Significant Features Determination for ATS Drug Identification

Y.C.Saw, A.K.Muda and Z.I.M.Yusoh
Computational Intelligence and Technologies Lab (CIT Lab), Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Durian Tunggal, 76100 Melaka, Malaysia
azah@utem.edu.my

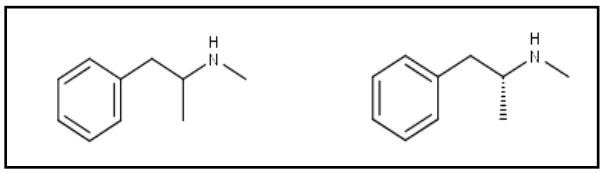
***Abstract*— Laboratory testing for ATS drug identification is a costly and lengthy process. In this paper, we propose a computational analysis approach as an alternative solution in identifying the ATS drugs. High dimensional dataset is one of the key challenges for computational analysis. This paper will investigate the effectiveness of several feature selection algorithms in identify the significant features and filter out the irrelevant features in the dataset. Specifically, four filters feature selection techniques (Information Gain (IG), Gain Ratio (GR), Symmetrical Uncertainty (SU), and ReliefF) and two embedded feature selection techniques (Support Vector Machine based Recursive Elimination Method (SVM-RFE) and Variable Importance based Random Forest (VIRF)) have been explored. The main fundamental perspective that is taken into consideration in performance analysis is to identify which feature selection technique can return minimal features while achieving a higher identification performance. The experimental evaluation on the ATS drugs 3D molecular structure representation dataset is performed using five classifiers, which are Random Forest (RF), Naïve Bayes (NB), IBK, SMO and J48 decision trees. The findings show that ReliefF and VIRF can select a smaller feature subset with the highest identification accuracy than the other feature selection techniques.**

***Index Terms*—ATS Drug; 3D Molecule Structure; Feature Selection; Filter-Embedded.**

I. INTRODUCTION

Amphetamine-type stimulant (ATS) drug is considered as a psychoactive drug that will stimulate the central nervous system (CNS) by increasing the concentration of dopamine. Dopamine is a neurotransmitter which acts as a chemical messenger in the body. Thus, the abuse of ATS drug may lead to addiction. It will elevate the mood, blood pressure, heart rate and increase the alertness of the users [1]. Today, the number of illicit manufacturers, trafficking and abuse of ATS drugs has become rampant worldwide. The investigation of these ATS drugs is essential, as an effort to prevent the international criminal illicit activities of ATS drugs [2].

The main goal of the investigation is to identify any controlled substance candidate present in the exhibit drug. In this case, the two most popular methods that are employed in laboratory investigation and testing for drugs are immunoassay and Chromatography [3]. These applications are powerful enough to provide sufficient information in identifying the compound that is present in the exhibit sample. However, it does present a few limitations. The key limitations of Immunoassay are due to its low specificity and high cross-reactivity, which may produce results with high false-positive test [3]. As for Chromatography test such as Gas Chromatography/Mass Spectrometry (GC/MS) is unable to identify drugs that pose diastereomers and positional isomers such as ATS drugs [4]. This is clearly shown in methamphetamine, which has two stereoisomers (l-methamphetamine and d-methamphetamine) as depicted in Figure 1. Besides, lab testing investigation is a laborious, expensive and time-consuming process. Hence, this work is aimed to propose a simple computational approach to overcome these limitations and eventually facilitate the activity of forensic analyst.



(a) (b)

Figure 1: (a) d-methamphetamine, and (b) l-methamphetamine

In general, a chemical substance can be identified by its characteristics, including its shape, colour, smell, etc. However, the key problem is how to represent the ATS drug to carry out computational analysis. The most well-known method is through visualisation method by capturing the shape of ATS drug chemical molecular structure. Visualization of ATS drug will be stored as either 2D or 3D object. In this work, we will only focus on visualising the ATS drug in the 3D object due to its better performance in conserving the bioactivity information of the compounds [5]–[7].

