Predicting successes and failures of clinical trials with an ensemble LS-SVR A PREPRINT

PREDICTING SUCCESSES AND FAILURES OF CLINICAL TRIALS WITH AN ENSEMBLE LS-SVR

A PREPRINT

Zhen-Yu Hong¹, Jooyong Shim², Woo Chan Son³, Changha Hwang^{4*}

¹ Arontier, Seoul, Korea

² Institute of Statistical Information, Department of Statistics, Inje University, Gimhae, Korea

³ Department of Pathology, College of Medicine, University of Ulsan, Asan Medical Center, Seoul, Korea

⁴ Department of Applied Statistics, Dankook University, Yongin, Korea

January 22, 2020

ABSTRACT

For a variety of reasons, most drug candidates cannot eventually pass the drug approval process. Thus, developing reliable methods for predicting clinical trial outcomes of drug candidates is crucial in improving the success rate of drug discovery and development. In this study, we propose an ensemble classifier based on weighted least squares support vector regression (LS-SVR) for predicting successes and failures of clinical trials. The efficacy of the proposed ensemble classifier is demonstrated through an experimental study on ProCTOR dataset, which consists of informative chemical features of the drugs and target-based features. Comparing random forest and other models, the proposed ensemble classifier obtains the highest value for the area under the receiver operator curve (AUC). The results of this study demonstrate that the proposed ensemble classifier can be used to effectively predict drug approvals.

Keywords Area under the receiver operator curve · Imbalanced data · Least squares support vector machine

1 Introduction

In the analysis of the drug development costs of 98 companies in the past 10 years, the average cost of each drug developed and approved by a single-drug company was 350 million US dollars [1]. Although the extensive efforts and resources invested in identifying and elaborately designing new compounds as well as the systematic examination of all steps in the early development, even the most promising high-value compounds often fail in clinical trials. The probability of launch from entry to phase I has also remained below 10% [2]. An analysis of the causes of clinical failure in 2016-2018 shows that these are largely unchanged over the past 3 years [3]: 79% were due to safety or effectiveness; 1% were attributable to operational or technical shortcomings; 13% were the result of strategic realignment; and 7% were for commercial reasons [4, 5, 6]. It is one of the main challenges of the pharmaceutical industry to find innovative solutions for reducing the clinical failure rates, because failure in later clinical trials will lead to huge losses. At present, it is general practice to apply a variety of computational modeling approaches and simulation tools to upgrade and speed up drug discovery, design and other steps in the early development phase [7]. Moreover, the recently adopted computational-based strategy is to develop a set of criteria that can be applied to predict the results of clinical trials before they begin. A major focus of the criteria consists in the ability to identify compounds with adverse toxicity properties.

In fact, the prediction of successes and failures of drug candidates in clinical trials is a binary classification problem. In addition, this is imbalanced data classification problem, which is inevitable in this area. The group of data that has a larger number of examples is called the majority class or negative class, whereas the group of data that has a smaller number of examples is called the minority class or positive class. Most of the data in biometrics as well as clinical trials are usually imbalanced and have high-dimensional multimodal feature. This seriously affects the classification

*corresponding author: chwang@dankook.ac.kr

performance of the model. [8] lately developed a classifier named ProCTOR to deal with clinal drug toxicity on the basis of random forests that not only considers the bioavailability related properties of the drugs, but also target-related properties. The properties include established informative chemical features of the drugs with target-based features and tissue selectivity. In their study, the authors proved that some of these properties, even when considered separately, have a significant discriminative power. In addition, employing a feature importance analysis, they have exhibited that both the target-based and chemical features contribute to effective classification. In this paper, we develop an ensemble classifier based on weighted LS-SVR for bimodal imbalanced data, and then compare this model with random forest model on almost the same dataset as ProCTOR dataset used in [8]. This dataset was provided by the corresponding author of [8].

Developing an effective classification method for imbalanced and multimodal data is a challenging task. The rest of this paper is organized as follows. In Section 2 we review the standard LS-SVR and study variants of LS-SVR for bimodal data. In Section 3 we study variants of LS-SVR for imbalanced data. Section 4 and section 5 present experimental study and conclusion, respectively.

