Chapter 8

Conclusions and Future Scope of Work

8.1 Conclusions

Following the overall conclusion of the thesis are drawn from the work reported in the various chapters. The method of eGA used in Chapter 2 produces the optimal feature set and using the proposed method of feature selection the redundant and irrelevant/noisy features are identified. Results of the experiments show that the proposed method is able to select the most informative features in terms of classification accuracy. In this Chapter 4, UCI datasets have been used to validate the efficacy of the proposed approach of feature selection using eGA. An analysis of irrelevant/noisy features has also been performed. In later chapters, fuzzy rough set based fitness functions have been used, hence to compare the performance of fuzzy rough set with respect to the feature selection method which use classification accuracy as fitness function, this study has been done. A decision tree method with feature selection is presented for predicting the electricity prices. The method uses eGA and decision tree classifier for feature selection. The result explains the efficacy of the feature selection method. It is established from the results that feature selection method provides better forecast accuracy of electricity prices in comparison to that of using full set.

Chapter 3 covered the preliminaries of rough set and fuzzy rough set along with an example explaining calculation procedure to obtain dependency measures for both the methods. Following motivations were drawn in favour of using L-FRFS measure in the present work.

1. For the case of real valued features, for applying rough set method, we need to discretize the feature values. The discernibility of features are affected by the quantization and therefore becomes dependent on the quantization of feature values.

2. In case of fuzzy rough sets, the real feature values are taken as it is and therefore no such quantization is done and discernibility of features are therefore more accurate as well as meaningful.

3. For real valued features the number of quantization levels can be large or infinite. Thus to capture the feature values in discrete sense is not simply possible.

4. If a real valued feature based problem is solved using rough set through discretization, it is highly possible that while application, a newer intermediate value for which the quantization level was not fixed may arise, making the rough set based method of no use since the nominations to values are predefined. In other words, new value is a new nomination and the rough set based feature reduction has to be done once again including the new nominations.

Chapter 4 developed a new initialization method for PSO and IDS algorithms. DS initialized swarm algorithms always produce smaller reducts, as they are able to pick the appropriate reducts in early generation. Following conclusion can be drawn with respect to the RST and L-FRFS measures.

1. Using RST measure, PSO-DS and IDS-DS, in general, achieve better performance as compared to PSO-RANDOM and IDS-RANDOM. PSO-DS and IDS-DS, outperform PSO-RANDOM and IDS-RANDOM in such cases where feature reduction is actually sought after, i.e. datasets having large number of features. DS-initialized PSO and IDS outperform random initialized PSO and IDS for large datasets, without compromising with classification accuracy.

2. Using L-FRFS measure, PSO-DS and IDS-DS, in general, achieve better performance as compared to PSO-RANDOM and IDS-RANDOM, which has been established using a t-test, Wilcoxon test and Friedman test.

3. It is observed from the Friedman test, that for RST measure, rank of PSO-RANDOM method is better than that of IDS-DS, but in L-FRFS measure, rank of IDS-DS method is better than that of PSO-RANDOM.

In Chapter 5, new feature selection methods viz. Hybrid-P and Hybrid-S have been proposed. The methods use RST and L-FRFS measures as their fitness function. These methods were developed employing hybridization of PSO and IDS to take the advantage of random exploration of IDS and guided search of PSO. These two hybrid methods were tested for different datasets and the effectiveness of the methods was established in terms of reducts achieved. The proposed methods have shown their effectiveness on large and practical datasets where feature selection are relevant and significant. It is observed that in most of the datasets, Hybrid-P method performs better than Hybrid-S method. It is also observed from Friedman ranking test, that with RST measure, rank of Hybrid-P method is better than that of Hybrid-S, but with L-FRFS measure, rank of Hybrid-S method is better than that of Hybrid-P.

Chapter 6 proposes new feature selection method based on BO using RST and L-FRFS measures as fitness function. Effectiveness of the BO was established in terms of number of reduct size, classification accuracy, t-test, Wilcoxon test and Friedman test. The BO has shown its effectiveness on large and practical datasets. It was demonstrated that the proposed feature selection method using BO is successful in identifying the irrelevant and redundant features. It is observed that BO ranks the best as compared to Hybrid-P and Hybrid-S methods, in case of RST as well as L-FRFS measures.

A new evolutionary optimization algorithm, AGA-MO, has been proposed in Chapter 7. The proposed AGA-MO optimization method has been applied for feature selection using RST and L-FRFS measures. AGA-MO utilizes DS-initialization and gives better reducts without compromising with classification accuracy. Superiority of AGA-MO over BO and eGA has been established in terms of reduct size, classification accuracy, t-test, Wilcoxon test and Friedman test. Further, AGA-MO has the edge, i.e. its performance is better than BO and eGA, in case of RST as well as L-FRFS measures, as suggested by Friedman ranking test. The AGA-MO has shown its effectiveness on large and practical datasets where feature selection is significant.