However, due to the complexity of the 3D molecular structure, the extracted features are still in high dimensionality and required more space to store it which presents a challenge to perform identification process. For example, the dataset that has been used in this work contains 7212 samples of ATS drug and non-ATS drug. Each of the samples consists of 1185 features. With this high dimensional of the dataset, it may contain irrelevant and redundant information which may degrade the performance of learning algorithm [8]. Hence, it is essential to identify the most potential feature that can denote and represent the identity of ATS drug from these thousands of features, and used for further knowledge inference. To deal with this issue, feature selection is a prominent method for dimension reduction and good in identifying the most distinguishing features [9]. Motivated by these factors, the feature selection technique has been employed in this work to identify the most representative and significant features to indicate the presence of ATS drug. This will indirectly reduce the

complexity for further processing in ATS drug identification.

The overall goal of this paper is to investigate the effectiveness of six feature selection methods which composed of four filter feature selection techniques (Information Gain (IG), Gain Ratio (GR), Symmetrical Uncertainty (SU), and ReliefF) and two embedded feature selection techniques (Support Vector Machine based Recursive Elimination Method (SVM-RFE) and Variable Importance based Random Forest (VIRF)) on the ATS drug dataset. The resulted feature subset of each feature selection techniques will be compared among each other's as well as the original dataset using five common classifiers namely Random Forest (RF), Naïve Bayes (NB), IBK, SMO and J48 decision trees.

A. Related Works

Feature selection has been widely applied in various forensic domains such as biomedical, signature or handwriting comparison, chemical profiling, face recognition, glass identification, footwear pattern classification, etc.

Research that investigated the probability of the existence of online digital fingerprint based on thinking style signature that used to determine the online users is done in [10]. 43 respondents of server-side web data were used and tested in this study. Five thinking styles which include Judiciary, Oligarchic, Legislative, Hierarchical, and External thinking style were then extracted and clustered into five dichotomies. Various supervised machine learning techniques were explored in this study to distinguish the individuals from each dichotomy. Based on the experimental results, it was noted that Meta classifier of a Logistic model tree (J48) with bagging technique produces better accuracy. In addition, the observed signature was further used in the digital forensic process.

Alshaikhdeeb et al. [11] conducted an extensive review of Biomedical Named Entity Recognition (BNER). The authors have studied three common features that used in BNER: morphological features, dictionary-based features, lexical features and distance-based features. Then, these features can be further classified into Numeric, Nominal and Boolean features. Throughout the discussion, the authors demonstrated that Morphological Boolean features outperformed the other features in BNER process. Thereby, this paper suggests that this feature should be used in facilitating the process of identifying extracting biomedical entities.

In [12], a new heuristic nucleotide physicalechemical property selection (HPCS) algorithm have been proposed to select the most representative nucleotide physicalechemical properties for N6-methyladenosine (m6A) site prediction. The experimental results proved that it is a promising approach to achieve a higher success rate compared to the existing state-of-the-art sequence-based m6A site predictors.

The use of machine learning methods had been explored in [13] for developing an automated system to detect colorectal cancer using near-infrared Raman spectroscopy together with feature selection technique. The goal of this paper is to identify the characteristic of Raman bands associated with biochemical components with swarm intelligence ACO-SVM technique, which will be used to classify colorectal cancer from normal tissue. Their experiment indicates that feature selection was necessary to select the important features of tissue Raman spectra and achieve a good result in classification performance.

In [14], a sequential forward search and a depency based evaluation criterion are used to improve the classification performance in forensic handwriting identification. It shows a promising result of the prediction performance, despite a major reduction of the features, from 58% (original dataset) to 80% (after feature selection). This result is comparable to the others work in literature where graphometric features were taken into account as well.

In addition, feature selection also used in classify Alzheimer's disease with Raman spectra [15]. The result gained from this research show that the proposed method can effectively select the most discriminative peak from the preprocessed spectrum. The selected features will then be used for the diagnosis of Alzheimer's disease.

Another application of feature selection is applied in recognition of model-free gait. Paper [16], explore a feature selection method based on Random Forest feature rank algorithm to extract the most informational features from the gait sequence. The main goal of this work is to describe the human walking by using probabilistic based gait modelling. The result showed that the proposed method could reduce the complexity of the learning task and achieve a better classification performance.