2 LS-SVRs for bimodal data

Least squares support vector machine (LS-SVM), which has been successfully applied to a number of real problems of classification and function estimation, is least squares version of SVM [9] and was proposed by [10]. LS-SVM has been proved to be a very appealing and promising method. There are some strong points of LS-SVM. One is that LS-SVM uses the linear equation which is simple to solve and good for computational time saving. Another is that LS-SVM classification is actually equivalent to LS-SVM regression in binary classification case [11]. For brevity, LS-SVM regression is called LS-SVR. For this reason, we will develop classification method for bimodal data using LS-SVR technique.

2.1 LS-SVR

Given a training data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ with each input $\mathbf{x}_i$ and corresponding response y_i , LS-SVR optimization problem in primal weight space is given as follows:

$$L(\mathbf{w}, b, \mathbf{e}) = \frac{1}{2} \mathbf{w}^t \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (1)$$

subject to equality constraints

$$y_i - \mathbf{w}^t \Phi(\mathbf{x}_i) - b = e_i, \quad i = 1, \dots, n \quad (2)$$

with a penalty parameter $\gamma > 0$, $\Phi : R^d \rightarrow R^{d_f}$ a function which maps the input space into a higher dimensional (possibly infinite dimensional) feature space, weight vector $\mathbf{w} \in R^{d_f}$ in primal weight space, error variables $e_i \in R$ and bias term b . To find minimizers of the objective function, we can construct the Lagrangian function as follows,

$$L(\mathbf{w}, b, \mathbf{e}; \alpha) = \frac{1}{2} \mathbf{w}^t \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (\mathbf{w}^t \Phi(\mathbf{x}_i) + b + e_i - y_i) \quad (3)$$

where α_i 's are the Lagrange multipliers. Then, the conditions for optimality are given by

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} = 0 & \rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i \Phi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 & \rightarrow \sum_{i=1}^n \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 & \rightarrow e_i = \frac{1}{\gamma} \alpha_i, \quad i = 1, \dots, n \\ \frac{\partial L}{\partial \alpha_i} = 0 & \rightarrow y_i - b - \mathbf{w}^t \Phi(\mathbf{x}_i) - e_i, \quad i = 1, \dots, n \end{aligned} \quad (4)$$

The estimation of parameters with equations in the above conditions requires the computations of inner products $\Phi(\mathbf{x}_i)^t \Phi(\mathbf{x}_j)$ in a potentially higher dimensional feature space. Under certain conditions these demanding computations can be reduced significantly by introducing a kernel function K such that $\Phi(\mathbf{x}_i)^t \Phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$ [15].

Possible kernel functions are linear kernel and radial basis function (RBF). The linear and RBF kernel are the most frequently used kernel function for the linear and the nonlinear case, respectively. After eliminating e_i and w , we can obtain the optimal values of (α, b) from the following linear equations

$$\begin{bmatrix} \hat{\alpha} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \mathbf{K} + \frac{1}{\gamma} \mathbf{I} & \mathbf{1} \\ \mathbf{1}^t & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}, \quad (5)$$

where $\mathbf{K} = \{K_{ij}\}$ is the kernel matrix with $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, \dots, n$.

Finally, for a given $\mathbf{x}_t$ the predicted value of response is given as

$$\hat{y}(\mathbf{x}_t) = \sum_{i=1}^n K(\mathbf{x}_t, \mathbf{x}_i) \hat{\alpha}_i + \hat{b}. \quad (6)$$

In particular, for the given training data set, we obtain

$$\hat{\mathbf{y}} = \mathbf{K} \hat{\alpha} + \hat{b} \mathbf{1}. \quad (7)$$

The functional structures of LS-SVR is characterized by the hyperparameters such as penalty parameter and kernel parameter. To choose optimal values of hyperparameters of the model we define a leave-one-out cross validation (LOOCV) function as follows:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(y_i - \hat{y}_i^{(-i)}(\lambda) \right)^2, \quad (8)$$

where $\lambda = \{\sigma, \gamma\}$ is the set of hyperparameters and $\hat{y}_i^{(-i)}$ is the predicted value of y_i obtained from the data without i th observation. Since for each candidate of hyper-parameters, $\hat{y}_i^{(-i)}(\lambda)$ for $i = 1, \dots, n$ should be evaluated, selecting parameters using CV function is computationally formidable. Generalized cross validation (GCV) function is obtained by applying the leaving-out-one lemma [12] as follows:

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{y}_i(\lambda))^2}{(1 - n^{-1} \text{tr}(\mathbf{S}(\lambda)))^2}, \quad (9)$$

where $\mathbf{S}(\lambda) = \mathbf{K} \left(\mathbf{Z}^{-1} - \mathbf{Z}^{-1} \frac{\mathbf{J}}{c} \mathbf{Z}^{-1} \right) + \frac{\mathbf{J}}{c} \mathbf{Z}^{-1}$. Here $c = \mathbf{1}^t \left(\mathbf{K} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \mathbf{1}$, $\mathbf{Z} = \left(\mathbf{K} + \frac{1}{\gamma} \mathbf{I} \right)^{-1}$ and $\mathbf{J}$ is a square matrix with all elements equal to 1. See for details [13].