Table 8.1: Best known result for each dataset along with their description

Dataset	Data Object	No. of classes	Total no. of Features	Best known result using RST (for discrete dataset)	Best known result using L-FRFS (for continuous dataset)
Cleveland	303	5	13	3	6
Ecoli	336	8	7	3	5
Glass	214	6	9	2	8
Ionosphere	351	2	34	2	6
Lung	32	2	56	3	3
Soybean small	47	4	35	2	2
Wine	178	3	13	2	4
LSVT	126	2	310	1	5

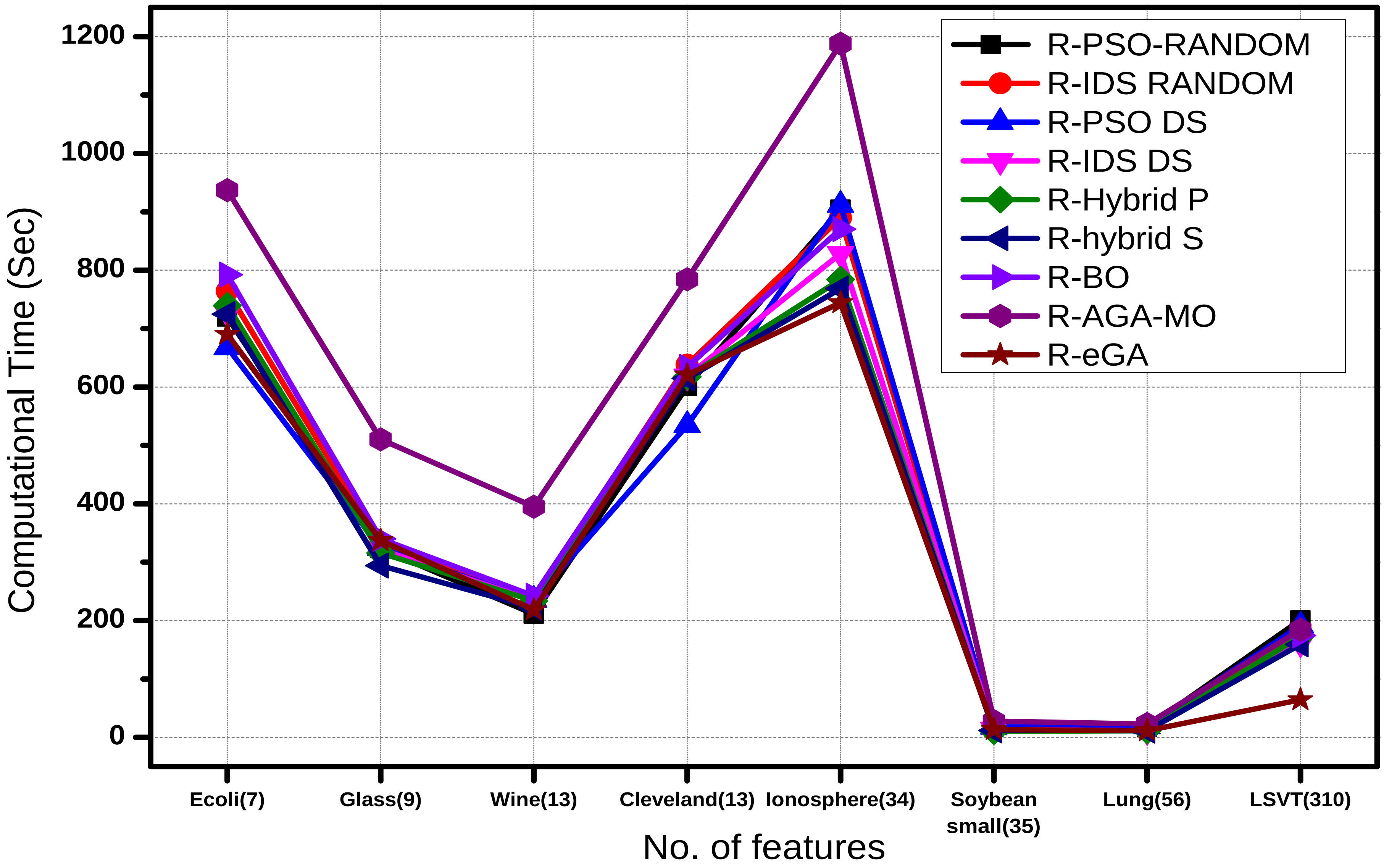