In short, feature selection has been widely applied in various areas and has offered a new opportunity to solve different issues. In the case of our work domain, ATS drugs identification requires the knowledge from the experts; the equipment is that used for laboratory testing and the budget to acquire materials. It is a lengthy and costly procedure in real life. In this sense, a more user-friendly alternative is necessary. Hence, we investigate various feature ranking algorithms to confront this situation. Specifically, we focused on four filter methods and two embedded method that follows the feature ranking procedure. Both methods are considered less computationally expensive as compared to wrapper method. Five different state-of-the-art classifiers are adopted to ensure robustness and reliability of the obtained features subset.

B. Feature Selection

Several feature selection algorithms are available in the literature. Generally, feature selection can be further grouped into three types: filter, wrapper, embedded [14]. Each of these has their advantages and disadvantages. Filter method is selecting features based on information theory and statistical evaluation criteria without the aid of the classification algorithm. This method will evaluate each feature individually and produce a feature subset with the minimal amount of irrelevant and redundant features. This method is less computationally expensive and less time-consuming. However, the interaction with the classifier algorithm is not considered.

In contrast, wrapper method will produce a subset of features with the aid of the classification algorithm. By doing so, predictive performance will act as a criterion to assess the relevant features subset. This method is believed to have better predictive accuracy, but it also suffers from high computationally expensive cost [9]. The embedded method is somewhat similar to wrapper methods, but it consumes less computational cost.

The embedded method works by incorporating the selection of features in the training process. In short, filter method and embedded method are relatively computationally

effective compared to wrapper method. Both methods are commonly applied as a feature ranking procedure. Meanwhile, feature selection can be further categorised into two types, which are feature ranking methods and feature subset selection methods. Feature ranking works by sorting all the features based on their importance while subset selection chooses a subset of the most important features for classification. In [15], [16], it is shown that the feature ranking methods are more efficient than the feature subset selection methods.

Therefore, in this work, we are focusing on the investigation of the utility of four filter methods and two embedded method based feature ranking techniques in the work of ATS drugs identification.

The rest of the paper is organised as follows: in the next section, we presented the material and method used in this paper. In Section 3, the experimental findings are discussed and evaluated. A conclusion is presented in Section 4.

II. THE MATERIALS AND METHODS

A. Data Collection

The dataset that used in this analysis is from 3D Exact Legendre descriptor (3D-ELD) [17]. This data source contained 7212 sample records, which contains 3602 of non-ATS drug molecular structure and 3610 of ATS drug molecular structure. Each instance is described by a fixed number of features, along with a class label. The features are recorded in voxel which aims to maintain the realistic properties of the 3D ATS molecular structure. This data source is used to train and test the proposed feature selection algorithm in this work. The characteristics of these datasets are presented in Table 1.

Table 1
Description of Dataset Used

MolID	Features								Class
	I1	I2	I3	.....	.....	I1183	I1184	I1185	
1	-0.02453	0.0175	-0.01403	.....	.....	-2.2780	-0.1131	8.243774	nATS
2	-0.01807	0.0223	-0.0166	.....	.....	-6.8526	-0.6992	21.38265	nATS
3	-0.01276	0.0101	-0.0153	.....	.....	-7.2899	-0.5213	23.4808	nATS
:::	::::	::::	::::	:::	:::	::::	::::	::::	::::
:::	::::	::::	::::	:::	:::	::::	::::	::::	::::
:::	::::	::::	::::	:::	:::	::::	::::	::::	::::
7210	-0.01884	0.0144	-0.0131	.....	.....	8.936222	-0.2686	-7.5110	ATS
7211	-0.01652	0.0276	-0.0168	.....	.....	0.9267	0.3648	-16.986	ATS
7212	-0.0135	0.0120	-0.0103	.....	.....	-19.705	0.3648	108.26	ATS

B. Performance Measurements

Several metrics are used to evaluate the classifier performance. All of the performance metrics are evaluated based on four possible measurements:

i. True positive (TP): ATS drug sample correctly classified
ii. False positive (FP): ATS drug sample incorrectly classified
iii. True negative (TN): Non-ATS drug sample correctly identified
iv. False negative (FN): Non -ATS drug sample incorrectly classified

The number of features selected for each feature ranking techniques will be considered as criteria to measure their performance. The main concern of this work is how well the predictive model will perform based on the selected features. Therefore, in this work we perform comparison analysis based on accuracy (ACC), in percentage, as defined as follows:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

C. Proposed Method

The experimental design composed of three main steps, which starts by ranks all the features based on different criteria by IG, GR, SU, ReliefF, SVM-RFE, and VI-RF. Next, the top-ranked features return by each of the feature ranking techniques is assessed by using the different dimension of feature subset. All these feature subsets are examine using RF, NB, IBK, SMO and J48 decision trees classifiers. Figure 2 depicts an overview of the experimental design.