2.2 Bimodal LS-SVR

Combining two different types of data for improving performance seems intuitively appealing task. We now devise LS-SVR model for bimodal data. For simplicity, it is named bLS-SVR. For bLS-SVR we consider two separate feature spaces, each of which is the molecular property related feature space and the target-based property related feature space. Then the optimization problem in primal weight space is written as follows:

$$L(\mathbf{w}_1, \mathbf{w}_2, b, e) = \frac{1}{2} \mathbf{w}_1^t \mathbf{w}_1 + \frac{1}{2} \mathbf{w}_2^t \mathbf{w}_2 + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (10)$$

subject to equality constraints

$$y_i - \mathbf{w}_1^t \Phi_1(\mathbf{x}_i) - \mathbf{w}_2^t \Phi_2(\mathbf{z}_i) - b = e_i, \quad i = 1, \dots, n, \quad (11)$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are the molecular property related input and the target-based property related input, respectively.

To find minimizers of the objective function, we can construct the Lagrangian function as follows,

$$L(\mathbf{w}_1, \mathbf{w}_2, b, e; \alpha) = \frac{1}{2} \mathbf{w}_1^t \mathbf{w}_1 + \frac{1}{2} \mathbf{w}_2^t \mathbf{w}_2 + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (\mathbf{w}_1^t \Phi_1(\mathbf{x}_i) + \mathbf{w}_2^t \Phi_2(\mathbf{z}_i) + b + e_i - y_i) \quad (12)$$

where α_i 's are the Lagrange multipliers. The optimal values of (α, b) from the following linear equations

$$\begin{bmatrix} \hat{\alpha} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_1 + \mathbf{K}_2 + \frac{1}{\gamma} \mathbf{I} & \mathbf{1} \\ \mathbf{1}^t & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}, \quad (13)$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are the kernel matrices with elements $K_1(\mathbf{x}_i, \mathbf{x}_j)$ and $K_2(\mathbf{z}_i, \mathbf{z}_j)$, respectively.

Thus, for a given $(\mathbf{x}_t, \mathbf{z}_t)$ the predicted value of response is given as

$$\hat{y}(\mathbf{x}_t, \mathbf{z}_t) = \sum_{i=1}^n (K_1(\mathbf{x}_t, \mathbf{x}_i) + K_2(\mathbf{z}_t, \mathbf{z}_i)) \hat{\alpha}_i + \hat{b}. \quad (14)$$

2.3 Stacked LS-SVR

To devise an LS-SVR reflecting the attribute of bimodal data, we borrow the idea of stacking which is an ensemble technique. For simplicity, it is named sLS-SVR. Because there two separate feature sets, we consider the ensemble algorithm to obtain better performance as follows:

1. Train LS-SVR using $\{(x_i, y_i)\}_{i=1}^n$ with the optimal values of hyperparameters and obtain $\hat{y}(x_i)$ and $\hat{y}(x_t)$ for given x_t .
2. Train LS-SVR using $\{(z_i, y_i)\}_{i=1}^n$ with the optimal values of hyperparameters and obtain $\hat{y}(z_i)$ and $\hat{y}(z_t)$ for given z_t .
3. Find regression coefficients $\beta_0, \beta_1, \beta_2$ from the linear regression model with the input $(\hat{y}(x_i), \hat{y}(z_i))$ and the corresponding output y_i .
4. Obtain the predicted value $\hat{y}(x_t, z_t) = \beta_0 + \beta_1 \hat{y}(x_t) + \beta_2 \hat{y}(z_t)$.

3 LS-SVR for imbalanced data

Since the ratio of approved drugs to those that failed for toxicity in clinical trials is imbalanced, the associated dataset is highly imbalanced. Thus, we now study some variants of LS-SVR for imbalanced data.

