Ensemble-support vector machine-random undersampling: Simulation study of multiclass classification for handling high dimensional and imbalanced data

Nur Silviyah Rahmi
Departmen Statistics, Universitas Brawijaya, Malang, Indonesia
E-mail: silviyahrahmi@ub.ac.id

Abstract. Microarray technology measures on a large and parallel scale to express tens of thousands of genes. It has widely applied to predict gene function, new subtypes of specific tumors and cancer classification. However, microarray data are known has feature characteristics such as high dimension, small sample, high noise, and imbalanced class distribution. Support Vector Machine (SVM) has been widely used and shows the success in major applications to improve classification performance. To overcome the high dimension, we applied the Ensemble-SVM method. This method classifies features use clustering hierarchy and each group will be classified. While the condition of imbalance data becomes a problem in classification because the classifier will tend to predict the majority class compared to the minority class. Therefore, a Random Undersampling or EnSVM-RUS method is used to balance the size of the majority class into the minority class. We uses threefold cross-validation with a feature selection method that is Fast Correlation Based Filter (FCBF). The multiclass method used is SVM One Against One (OAO). While the evaluation criteria of performance classification based on the value of accuracy, F-score and G-mean and running time. We performs a simulation study with various scenario level of ratio imbalance (IR) that is ratio 1, 5, and 8 to know the performance of the proposed method. While the application on real data using Microarray DNA data with IR 4.22, 15.00 and 23.17 The results showed that EnSVM-RUS-OAO method with 2 clusters had higher performance than the EnSVM-OAO and EnSVM-OAO methods. Increasing the ratio imbalance doesn’t affect the advantage of the EnSVM-RUS-OAO method when compared to EnSVM-OAO and EnSVM-OAO methods. While on the use of the kernel, RBF kernel and polynomials produce higher performance and shorter computation time than linear kernels.

1. Introduction
DNA microarray is a data containing gene expression with a small sample size, but has a very large number of features. In addition, the problem of imbalance classes is a common problem in microarray data. Therefore, it is necessary to classify the problems of high dimensional and also imbalance problems. Microarray technology remarks that Support Vector Machine (SVM) method has been widely applied to classification. Seeja and Shweta classified gene expression in DNA microarrays using SVM and found out that SVM is better than Neural Network [1]. Support Vector Machine (SVM) is one of the most well-studied data mining techniques and has demonstrated success in many of its major applications in improving classification performance [2]. Simply put, SVM concept search for the best hyperplane that serves as a separator of two classes in the input space [3].
SVM was originally designed for binary classification problems. Therefore, to apply SVM to multiclass issues, this should be reconfigured using a coding strategy that is the first strategy for addressing data imbalance issues [4]. Trapsilaswiti states that the classification of multiclass using One Against One approach (OAO) is better than One Against All (OAA) approach [5]. Statnikov do systematically that strategy by some experiment and show that One Against One approach (OAO) often produces higher accuracy [6]. Thus, the authors use SVM OAO to overcome multiclass cases.

The second strategy is the ensemble learning method. The ensemble classifier method can be applied when we uses more than one feature to creates a classification model. Each feature used contains different information, so it needs to be trained with a different classifier. Ensemble learning has been used to improve the accuracy of gene selection features and cancer classification [7]. We used an ensemble feature subset method by clustering feature of the dataset. Then each feature set which has been formed is modeled using SVM. We also use a random undersampling (RUS) algorithm to reduce and balancing the number of features. Random Undersampling (RUS) is a method to calculate the difference between the majority and minority classes and then recalculate the difference of calculation results, as long as the majority of data loops are randomly deleted, so that the majority class is equal to the minority [8]. This process is expected to move the chance of classification errors from minority classes into the majority class so as to improve accuracy in minority classes.

This work aim to provide an empirical study of the performance of SVM-OAO, EnSVM-OAO and EnSVM-RUS-OAO algorithms. We uses simulation study with various scenario level of ratio imbalance (IR) that is ratio 1, 5, and 8 to know the performance of the proposed method. The expected result of this research is to get the classification method which suitable to multiclass condition and imbalance then improve the performance.