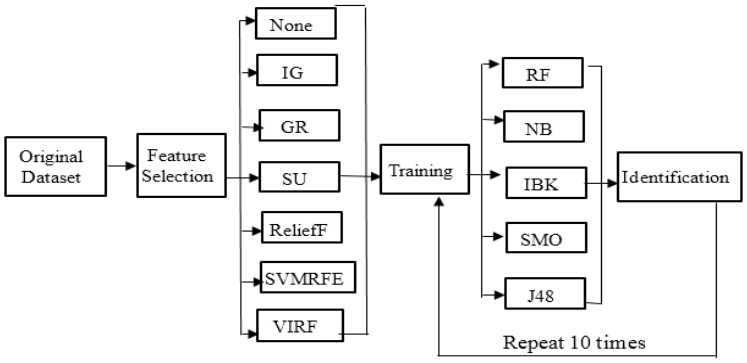


Figure 2: Overview of the experimental design

The feature ranking techniques that employ in this work are further elaborate as follows:

1) Information Gain

Information gain is a symmetrical measurement technique which is based on entropy concept from information theory [18]. Given the entropy is a criterion to measure the impurity of a training set. Given the value of another attribute X, IG measures the additional information provided by class attribute Y [19]. The decrease in the amount by which the weighted average impurity of Y, compared to the original dataset will be calculated. The formula of IG is given by:

$$IG = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad (2)$$

A significant limitation of IG is that it tends to bias to the feature with majority number of positive values, rather than the feature with fewer values even though the features with fewer values are more informative.

2) Gain Ratio (GR)

The Gain Ratio is the non-symmetrical measurement which will penalise the multi-valued features to compensate for the bias of the IG towards features with more values[19]. GR works by dividing IG with the entropy of X to normalise the value of IG. The result from the normalisation process will be in the range of [0, 1]. When GR=1 means the attribute X is informative to predict class attribute Y, while GR=0 means the attribute X is non-informative to Y. In opposition to IG, GR tends to bias to favour the attribute with fewer values. In addition, GR The formula of GR is given by:

$$GR = \frac{IG}{H(X)} \quad (3)$$

3) Symmetrical Uncertainty (SU)

Symmetrical Uncertainty is a symmetrical measurement technique that tends to evaluate the association between the feature and the targeted class. Furthermore, it is used to compensate for the bias in IG by dividing the IG with the sum of the entropies of X and Y [19]. SU treated a pair of features symmetrically and used to measure the correlation between the features. Similar to GR, the value of the result is normalising within the range of [0, 1]. When SU=1 means the attribute X is informative to predict class attribute Y, while SU=0 means the attribute X is non-informative to Y. Similar to GR, SU tends to bias to favour the attribute with fewer values. The formula of SU is given by:

$$SU = 2\frac{IG}{H(Y)+H(X)} \quad (4)$$

4) ReliefF

ReliefF is an approach which was extended by Kononenko in the year 1994 to cater to the limitation of noisy and incomplete data and two-class classification problems[20]. The original Relief was introduced by Kira and Rendell in 1992. The basic idea of ReliefF is to select a sample instance at random and search two nearest neighbours: one from the same class (nearest hit) and one from the different class (nearest miss). Then it will update the feature weighting vector according to the two nearest neighbours. The quality estimation of the selected features is based on the weight computation of the probability between the selected instances and their two nearest neighbours (nearest hit and nearest miss).

The rationale of this idea is a feature is considered good when the probability of two nearest neighbours from the same class having the same value. Meanwhile, the features with their nearest neighbours from two different classes should have different values. Therefore, the larger the difference between this probability, the better the features to be. The final output of ReliefF is by return a ranked list whose weight exceeds the user-defined threshold. The list will be sorted in descending order, and the top-ranked features are selected as the optimal features for the candidate solution. In short, this technique provides a good capability in dealing with incomplete and noisy data [21].

