



# Article Comparison of Selection Criteria for Model Selection of Support Vector Machine on Physiological Data with Inter-Subject Variance

Minho Choi<sup>1</sup> and Jae Jin Jeong<sup>2,\*</sup>

- <sup>1</sup> Department of Creative IT Engineering, Pohang University of Science and Technology, Jigok-ro 80, Pohang 37673, Korea; minho17@postech.ac.kr
- <sup>2</sup> School of Electronic and Electrical Engineering, Daegu Catholic University,
- Hayang-ro 13-13, Gyeongsan 38430, Korea
- Correspondence: jjeong@cu.ac.kr

**Abstract:** Support vector machines (SVMs) utilize hyper-parameters for classification. Model selection (MS) is an essential step in the construction of the SVM classifier as it involves the identification of the appropriate parameters. Several selection criteria have been proposed for MS, but their usefulness is limited for physiological data exhibiting inter-subject variance (ISV) that makes different characteristics between training and test data. To identify an effective solution for the constraint, this study considered a leave-one-subject-out cross validation-based selection criterion (LSSC) with six well-known selection criteria and compared their effectiveness. Nine classification problems were examined for the comparison, and the MS results of each selection criterion were obtained and analyzed. The results showed that the SVM model selected by the LSSC yielded the highest average classification accuracy among all selection criteria in the nine problems. The average accuracy was 2.96% higher than that obtained with the conventional K-fold cross validation-based selection criterion. In addition, the advantage of the LSSC was more evident for data with larger ISV. Thus, the results of this study can help optimize SVM classifiers for physiological data and are expected to be useful for the analysis of physiological data to develop various medical decision systems.

Keywords: support vector machine; model selection; physiological data; inter-subject variance

# 1. Introduction

Deep-learning techniques, including convolutional and recurrent neural networks, are being actively researched. Several related studies have reported successful results in many fields [1–5]. Nevertheless, the support vector machine (SVM)—one of the well-known machine learning techniques—still constitutes an option for the construction of classification systems [6–8]. The SVM has superior generalization ability compared with other classifiers and can be used in instances wherein the training data are not extensive [9,10]. Additionally, the complexity of the SVM is also relatively lower than those of deep-learning techniques [11]. Correspondingly, SVM is an effective solution given the limited computational resources.

However, the techniques using SVM rely on hyper-parameters to construct the kernel function and to penalize slack variables [12,13]. The process used to select the appropriate hyper-parameters is essential to the design of the SVM classifier with a preferred performance. This hyper-parameter setting is called model selection (MS). Gradient-based methods comprise one of the categories of the MS. Gradient-based methods calculate the gradient of a specific generalization error function and use the gradient descent approach to determine the appropriate model [14,15]. Various differentiable generalization error functions have been used in gradient-based methods, but the local minima problem and sensitivity to initialization are the major drawbacks of these methods [16]. Grid-search methods are alternative approaches that can be used for MS. These methods construct a grid on the hyper-parameters and evaluate each model that corresponds to each grid point



Citation: Choi, M.; Jeong, J.J. Comparison of Selection Criteria for Model Selection of Support Vector Machine on Physiological Data with Inter-Subject Variance. *Appl. Sci.* 2022, *12*, 1749. https://doi.org/ 10.3390/app12031749

Academic Editor: Federico Divina

Received: 10 December 2021 Accepted: 3 February 2022 Published: 8 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). to select the optimal SVM model [17,18]. The grid-search-based methods are simple and are not prone to the local minima problem, but only the models that are based on fixed grid points can be tested for MS. To resolve this weakness and facilitate more flexibility for model searching, uniform design and evolutionary-algorithm-based methods have been introduced [19,20]. A more detailed review on the recent MS methods can be found in [21]. There are some differences in the manner in which candidate SVM models are chosen for testing. The basic ideas of the grid search, uniform design, and evolutionary-algorithm-based methods are similar as these methods assess the various candidate SVM models and select the model with the optimum performance.