2. Background Study
Several feature selection techniques for dimensionality reduction in microarray data are available in the literature. This section deals with the feature selection techniques used for microarray data.

2.1 Feature Selection Techniques Based On Fast Correlation Based Filter and Minimum Description Length Principle for Microarray Data
Data usually comes in a mixed format: nominal, discrete, and/or continuous. We use discretization to quantize continuous attributes. Discretization makes learning more accurate and faster. A minimum description length principle (MDLP) is used to choose useful cut-points. The success of discretization can significantly extend the borders of many learning algorithms [9]. After discretization process, we use FCBF as feature selection algorithm. FCBF is one multivariate feature selection algorithm and measures feature class and correlation between features [10]. FCBF is designed to address high-dimensional data cases that are expected to reduce or select unnecessary features and redundant features.

The entropy of variable $X$ is defined in Equation (1)

$$H(X) = - \sum_{i=1}^{n} P(x_i) \log_2(P(x_i))$$

The entropy of variable $X$ if known variable $Y$ is defined in Equation (2)

$$H(X \mid Y) = - \sum_{i=1}^{n} P(y_i) \sum_{j=1}^{m} P(x_i \mid y_i) \log_2(P(x_i \mid y_i))$$

$P(x_i)$ is prior probabilities for all $X$ and $P(x_i \mid y_i)$ is the posterior probabilities of $X$ if $Y$ is known. From the entropy can be obtained Information Gain as follows.

$$IG(X \mid Y) = H(X) - H(X \mid Y)$$

(3)

To measure the correlation between features, then use symmetrical uncertainty. Symmetrical uncertainty values range from 0 to 1. Symmetrical uncertainty is formulated as follows.

$$SU(X,Y) = 2 \frac{IG(X \mid Y)}{H(X)+H(Y)}$$

(4)
So that will be obtained features relevant to the class with a predefined threshold value. From the relevant features, redundant features will be removed from the FCBF feature selection results.

2.2 Base Support Vector Machine (SVM) Training

SVM was first introduced by Vapnik in 1992 with the aim of finding the best hyperplane that separates two classes on the input space [11]. Linear SVM Classification; SVM works with the principle of a linear classifier, then developed to work on non-linear cases using kernel concepts in high-dimensional workspaces.

![Figure 1. Hyperplane Concept on Linear SVM](image)

Figure 1 shows that some patterns are members of two classes of +1 and -1. Patterns incorporated in class -1 is symbolized by red circles, while the pattern in the +1 class is symbolized by a green sign. SVM is a classification method that can be separated linearly with functions as follows.

\[
\hat{f}(x_{new}) = \text{sign} \left( x_{new}^T \hat{w} + \hat{b} \right),
\]

where \( x, w \in \mathbb{R}^d \) and \( b \in \mathbb{R} \). Hyperplane the best separator between the two classes can be found by measuring the hyperplane's margins and searching for the maximal point. Linear Separable Classification; The linear classification of SVM is divided into 2 types that are separable and non-separable. In SVM linear separable can be defined the input \( x = (x_1, x_2, \ldots)^T \) is expressed as positive class if \( f(x) \geq 0 \) and others belong to the negative class. Hard margin classification is used in equation (5). In general, linearly separated data will satisfy the following equation.

\[
\max_a L_{opt}(a) = \max \left( \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \right)
\]

Non-Separable Linear Classification; The second SVM linear classification is a non-separable type. The following is a modified limit for non-separable cases.

\[
y_i(x_i^T w + b) \geq 1 - \xi_i
\]

with \( \xi_i \geq 0, C > 0 \) where \( C \) is the parameter that determines the cost due to misclassification of the training data during the learning process and the value is determined by the researcher. The lagrange function for the primal problem is.

