International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15 (2023) pp. 566-574

Enhancement of Embedded Feature Selection Method for 3D Molecular Structure of Amphetamine-Type Stimulants (ATS) Drugs

Phoebe E. Knight[1], Azah Kamilah Muda[2], Norfadzlia Mohd Yusof[3] and Noor Azilah Muda[4]

[1,2,4] Faculty of Information and Communication Technology,
Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, Durian Tunggal,
76100 Melaka, Malaysia.
phoebeellyin@gmail.com, azah@utem.edu.my, azilah@utem.edu.my

[3] Faculty of Electrical and Electronic Engineering Technology,
Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, Durian Tunggal,
76100 Melaka, Malaysia.
norfadzlia@utem.edu.my

***Abstract*: The fundamental of this research paper is to propose the enhancement of the embedded feature selection method between Sequential Forward Floating Selection (SFFS) and Support Vector Machine-Recursive Feature Elimination (SVM- RFE). These two feature selection methods were primarily embedded to enhance the effectiveness and quality of data identification. In this research, three feature selection variables from MATLAB were used as a mechanism to compare and evaluate the best feature throughout the embedded feature selection process. Those features are fscmrmr, relieff and chi-square. Lastly, a standard evaluation technique, a cross-validation process in WEKA, is used to systematically run repeated percentage splits. This study ran selected classifiers with 10 times cross-validations to capture each experiment's accuracy.**

Keywords: Amphetamine-Type Stimulants (ATS), Sequential Forward Floating Selection (SFFS), 3D Molecular Structure, Embedded Feature Selection, Drug Image Recognition, Support Vector Machine-Recursive Feature Elimination (SVM- RFE)

I. Introduction

In the previous research paper [1], the analysis of different feature selection methods, which were composed of 4 features selection Sequential Forward Floating Selection (SFFS), Sequential Backward Floating Selection (SBFS), Sequential Forward Selection (SFS) and Sequential Backward Selection (SBF) that was compared with 6 different classifiers Random Forest (RF), Naive Bayes (NB), IBK, Filtered Classifier (FC), Decision Tables (DT), and J48 and at the same time with different size of train and test set have been evaluated. Meanwhile, this research paper aims to investigate the performance of an enhancement-embedded feature selection method where the feature selection was chosen from previous experiments.

Due to WEKA tools' user-friendly interface, which makes it simpler for researchers to use, they were employed in this study to run selected classifiers with 10 times cross-validations to capture each experiment's accuracy. In addition, since MATLAB offers a sophisticated machine-learning method, it is also utilised in this research. Three crucial factors were compared in MATLAB to choose the optimal feature for improved feature selection techniques, which are fscmrmr, relieff and chi-square. In order to reduce the amount of data needed to execute the experiment, only the top 50 selected features will be chosen.

II. Related Work

Embedded feature selection method have been long used in data analysis due to the effectiveness and improvement of quality data. Researchers were struggles with handling massive and unfiltered data that contain noises and unnecessary features while analysing these data. One of the primary cause for researcher prefer embedded feature selection is it ease them to aim their desire and targeted results which where they preferred to focus on.

In 2022, a research by Knight et al., proposed a few feature selection techniques, including Sequential Forward Floating Selection (SFFS), Sequential Forward Selection (SFS), Sequential Backward Floating Selection (SBFS), Sequential Backward Selection (SBS), and Support Vector Machine-Recursive Feature Elimination (SVM-RFE), were used to analyse the effectiveness of ATS drugs identification. This paper aims to assess which feature selection techniques perform better regarding classification accuracy and identification for a big dataset. The performance of categorization accuracy is assessed through a thorough

WEKA verification. This is accomplished by contrasting various classifiers with all features and with only a subset of features (using feature selection methods). According to the experimental work, the classification accuracy performance with a subset of features has a similar accuracy to the classification accuracy performance with all features. This demonstrates how feature selection techniques aid in accelerating and improving accuracy performance. The outcome also shows that J48, IBk, and Random Forest (RF) are the top three classifiers to employ for further evaluation, while SFFS are the best feature selection methods to use to embed with SVM-RFE.

In research by Peng et al.2010), a classical Sequential Forward Floating Selection (SFFS) method, starting from X0, continuously performs the loop of feature inclusion, conditional exclusion and continuation of conditional exclusion on the features in Xk, based on a specifically defined evaluation function J(Xk). In conventional wrapper approaches, classification accuracy is normally used to define the J(Xk). This paper defines the J(Xk) by the average AUC of the classification cross-validation using the associated feature subset Xk and the SVM. For k-fold cross-validation, a training dataset D is divided into k data subsets, denoted as Di, i = 1 ~ k. Each of these data subsets (e.g. Di) is used as the validation data, and the rest (D − Di) is then used to train an SVM classifier. The evaluation function J(X) for the associated feature subset X is defined by the average of the AUCs:

$$J(X) = \frac{\sum_{i=1}^{k} AUC_i}{k}$$

where AUCi is the area under the testing ROC curve of the classifier trained by the data (D − Di) and tested by Di.

