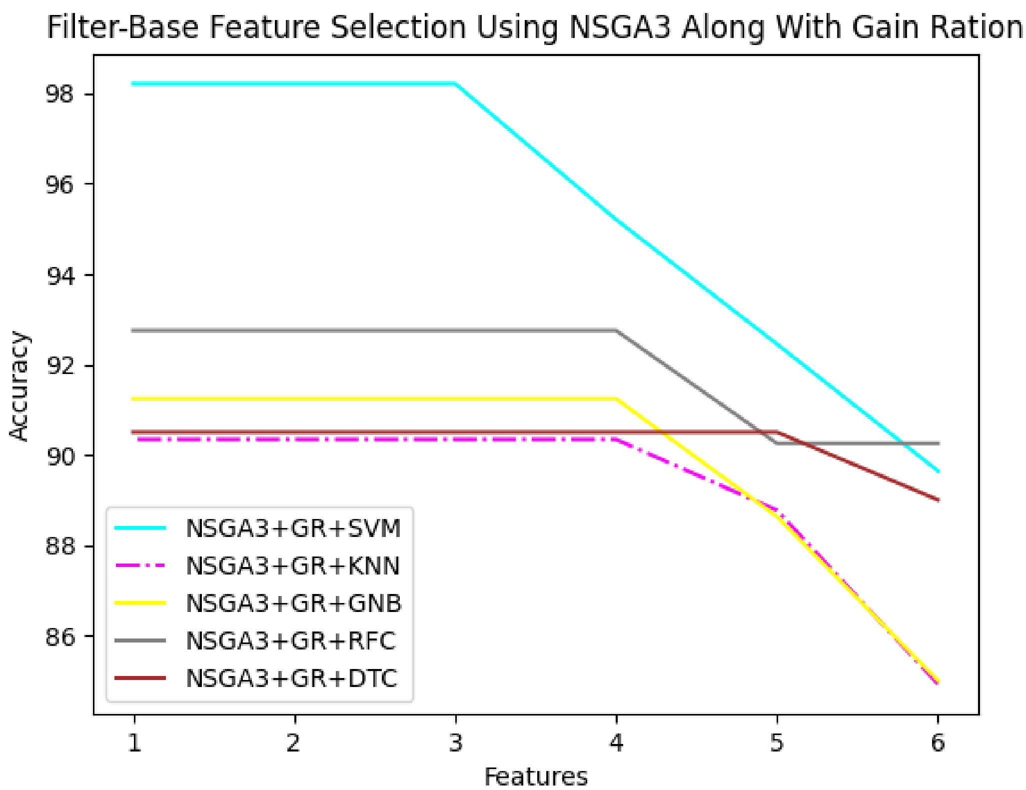

Figure 5. NSGA3 + GR with various classifiers.