\[
L_{primal}(w, b, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left( y_i (x_i^T w + b) - 1 + \xi_i \right) - \sum_{i=1}^{n} \mu_i \xi_i
\]

where \( \alpha_i \geq 0 \) and \( \mu_i \geq 0 \) are Lagrange Multiplier. The first-order condition is

\[
\frac{\partial L_{primal}(w, b, \alpha)}{\partial w} = 0; w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i
\]
\[ \frac{\partial L_{primal}(\mathbf{w}, b, \alpha)}{\partial \mathbf{b}} = 0 \Leftrightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \quad (10) \]
\[ \frac{\partial L_{primal}(\mathbf{w}, b, \alpha)}{\partial \alpha_i} = 0 \Leftrightarrow C - \alpha_i - \mu_i = 0 \Rightarrow \alpha_i = C - \mu_i \quad (11) \]

by substituting the value \( \hat{\mathbf{w}} = \sum_{i=1}^{n} \hat{\alpha}_i y_i \mathbf{x}_i \) into the primal problem becomes the dual problem equation as follows then obtained equation (12).

\[ \max L_0(\alpha) = \max \left( \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \quad (12) \]

with the following constrain.

\[ 0 \leq \alpha_i \leq C ; \sum_{i=1}^{n} \alpha_i y_i = 0 \quad (13) \]

Sample \( \mathbf{x}_i \) for \( \alpha > 0 \) (support vector) is the point above the margin or in the margin when soft margin is used. Support vector is often spread and its spread level is at the upper limit for misclassification rate [12]. Classification of Non-Linear SVM: To solve nonlinear problems, SVM is modified by entering the Kernel function. In this new vector space, the hyperplane that separates two classes can be constructed.

![Hyperplane on Non-Linear SVM](image)

**Figure 2.** Hyperplane on Non-Linear SVM

Figure 2 on the left side shows the data in the green class and the red class is in the two-dimensional input space. On the right side shows that the function \( \Phi \) maps each data in the input space to a new, higher-dimensional vector space separated by the hyperplane [13]. Furthermore, SVM will find vector support points in the form of a dot product from data that has been transformed to new space with higher dimension, that is \( \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}') \). Because it is generally a transformation \( \Phi \) this is unknown, then the dot product calculation according to Mercer theory is replaced with kernel function \( \langle \mathbf{x}_i, \mathbf{x}_j \rangle \) which defines implicitly transformations \( \Phi \). This is called the kernel trick [14], which is formulated as follows.

\[ k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (14) \]

Commonly used kernel functions are.

1. **Linear Kernel**
   \[ k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (15) \]
2. **Polynomial Kernel**
   \[ k(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x} + r)^p, \quad \gamma > 0 \quad (16) \]
3. **Radial Basis Function (RBF) Kernel**
   \[ k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \gamma > 0 \quad (18) \]
4. **Kernel Sigmoid**
   \[ k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x} + r), \quad \gamma > 0 \quad (19) \]

where \( \gamma, r \) and \( d \) is kernel parameter.

Determining the kernel and error penalty parameters for support vector machines (SVM) is very problem-dependent in practice. A popular method to deciding the kernel parameters is the grid search method. In this paper we propose using the inter-cluster distances in the feature spaces to choose the
kernel parameters. Calculating such distance costs much less computation time than training the corresponding SVM classifiers; thus the proper kernel parameters can be chosen much faster [15].

2.3 Ensemble SVM Feature Clustering With Random Undersampling

Ensemble learning is an intensively studies technique in machine learning and pattern recognition. Recent work in computational biology has seen an increasing use of ensemble learning methods due to their unique advantages in dealing with small sample size, high-dimensionality, and complexity data structures [16]. We utilized one-against-all coding strategy to transform multiclass to multiple binary classes, each of them carrying out feature subspace, which is an evolving version of random subspace that generates multiple diverse training subsets. Next, we introduced one correction technologies, namely random undersampling into each training subset to alleviate the damage of class imbalance [17]. This is the frame diagram of ensemble learning using feature subspace clustering and random undersampling process.