Another related research about hybrid wrappers that were studied by Hui et al. in 2017 suggested a better Wrapper-based Feature Selection approach for carrying out the feature selection task. The SVM was used as a classifier in feature selection by the suggested WFS approach. The SVM classifier's training accuracy determined each feature's effectiveness following a multi-fold cross-validation evaluation that sought to maximise model consistency while reducing bias and overfitting. By eliminating repetitive computations of identical and undesired feature combinations, the suggested WFS decreased execution time. As a result, the suggested WFS method only examined distinct feature combinations using two ways for each iteration. It is noticed by ignoring the repetitive evaluation of identical feature combinations that take place during the process of feature combinations being generated randomly, as well as unfavourable low-quality solutions from previous recursive simulations. Additionally, the proposed WFS technique created feature combinations for the subsequent level based on the effectiveness of the preceding level. The proposed WFS algorithm's methodology is shown in Figure 1. The algorithm assessed each individual feature during first-level selection. The computer then created the second-level feature combinations by fusing the characteristics that performed better than average with the unselected individual features (red-outlined rectangle in Figure 1). When the feature combination had fully exploited all of the retrieved characteristics, this process ended. In the end, the algorithm determined that the feature combinations with the fewest features and the best training accuracy (shown by the yellow-filled rectangle in Figure 1), were the ones that best represented the complete dataset. The feature selection process lowered the feature dimensionality for machine learning algorithms in addition to choosing the dataset's most representative features. The skewness factor and form factor, or characteristics A and D, were chosen as a consequence.

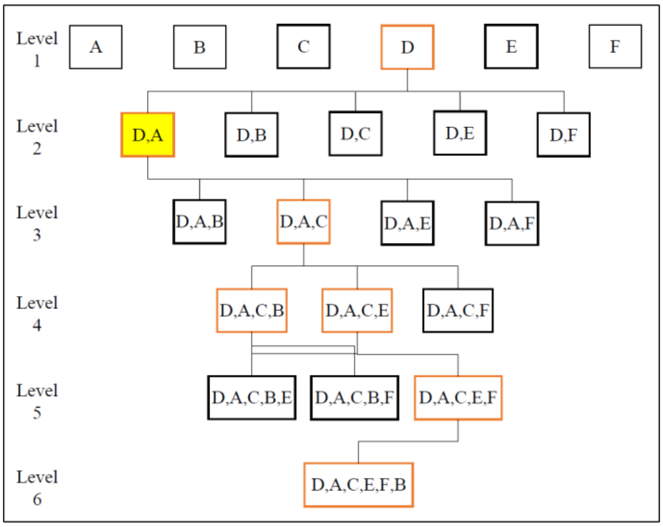


Figure 1. Methodology of WFS algorithm

A later study by Chen and Chen in 2015, incorporating the cosine distance into support vector machines (SVM), presents a wrapper method known as cosine similarity measure support vector machines (CSMSVM) to remove pointless or unnecessary features during classifier generation. In the past, feature selection methods had extracted features and learned SVM parameters separately or in the attribute space,. This may have led to the loss of information about the classification process or increased classification error when the kernel SVM was introduced. To remove low-relevance features, the proposed CSMSVM system simultaneously performs feature selection, and SVM parameter learning, and optimises the shape of an anisotropic RBF kernel in feature space. Additionally, the suggested method is built on a strong theoretical foundation thanks to the Bayesian interpretation of the unique methodology, and the iteration algorithm, which is proposed to maximise the feature weight, has proven successful in maximising the maximum posterior (MAP). CSMSVM outperformed the other procedures in trials, comparing the novel method with well-known feature selection strategies, boosting pattern recognition accuracy with fewer features.

III. Data Collection

In this research, the raw data were originally collected from ATS Drugs 3D Molecular Structure Representation Dataset by Pratama et al., 2017. This data source has an enormous number of features (1185 features), with 3595 ATS drugs and 3595 non-ATS drugs among its total 7190 sample records. A computational numerical dataset for 3D Molecular Structures of ATS Drugs is the result of this computation method This dataset will be implemented in this study to comprehend 3D molecule structure and identify important aspects for conformational properties. The data files are saved in Excel format, and the characteristics are specified in decimals can see the original author's work in [2] for further information on the dataset's data-collecting method. Table 1 below shows the summary of datasets that were originally collected.

Dataset's Name	Number of Features	Number of Instances	Number of Classes	Class Distribution
ATS Drugs 3D Molecular Structure Representation Dataset (Pratama et al., 2017)	1185	7190	2	ATS: 3595 Non-ATS: 3595

Table 1. Summary of ATS Drugs 3D Molecular Structure Representation Dataset

From the dataset that has been collected, the dataset is then continue being further analyzed as written in [2]. A train set and a test set will be created from the dataset. These two sets will be divided into five train-test partitions with the following data size ratios: 90:10, 80:20, 70:30, 60:40, and 50:50. By comparing 5 feature selection approaches and 6 classifiers, the partitioning of data size is intended to gather a deeper accuracy. The five feature selection techniques that were involved are Sequential Forward Floating Selection (SFFS), Sequential Backward Floating Selection (SBFS), Sequential Forward Selection (SFS), Sequential Backward Selection (SBF) and Support Vector Machine-Recursive Feature Elimination (SVM-RFE). Whereas the 6 classifiers include Random Forest (RF), IBK, Naive Bayes, J48, Decision Table and Filtered Classifier. Table 2 below shows the summary of details involved in the analysis of feature selection using the dataset collected.