5) Support Vector Machine-based Recursive Feature Elimination (SVMRFE)

SVM-RFE is a feature ranking algorithm that proposed by[22]. It is an algorithm that eliminates feature recursively based on the weighting provided by SVMs. The algorithm will begin with all the feature and remove the least important features recursively for the classification in a backward elimination manner. For a better understanding of SVM-RFE, its algorithm is presented in the algorithm 4. The basic procedure of SVM-RFE is based on the following intuition:

i. Train a classifier.
ii. Compute a ranking criterion for all the features.
iii. Remove the feature with the smallest ranking criterion.

6) Variable Importance based Random Forest (VIRF)

VI-RF is an embedded feature selection technique, which selects the relevant features based on the variable importance yielded by random forest. In the context of random forest which made of an ensemble of decision trees, Breiman in the year of 2001 proposed a permutation test procedure in order to compute variable importance based on the classification error [23]. The difference in classification accuracy caused by the permutation is taking into account to define the variable importance. The prediction accuracy will not be affected by permuting the values of the variable that consists of purely random noise. Formally, the variable importance using random forest is computed based on two main principle, which is: randomisation and out-of-bag error (OOB) estimates. Let $\bar{B}^{(t)}$ be the out-of-bag (OOB) sample for a tree, with t ε {1, …, ntree}. The importance measure for variable X_j in tree t is precisely defined as follows:

$$VI^{(t)}(\mathrm{x_j}) = \frac{\sum_{i\varepsilon\bar{B}^{(t)}} I(y_i = \hat{y}_i^{(t)})}{|\bar{B}^{(t)}|} - \frac{\sum_{i\varepsilon\bar{B}^{(t)}} I(y_i = \hat{y}_{i,\pi_j}^{(t)})}{|\bar{B}^{(t)}|} \quad (5)$$

where $\hat{y}_i^{(t)} = f^{(t)}(x_i)$ is the predicted class for observation i before, and $\hat{y}_{i,\pi_j}^{(t)} = f^{(t)}(x_{i,\pi_j})$ the predicted class for observation i after permuting its value of variable X_j, i.e. with $x_{i,\pi_j} = (x_{i,1}, \dots, x_{i,j-1}, x_{\pi_j(i),j}, x_{i,j+1}, \dots, x_{i,p})$. (Note that $VI^{(t)}(\mathrm{x_j}) = 0$ by definition, if variable X_j is not in tree *t*.) The raw variable importance score for each variable is then computed as the mean importance over all trees: $\frac{\sum_{t=1}^{ntree} VI^{(t)}(\mathrm{x_j})}{ntree}$.

III. RESULT AND DISCUSSION

In this work, six feature ranking techniques have been applied. Each of the technique will return a feature subset which sorted in descending order based on their ranking criteria. Each feature subset will employ Recursive Feature Elimination (RFE) method to eliminate a chunk of features at a time [22]. At each elimination iteration, half of the features will be removed. The features will be divided into seven partitions: 592, 296, 148, 74, 37, 19, 9 feature sizes. The quality of each feature subset was then examined by training various classifiers including RF, NB, IBK, SMO and J48 decision tree. This is motivated by the "No Free Lunch Theorem", which means no one algorithm can guarantee works best for every problem [24]. All the classifiers were executed with the default parameter setting in WEKA [18]. As this research does not focus on building the best prediction system for ATS drug identification, hence testing all the possible parameter in the configuration setting for each classifier is not within the scope of this research. The predictive performance is evaluated by the average results of 10-fold cross-validation, and the classification accuracy and AUC are taking into consideration. This section will present a summary of the comparison of classification performance by using different feature subset sizes that selected by the five chosen feature selection techniques as well as the original dataset.

A. Comparison between Different Feature Subset Sizes

A summarisation of the classification performance of five classifiers affects by different feature subset sizes selected by six feature ranking techniques is provided in Table 2. As it can be observed in most cases, by comparing the average of classification performance from each chosen learner, using smaller feature subset size generally further improved the classification performance than using all the original features. The only exception is feature subset sizes with 19, and 9 features returned the worst classification performance. This is mainly because some individually relevant features are not included in the top 9 and 19 features, and they are not sufficient features to classify ATS drugs. The best classification accuracy and related feature subset sizes are marked and bolded. By further inspection of the average of classification performance from each chosen learner, the partitions among the seven different feature subset sizes, the results suggesting that the feature subset size of 296 is optimal to build a classification model. This suggests that most of the discriminating features fall in the range of 296 feature subset. However, exceed that optimal range, the feature subset may contain higher noise rate, which results in significantly lower classification performance.