In this respect, it is important to evaluate the performance of each SVM model precisely for MS, and selection criteria are used for the purpose. Generalization error obtained by leave-one-out cross validation (LOOCV) is a popular selection criterion used to evaluate the SVM models [22]. Herein, when the training set contains N data, LOOCV trains the SVM using N - 1 data and tests it with the remaining one datum to evaluate the classification error. The generalization error is calculated from the mean of the classification errors for the N sets of data. These generalization errors are then compared for each model, and the model with the lowest generalization error is selected in the MS step. The LOOCV is commonly known as an unbiased estimator of the generalization error but incurs a heavy computational load because the SVM is trained N times for each model. However, the K-fold cross validation (KCV) is associated with a smaller computational load, and generalization error calculated by the KCV has greater practical use to be considered as a suitable alternative selection criterion. In this method, the training data are divided into K segments, and the segments are individually cross validated. The generalization error is obtained as the average error from the K tests. The mean value of sensitivity and specificity can also be used as a selection criterion in the KCV [18]. The radius margin bound (RMB), maximal discrepancy (MD), xi-alpha bound (XAB), generalized approximate cross validation (GACV), distance between two classes (DBTC), and expected square distance ratio (ESDR) are some of the other selection criteria for MS [23–25]. The effectiveness of these selection criteria was compared in previous studies. Duan used five datasets to analyze the benefits of seven different selection criteria, including the KCV-based selection criterion (KSC), XAB, and GACV; in their study, the test errors were calculated for the SVM models that used each of these selection criteria [26]. The KSC approach yielded the lowest error when used for MS, followed by the XAB and GACV approaches. Duarte simulated the MS for SVM using the KSC, XAB, GACV, DBTC, MD, XAB, and RMB approaches for 110 public datasets [27]. Similar to the results of Duan, KSC outperformed the other approaches in terms of classification accuracy, followed by the DBTC approach.

The effectiveness of the selection criteria was verified for many datasets in previous studies, but the results may differ for physiological data which exhibit inter-subject variance (ISV). The physiological data are obtained from physiological signals such as electrocardiogram (ECG), electrodermal activity (EDA), and electromyogram (EMG). The signals are generated by the physical and mental activity of the human, and they can be used to know the information on our body. Responses to external stimuli and body conditions are not the same for every person, and the characteristics of physiological signals are different for each person [28]. The characteristic differences are reflected to physiological data as ISV, and it is one of factors that makes the analysis of the physiological data difficult. Owing to the ISV, the performance of a classifier can be degraded for the data of an unseen subject that are not included in the training set because there is a difference between the test and training data with respect to characteristics. Considering the difficulty, it is recommended to verify the performance of a classifier through leave-one-subject-out cross validation (LOSOCV) when physiological data are analyzed. The LOSOCV is similar to KCV but divides the entire data into datasets for each of the subjects for cross validation to know generalized performance for the data of an unseen subject. Indeed, Gholamiangonabadi recently showed that the LOSOCV is useful for evaluating the generalized ability of deep learning models for the data of unseen subjects, and it is necessary to tune the model based on LOSOCV [29]. These ISV and validation processes using LOSOCV must be considered

in the MS process of SVM as well. Thus, results verified through a classification problem that contained training and test data with relatively similar characteristics should be re-examined for the MS on physiological data. Rojas-Domínguez compared several MS methods for medical datasets [30]. However, to our knowledge, to-this-date, no study has analyzed the effectiveness of selection criteria in MS cases associated with physiological data. Therefore, this study uses a LOSOCV-based selection criterion (LSSC) together with other well-known selection criteria, including KSC, DBTC, ESDR, XAB, and GACV for the MS of SVM on physiological data and compares their effectiveness. The LSSC reflects the principle of the validation process via LOSOCV to MS, and it evaluates the performance of an SVM model using the LOSOCV results of the data available for MS. Nine classification problems are defined using physiological data to simulate the MS of the SVM according to each of the selection criterion. Results show that the LSSC outperforms others in terms of classification accuracy. A detailed comparison of selection criteria for physiological data with ISV and a discussion on the case enhancing the advantages of LSSC are also primary contributions of this research.

## 2. Materials and Methods

# 2.1. MS of SVM

The SVM chooses the classification boundary between two classes by solving the optimization problem,

minimize 
$$\frac{1}{2} \|W\|^2 + C \sum_{i=1}^N \xi_i$$
 (1)  
s.t.  $d_i \{\varphi(X_i) \cdot W + w_0\} \ge 1 - \xi_i,$   
 $\xi_i \ge 0, \quad i = 1, ..., N,$ 

where  $X_i \in \mathbb{R}^L$  is an *L*-dimensional datum, and  $d_i \in \{-1, 1\}$  is a label in the training set [28,31]. The influence of misclassified data is represented by the slack variable,  $\xi_i$ , and *C* determines the degree of penalty for  $\xi_i$ . The function  $\varphi(X_i)$  is the high-dimensional mapping of  $X_i$ . The SVM uses a kernel trick to simplify the optimization problem and converts the inner product,  $\varphi(X_i) \cdot \varphi(X_j)$  to a kernel function,  $K(X_i, X_j)$ . There are several available kernel functions for the SVM as follows:

Linear: 
$$X_i \cdot X_j$$
, (2)

Polynomial:  $(\gamma X_i \cdot X_j + a)^b$ , (3)

Radial basis function (RBF):  $\exp(-\gamma ||X_i - X_j||^2)$ , (4)

Sigmoid: 
$$tanh(\gamma X_i \cdot X_j + a)$$
. (5)

Among these, the RBF is used for the kernel function of the SVM in this study because of its performance and popularity [32]. The two hyper-parameters, namely *C* and the kernel parameter  $\gamma$ , are designated via MS.

## 2.2. LSSC

This study mainly focuses on LSSC, and its detailed description is presented in this section. The explanation on the other selection criteria can be found in [26,27]. The conventional selection criteria for the MS of SVM assume that all the data in the training set have similar characteristics. However, there are certain degrees of differences in the characteristics among data from different subjects because of ISV. This study utilizes the LSSC to consider the ISVs of the physiological data and to measure the performance of the SVM model during MS. Algorithm 1 shows the detailed procedure used to calculate the LSSC, which is obtained for each tested SVM model, that is, the pair of *C* and  $\gamma$  values designated by the MS method. In the procedure, available data for MS are normalized

and separated into datasets for each of the subjects. If the data were measured from *N* subjects, then a total of *N* datasets are considered. Subsequently, the divided datasets are cross validated using the LOSOCV approach; this means that the dataset of a single subject is selected as the test data (*TE*), and the datasets of the remaining subjects are used as the training data (*TR*). The SVM is trained using the *TR* with the designated *C* and  $\gamma$  values, and the classification accuracy of the SVM is evaluated with the *TE*. This evaluation process is repeated *N* times, and the LSSC is obtained as the averaged result from the *N* evaluation processes. This selection criterion is mainly used to represent the difference in data characteristics among the data of each subject to MS. Therefore, each SVM model

is evaluated for the data of an unseen subject, and the results are averaged to derive the performance of the tested SVM model for data that had specific characteristic differences with the training data. A sample MATLAB code (version R2017b, MathWorks, Natick, MA, USA) that obtains the LSSC can be observed online (https://github.com/minho17/paper-Comparison-of-performance-measures (accessed on 3 December 2021)).

Algorithm 1: Pseudocode used for LSSC in MS
Input:
- <i>C</i> and $\gamma$ values to be tested
- Data of $N$ subjects that can be used for MS
<b>Output:</b> evaluated performance ( <i>P</i> ) for the tested model
P = 0
Normalization for the data of N subjects
For $i = [1, N]$
TE = Dataset of subject <i>i</i>
$TR = \{ \text{Dataset of subject } j \mid j = 1,, N, j \neq i \}$
Train SVM using <i>C</i> , $\gamma$ , and <i>TR</i>
ACC = classification accuracy of the SVM for $TE$
P = P + ACC
End For
$P = \frac{P}{N}$
Return P

## 2.3. Classification Problems on Physiological Data

Physiological data were extracted from several datasets containing physiological signals, and nine classification problems were constructed using the physiological data to simulate LSSC and compare its effectiveness with other conventional selection criteria. Table 1 shows the sample size of each class and the number of subjects and features for the nine problems. The methods used to extract features were the same as previous studies [9,33–42], and details of the extracted features can be found in Appendix A. The first problem here is to detect medium or high stress from low-stress data [33,34]. These data contain features that were extracted from several physiological signals, such as the ECG, EDA, EMG, and respiration. The SWELL database provides the ECG and EDA data for the normal and stress states of subjects engaged in office work [35]. The stress data are detected in problem 2 using the heart rate variability (HRV) features of the ECG [36]. Problems 3–5 are based on our previous work, which utilized photoplethysmogram (PPG) and EDA data to detect the stress, drowsiness, and fatigue of drivers [9,28]. The brain–computer interface competition IV-2b database provides electroencephalography (EEG) signals recorded when subjects imagined the movement of one of their hands (left or right) [37]. Problem 6 uses the database to design a classification problem to distinguish between EEG data for movements of the left and right hands. The EEG data are feature vectors that were extracted from EEG signals (with lengths = 9 s) with the use of a common spatial pattern algorithm in the problem [41]. Lopez obtained the heart rate (HR) data from subjects in three thermal conditions: very hot (VH) (32 °C), VH with adjustable neck-coolers (18–28 °C), and hot environment (29 °C) [38]. Nkurikiyeyezu proposed HRV-based features to distinguish the data for each of these three