Train : Test	90:10	80:20	70:30	60:40	50:50	-
Classifier	Random Forest (RF)	Naïve Bayes (NB)	Decision Table (DT)	Filtered Classifier (FC)	IBk	J48
Feature Selection	SFS	SBF	SFFS	SBFS	SVM-RFE	-

Table 2. Summary of Analysis of Feature Selection of ATS Drugs 3D Molecular Structure

There were two experiments involved in this experiment, classification without feature selection and classification with feature selection. The illustrations of both experiments are illustrated in Figures 2 and 3 below.

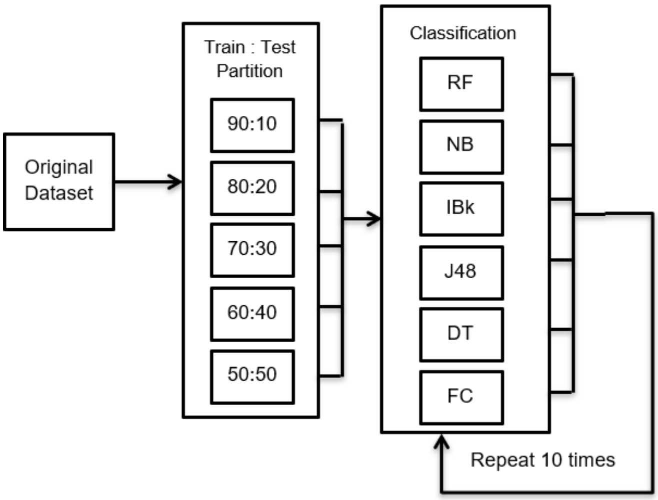


Figure 2. Classification without Feature Selection

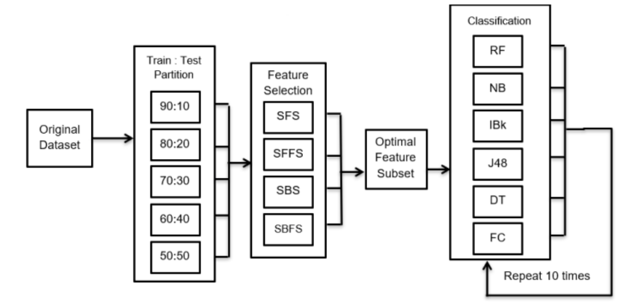
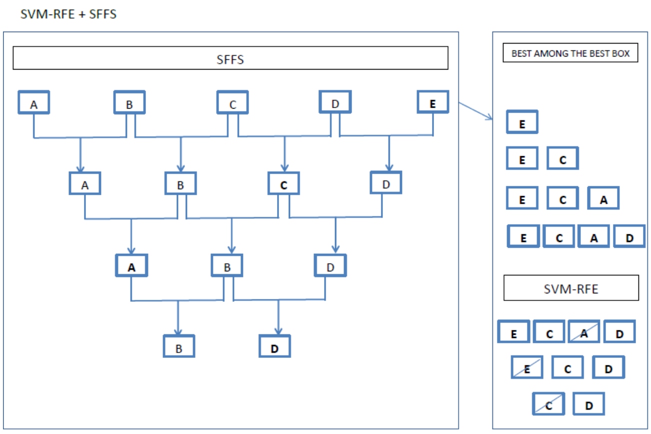


Figure 3. Classification with Feature Selection

The result of the experiments above shows that J48, IBk and Random Forest (RF) are the best three classifiers to use for future evaluation, while Sequential Forward Floating Selection (SFFS) are the best feature selection methods to use to embed with SVM- RFE for next experiment.

IV. Proposed Method

To shorten the process, this experiment will use the overall results data collected from the previous experiments as guidance for the experiment. Based on previous experiments' results, Sequential Forward Floating Selection (SFFS) might be a good companion for SVM- RFE for this enhancement experiment. To better understand the experiment, illustrations of experiments were illustrated in Figures 4 and 5 below.



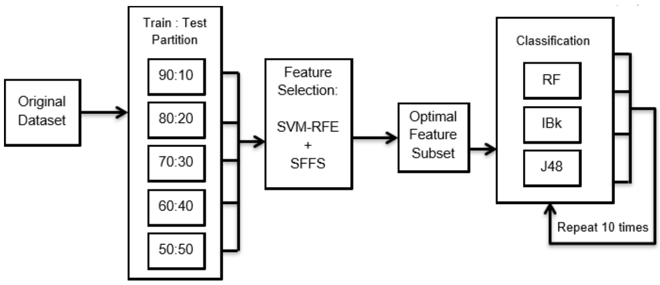


Figure 4. Classification with Enhanced Feature Selection
Figure 5. Illustration of Enhancement Feature Selection

In this experiment, the feature selection from enhancement experiment is extended into a detailed process by enhancing it with the help of 3 different feature selection variables, fscmrmr, relieff and chi-square, as a mechanism to select the best features in the enhancement of feature selection between SVM- RFE and Sequential Forward Floating Selection (SFFS). Table 3 below shows the list of information that is involved in the experiments.