Table 2
Average Overall Classification Accuracy (%) on ATS Drug Dataset based on Different Number of Features

	RF	NB	IBK	SMO	J48	Average
1185	82.169	68.968	74.265	81.683	74.986	76.414
592	82.21	70.958	74.873	81.542	75.504	77.017
296	82.03	71.49	74.982	81.092	76.225	**77.164**
148	81.739	71.217	74.67	80.662	76.717	77.001
74	81.388	71.421	74.471	80.093	76.959	76.866
37	80.986	71.933	73.937	79.666	76.877	76.68
19	80.107	72.248	73.165	78.984	76.617	76.224
9	79.430	73.375	72.691	78.441	76.770	76.141

B. Comparison between Different Feature Rankers

Use As a further step, investigation of classification performance for each feature ranking using the optimal feature subset has been done. Five classifiers were used: RF, NB, IBK, SMO, and J48. Each of these classifiers returns a deterministic result, in order to simplify the comparison, averaging the results are taken into account. The average of overall classification accuracy is presented in Table 3. As shown in Table 2, comparison of the six different feature rankers using the 296 optimal feature subset, ReliefF and VI-RF are the most stable and effective methods. As observed from the table below, both of their average of the classification accuracy is very close, which is 77.401% and 77.608%, with the mild difference of 0.2%. The two most ineffective feature ranking techniques are SVMRFE and IG. They are the two techniques with the lowest classification accuracy with the average of 74.788% and 76.642% as demonstrated in the table below.

Among these six-feature ranking methods, it can further categorise into univariate approaches (IG, SU, GR) and multivariate approaches (ReliefF, SVM-RFE, and VI-RF). It is clearly shown that univariate approaches exhibit in almost similar trend, with a slight superiority of GR in ATS drugs dataset. Among the multivariate approaches, ReliefF and VI-RF outperform the univariate approaches in most cases. However, an exception of this case, though SVM-RFE is well known in the literature Guyon et al., as an effective feature selection technique, it exhibits the worst classification performance among the others six feature selection techniques. This problem is partly answered by Hardin et al. [25], who claimed that linear SVM may assign higher weights to weakly relevant features, while assigned zero-weights to highly relevant features. Furthermore, they also pointed out that SVM-RFE will not necessary works together with the classification models.

Table 3
Average Overall Classification Accuracy on ATS Drug Dataset Based on Different Feature Rankers

	IG (%)	GR (%)	SU (%)	ReliefF (%)	SVM-RFE (%)	VI-RF (%)
RF	80.772	81.107	80.948	81.986	80.352	81.596
NB	71.676	72.221	72.009	72.312	68.964	73.653
IBK	74.02	73.97	74.156	75.048	72.417	75.066
SMO	79.833	80.236	80.257	80.647	78.823	80.616
J48	76.907	77.327	77.407	77.013	73.382	77.11
Average	76.642	76.972	76.955	77.401	74.788	77.608

IV. CONCLUSION

This paper presents a methodology that sought to compare the effectiveness of four filters feature selection techniques (IG, GR, SU, ReliefF) and two embedded feature selection techniques (SVM-RFE and VI-RF) in the context of ATS drug identification.

The results of classification accuracy demonstrate the effectiveness of feature selection methods. In general, ReliefF and VI-RF are outperforming the other techniques as well as the original dataset. The result shows that multivariate approaches clearly outperformed the univariate approaches, except for SVM-RFE. In addition, the goal of this work is attempted to empirically estimate the number of needed features to achieve the desired classification performance. In this case, only a setting using 296 selected features has been selected.

ACKNOWLEDGEMENT

This work is funded by Universiti Teknikal Malaysia Melaka (UTeM) through the PJP High Impact Research Grant [PJP/2016/FTMK/HI3/S01473].

REFERENCES

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

[2] U. N. O. on D. and C. UNODC, *WORLD DRUG REPORT 2016*. 2016.