3.1 zLS-SVR

In SVM [9] the classification decision function can be written as follows:

$$\hat{y}(x) = \sum_{i \in SV_+} K(x, x_i) y_i \hat{\alpha}_i + \sum_{i \in SV_-} K(x, x_i) y_i \hat{\alpha}_i + \hat{b}, \quad (15)$$

where $y_i = 0$ or 1 , SV_+ and SV_- are the index sets of support vectors corresponding to positive and negative y_i 's, respectively. To reduce the bias of a learned SVM to majority class for imbalanced data, [14] introduced a multiplicative weight z , associated with each of the positive class support vectors. Under the assumption that minority class is considered as positive class, they reformulated SVM classification decision function as follows:

$$\hat{y}(x, z) = \sum_{i \in SV_+} z K(x, x_i) y_i \hat{\alpha}_i + \sum_{i \in SV_-} K(x, x_i) y_i \hat{\alpha}_i + \hat{b}, \quad (16)$$

This can be regarded as weighting the Lagrange multipliers $\hat{\alpha}_i$'s of positive class so that the minority classification can be improved.

We now apply this idea to LS-SVR. Thus, the classification decision function associated with LS-SVR can be rewritten as follows:

$$\hat{y}(x, z) = \sum_{i: y_i > 0} z K(x, x_i) \hat{\alpha}_i + \sum_{j: y_j < 0} K(x, x_j) \hat{\alpha}_j + \hat{b}, \quad (17)$$

where $\hat{\alpha}_i$ and $\hat{\alpha}_j$ are the Lagrange multipliers associated with positive and negative classes, respectively. The optimal value of a multiplicative weight z is obtained by grid search method, which minimizes the quadratic loss function associated with the training data. For simplicity, it is named zLS-SVR. The algorithm of zLS-SVR is given as follows:

1. Train LS-SVR using the training data with the optimal values of hyperparameters.
2. Find the optimal value of multiplicative weight z , which minimizes quadratic loss function of the training data.
3. Find the classification decision function $\hat{y}(x, z)$ for the test data using the optimal value z .
4. Classify the test data by using the classification decision function obtained in 3.

3.2 weighted LS-SVR

Under the assumption that minority class is considered as positive class, we set the weight $\xi_i = \frac{n_+}{n} I(y_i > 0) + \frac{n_-}{n} I(y_i < 0)$, n_+ is the number of positive y_i 's and n_- is the number of negative y_i 's. Then, the LS-SVR optimization problem in primal weight space is given as follows:

$$L(w, b, e) = \frac{1}{2} w^t w + \frac{\gamma}{2} \sum_{i=1}^n \xi_i e_i^2 \quad (18)$$

subject to equality constraints

$$y_i - \mathbf{w}^t \Phi(\mathbf{x}_i) - b = e_i, i = 1, \dots, n. \quad (19)$$

The optimal values of Lagrange multipliers and bias are obtained from the following linear equations

$$\begin{bmatrix} \hat{\alpha} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \mathbf{K} + \frac{1}{\gamma} \mathbf{W} & \mathbf{1} \\ \mathbf{1}^t & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}, \quad (20)$$

where $\mathbf{W}$ is the diagonal matrix of $1/\xi$'s. Finally, for a given $\mathbf{x}_t$ the predicted value of response is given as

$$\hat{y}(\mathbf{x}_t) = \sum_{i=1}^n K(\mathbf{x}_t, \mathbf{x}_i) \hat{\alpha}_i + \hat{b}. \quad (21)$$

For simplicity, this method is named wLS-SVR.

4 Experimental study

In this section we assess the performance of the proposed ensemble classifier through ProCTOR dataset, which consists of 68 failed drugs for positive class and 708 approved drugs for negative class. The feature vector is divided into $\mathbf{x} \in R^{17}$ (molecular property related) and $\mathbf{z} \in R^{30}$ (target-based property related). We calculate means and standard errors of AUCs for LS-SVR, bLS-SVR, sLS-SVR and random forest (RF) with 100 trees via 5-fold and 10-fold cross validation techniques. We iterate the above procedures 100 times to obtain reliable results. We first investigate how

Table 1: AUC results for RF and three types of LS-SVR via 5-fold and 10-fold cross validation techniques without considering class-imbalance

	RF	LS-SVR	bLS-SVR	sLS-SVR
5-fold CV	0.6738 (0.0017)	0.6899 (0.0017)	0.6978 (0.0014)	0.7098 (0.0018)
10-fold CV	0.6785 (0.0013)	0.6894 (0.0013)	0.7016 (0.0010)	0.7157 (0.0011)

three types of LS-SVR model work for bimodal classification without coping with class-imbalance. Table 1 shows the results for the means and standard errors of 100 AUCs for original dataset. Standard errors are given in parenthesis. Boldfaced values indicate best performance result. As seen from Table 1, stacked LS-SVR model shows the best performance on original dataset.