![Figure 3. Hyperplane on Non Linear SVM](image)

Ensemble Method Subspace feature has performance related to two factors that are accuracy and diversity of the base cluster. The cluster used in this ensemble is a hierarchical clustering complete linkage that uses euclidian distance as a measure of distance to remove redundancy and noise ratios [18]. This ensemble method clustering of variables. Suppose there are variables \( \mathbf{x} = (x_1, x_2, \ldots, x_n) \) and \( \mathbf{z} = (z_1, z_2, \ldots, z_n) \) then the distance value between variables is as follow.

\[
d(x, z) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2 + \cdots + (x_n - z_n)^2}. \tag{20}
\]

Clustering is challenging in high-dimensional space, where the distance between two objects is often difficult to measure. This is because such a distance is dominated by the different sets of dimensions in which the objects are residing. Thus, instead of clustering objects in their full high-dimensional spaces, it can be more meaningful to find clusters in certain subspaces. Recently, researchers have developed subspace-based pattern growth methods that cluster objects based on their common frequent patterns. They have shown that such methods are effective for clustering microarray-based gene expression data [19].

RUS method is a method to calculate the difference between the majority class and minority then do the looping of the calculation result. As long as the majority of data loops are randomly deleted, the majority of the class is equal to the minority [8]. The first step in the RUS method is the selection of the dataset and then calculated the difference between the majority and minority classes, if there is still a
difference between the number of classes then the majority class datasets will be randomly deleted until the majority class is equal to the minority class. RUS can be more effective and faster in the process of predict class imbalance training in a microarray case.

2.4 Multiclass SVM Using One Against One (OAO) Strategy
The OAO method will find a separation function of \( k(k - 1)/2 \), where each separating function is trained with samples of two classes. For example, there is a classification problem with 3 classes means that 3 for \( p \) separating functions can be defined, \( p^{12}, p^{13}, \) and \( p^{23} \). When \( p^{12} \) is trained, all samples in class 1 are labeled positive (+1) and all samples in class 2 are labeled negative (-1). This is also on \( p^{13} \) and \( p^{23} \). As an illustration, given the training data of classes \( i \) and \( j \), for this problem will be solved by the following equation.

\[
\min_{w, b} \frac{1}{2} w^T w + \frac{1}{2} \sum_{i,j} C \xi_{ij} + \sum_{i} \frac{1}{2} \sum_{j} \xi_{ij} \]

where \( r \) denotes the data index of each class. Based on the max voting strategy for the \( p^{ij} \) separator. If the positive sign of a data \( x \) is in class \( i \), then the voting for class \( i \) plus one. Conversely, if the positive sign of a data \( x \) is in class \( j \), then the vote for class \( j \) plus one. Then, put in which class is based on the highest voting value [20].

Stratified K-Fold Cross Validation; Stratification is the process of rearranging data to make sure each fold is a good representation of the entire data. This process is repeated as much as k-subsets and the result of classification accuracy is the average result of each training and testing data [21]. In k-fold cross-validation, the data is randomly split into \( k \) mutually exclusive subsets of approximately equal size. A learning algorithm is trained and tested \( k \) times; each time it is tested on one of the \( k \) folds and trained using the remaining \( k - 1 \) folds. The cross-validation estimate of accuracy is the overall number of correct classifications, divided by the number of examples in the data. [22]

3. Experimental Setting
A Simulation study was conducted to determine the performance of EnSVM-RUS-OAO, EnSVM-OAO, and SVM-OAO methods in multiclass imbalance data classification with various ratios. Then compare the performance of the three models based on the criteria of Accuracy, F-score, and G-mean. In this research, EnSVM-RUS-OAO method will be conducted simulation study with stage of analysis is as follows.

a. Set leukemia dataset (http://www.gems-system.org/) as the reference to obtain characteristic vector and covariance variance matrices. Leukemia data is defined as a reference because it has a low IR of 4.22 and the number of classes is 3, according to the limits of the research problem.

b. Get data characteristics that are mean vectors and covariance variance matrices based on leukemia datasets. The leukemia dataset has 3 classes and 5327 features.

c. Generate normal multivariate data according to the scenario. The researcher designs 3 scenarios based on the ratio of imbalance and the number of features. The number of features as many as 1000 and 5000 with the number of samples is 100.