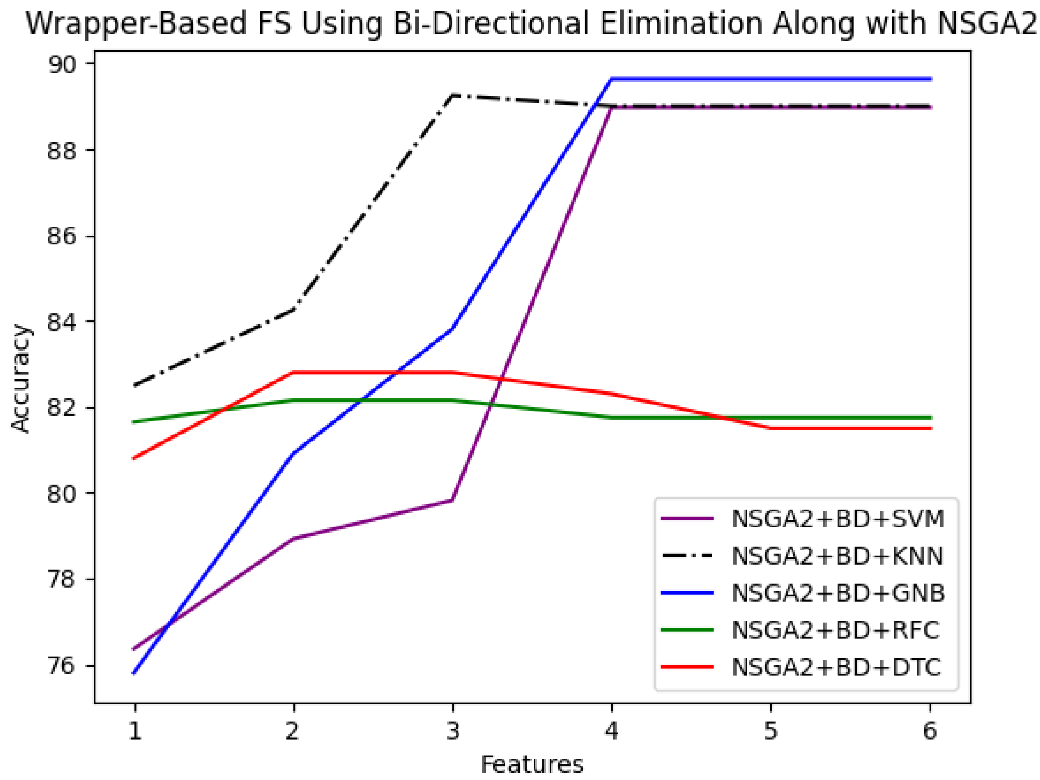

Figure 6. Performance of the five classifiers developed with subset features selected by NSGA2 + Bi-directional.

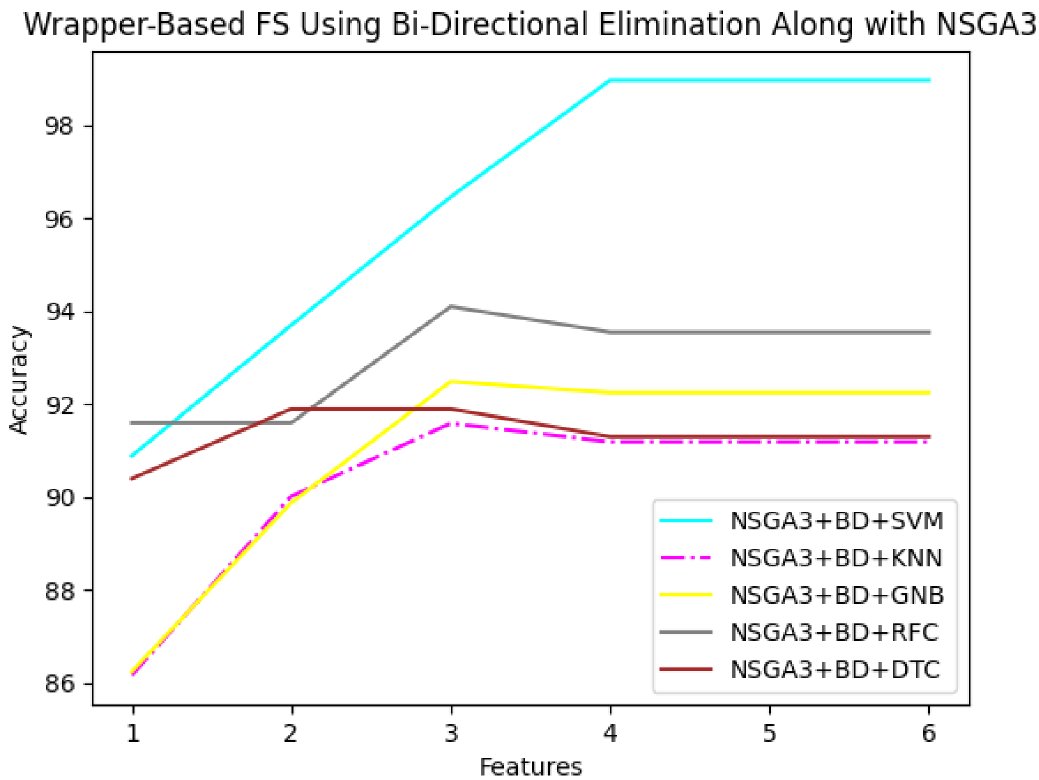

Figure 7. Wrapper-based NSGA3 + Bi-directional elimination for the classifiers.