[3] D. Rouen and K. Dolan, "A Review of Drug Detection Testing and an Examination of Urine , Hair , Saliva and Sweat," 2001.

[4] R. Shipman, T. Conti, T. Tighe, and E. Buel, "Forensic Drug Identification by Gas Chromatography – Infrared Spectroscopy," 2013.

[5] W. Shin, X. Zhu, M. G. Bures, and D. Kihara, "Three-Dimensional Compound Comparison Methods and Their Application in Drug Discovery," *Mol. 20, no. 7*, pp. 12841–12862, 2015.

[6] M. Awale, X. Jin, and J. Reymond, "Stereoselective virtual screening of the ZINC database using atom pair 3D-fingerprints," *J. Cheminform.*, pp. 1–15, 2015.

[7] M. Aldeghi, S. Malhotra, L. David, A. Wing, and E. Chan, "Two- and Three-dimensional Rings in Drugs," *Chem. Biol. Drug Des.*, vol. 4, no. 83, pp. 450–461, 2014.

[8] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution," *Yu, Lei, Huan Liu. "Feature Sel. high-dimensional data A fast Correl. filter Solut. Proc. 20th Int. Conf. Mach. Learn.*, pp. 856–863, 2003.

[9] I. Guyon, "An Introduction to Variable and Feature Selection Introduction," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[10] I. R. Adeyemi, S. A. Razak, M. Salleh, and H. S. Venter, "Leveraging Human Thinking Style for User Attribution in Digital Forensic Process," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 1, pp. 198–206, 2017.

[11] B. Alshaikhdeeb and K. Ahmad, "Biomedical Named Entity Recognition : A Review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 889–895, 2016.

[12] M. Zhang, J. Sun, Z. Liu, M. Ren, and H. Shen, "Improving N 6 - methyladenosine site prediction with heuristic selection of nucleotide physical e chemical properties," *Anal. Biochem.*, vol. 508, pp. 104–113, 2016.

[13] S. Li *et al.*, "Identification and characterization of colorectal cancer using Raman spectroscopy and feature selection techniques.," *Opt. Express*, vol. 22, no. 21, pp. 25895–908, 2014.

[14] A. M. M. M. Amaral, C. O. D. A. Freitas, and F. Bortolozzi, "Feature selection for forensic handwriting identification," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 922–926, 2013.

[15] A. Park, S. J. Baek, A. Shen, and J. Hu, "Detection of Alzheimer's disease by Raman spectra of rat's platelet with a simple feature selection," *Chemom. Intell. Lab. Syst.*, vol. 121, pp. 52–56, 2013.

[16] Y. Dupuis, X. Savatier, and P. Vasseur, "Feature subset selection applied to model-free gait recognition," *Image Vis. Comput.*, vol. 31, no. 8, pp. 580–591, 2013.

[17] ICGEB CRP Research Grant Programme Projects, "CRP - ICGEB Research Grants Completed in 2016 (A New 3D Descriptor of Synthetic Drug Molecular Structure for Drug Analysis - CRP/13/010)," 2012.

[18] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*, vol. 54, no. 2. 2011.

[19] M. Hall and L. a Smith, "Feature Selection for Machine Learning : Comparing a Correlation-based Filter Approach to the Wrapper CFS : Correlation-based Feature," *Int. FLAIRS Conf.*, p. 5, 1999.

[20] I. Kononenko, "ReliefF for estimation and discretization of attributes in classification, regression, and ILP problems," *Artif. Intell. Methodol. Syst. Appl.*, pp. 1–15, 1996.

[21] M. R.- Sikonja, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Mach. Learn. J.*, pp. 23–69, 2003.

[22] I. Guyon, J. Weston, S. Barnhill, T. Labs, and R. Bank, "Gene Selection for Cancer Classification using Support Vector Machines," *Mach. Learn.*, vol. 46, no. 1–3, pp. 389–422, 2002.

[23] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[24] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," *New York: John Wiley, Section*. p. 680, 2001.

[25] D. Hardin, I. Tsamardinos, and C. Aliferis, "A theoretical characterization of linear SVM-based feature selection," *The Twenty-First International Conference on Machine Learning (ICML 2004)*. pp. 48–55, 2004.