conditions [42], and these feature data are utilized for problems 7 and 8. Ventricular ectopic beats (VEBs) in the ECG are detected in problem 9; the RR interval, higher order statistics, wavelet transform, and morphological descriptor-based features were obtained from the modified-lead II ECG data of the MIT-BIH arrhythmia database, and the data of a subject that had at least 100 VEB data were used for the problem [39,40]. In addition, the data size was limited to 100 for each class and subject in problems 2 and 7–9 to simplify the classification and to allow considerations of imbalanced data sizes.

**Table 1.** Nine classification problems for MS simulations. #S and #F, respectively, denote the numbers of subjects and features.

Problem	#S	#F	Data (Sample Size)
1. Stress detection 1	10	22	low (166) / medium + high stress (217)
2. Stress detection 2	18	20	normal (1800)/stress (1800)
3. Stress detection 3	28	5	normal (1692)/stress (1548)
4. Fatigue detection	28	7	normal (1692)/fatigue (3073)
5. Drowsiness detection	28	8	normal (1692)/drowsiness (3110)
6. Motor imaginary	9	18	left (1706)/right hand (1717)
7. Thermal comfort 1	11	29	VH with cooler (1100)/VH (1100)
8. Thermal comfort 2	11	29	VH with cooler (1100)/hot (1100)
9. VEB detection	16	45	normal (1600)/VEB (1600)

#### 3. Results

Simulations were conducted to analyze the effectiveness of LSSC and compare it with other well-known selection criteria, namely KSC, KSC2, DBTC, ESDR, XAB, and GACV. These selection criteria were selected in this study because they have been extensively used for MS, and their effectiveness has been validated in several studies [25-27]. Among these, the KSC2 is a type of KSC and uses the mean value of the sensitivity and specificity rather than classification accuracy to evaluate a model. The KSC and KSC2 methods divided the entire data into five segments for cross validation. A simple grid-search method with *C* and  $\gamma \in \{2^{-6}, 2^{-4}, ..., 2^{10}, 2^{12}\}$  was used as the MS method for all selection criteria. The performance of all models at the grid points was evaluated by the seven selection criteria, and the best model was selected for each selection criterion. SVM was then trained using the selected model for each selection criterion, and the classification accuracies of the SVMs were compared. The LOSOCV was used to obtain the selection criterion for the MS in this study, but the LOSOCV is commonly used to test classifiers for the data of unseen subjects who were excluded from the training set, similar to real-world scenarios. Therefore, the aforementioned nine problems were validated and classification accuracies were obtained by using the LOSOCV in the simulations.

Tables 2 and 3 show the simulation results for the classification accuracies and their rankings for the nine problems, respectively. The Optimal values are the results for the best cases with the highest grid-point accuracies in each test, and they are the upper limits of the classification accuracies that can be achieved by MS. The accuracy for the LSSC was 81.09% on average for the nine problems in the simulations. This value was lower than the **Optimal** value but outperformed other selection criteria. It was 2.38% higher than that of the DBTC, which yielded the second highest accuracy, and 2.96% higher than the KSC, which is the most commonly used selection criterion. The ranking of the LSSC was 1.44 on average, and it was superior to the rankings of the other selection criteria. The DBTC and GACV were the second-best selection criteria, but the variation in their rankings was severe. The standard deviation (SD) values of DBTC and GACV were 1.80 and 2.35 in the ranking, and they were larger than that of the LSSC, which had an SD value of 0.73. The ranking of the LSSC was not more than three and was equal to one in most cases. Therefore, the LSSC was more effective and less sensitive to the classification problem compared with the other selection criteria. Among the other selection criteria, the DBTC was the second most useful. The GACV had the same ranking on average with the DBTC,

but it was less effective than some of the other selection criteria for some of the problems, such as problems 2 and 6. The DBTC was generally more stable and yielded smaller SD values compared with GACV in terms of classification accuracy and ranking, and the DBTC had a higher mean accuracy. In the case of KSC, the mean of ranking score was worse than that of GACV; however, its mean accuracy was better than that of the GACV and ranked after that of the DBTC because the KSC yielded a minor variation in accuracy according to the classification problem compared with the GACV.