Figure 8.1: Computational time vs Number of features with RST technique

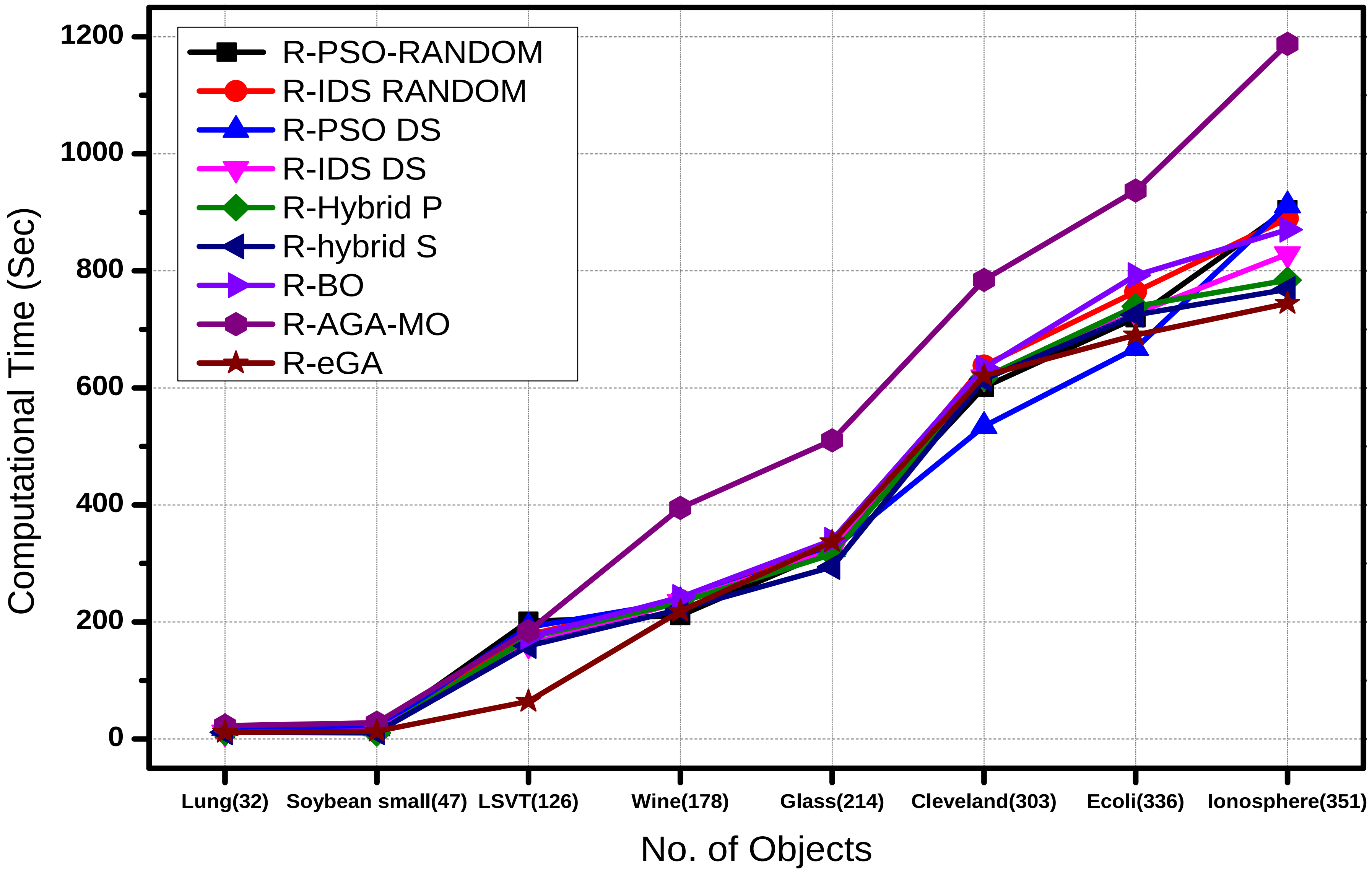

Figure 8.2: Computational time vs Number of Objects with RST technique

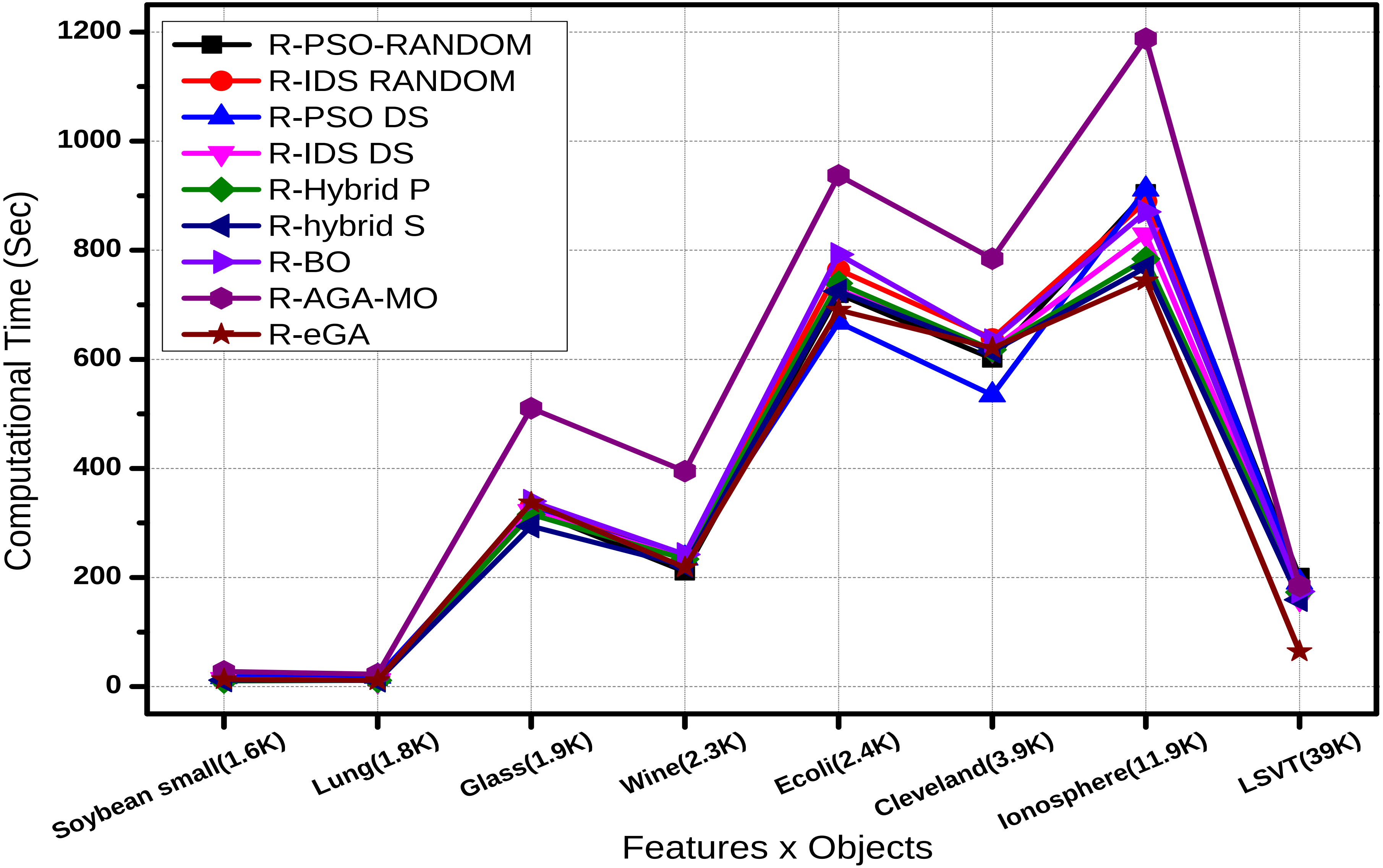

Figure 8.3: Computational time vs (Number of Features x Objects) with RST technique

The graph of computational time vs number of features is shown in Figure 8.1. It is observed that the computational time required by a method does not have a discernible dependency on the number of features. However, a clear dependency of computational time on the number of objects in the dataset can be observed from Figure 8.2. Also the dependency of computational time can be approximated to be linearly related to the product of number of objects and number of features. For a given dataset, all the methods take almost similar amount of computational time, as evident from Figure 8.3.