- Scenario 1: The data is divided into 3 classes that is 33:33:34
- Scenario 2: The data is divided into 3 classes that is 10:40:50
- Scenario 3: The data is divided into 3 classes that is 10:10:80

d. Divide data into training data and testing data with threefold cross-validation stratification.

e. Calculate performance using SVM-OAO. The kernels used are linear, polynomial and RBF. To get the optimum parameter, used grid search method with \( C = 2^{-5} \) to \( 2^{15} \) and \( \gamma = 2^{-15} \) to \( 2^{1} \) [23]. Here is SVM-OAO algorithm (Algorithm 1).
**Figure 4. SVM-OAO Algorithm**

f. Perform feature selection using FCBF (Fast Correlation Based Filter) algorithm. Then calculate the SVM-OAO performance that has been in FCBF.

g. Clustering feature with Ensemble Feature Subspace (FSS) method of cluster number of 2 and 5 cluster. Each cluster contained SVM classifier to get EnSVM-OAO performance (Algorithm 2).

**Figure 5. Feature Subspace Algorithm**

h. Delete the major class data so that it same as the minor class data with Random Undersampling (RUS) then calculate EnSVM-RUS-OAO performance (Algorithm 3).

**Figure 6. Random Undersampling Algorithm**

i. Compare the classification performance results for all scenarios and methods, and draw conclusion.
4. Results and Discussion

The application of En-SVM-RUS-OAO method is applied to simulation data with various major and minor class comparison ratios. The simulated data is generated by following the Leukemia data pattern that takes mean and covariance for each class. The number of classes is limited to 3 classes showing multiclass. Then the feature size is raised as many as 1000 and 5000 features. These following tables are the average result of threefold cross-validation for scenario 1 with the division of the number classes is 33:33:34 which is further categorized in ratio 1. Scenario 2 with the division of the number of classes is 10:40:50 which is further categorized in the ratio 5. And scenario 3 with the division of the number of classes is 10:10:80 which is further categorized in Ratio 8. Then the number of clusters is limited to 2 clusters and 5 clusters based on the result of dendogram graphic in clustering hierarchy complete linkage process. The results are shown in Table 1,2, and 3 as follows.

| Table 1. Classification Performance of Ratio 1 (Scenario 1; 33:33:34) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Kernel          | Metode           | 1000 feature     | 5000 feature     |
|                 | Accuracy | F-score | G-mean | Accuracy | F-score | G-mean | Accuracy | F-score | G-mean |
| Linear          | SVM-OAO         | 95.00 | 95.04 | 94.93 | 95.5  | 95.01 | 95.5  |
|                 | SVM-OAO (FCBF)  | 95.18 | 95.16 | 95.12 | 96.18 | 95.83 | 96.66 |
|                 | EnSVM-OAO       | 95.31 | 95.01 | 95.35 | 96.45 | 96.42 | 97.03 |
|                 | SVM-OAO (FCBF)  | 95.33 | 95.05 | 95.33 | 96.40 | 96.34 | 97.05 |
| Polynomial      | SVM-OAO         | 94.73 | 95.17 | 95.25 | 95.32 | 94.79 | 94.98 |
|                 | SVM-OAO (FCBF)  | 95.69 | 96.01 | 96.06 | 95.96 | 95.04 | 96.28 |
|                 | EnSVM-OAO       | 95.74 | 97.20 | 96.42 | 96.22 | 95.10 | 96.65 |
|                 | EnSVM-RUS-OAO   | 95.72 | 97.19 | 96.45 | 96.25 | 95.08 | 96.62 |
| RBF             | SVM-OAO         | 95.17 | 95.14 | 94.96 | 95.12 | 95.15 | 94.64 |
|                 | SVM-OAO (FCBF)  | 95.81 | 95.84 | 95.2  | 95.81 | 96.06 | 95.14 |
|                 | EnSVM-OAO       | 96.83 | 96.29 | 96.62 | 96.67 | 96.61 | 96.89 |
|                 | EnSVM-RUS-OAO   | 96.82 | 96.28 | 96.63 | 96.68 | 96.68 | 96.91 |