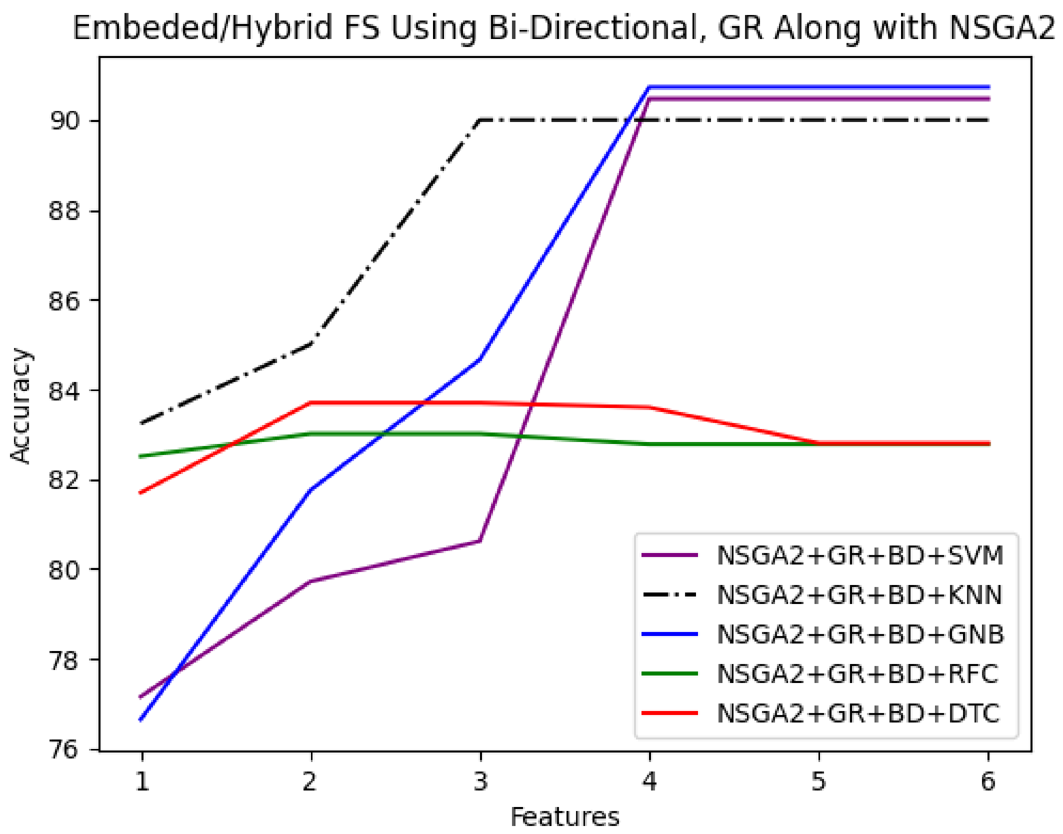

Figure 8. Hybrid NSGA2 + Bi-directional + GR feature selection performance.

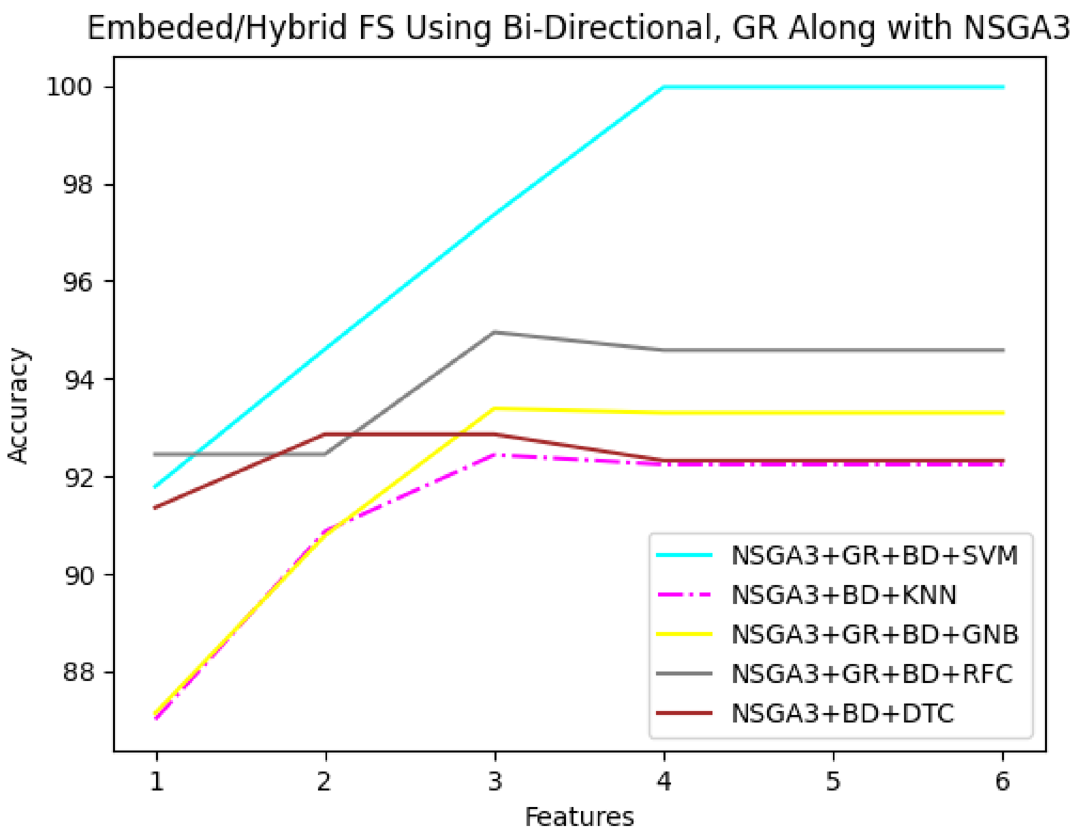

Figure 9. Hybrid NSGA3 + Bi-directional + GR feature selection performance.