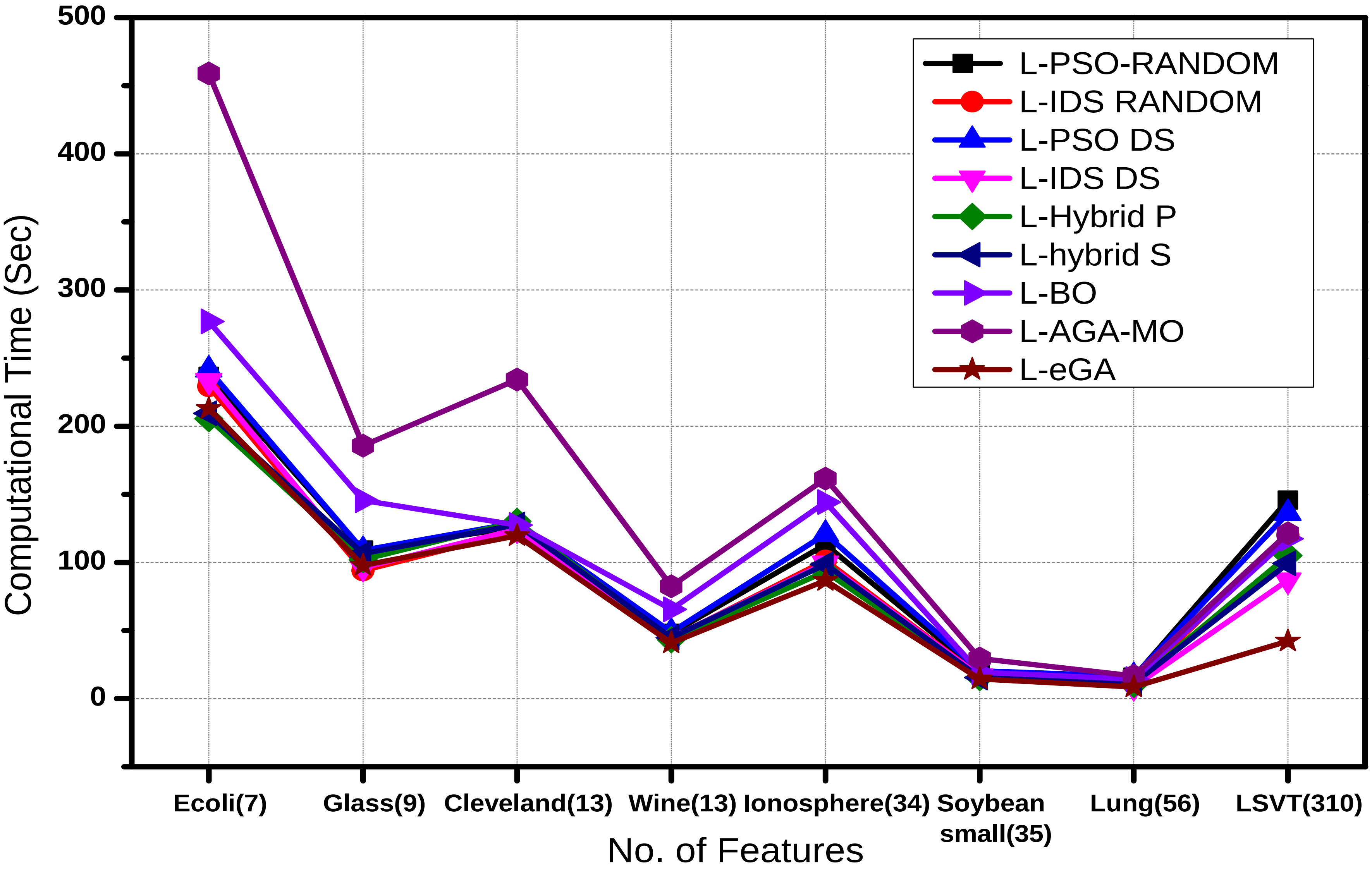

Figure 8.4: Computational time vs Number of features with L-FRFS technique

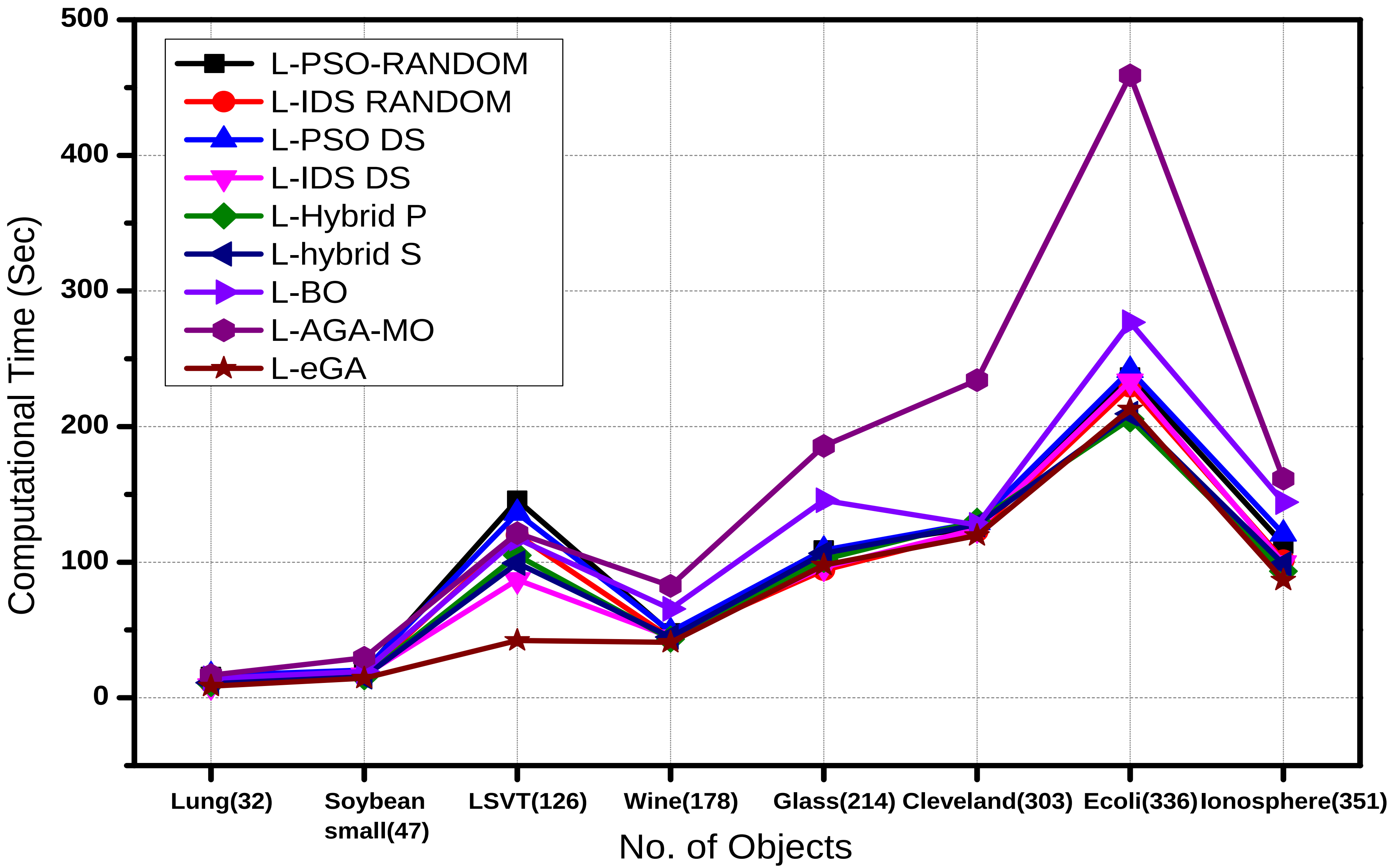

Figure 8.5: Computational time vs Number of Objects with L-FRFS technique

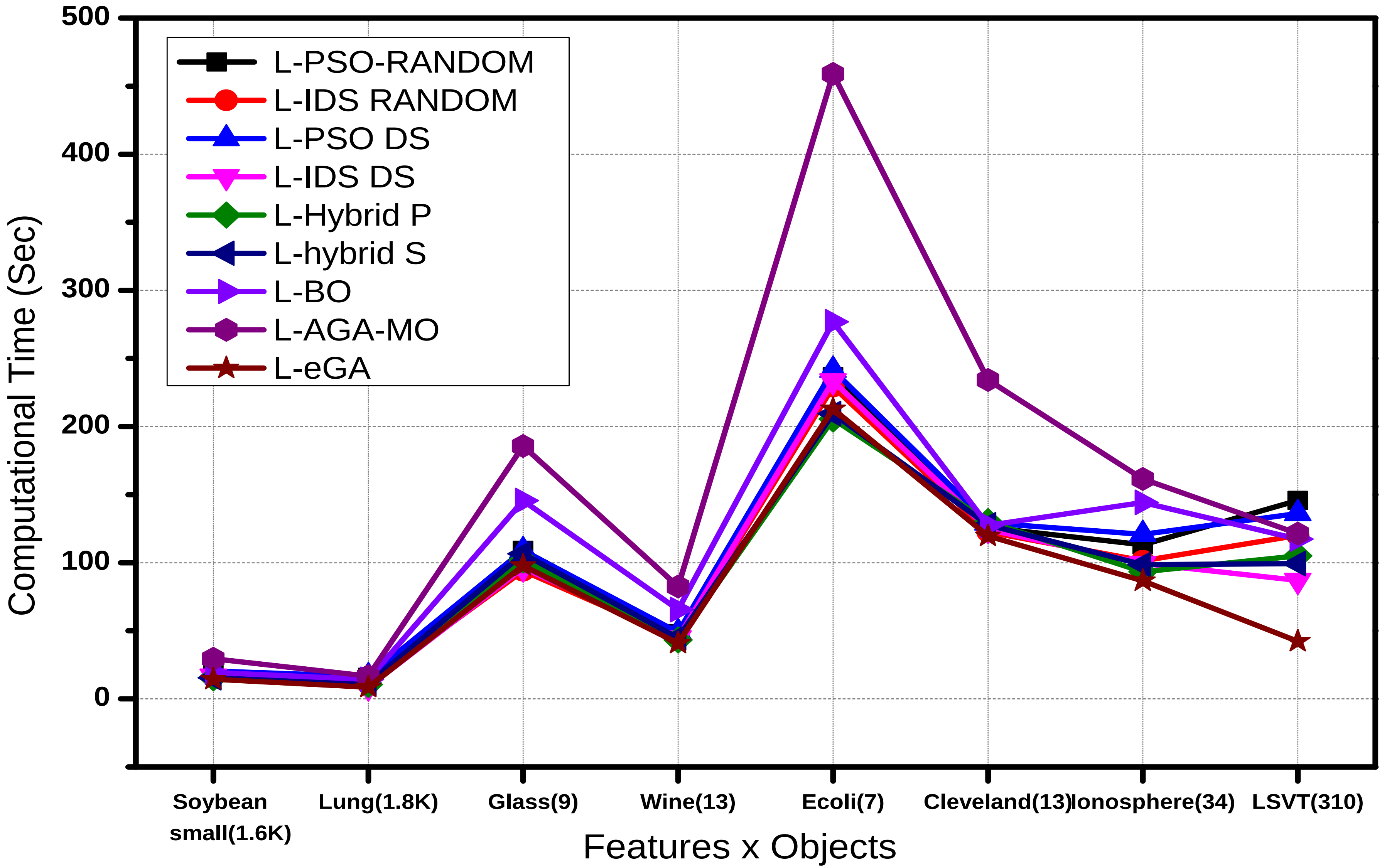

Figure 8.6: Computational time vs (Number of Features x Objects) with L-FRFS technique

Similar trend related to computational time was observed for L-FRFS based methodology. A corresponding plot of computational time vs object count is shown in Figure 8.4, similarly plot of computational time vs object count is shown in Figure 8.5. Also the dependency of computational time can be approximated to be linearly related to the product of number of objects and number of features. For a given dataset, all the methods take almost similar amount of computational time, as evident from Figure 8.6.