Train : Test	90:10	80:20	70:30	60:40	50:50
Classifiers	Random Forest (RF)		IBk		J48
Feature Selection	SFFS		SVM- RFE		
Feature Selection Mechanisms	fscmrmr		chi-square		relieff

Table 3. Summary Details of an Enhanced Feature Selection Experiment

Therefore, three experiments were involved: classification with enhanced feature selection using fscmrmr feature selection, classification with enhanced feature selection using relieff feature selection and classification with enhanced feature selection using chi-square feature selection. All three experiments were illustrated in Figure 6, Figure 7 and Figure 8.

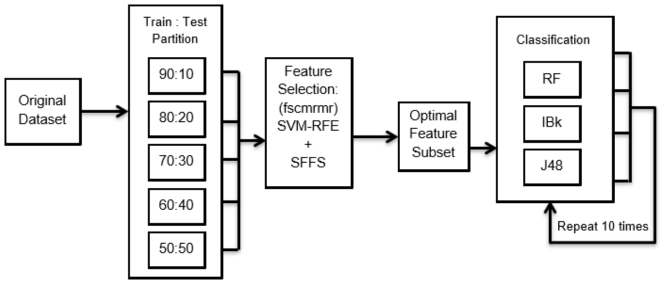


Figure 6. Classification with Enhanced Feature Selection using fscmrmr Feature Selection Experiment

Based on Figure 6, the experiment of the classification with enhanced feature selection using fscmrmr feature selection means the implementation of hybrid between SVM- RFE and Sequential Forward Floating Selection (SFFS) is using fscmrmr algorithm to identify the worst and best selection data to undergo the hybrid process.

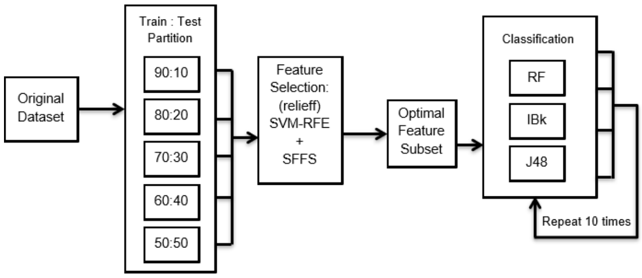


Figure 7. Classification with Enhanced Feature Selection using relieff Feature Selection Experiment

Based on Figure 7 above, the experiment of the classification with enhanced feature selection using relieff feature selection means the implementation of hybrid between SVM- RFE and Sequential Forward Floating Selection (SFFS) is using relieff algorithm to identify the worst and best selection data, to undergo the hybrid process.

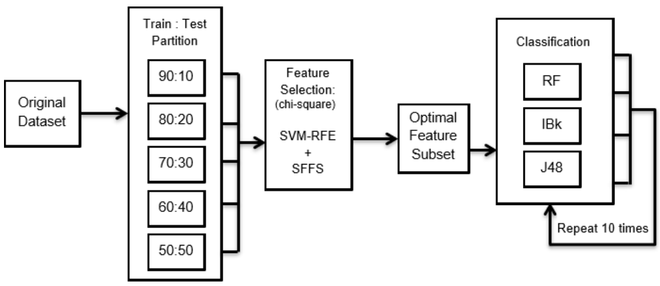


Figure 8. Classification with Enhanced Feature Selection using chi-square Feature Selection Experiment

Based on Figure 8, the experiment of the classification with enhanced feature selection using chi-square feature selection means the implementation of a hybrid between SVM- RFE and Sequential Forward Floating Selection (SFFS) uses the chi-square algorithm to identify the worst and best selection data, to undergo the hybrid process.

These three experiments were proposed to further the research on the enhancement of feature selection. In conclusion, three features selection (fscmrmr, relieff and chi-square) were used to compare and select the best features in the enhanced feature selection methods between Sequential Floating Forward Selection (SFFS) and SVM- RFE. This process was then classified with three of the best classifiers Random Forest (RF), IBk and J48 and at the same time, with different sizes of train and test sets.

V. Performance Measurement

A. Cross-Validation in WEKA

Repeated percentage splits are done systematically using cross-validation, a common evaluation approach. Divide a dataset into ten equal pieces ("folds"), test one at a time, then train on the remaining nine pieces simultaneously. The average of the resulting 10 evaluation findings is obtained. When doing the initial division in "stratified" cross-validation, we make sure that each fold has roughly the right amount of the class values. Weka runs the learning algorithm a final (11th) time on the complete dataset after doing 10-fold cross-validation and computing the evaluation findings to produce the model that it prints out.

B. Feature Selection variables in MATLAB

There are three feature selection variables that were used in MATLAB as a mechanism to compare and evaluate the best feature throughout the embedded feature selection process. Those feature selections are fscmrmr, relieff and chi-square.

1) Features Selection Classification using Minimum Redundancy Maximum Relevance (fscmrmr)

The MRMR method [23] identifies an ideal set of characteristics that may accurately describe the response variable and are mutually and maximally different. The method maximises the relevance of a feature set to the response variable while minimising its redundancy. The mutual information of variables—pairwise mutual information of features and mutual information of a feature and the response—is used by the algorithm to quantify the redundancy and relevance. This algorithm can be applied to classification issues.