| Table 2. Classification Performance of Ratio 5 (Scenario 2; 10:40:50) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Kernel          | Metode           | 1000 feature     | 5000 feature     |
|                 | Accuracy | F-score | G-mean | Accuracy | F-score | G-mean | Accuracy | F-score | G-mean |
| Linear          | SVM-OAO         | 89.87 | 89.51 | 90.44 | 90.24 | 90.14 | 90.14 |
|                 | SVM-OAO (FCBF)  | 90.41 | 89.93 | 90.69 | 90.51 | 90.32 | 90.26 |
|                 | EnSVM-OAO       | 90.72 | 90.4  | 90.78 | 90.77 | 90.88 | 90.72 |
|                 | EnSVM-RUS-OAO   | 90.83 | 90.61 | 90.98 | 91.05 | 91.13 | 90.99 |
| Polynomial      | SVM-OAO         | 88.49 | 91.98 | 90.36 | 90.24 | 90.14 | 90.14 |
|                 | SVM-OAO (FCBF)  | 89.95 | 91.96 | 90.57 | 90.51 | 90.32 | 90.78 |
|                 | EnSVM-OAO       | 91.31 | 91.99 | 90.97 | 90.77 | 90.88 | 91.10 |
|                 | EnSVM-RUS-OAO   | 91.73 | 92.19 | 91.36 | 91.84 | 91.13 | 91.21 |
| RBF             | SVM-OAO         | 91.92 | 92.37 | 91.74 | 90.67 | 92.06 | 91.13 |
|                 | SVM-OAO (FCBF)  | 92.22 | 92.69 | 91.90 | 91.14 | 92.17 | 92.44 |
|                 | EnSVM-OAO       | 92.62 | 93.36 | 92.03 | 92.01 | 93.43 | 93.52 |
|                 | EnSVM-RUS-OAO   | 93.02 | 93.83 | 93.94 | 93.50 | 94.61 | 94.21 |

| Table 3. Classification Performance of Ratio 8 (Scenario 3; 10:10:80) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Kernel          | Metode           | 1000 feature     | 5000 feature     |
|                 | Accuracy | F-score | G-mean | Accuracy | F-score | G-mean | Accuracy | F-score | G-mean |
| Linear          | SVM-OAO         | 89.84 | 89.3  | 89.39 | 88.78 | 89.11 | 88.83 |
|                 | SVM-OAO (FCBF)  | 90.65 | 89.92 | 90.14 | 90.56 | 90.37 | 90.13 |
|                 | EnSVM-OAO       | 91.00 | 90.94 | 90.26 | 91.00 | 90.89 | 90.22 |
|                 | EnSVM-RUS-OAO   | 91.50 | 91.26 | 91.16 | 91.36 | 91.14 | 91.48 |
| Polynomial      | SVM-OAO         | 88.97 | 88.94 | 88.44 | 89.38 | 88.44 | 89.27 |
|                 | SVM-OAO (FCBF)  | 89.61 | 89.99 | 89.89 | 90.05 | 89.69 | 90.06 |
|                 | EnSVM-OAO       | 90.25 | 90.44 | 90.40 | 90.21 | 89.94 | 90.29 |
|                 | EnSVM-RUS-OAO   | 91.70 | 91.72 | 91.87 | 92.09 | 92.06 | 92.31 |
| RBF             | SVM-OAO         | 88.87 | 88.95 | 88.69 | 89.40 | 88.65 | 89.96 |
|                 | SVM-OAO (FCBF)  | 89.93 | 90.16 | 90.26 | 89.96 | 89.53 | 90.26 |
|                 | EnSVM-OAO       | 90.06 | 90.19 | 90.85 | 90.37 | 90.37 | 90.39 |
|                 | EnSVM-RUS-OAO   | 92.45 | 92.53 | 92.29 | 92.78 | 92.53 | 92.62 |
A simulation study with 1000 features and 5000 feature shows the result that the EnSVM-RUS-OAO method has higher performance than the EnSVM-OAO and SVM-OAO methods. These are the best parameter selected using grid search algorithm. In the linear kernel is default parameter cost = 1, in polynomial kernels with $\gamma = 0.001$, cost = 1 and $p = 3$ and RBF kernel, with $\gamma = 2^{10}$, cost = $2^{13}$. Overall, it can be concluded that the EnSVM-RUS-OAO method using the RBF kernel gives higher results than linear kernels and polynomials. Overall, it can be concluded that the EnSVM-RUS-OAO method using the RBF kernel gives higher results than linear kernels and polynomials.