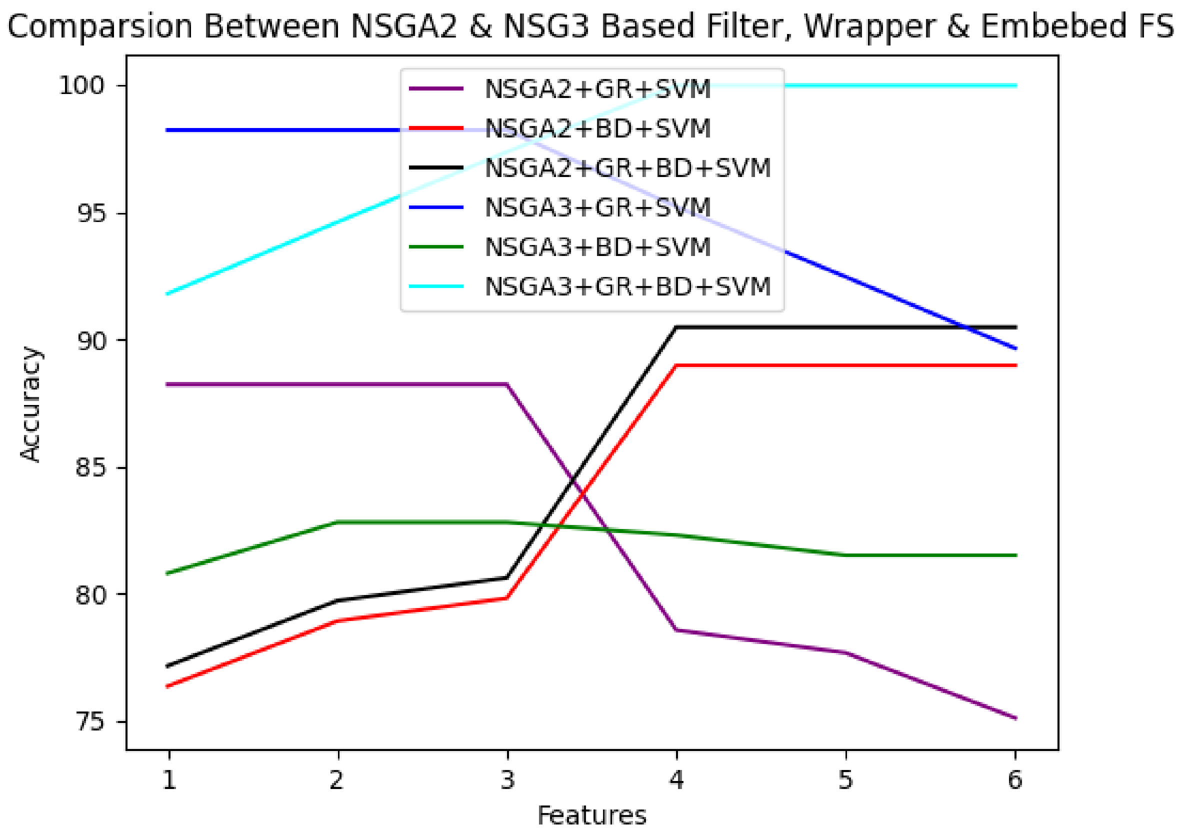

Figure 10. Performance comparison for NSGA2 and NSGA3 across filter-, wrapper- and hybrid-based feature selection methods.

Table 2 indicates that the NSGA2 +GR with the five different classifiers shows steady results when three features are selected by NSGA2 + GR to build the SVM and the GNB classifiers, with the GNB recording the highest accuracy of 88.51 on the three selected IoV vehicle collision detection subset features. On the other hand, KNN recorded the highest accuracy in the category without feature selection compared to the SVM and the GNB. The four subsets of IoV vehicle collision detection features are consistent with the KNN, while five features are consistent with the RFC classifier. Comparing the performance of all classifiers, it can be observed that those with the highest accuracy selected only three features, followed by the classifier with four features. Even though RFC is stable on four features, it recorded a low accuracy. Hence, it can be deduced that GNB can produce better accuracy in detecting collision in the IoV with only three features selected by NSGA2 + GR from the available IoV vehicle collision features. It is observed that the feature selection results in all cases outperformed the classifiers developed with all the IoV vehicle collision detection features without feature selection. Figure 4 depicts the curve for the different classifiers with selected features based on NSGA2 + GR.