The MRMR algorithm's objective is to identify an ideal set of features S that maximises V_S, the relevance of S with respect to a response variable y, and minimises W_S, the redundancy of S, where V_S and W_S are defined with mutual information *I*:

$$V_S = \frac{1}{|S|} \sum_{x \in S} I(x, y),$$

$$W_S = \frac{1}{|S|^2} \sum_{x,z \in S} I(x, z). \qquad (1)$$

|S| is the number of features in S.

Finding an optimal set S requires considering all $2^{|\Omega|}$ combinations, where Ω is the entire feature set. Instead, the MRMR algorithm ranks features through the forward addition scheme, which requires O(|Ω|·|S|) computations, by using the mutual information quotient (MIQ) value.

$$\mathrm{MIQ}_x = \frac{V_x}{W_x}, \qquad (2)$$

where V_x and W_x are the relevance and redundancy of a feature, respectively:

$$V_x = I(x, y),$$

$$W_x = \frac{1}{|S|} \sum_{z \in S} I(x, z). \qquad (3)$$

Using the MRMR method, the fscmrmr function ranks all features in Ω and returns idx (the indices of features ordered by feature importance). As a result, the computation's cost is O(|Ω|2). The function uses a heuristic technique to quantify the significance of a feature and then produces a score (or scores). A high score value denotes the significance of the corresponding predictor. A decline in the feature significance score also indicates confidence in feature selection. For instance, the score value of the second most essential attribute is substantially smaller than the score value of x if the software is confident in choosing it. The results can be used to identify an ideal set S for a particular collection of features.

fscmrmr ranks features as follows:

1. Select the feature with the largest relevance, $\max_{x \in \Omega} V_x$. Add the selected feature to an empty set S.
2. Find the features with nonzero relevance and zero redundancy in the complement of S, S^c.
   - If S^c does not include a feature with nonzero relevance and zero redundancy, go to step 4.
   - Otherwise, select the feature with the largest relevance, $\max_{x \in S^c, W_x = 0} V_x$. Add the selected feature to the set S.
3. Repeat Step 2 until the redundancy is not zero for all features in S^c.
4. Select the feature that has the largest MIQ value with nonzero relevance and nonzero redundancy in S^c, and add the selected feature to the set S.
$$\max_{x \in S^c} \mathrm{MIQ}_x = \max_{x \in S^c} \frac{I(x, y)}{\frac{1}{|S|} \sum_{z \in S} I(x, z)}.$$
5. Repeat Step 4 until the relevance is zero for all features in S^c.
6. Add the features with zero relevance to S in random order.

If the software is unable to locate a feature that meets the step's requirements, it may skip any phase altogether.

2) ReliefF

When y is a multiclass categorical variable, ReliefF determines the weights of the predictors. The method rewards predictors who assign different values to neighbours in various classes while penalising those who assign different values to neighbours in the same class.

All predictor weights W_j are initially set to 0 by ReliefF. The algorithm then chooses a random observation x_r, finds the k-nearest observations to x_r for each class, and updates all of the weights for the predictors F_j for each nearest neighbour x_q as follows:

If x_r and x_q are in the same class,

$$W_j^i = W_j^{i-1} - \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq}.$$

If x_r and x_q are in different classes,

$$W_j^i = W_j^{i-1} + \frac{p_{y_q}}{1 - p_{y_r}} \cdot \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq}$$

- W_j^i is the weight of the predictor F_j at the *i*th iteration step.
- p_{yr} is the prior probability of the class to which x_r belongs, and p_{yq} is the prior probability of the class to which x_q belongs.
- *m* is the number of iterations specified by 'updates'.

- $\Delta_j(x_r,x_q)$ is the difference in the value of the predictor F_j between observations x_r and x_q. Let x_{rj} denote the value of the *j*th predictor for observation x_r, and let x_{qj} denote the value of the *j*th predictor for observation x_q.
    - For discrete F_j,

$$\Delta_j(x_r, x_q) = \begin{cases} 0, & x_{rj} = x_{qj} \\ 1, & x_{rj} \neq x_{qj} \end{cases}.$$

    - For continuous F_j,

$$\Delta_j(x_r, x_q) = \frac{|x_{rj} - x_{qj}|}{\max(F_j) - \min(F_j)}.$$

- d_{rq} is a distance function of the form

$$d_{rq} = \frac{\tilde{d}_{rq}}{\sum_{l=1}^{k} \tilde{d}_{rl}}.$$

The distance is subject to the scaling
~drq=e$^{-(rank(r,q)/sigma)2}$

where rank(r,q) is the location of the *q*th observation among the *r*th observation's closest neighbours, sorted by distance. The number k indicates the nearest neighbours. By specifying 'sigma', the scaling can be changed.