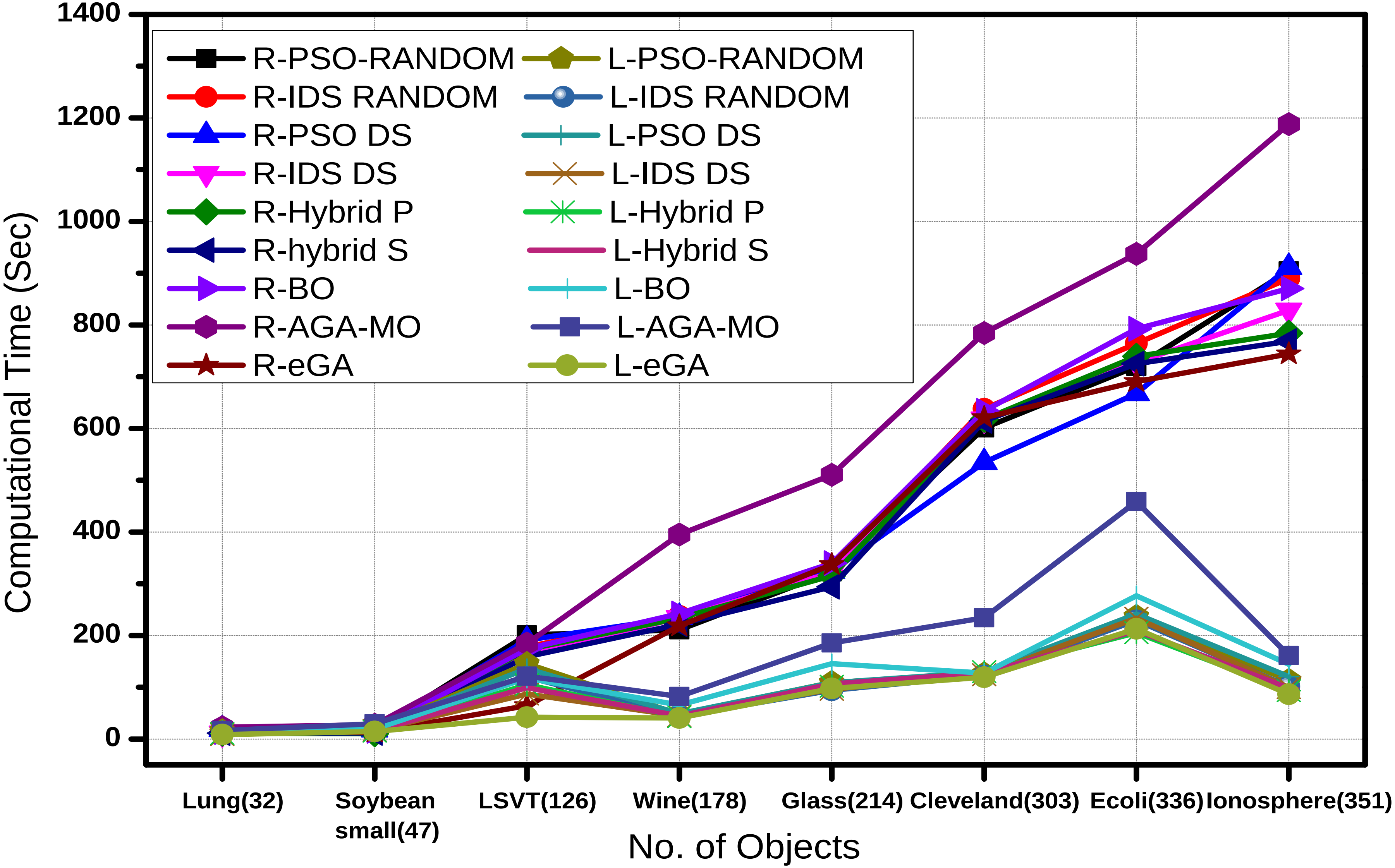

Figure 8.7: Computational time vs Number of Objects with RST and L-FRFS technique

It is clearly observed from Figure 8.7 that the computational time taken by the RST technique for all the search methods is more than those of L-FRFS technique, as we go for the higher number of objects. In Figure 8.7, prefix R in R-PSO-RANDOM, R-IDS-RANDOM, R-PSO-DS, R-IDS-DS, R-Hybrid-P, R-Hybrid-S, R-BO, R-AGA-MO and R-eGA corresponds to RST and prefix L in L-PSO-RANDOM, L-IDS-RANDOM, L-PSO-DS, R-IDS-DS, L-Hybrid-P, L-Hybrid-S, L-BO, L-AGA-MO and L-eGA corresponds to L-FRFS.

At last let us have the best results out of this work, for all the dataset are shown in Table 8.1. Here we have best result for both RST and L-FRFS, and we observe that the reduct produced by RST method is smaller to L-FRFS method.

The following qualitative reasons explains the smaller feature subset of RST method compared to L-FRFS approach.

1. L-FRFS is not used for improved performance in terms of accuracy or reduct size, rather it was used for capturing fuzziness in description of datasets. Hence, in the effort to capture fuzziness, reduct may not be smaller as compared to a rough set approach.

Fuzzy lower approximation maintains dependency of data which will never be zero, whereas, rough set is crisp and hard-limits the approximation. Due to this, RST may ignore or preserve features in a crisp manner resulting in loss of dependency and therefore reduct produced may be smaller in size.

2. In RST, there is a need to discretize the feature values. The discernibility of features are affected by the quantization and therefore become dependent on the quantization of feature value. Whereas, in L-FRFS the real feature values are taken as it is and no quantization is required and discernibility of features are therefore more accurate as well as meaningful. This is the reason why the RST approach may provide smaller size features compared to L-FRFS approach.

For the dataset Cleveland, Ecoli, Glass and Wine all methods proposed in this work produce stable results i.e. standard deviation in the number of features is zero. For Ionosphere and Lung, BO and AGA-MO produce the smallest and stable reduct (i.e. standard deviation is zero) with high and stable accuracy (i.e. standard deviation is zero). For Soybean small and Lung the reduct size is same in both method i.e. RST and L-FRFS.

While comparing results in Tables I and II, it be observed that the results are superior to that of the state-of-the-art methods reported in the literature in terms of both, reduct size as well as classification accuracy.

8.2 Future Scope of Work

Following are the future scope of the work. DS-initialization proposed in this thesis can be further investigated for its applicability in other swarm and evolutionary algorithms. The methods proposed in this thesis can be extended for other datasets. Boundary approximation of fuzzy rough set can also be investigated for feature selection problems. Studies similar to this thesis can be performed for unsupervised datasets, where labels are unknown. The work can be extended by exploring further hybridization of the algorithms proposed in this thesis with other algorithms. The proposed methods can be extended to the other real life application areas.