Similarly, Table 2 indicates that the results of the NSGA3 + GR feature selection approach show that the SVM outperforms the KNN and the GNB both in terms of the number of selected features and accuracy. Thus, comparing the performances of both NSGA2 + GA and NSGA3 + GA, it is found that the NSGA3 + GA with SVM performs better in terms of both the number of IoV vehicle collision selected features, as well as collision detection accuracy. As observed in the case of NSGA2 + GA, similarly, NSGA3 + GA shows better performance with selected IoV vehicle collision detection features compared to the result for all the features without selection. Figure 5 shows the performance of each classifier. NSGA3 + GR with SVM recorded the highest detection accuracy and was stable on three subsets of features. NSGA3 + GR with KNN, GNB and RF is stable on four subset features, with RFC attaining the highest accuracy among them. On the other hand, DTC is stable on five features, but recorded the lowest accuracy apart from the KNN. Comparing the performance of all the five classifiers in terms of accuracy clearly shows that SVM outperforms the compared classifiers and selects the least number of features.

The results of wrapper-based NSGA2 + Bi-directional indicated that it is steady when three features are selected by the wrapper-based NSGA2 + Bi-directional and it achieved the best performance with the SVM classifier compared to the other classifiers. The wrapper-based NSGA2 + Bi-directional recorded 91.44 as the best accuracy with three subset features. With all the features, the highest accuracy without feature selection was recorded by wrapper-based NSGA2 + Bi-directional with the SVM classifier, showing that feature selection by NSGA2 + Bi-directional has improved the accuracy in detecting vehicle collision in the IoV. Figure 6 presents the learning curve for the different classifiers with subset features selected by wrapper-based NSGA2 + Bi-directional.

Figure 7 presents the performance of the classifiers developed based on wrapper-based NSGA3 + Bi-directional. The wrapper-based NSGA3 + Bi-directional with the SVM has the highest vehicle collision detection accuracy in the IoV and it is stable on three feature subsets. The wrapper-based NSGA3 + Bi-directional with the DT, GNB and RF is stable on three subsets of the features, whereas the KNN is stable on four features. Comparing the performance of all the five classifiers in terms of accuracy clearly shows that the SVM outperforms all the other classifiers with 98.97% vehicle collision detection accuracy in the IoV.

Table 2 indicated that the hybrid NSGA2 + Bi-directional + GR with the five classifiers is steady for three subset features selected by the hybrid NSGA2 + Bi-directional + GR to build the SVM classifier, recording the highest accuracy of 93% vehicle collision detection in the IoV. Three subsets of IoV collision detection features are consistent with the SVM, GNB, DT and RF classifiers, except for KNN with four subset features. Hence, it can be deduced that SVM can produce better accuracy in detecting vehicle collision in the IoV, with only three features selected by the hybrid NSGA2 + Bi-directional + GR. It is observed that the feature selection results by each of the classifiers outperformed the classifiers developed with the maximum IoV vehicle collision detection features. Figure 8 depicts the different classifiers with selected features based on hybrid NSGA2 + Bi-directional + GR.

The results of the hybrid NSGA3 + Bi-directional + GR feature selection approach in Table 2 show that the SVM outperforms the compared classifiers based on the subset features selected by the hybrid NSGA3 + Bi-directional + GR. Thus, by comparing the performances of both hybrid NSGA2 + Bi-directional + GR and hybrid NSGA3 + Bi-directional + GR, it is found that the hybrid NSGA3 + Bi-directional + GR with SVM performs better in terms of vehicle collision detection accuracy. As observed in the case of hybrid NSGA2 + Bi-directional + GR, hybrid NSGA3 + Bi-directional + GR shows better performance with subset selected IoV vehicle collision detection features compared to the result with features without selection. Figure 9 shows the performance of each classifier build based on subset features selected by hybrid NSGA3 + Bi-directional + GR. The hybrid NSGA3 + Bi-directional + GR was stable on three subsets of features. Hybrid NSGA3 + Bi-directional + GR with KNN, DT, GNB, SVM and RF are all stable on three subset features. Comparing the performance of all the five classifiers in terms of accuracy clearly shows that the SVM outperforms all other algorithms.

Figure 10 depicts the NSGA2 + GR and NSGA3 + GR performance across filter-, wrapper- and hybrid-based feature selection approaches. The SVM in each of the categories of feature selection methods was selected for comparison, because in each case it performed the best, except in the case of NSGA2 + GR where GNB outperformed the SVM. It is found that the best performing feature selection algorithm is the hybrid NSGA3 + Bi-directional + GR, as it successfully minimizes the subset features and maximizes the accuracy in detecting vehicle collision in the IoV environment better than the compared algorithms: hybrid NSGA2 + Bi-directional + GR, filter-based NSGA2 + GR, filter-based NSGA3 + GR, wrapper-based NSGA2 + Bi-directional and wrapper-based NSGA3 + Bi-directional. However, regarding the issue of minimum subset features, most of the algorithms in different scenarios selected three subset features followed by four, five and two as shown in Table 2.