Table 2: AUC results for RF and nine types of LS-SVR via 5-fold and 10-fold cross validation techniques for with considering class-imbalance

	RF	LS-SVR	bLS-SVR	sLS-SVR
5-fold CV	0.7142 (0.0017)	0.6844 (0.0017)	0.6921 (0.0016)	0.7040 (0.0016)
10-fold CV	0.7172 (0.0013)	0.6888 (0.0013)	0.7010 (0.0012)	0.7119 (0.0011)
	zLS-SVR	bzLS-SVR	szLS-SVR	
5-fold CV	0.6981 (0.0017)	0.6978 (0.0014)	0.7098 (0.0017)	
10-fold CV	0.6891 (0.0013)	0.7016 (0.0010)	0.7157 (0.0011)	
	wLS-SVR	bwLS-SVR	swLS-SVR	
5-fold CV	0.7111 (0.0017)	0.7130 (0.0016)	0.7204 (0.0016)	
10-fold CV	0.7143 (0.0070)	0.7093 (0.0072)	0.7236 (0.0006)	

We next investigate how various types of LS-SVR model work for bimodal classification with coping with class-imbalance. We apply synthetic minority over-sampling technique (SMOTE) [16] in advance before training RF, LS-SVR, bLS-SVR and sLS-SVR which are not able to cope with class-imbalance problem. We utilize zLS-SVR and

wLS-SVR to overcome imbalance of the data. Table 2 shows the results for the means and standard errors of 100 AUCs. Here, bzLS-SVR, szLS-SVR, bwLS-SVR and swLS-SVR stand for binomial zLS-SVR, stacked zLS-SVR, binomial wLS-SVR and stacked wLS-SVR, respectively. As seen from Table 2, stacked wLS-SVR shows the best performance when considering the class-imbalance. Thus, we recognize that stacked wLS-SVR performs best for bimodal imbalanced ProCTOR data.

Figure 1 shows AUC results via 5-fold CV for LS-SVR, LS-SVR(SMOTE), zLS-SVR and wLS-SVR. Figure 2 shows AUCs results via 5-fold CV for bLS-SVR, bLS-SVR(SMOTE), bzLS-SVR and bwLS-SVR. Figure 3 shows AUC results via 5-fold CV for sLS-SVR, sLS-SVR(SMOTE), szLS-SVR and swLS-SVR. Figure 4 shows AUC results via 10-fold CV for LS-SVR, LS-SVR(SMOTE), zLS-SVR and wLS-SVR. Figure 5 shows AUCs results via 10-fold CV for bLS-SVR, bLS-SVR(SMOTE), bzLS-SVR and bwLS-SVR. Figure 6 shows AUC results via 10-fold CV for sLS-SVR, sLS-SVR(SMOTE), szLS-SVR and swLS-SVR. Figure 7 shows AUC results via 5-fold and 10-fold CV for RF and RF(SMOTE). From tables and figures we can see that the proposed ensemble classifier shows the best performance for bimodal imbalanced ProCTOR data.

Figure 1: **ROC curves for 5-fold CV.** LS-SVR (upper left), LS-SVR(SMOTE) (upper right), zLS-SVR (lower left) and wLS-SVR (lower right)

Figure 2: **ROC curves for 5-fold CV.** bLS-SVR (upper left), bLS-SVR(SMOTE) (upper right), bzLS-SVR (lower left) and bwLS-SVR (lower right)

Figure 3: **ROC curves for 5-fold CV.** sLS-SVR (upper left), sLS-SVR(SMOTE) (upper right), szLS-SVR (lower left) and swLS-SVR (lower right)

Figure 4: **ROC curves for 10-fold CV.** LS-SVR (upper left), LS-SVR(SMOTE) (upper right), zLS-SVR (lower left) and wLS-SVR (lower right)

Figure 5: **ROC curves for 10-fold CV.** bLS-SVR (upper left), bLS-SVR(SMOTE) (upper right), bzLS-SVR (lower left) and bwLS-SVR (lower right)

Figure 6: **ROC curves for 10-fold CV.** sLS-SVR ensemble (upper left), sLS-SVR(SMOTE) (upper right), szLS-SVR (lower left) and swLS-SVR (lower right)