The Random undersampling correction technology helps SVM-OAO classifiers has a better classification performance in these sensitive data sets. Best Performance of the method is better reflected by the $F$-score and $G$-mean metric, which is used to evaluate the level of balance of the classification results. Thus, correction technology is useless when classifying tasks are strong for class imbalances than balance class.

In contrast with SVM-OAO, the ensemble version of EnSVM-OAO helps to slightly improve the overall classification accuracy, $F$-score and $G$-mean with possible sacrifice of two other evaluation metrics on most datasets. It means that classification accuracies between majority and minority classes are further increased. Feature subspace with clustering hierarchy process will cluster some feature into the similar characteristics by euclidian distance. Then each cluster that had been formed, will be classified by multiple binary SVM-OAO and majority voting used for the decision value. It is proven by simulation study that ensemble version EnSVM-OAO outperforms SVM-OAO.

![Figure 7. Computational Time (a) Ratio 1 (b) Ratio 5 and (c) Ratio 8](image-url)
5. Comparison of All Simulation Data Scenarios
The following is a comparison of simulation study results using EnSVM-RUS-OAO, EnSVM-OAO and SVM-OAO methods. The result of accuracy, $F$-score and $G$-mean can be shown in graph 7 as follows.

![Graph 7: Comparison of Accuracy, F-score, and G-mean for different methods](image)

**Figure 7.** Comparison of the Performance of All Simulation Data Scenarios

Based on Figure 8 it can be seen that, in condition ratio 1, the comparison of 33:33:34 class shows no significant difference between EnSVM-RUS-OAO, EnSVM-OAO and SVM-OAO methods. This is because during the RUS process, only delete 1 data only, the difference lies in the FCBF process. At the time of condition ratio 5 that is comparison of class 10:40:50 and ratio 8 that is comparison of class 10:10:80, look difference between main method after done feature of FCBF selection. So the EnSVM-RUS-OAO method is appropriate if applied to the classification of imbalance data. The higher the imbalance ratio makes the performance decrease. However, the EnSVM-RUS-OAO method provides superior performance compared to the EnSVM-OAO and SVM-OAO methods.

6. Conclusion
Ensemble methods have been increasingly applied to microarray problems in dealing with small sample size, high-dimensional, and complexity data structure. Then ensemble algorithm generally offer higher accuracy and stability than a single algorithm. In addition, ensemble algorithms can often alleviate the problems of small sample size with class imbalance and high dimensionality which commonly occur in many microarray data. It is worth mentioning that the increased accuracy is often accompanied with increased model complexity which causes decreased model interpretability and higher computational intensity. Nevertheless, the theoretical studies of ensemble approaches and the increase of computational power may counter those difficulties. Ensemble based feature selection is a promising approach for feature selection and biomarker identification in microarray problems because it can reduce time consuming for training classification time. Linear kernel has less computation time compared to the polynomial kernel and RBF, based on those simulations. It occurs because the number of a hyperplane by 2 dimensions is easier in separating process than 3 dimensions.

Acknowledgments
This study was supported under DIKTI GRANT National Competition Research (HIKOM) Scheme 527/PKS/ITS/2017.
References

[1] Seeja K R and Shweta S 2011 Microarray Data Classification Using Support Vector Machine *International Journal of Biometrics and Bioinformatics (IJBB)* 5(1) pp 10-5

[2] Purnami S W and Andari S 2015 High-Dimensional Data Classification Based on Smooth Support Vector Machines *Procedia Computer Science* 72 pp 477-84

[3] Rahman F and Purnami S W 2012 Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi LogistikOrdinal Dan Support Vector Machine (SVM) Jurnal SAINS dan Seni ITS 1 (2301-928X) 1.