3) Chi-Square

When comparing categorical variables from a random sample, the statistical test known as chi-square is used to assess the degree of fit between the predicted and actual results. Chi-square is most frequently employed by academics who are analysing survey response data because it pertains to categorical variables. Chi-square tests come in two different varieties. Both have different uses for the chi-square statistic and distribution:

- A chi-square goodness of fit test establishes if sample data correspond to the population.
- A chi-square test for independence examines the relationship between two variables in a contingency table. It examines whether the distributions of categorical variables diverge from one another in a more generic sense.

The formula for the chi-square statistic used in the chi square test is:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \qquad (4)$$

where:
c = degree of freedom, O = Observed value(s), E= Expected value(s)

You can obtain the *p*-value using a chi square test. You can determine the significance of your test results using the *p*-value. You require two pieces of data to run a chi square test and determine the *p*-value:

1. Degrees of freedom. Simply the number of categories minus one results in that.
2. The alpha level (α). You or the researcher choose this. There are various levels, such as 0.01 or 0.10, besides the standard alpha level of 0.05 (5%).

Both the degrees of freedom (df) and the alpha level are typically provided to you in a question in elementary statistics or AP statistics. Normally, you don't need to figure out what they are. The df is rather straightforward, but you might have to figure it out for yourself: After counting the categories, take away 1. Following the chi-square (X^2) sign are the degrees of freedom. For instance, the chi square below displays 6 df: X^2_6. The chi square also displays 4 df: X^2_4.

VI. Result and Discussion

In this research, three experiments were composed where the enhanced feature selection process used three different variables as a mechanism to select the best feature. The three variables that involve selecting the best features are fscmrmr, relieff and chi-square. Below are the details of three experiments with their accuracy that have been collected in tables.

A. Classification with Enhanced Feature Selection using fscmrmr Feature Selection

The first experiment was enhanced by fscmrmr as a method to select the best feature. Then it was classified with 3 best classifiers chosen from previous experiments. Table 9 shows the accuracy data that have been collected.

Train:Test\ Classifiers	J48	IBk	RF
90:10	63.3596	60.1221	67.3312
80:20	60.3964	59.9183	66.3770
70:30	63.9248	59.8583	67.0508
60:40	62.1523	59.2837	66.0756
50:50	61.6273	58.3450	65.8276
Average (%)	**62.29**	**59.51**	**66.53**

Table 9. Classifications Accuracy with Enhanced Feature Selection using fscmrmr Feature Selection.

Based on the table above, it shows that Random Forest (RF) has the highest number of accuracy after 10 times cross-validation with different sizes of train and test set partitions after using fscmrmr algorithm to enhanced feature selection between SVM- RFE and Sequential Floating Forward Selection (SFFS).

B. Classification with Enhanced Feature Selection using relieff Feature Selection

Next, the experiment was enhanced by relieff as a method to select the best feature. Then it was classified with 3 best classifiers chosen from previous experiments. Table 10 shows the accuracy data that have been collected.

Train:Test\ Classifiers	J48	IBk	RF
90:10	64.3873	62.7029	68.8630
80:20	63.7692	63.6735	69.0021
70:30	65.2361	62.5240	69.0609
60:40	64.6964	62.5696	68.2198
50:50	63.5605	62.5591	68.8456
Average (%)	64.33	62.81	**68.80**

Table 10. Classifications Accuracy with Enhanced Feature Selection using relieff Feature Selection.

Based on the table above, it shows that random forest (RF) has the highest number of accuracy after 10 times cross-validation with different sizes of train and test set partitions after using relieff algorithm to enhanced feature selection between SVM-RFE and Sequential Floating Forward Selection (SFFS).

C. Classification with Enhanced Feature Selection using chi-square Feature Selection

Lastly, the experiment was enhanced by chi-square as a method to select the best feature. Then it was classified with 3 best classifiers chosen from previous experiments. Table 11 shows the accuracy data that have been collected.

Train:Test\ Classifiers	J48	IBk	RF
90:10	60.9952	58.7622	64.4491
80:20	59.6054	59.3881	65.3425
70:30	62.9744	58.4840	64.9679
60:40	61.3816	60.3037	65.7685
50:50	62.2114	59.1934	65.7858
Average (%)	**61.43**	**59.23**	**65.26**

Table 11. Classifications Accuracy with Enhanced Feature Selection using Chi-square Feature Selection.

Table 11 shows that Random Forest (RF) has the highest accuracy after 10 times cross-validation with different sizes of train and test set partitions after using the chi-square algorithm to enhance feature selection between SVM- RFE and Sequential Floating Forward Selection (SFFS).

VII. Conclusion

This concludes that 3 features selection (fscmrmr, relieff and chi- square) was used to compare and select the best features in the enhanced feature selection methods between Sequential Floating Forward Selection (SFFS) and SVM- RFE. This process was then classified with 3 of the best classifiers Random Forest (RF), IBk and J48 and at the same time with different sizes of train and test sets. The result shows that Random Forest (RF) is the best classifier, while Relieff is the best feature selection as a medium to select the important features to use for future evaluation.

This research is unique in solving the high dimensionality of the ATS drug dataset by sorting out the important features that have high discriminative power in identifying ATS drugs using the proposed embedded feature selection method between Support Vector Machine-Recursive Feature Elimination (SVM- RFE) and Sequential Forward Floating Selection (SFFS). The effectiveness of the proposed method is verified by comparing the classification performance with the individual feature selection technique and multiple sets of partition data, training and test sets. The results show that the proposed method performed best in selecting significant features, proven by a series of experiments conducted in this research.