Figure 7: **ROC curves for 5-fold & 10-fold CV.** RF (5-fold CV, upper left), RF(SMOTE) (5-fold CV, upper right), RF (10-fold CV, lower left) and RF(SMOTE) (10-fold CV, lower right)

5 Conclusions

In this paper, we dealt with predicting successes and failures of clinical trials with some variants of LS-SVR model. We found that an ensemble classifier based on weighted LS-SVR provides good results in predicting clinical trial outcomes. In general, the LS-SVR requires knowledge of the related hyperparameters. By the way, the model selection process of LS-SVR makes it possible to obtain the estimates of hyperparameters. Thus, the proposed ensemble classifier appears to be useful in predicting clinical trial outcomes.

To conclude, the proposed ensemble classifier basically has two advantages. One is that this model takes over advantages that SVM works very well for a number of real world problems and overcomes the curse of dimensionality. Thus, the proposed ensemble classifier can be applied easily and effectively to the prediction problem with high dimensional input vector. The other is that this method can predict clinical trial outcomes without knowledge of hyperparameters in advance, because the estimates of hyperparameters are obtained during the model selection process.

Acknowledgements

We thank Olivier Elemento for providing us access to his data. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07042349). This work was supported by "Human Resources Program in Energy Technology" of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20174030201740). This research was supported by the Bio Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2019M3E5D4066897).

References

- [1] Herper, M. (11 August 2013). The cost of creating a new drug now \$5 billion, Pushing Big Pharma To Change. Forbes, Pharma Healthcare. Retrieved 17 July 2016.
- [2] Helen D., Jamie M. Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Discov.* 2019 Jul; 18: 495–496.
- [3] Richard K.H. Phase II and phase III failures: 2013–2015. *Nat. Rev. Drug Discov.* 2019 Jul; 15: 817–818.
- [4] Graul A.I, Dulsat C., Tracy M., Cruces E. The year's new drugs biologics 2016: Part II - Trends and highlights of an unforgettable year. *Drugs Today (Barc)*. 2017 Feb; 53(2): 117–158.
- [5] Graul A.I, Dulsat C., Pina P., Tracy M., D'Souza P. The year's new drugs & biologics, 2017, part II - News that shaped the industry in 2017. *Drugs of Today (Barc)*. 2018 Feb; 54(2): 137–167.
- [6] Graul A.I, Dulsat C., Pina P., Cruces E., Tracy M. The year's new drugs and biologics 2018: Part II - News that shaped the industry in 2018. *Drugs Today (Barc)*. 2019 Feb; 55(2): 131–160.
- [7] Artem V.A., Evgeny P., Quentin V., Alexander A., Ivan V.O., Alex Z. Integrated deep learned transcriptomic and structure-based predictor of clinical trials outcomes. *bioRxiv preprint first posted online Dec. 20, 2016*; doi: <http://dx.doi.org/10.1101/095653>.
- [8] Gayvert K.M., Madhukar N.S., Elemento O. A data-driven approach to predicting successes and failures of clinical trials. *Cell Chem Biol.* 2016 Oct 20;23(10): 1294–1301.
- [9] Vapnik V.N. The nature of statistical learning theory. Springer, New York; 1995.
- [10] Suykens J.A.K., Vanderwalle, J. Least square support vector machine classifier. *Neural Process. Lett.* 1999 Jun; 9(3): 293–300.
- [11] Shim J., Bae J., Hwang, C. Multiclass classification via least squares support vector machine regression. *Commun. Stat. Appl. Methods.* 2008 May; 15(3): 441–450.
- [12] Kimeldorf G.S., Wahba G. Some results on Tchebycheffian spline functions. *J Math Anal Appl.* 1971 Jan; 33(1): 82–95.
- [13] De Brabanter J., Pelckmans K., Suykens J.A.K., Vandewalle J., De Moor B. Robust cross-validation score functions with application to weighted least squares support vector machine function estimation. Tech. Report, Department of Electrical Engineering, ESAT-SISTA 2003.

- [14] Imam T., Ting K.M., Kamruzzaman, J. (2006). z-SVM: An SVM for improved classification of imbalanced data. In *Advances in Artificial Intelligence*, 19th Australian Joint Conference on Artificial Intelligence, Hobart, 4–8; Dec 2006, Proceedings, 264-273.
- [15] Mercer J. Function of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A* 1909; 415–446.
- [16] Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. (2002). SMOTE: synthetic minority over-sampling technique. *Artif Intell Res.* 2002; 16(1): 321–357.