5.1. Computational Time Complexity

The issue of processing time in the IoV environment is imperative because delay is critical. Therefore, the processing time of the proposed hybrid NSGA3 and compared algorithms were recorded and are presented in Table 3. It is clearly indicated that the proposed hybrid NSGA3 converges slower compared to the compared algorithms within the family of the NSGA. This is expected in view of the fact that the filter approach is always faster than the other feature selection methods, as the filter method mainly focuses on the data characteristics and it is independent of the learning algorithm, as reported by Aggrawal et al. [3]). The time difference between the compared algorithms and the proposed hybrid NSGA3 was very close; the difference is minimal. Time matters significantly in the IoV as a small delay may cause serious vehicle collision within the IoV environment. However, the accuracy achieved by the proposed hybrid NSGA3 shows that it can detect and alert the driver ahead of time to avoid collision.

Table 3. Computational time for the feature selection algorithms.

Feature Selection Method		**Computational Time (s)**					**Average Time**
		SVM	**KNN**	**GNB**	**RFC**	**DTC**	
Filter	NSGA2 + GR	0.742	0.891	0.811	0.742	0.749	0.787
	NSGA3 + GR	**0.681**	**0.723**	**0.743**	**0.681**	**0.749**	**0.715**
Wrapper	NSGA2 + Bi-directional	0.779	0.936	0.852	0.779	0.786	0.826
	NSGA3 + Bi-directional	0.712	0.756	0.776	0.712	0.783	0.748
Hybrid	NSGA2 + Bi-directional + GR	0.712	0.807	0.777	0.712	0.749	0.751
	NSGA3 + Bi-directional + GR	**0.696**	**0.739**	**0.76**	**0.696**	**0.766**	**0.731**
Average time		0.72	0.809	0.786	0.72	0.764	

5.2. Performance Comparison with Other Classes of Feature Selection Algorithms

The proposed hybrid NSGA3 performance is further compared with other algorithms that are different to the family of the NSGA to verify the advantage of the proposal. Table 4 presents the results of the comparison between the proposed hybrid NSGA3 and other meta-heuristic algorithms. The results indicated that the proposal perform better than the compared algorithms. This has proven the advantage of the hybrid NSGA3 in selecting optimum subset features influencing vehicle collision in the IoV while maximizing accuracy. Regarding the convergence time, the hybrid NSGA3 converges faster than the compared algorithms. Therefore, the proposed hybrid NSGA3 outperforms the compared algorithms in both accuracy and convergence time.

Table 4. Performance comparison of the proposed hybrid NSGA3 with other classes of algorithms.

Algorithm	**Accuracy**	**Time (s)**
Pareto envelope-based selection algorithm II	77.65	3.2010
Multi-objective evolutionary algorithm based on decomposition	89.65	1.2350
Strength Pareto evolutionary algorithm 2	82.62	1.8650
Niched Pareto genetic algorithm	78.66	2.7850
Multi-objective genetic algorithm	89.11	1.9870
Hybrid NSGA3 + GR + Bi-directional	**99.97**	**0.6963**

5.3. Implication of the Study to Theory and Practice

It was found in the study that it is possible to develop a vehicle collision alert system for detecting collisions of vehicles ahead of time in the IoV environment using three subset features with a high level of accuracy. The IoV vehicle collision detection features were reduced to a minimum and the accuracy of detecting collision in the IoV was maximized. This indicates that the quality of the data and performance of the algorithm were improved as a result of feature selection. When developing vehicle collision alarm systems to be deployed in the edge of the IoV to provide service to vehicles moving on roads in the IoV environment, a small vehicle collision alarm device requiring a small dataset for its smooth operation can adopt the three subset features without the need to acquire a large-scale dataset. This can avoid the utilization of costly computing resources in subsequent data collection. This is highly relevant, especially in the situation where computing resources and data are highly restricted. This study can help researchers in the future to avoid using large-scale computing resources in acquiring data to develop collision alert system in the IoV in view of the fact that only the subset features discovered in this study will be collected, as opposed to wasting time collecting all features that can come with irrelevant, redundant and noisy features. The complex nature of the alarm systems for detecting IoV vehicle collision can be reduced because the size of the data is at the minimal level, thereby requiring minimal memory space in the computing device. Thus, this can reduce the hardware size required for developing collision alert systems, thereby making the vehicle collision alert system devices cheaper compared to similar alert devices that require large-size hardware. This is so because the larger the physical size of hardware, the higher the cost of the device, as the size of the hardware device reduces the cost (the size of the hardware device is directly proportional to the cost).