Acknowledgment

The authors would like to acknowledge Universiti Teknikal Malaysia Melaka through the Fundamental Research Grant Scheme [FRGS/1/2020/FTMK-CACT/F00461] from the Ministry of Higher Education, Malaysia.

References

[1] Knight, P. E., Muda, A. K., & Pratama, S. F. (2021, December). Analysis of feature selection method for 3D molecular structure of amphetamine-type stimulants (ATS) drugs. In *International Conference on Soft Computing and Pattern Recognition* (pp. 118-135). Cham: Springer International Publishing.

[2] Pratama, S. F., Muda, A. K., Choo, Y.-H. and Muda, N. A., 2014. "A New Swarm-Based Framework for Handwritten Authorship Identification in Forensic Document Analysis". In: Muda, A. K., Choo, Y.-H., Abraham, A. and N. Srihari, S. (eds.), Computational Intelligence in Digital Forensics: Forensic Investigation and Applications, pp. 385-411. Cham, Switzerland: Springer International Publishing

[3] A. Bonnaccorsi. "On the Relationship between Firm Size and Export Intensity", *Journal of International Business Studies*, XXIII (4), pp. 605-635, 1992. *(journal style)*

[4] R. Caves. *Multinational Enterprise and Economic Analysis*, Cambridge University Press, Cambridge, 1982. *(book style)*

[5] M. Clerc, M. "The Swarm and the Queen: Towards a Deterministic and Adaptive Particle Swarm Optimization". In *Proceedings of the IEEE Congress on*

Evolutionary Computation (CEC), pp. 1951-1957, 1999. *(conference style)*

[6] H.H. Crokell. "Specialization and International Competitiveness", in *Managing the Multinational Subsidiary*, H. Etemad and L. S, Sulude (eds.), Croom-Helm, London, 1986. *(book chapter style)*

[7] K. Deb, S. Agrawal, A. Pratab, T. Meyarivan. "A Fast Elitist Non-dominated Sorting Genetic Algorithms for Multiobjective Optimization: NSGA II". *KanGAL report 200001*, Indian Institute of Technology, Kanpur, India, 2000. *(technical report style)*

[8] Sanz, Hector, Clarissa Valim, Esteban Vegas, Josep M. Oller, and Ferran Reverter. "SVM- RFE: selection and visualisation of the most relevant features through non-linear kernels." BMC Bioinformatics 19, no. 1 (2018): 432.

[9] Sewell, M., 2007. Feature Selection. [online]. Available at:http://machine-learning.martinsewell.com/feature-selection/feature- selection.pdf [Accessed on 25 October 2009].

[10] Tang, Y., Zhang, Y. and Huang, Z., 2007. Development of Two-Stage SVM- RFE Gene Selection Strategy for Microarray Expression Data Analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4(3), pp. 365-381.

[11] Wang, Yan, Keke Liu, Qin Ma, Yongfei Tan, Wei Du, Yidan Lv, Yuan Tian, and Hao Wang. "Pancreatic cancer biomarker detection by two support vector strategies for recursive feature elimination." Biomarkers in medicine 00 (2019).

[12] Yan, K. and Zhang, D., 2015. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sensors and Actuators B: Chemical, 212, pp. 353- 363.

[13] Yoon, S. and Kim, S., 2009. Mutual information-based SVM- RFE for diagnostic classification of digitised mammograms. Pattern Recognition Letters, 30(16), pp. 1489-1495.

[14] Zhang, Ying, Qingchun Deng, Wenbin Liang, and Xianchun Zou. "An efficient feature selection strategy based on multiple support vector machine technology with gene expression data." BioMed research international 2018 (2018).

[15] Witten, I. (2013, August 8). Cross-validation. FutureLearn. https://www.futurelearn.com/info/courses/data-mining-with-weka/0/steps/25384.

[16] Ding, C., and H. Peng. "Minimum redundancy feature selection from microarray gene expression data." Journal of Bioinformatics and Computational Biology. Vol. 3, Number 2, 2005, pp. 185–205.

[17] Darbellay, G. A., and I. Vajda. "Estimation of the information by an adaptive partitioning of the observation space." IEEE Transactions on Information Theory. Vol. 45, Number 4, 1999, pp. 1315–1321.

[18] [25] W. H. O. Geneva, "Neuroscience of psychoactive substance use and dependence," World Health Organization, Switzerland, 2004.

[19] [26] Ding, Y. and Wilkins, D., 2006."Improving the Performance of SVM- RFE to Select Genes in Microarray Data." BMC Bioinformatics, 7(2), p. S12.

[20] [27] Guyon, I. and Elisseeff, A., 2003. "An Introduction to Variable and Feature Selection." Journal of Machine Learning Research, 3, pp. 1157-1182.

[21] Guyon, I., Weston, J., Barnhill, S. and Vapnik, V., 2002. "Gene Selection for Cancer Classification using Support Vector Machines." Machine Learning, 46(1), pp. 389-422.