[4] Lorena A C, De Carvalho and Gama J M 2008 A review on the combination of binary classifiers in multiclass problems *Artificial Intelligence Review* 30 pp 19-37

[5] Trapsilasiwati R K and Purnami S W 2017 SMOTE-Least Square Support Vector Machine for Classification of Multiclass Imbalanced Data *Proceedings of the 9th International Conference on Machine Learning and Computing* 107-111

[6] Statnikov A, Aliferis C F, Tsamaridis L and Levy S 2005 A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis *Bioinformatics* 21(5) pp 631-43.

[7] Chen Y and Zhao Y 2008 A Novel Ensemble of Classifiers for Microarray Data Classification *Applied Soft Computing* 8(4) pp 1664-9

[8] Saifudin A and Wahono R S 2015 Penerapan Teknik Ensemble untuk Menangani Ketidaksimbangan Kelas pada Prediksi Cacat Software *Journal of Software Engineering* 1(2356-3974) 1

[9] Liu H, Hussain F, Tan C L and Dash M 2002 Discretization: An Enabling Technique *Data Mining and Knowledge Discovery* 6 pp 393–423

[10] Canedo V B, Marono N S and Betanzos A A 2015 Feature Selection For High-Dimensional Data *Artificial Intelligence: Foundations, Theory, And Algorithms* (New York: Springer)

[11] Vapnik V 1995 *The Nature of Statistical Learning Theory* (New York: Springer)

[12] Scholkopf B and Smola A J 2000 *Learning with Kernel: Support Vector Machines, Regularization, Optimization, and Beyond* Massachusetts (London: The MIT Press)

[13] Li K, Liu Z and Han Y 2012 Study of Selective Ensemble Learning Methods Based on Support Vector Machine *Proc. Of International Conference on Medical Physics and Biomedical Engineering* 33 pp 1518 -25.

[14] Umamaheswari J and Radhamani G 2012 Quadratic Program Optimization using Support Vector Machine for CT Brain Image Classification *International Journal of Computer Science Issues* 9 (4) pp 305-310.

[15] Wu K P and Wang S D 2009 Choosing the kernel parameters for support vector machines by the inter-cluster *Pattern Recognition* 42 pp 710-717

[16] Yang P, Yang Y H, Zhou B B and Zomaya A Y 2010 A review of ensemble methods in bioinformatics; Including stability of feature selection and ensemble feature selection methods *Current Bioinformatics* 5(4) pp 96-308.

[17] Yu H, Hong S, Yang X, Ni J, Dan Y and Qin B 2013 Recognition of Multiple Imbalanced Cancer Types Based on DNA Microarray Data Using Ensemble Classifiers *BioMed Research International* 2013 1.

[18] Prachi V, Kar U, Rana M, Das M and Mishra B 2015 Microarray Classification of Cancerous Cell Using Soft Computing Technique *Procedia Computer Science* 49 pp 66-73.

[19] Han J, Kamber M and Pei J 2011 *Data Mining Concepts and Techniques:Third Edition Advanced Pattern Mining* (United States of America: Elsevier)

[20] Hsu C W and Lin C J 2002 A Comparison of Methods for Multi-class Support Vector Machines *Machine Learning* 46 pp 291–314
[21] Sahu B, Dehuri S and Jagadev A K 2017 Feature selection model based on clustering and ranking in pipeline for microarray data Informatics in Medicine Unlocked 9 pp 107-122.

[22] Hall A M 1999 Correlation-based Feature Selection for Machine Learning Performance Evaluation Department of Computer Science (The University of Waikato: Ph.D Thesis) pp 16-18

[23] Hsu, C W, Chang C and Lin C J 2016 A Practical Guide to Support Vector Classification Department of Computer Science 1-16