6. Conclusions

The paper proposed the selection of optimum subset features influencing vehicle collision in the IoV environment. A multi-objective hybrid NSGA3 Bi-directional GR is proposed for IoV vehicle collision subset feature selection. For the purpose of performance evaluation, NSGA2 was used for the selection of IoV vehicle collision subset features across wrapper, filter and hybrid methods resulting in five different algorithms for the evaluation. A comparative study proves the superiority of the proposed multi-objective hybrid NSGA3 Bi-directional GR in minimizing features and maximizing the detection of vehicle collision accuracy in the IoV. When developing collision alarm systems to be deployed in the edge of the IoV to provide service to vehicles moving on roads in the IoV environment, a small collision alarm device requiring a small dataset for its smooth operation can adopt the subset features in this study without the need to acquire very-large-scale features.

Author Contributions: Conceptualization, M.S.A. and H.C.; methodology, M.S.A., K.A. and H.C.; formal analysis, H.C.; investigation, M.S.A. and K.A; resources, M.S.A. and H.C.; data curation, M.S.A. and K.A; writing—original draft preparation, M.S.A., K.A. and H.C.; writing—review and editing, H.C.; visualization, K.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to acknowledge the support of the Deanship of Scientific Research of the University of Hafr Al Batin (UHB), Hafr Al Batin, Saudi Arabia through Research Grant no. IFP-A-2022-2-1-11.

Institutional Review Board Statement: Not applicable to our study.

Informed Consent Statement: The study does not involve human subjects.

Data Availability Statement: The data are available at https://ieee-dataport.org/topic-tags/transportation (accessed on 1 January 2023).

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. Chiroma, H.; Abdulhamid SI, M.; Hashem, I.A.; Adewole, K.S.; Ezugwu, A.E.; Abubakar, S.; Shuib, L. Deep learning-based big data analytics for internet of vehicles: Taxonomy, challenges, and research directions. *Math. Probl. Eng.* **2021**, *2021*, 9022558. [CrossRef]
2. Liu, N. Internet of vehicles: Your next connection. *Huawei WinWin* **2011**, *11*, 23–28. Available online: http://www.binarytek.cn/page104?_l=en&article_id=145&brd=1 accessed on 15 November 2022.