[22] Hall, M. A., 1999. "Correlation-based Feature Subset Selection for Machine Learning."

[23] Doctor of Philosophy Dissertation, University of Waikato, Hamilton, New Zealand.

[24] Kennedy, J. and Eberhart, R. C., 1995. "Particle Swarm Optimization. In: Proceedings of IEEE International Conference on Neural Networks."

[25] Kohavi, R. and John, G. H., 1997. "Wrappers for Feature Subset Selection." Artificial Intelligence, 97(1-2), pp. 1-43.

[26] Li, Z., Xie, W. and Liu, T., 2018. "Efficient feature selection and classification for microarray data." PLoS ONE, 13(8), p. e0202167.

[27] Mundra, P. A. and Rajapakse, J. C., 2010. "SVM- RFE with MRMR Filter for Gene Selection." IEEE Transactions on NanoBioscience, 9(1), pp. 31-37.

[28] Portinale, L. and Saitta, L., 2002. "Feature Selection: State of the Art. Feature Selection", pp. 1-22. Alessandria: Universita del Piemonte Orientale.

[29] Pratama, S. F., Muda, A. K., Choo, Y.-H. and Muda, N. A., 2014. "A New Swarm-Based Framework for Handwritten Authorship Identification in Forensic Document Analysis". In: Muda, A. K., Choo, Y.-H., Abraham, A. and N. Srihari, S. (eds.), Computational Intelligence in Digital Forensics: Forensic Investigation and Applications, pp. 385-411. Cham, Switzerland: Springer International Publishing.

[30] Pudil, P., Novovicova, J. and Kittler, J., 1994. "Floating Search Methods in Feature Selection. Pattern Recognition Letters", 15, pp. 1119-1125.

[31] Rustam, Z., and N. Maghfirah. "Correlated based SVM-RFE as feature selection for cancer classification using microarray databases." In AIP Conference Proceedings, vol. 2023, no. 1, p. 020235. AIP Publishing, 2018.

[32] Saeys, Y., Inza, I. and Larranaga, P., 2007. "A review of feature selection techniques in bioinformatics. Bioinformatics", 23(19), pp. 2507-2517.

[33] Sanz, Hector, Clarissa Valim, Esteban Vegas, Josep M. Oller, and Ferran Reverter. "SVM- RFE: selection and visualisation of the most relevant features through non-linear kernels." BMC Bioinformatics 19, no. 1 (2018): 432.

[34] Vikash Raj Luhaniwal. (2020, October 24). "A comprehensive guide to feature selection using wrapper methods in Python."Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/10/a-comprehensive-guide-to-feature- selection-using-wrapper-methods-in-python/Tang, Y., Zhang, Y. and Huang, Z., 2007. Development of Two-Stage SVM-RFE Gene Selection Strategy for Microarray Expression Data Analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4(3), pp. 365-381.

Author Biographies

Phoebe E. Knight received her B.Sc in Computer Science from Universiti Teknikal Malaysia Melaka (UTeM in 2019. She is pursuing her M.SC in Artificial Intelligence, focusing on data analysis and image recognition. She now serves the Faculty of Information and Communication Technology (FTMK), Universiti Teknikal Malaysia Melaka (UTeM) as a Master's student in Computer Science and Communications Department. Her currarch interests include soft computing, pattern recognition, image processing, machine learning, computational intelligence and hybrid systems.

Azah Kamilah Draman @ Muda is a senior lecturer at the Faculty of Information and Communication Technology (FTMK), Universiti Teknikal Malaysia Melaka (UTeM). She received her degree and master's in Computer Science (Software Engineering) and her PhD in Computer Science from Universiti Teknologi Malaysia. Her research interests include soft computing, pattern recognition, image processing, machine learning, computational intelligence and hybrid systems. Besides research, Dr Azah is also involved in curriculum development at FTMK. She was appointed as the Head of Department for Software Engineering department for 4 years (2009-2012) and needs to in charge of two bachelor programs, which are Bachelor of Computer Science (Software Development) and Bachelor of Computer Science (Database Management). Those valuable experiences led her to get involved in curriculum development for another university as well. Dr Azah also has 15 years of experience teaching various subjects, including Programming, Data Structure and Algorithms, C++, Java, MFC, Object Oriented Programming, Neural Networks, Image Processing and Artificial Intelligence. She hopes she can contribute something, if not everything, to the paramount improvement omajor.

Norfadzlia Mohd Yusof is a senior lecturer at the Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka (UTeM). She received her degree (2004) and master's (2006) in Computer Science (Software Engineering) from Universiti Teknologi Malaysia (UTM). Her research interests include Artificial Intelligence, Software Engineering and Decision Support Systems. She is currently pursuing her PhDcial Intelligence at Universiti Teknikal Malaysia Melaka (UTeM).

Noor Azilah Muda received her B.Sc. and M.Sc. degrees in 1998 and 2001, respectively. She is pursuing her PhD in pattern recognition, focusing on music analysis. She is now serving the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia, as a Senior Lecturer in Software Engineering Department. Her current research interests include dning, artificial immune system, particularly the negative selection algorithm, feature selection, and the applications of data science and data analytics in different domains, including speech recognition, data and knowledge management.