3. Agrawal, P.; Abutarboush, H.F.; Ganesh, T.; Mohamed, A.W. Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019). *IEEE Access* **2021**, *9*, 26766–26791. [CrossRef]
4. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]
5. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the 2014 Science and Information Conference, London, UK, 27–29 August 2014; pp. 372–378.
6. Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification: A review. In *Data Classification: Algorithms and Applications*; CRC Press: Boca Raton, FL, USA, 2014; pp. 37–64.
7. 7Brik, B.; Ksentini, A. Toward optimal MEC resource dimensioning for a vehicle collision avoidance system: A deep learning approach. *IEEE Netw.* **2021**, *35*, 74–80.
8. Chang, C.C.; Ooi, Y.M.; Sieh, B.H. IoV-based collision avoidance architecture using machine learning prediction. *IEEE Access* **2021**, *9*, 115497–115505. [CrossRef]
9. Chang, W.J.; Chen, L.B.; Su, K.Y. DeepCrash: A deep learning-based Internet of vehicles system for head-on and single-vehicle accident detection with emergency notification. *IEEE Access* **2019**, *7*, 148163–148175. [CrossRef]
10. Wang, X.; Liu, J.; Qiu, T.; Mu, C.; Chen, C.; Zhou, P. A real-time collision prediction mechanism with deep learning for intelligent transportation system. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9497–9508. [CrossRef]
11. Chen, L.B.; Su, K.Y.; Mo, Y.C.; Chang, W.J.; Hu, W.W.; Tang, J.J.; Yu, C.T. An implementation of deep learning based IoV system for traffic accident collisions detection with an emergency alert mechanism. In Proceedings of the 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), Berlin, Germany, 2–5 September 2018; pp. 1–2.
12. Almutairi, M.S.; Almutairi, K.; Chiroma, H. Hybrid of Deep Recurrent Network and Long Short Term Memory for Rear-end Collision Detection in Fog based Internet of Vehicles. *Expert Syst. Appl.* **2023**, *213*, 119033. [CrossRef]
13. Chen, C.; Xiang, H.; Qiu, T.; Wang, C.; Zhou, Y.; Chang, V. A rear-end collision prediction scheme based on deep learning in the Internet of Vehicles. *J. Parallel Distrib. Comput.* **2018**, *117*, 192–204. [CrossRef]
14. Zhu, Y.; Liang, J.; Chen, J.; Ming, Z. An improved NSGA-III algorithm for feature selection used in intrusion detection. *Knowl.-Based Syst.* **2017**, *116*, 74–85. [CrossRef]
15. University of Pritoria. Feature selection problems. Chapter 3. Ph.D. Thesis, University of Pritoria, Pretoria, South Africa, 2011. Available online: https://repository.up.ac.za/bitstream/handle/2263/29486/02chapter3-.pdf?sequence=3&isAllowed=y (accessed on 21 December 2022).
16. Xue, Y.; Tang, Y.; Xu, X.; Liang, J.; Neri, F. Multi-objective feature selection with missing data in classification. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *6*, 355–364. [CrossRef]
17. Contreras-Castillo, J.; Zeadally, S.; Guerrero-Ibañez, J.A. Internet of vehicles: Architecture, protocols, and security. *IEEE Internet Things J.* **2017**, *5*, 3701–3709. [CrossRef]
18. Kaiwartya, O.; Abdullah, A.H.; Cao, Y.; Altameem, A.; Prasad, M.; Lin, C.T.; Liu, X. Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects. *IEEE Access* **2016**, *4*, 5356–5373. [CrossRef]
19. Yang, F.; Wang, S.; Li, J.; Liu, Z.; Sun, Q. An overview of internet of vehicles. *China Commun.* **2014**, *11*, 1–15. [CrossRef]
20. Niu, Z.; Shen, S.; Zhang, Q.Y. Space-air-ground integrated vehicular network for immersive driving experience. *Chin. J. IoT* **2017**, *2*, 17–27.
21. Ji, B.; Zhang, X.; Mumtaz, S.; Han, C.; Li, C.; Wen, H.; Wang, D. Survey on the internet of vehicles: Network architectures and applications. *IEEE Commun. Stand. Mag.* **2020**, *4*, 34–41. [CrossRef]
22. Lin, Z.; Niu, H.; An, K.; Wang, Y.; Zheng, G.; Chatzinotas, S.; Hu, Y. Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 3717–3724. [CrossRef]
23. Lin, Z.; An, K.; Niu, H.; Hu, Y.; Chatzinotas, S.; Zheng, G.; Wang, J. SLNR-based secure energy efficient beamforming in multibeam satellite systems. *IEEE Trans. Aerosp. Electron. Syst.* Early Access. **2022**. [CrossRef]
24. An, K.; Lin, M.; Ouyang, J.; Zhu, W.P. Secure transmission in cognitive satellite terrestrial networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 3025–3037. [CrossRef]
25. An, K.; Liang, T.; Zheng, G.; Yan, X.; Li, Y.; Chatzinotas, S. Performance limits of cognitive-uplink FSS and terrestrial FS for Ka-band. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *55*, 2604–2611. [CrossRef]
26. Zongker, D.; Jain, A. Algorithms for feature selection: An evaluation. In Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 25–29 August 1996; Volume 2, pp. 18–22.
27. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.A.M.T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
28. Ishibuchi, H.; Imada, R.; Setoguchi, Y.; Nojima, Y. Performance comparison of NSGA-II and NSGA-III on various many-objective test problems. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; pp. 3045–3052.
29. Dai, J.; Xu, Q. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Appl. Soft Comput.* **2013**, *13*, 211–221. [CrossRef]
30. Mkaouer, W.; Kessentini, M.; Shaout, A.; Koligheu, P.; Bechikh, S.; Deb, K.; Ouni, A. Many-objective software remodularization using NSGA-III. *ACM Trans. Softw. Eng. Methodol. (TOSEM)* **2015**, *24*, 1–45. [CrossRef]

31. Usman, A.M.; Yusof, U.K.; Naim, S. Filter-based feature selection: A comparison among binary and continuous Cuckoo optimization algorithms along with multi-objective optimization algorithms using gain ratio-based entropy. *Int. J. Bioinspired Comput.* **2022**, *20*, 183–192. [CrossRef]
32. Kumar, H.; Yadav, S.P. Using reference point-based NSGA-II to system reliability. *Constraints* **2017**, *1*, 2. [CrossRef]
33. PyGMO: A Python Parallel Global Multiobjective Optimizer a Scientific Library That Provides a Large Number of Optimisation Problems and Algorithms Settings Library. Available online: https://esa.github.io/pygmo/index.html (accessed on 20 December 2022).
34. Corne, D.W.; Jerram, N.R.; Knowles, J.D.; Oates, M.J. PESA-II: Region-based selection in evolutionary multiobjective optimization. In Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation, San Francisco, CA, USA, 7–11 July 2001; pp. 283–290.
35. Zhang, Q.; Li, H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [CrossRef]
36. Zitzler, E.; Laumanns, M.; Thiele, L. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK-Report* **2001**, *103*. [CrossRef]
37. Erickson, M.; Mayer, A.; Horn, J. Multi-objective optimal design of groundwater remediation systems: Application of the niched Pareto genetic algorithm (NPGA). *Adv. Water Resour.* **2002**, *25*, 51–65. [CrossRef]
38. Murata, T.; Ishibuchi, H. MOGA: Multi-objective genetic algorithms. In *IEEE International Conference on Evolutionary Computation*; IEEE: Piscataway, NJ, USA, 1995; Volume 1, pp. 289–294.