Problem	KSC	KSC2	DBTC	ESDR	XAB	GACV	LSSC	Optimal
1	81.53	82.24	87.29	79.47	76.30	83.83	83.96	96.23
2	74.00	73.97	72.97	73.42	71.19	64.19	79.33	89.22
3	82.06	81.94	81.95	82.17	82.85	83.36	83.49	91.17
4	88.90	88.85	89.10	89.02	90.42	90.67	90.51	96.57
5	90.84	90.42	90.99	90.73	92.85	92.44	91.82	97.37
6	72.80	72.80	72.95	72.74	70.22	50.65	73.41	76.21
7	63.32	63.36	63.64	60.77	59.05	59.68	69.23	77.00
8	64.05	64.00	64.55	61.77	59.91	65.00	70.73	79.50
9	85.72	85.22	84.97	86.44	84.16	86.06	87.38	94.34
Mean	78.13	78.09	78.71	77.39	76.33	75.10	81.09	88.62
SD	10.16	10.07	10.47	11.07	12.28	15.26	8.42	8.73

Table 2. Classification accuracy (%) for each selection criterion.

Table 3. Ranking of accuracy for each classification problem.

Problem	KSC	KSC2	DBTC	ESDR	XAB	GACV	LSSC
1	5	4	1	6	7	3	2
2	2	3	5	4	6	7	1
3	5	7	6	4	3	2	1
4	6	7	4	5	3	1	2
5	5	7	4	6	1	2	3
6	3	3	2	5	6	7	1
7	4	3	2	5	7	6	1
8	4	5	3	6	7	2	1
9	4	5	6	2	7	3	1
Mean	4.22	4.89	3.67	4.78	5.22	3.67	1.44
SD	1.20	1.76	1.80	1.30	2.28	2.35	0.73

Computational time was measured to compare the computational complexity of each selection criterion (Table 4). This study utilized ten C and ten  $\gamma$  values equal to  $2^{-6}$ ,  $2^{-4}$ , ...,  $2^{10}$ , and  $2^{12}$  in the MS. The time needed for evaluating the performance of SVM models on the grid points was obtained for each selection criterion with the use of MATLAB and a computer equipped with an Intel Core i7-6700 CPU and 24 GB RAM. The result will be changed according to the used MS method, the range of tested C and  $\gamma$  values, and the number of subjects and features, but the result will provide an insight on the computational complexity of selection criteria. In the simulation results, the time required for DBTC or ESDR was short. Other selection criteria test SVM models at all grid points, but the two selection criteria determined a  $\gamma$  value first by analyzing data, and a *C* value was selected. Therefore, the computational time was shorter than those of other selection criteria. In contrast, the time for the LSSC was the longest because it conducted cross validation N times for each grid point when the data of N subjects were used for MS. The time increased in the case of the classification problem, which contained many subjects, and the computational time of the LSSC was larger than others, particularly in problems 3–5, which contained 28 subjects. The computational time of the LSSC was also affected

by the number of features. The number of subjects was the same in problems 3–5, but the computational time increased in the case of the problem that used more features.

Problem	KSC	KSC2	DBTC	ESDR	XAB	GACV	LSSC
1	2.21	2.22	0.20	0.21	2.12	1.31	4.91
2	181.85	181.47	12.22	15.80	141.15	89.84	738.46
3	205.72	207.95	38.16	9.90	104.05	82.10	1309.69
4	203.88	205.48	14.84	19.58	142.74	95.23	1467.00
5	209.10	210.07	13.35	20.86	150.53	98.61	1529.19
6	313.58	316.27	42.36	16.06	169.13	129.10	577.19
7	102.80	103.30	10.85	16.87	75.34	49.74	224.57
8	90.65	90.68	8.18	9.35	69.58	44.45	199.50
9	191.61	191.53	11.51	15.64	483.72	203.65	654.59
Mean	166.82	167.66	16.85	13.81	148.71	88.23	745.01

Table 4. Computational time (s) for each selection criterion.

The comparison results on classification accuracy and computational time will provide valuable information when selecting a selection criterion for a classification problem on physiological data. For example, the LSSC had the highest classification accuracy but needed a long computational time. Therefore, it could be considered in a preferential manner in the common situation wherein MS is conducted offline. DBTC ranked second in terms of classification accuracy, but the computational time was short. It could be an alternative to the LSSC in situations where MS and the training process of SVM are implemented online.

## 4. Discussion

#### 4.1. Statistical Analysis on Classification Accuracy

Statistical analysis was conducted to verify the effectiveness of LSSC compared with other selection criteria. The *t*-test was used as a statistical analysis tool to identify whether the difference between classification accuracies by the LSSC and others was significant [43]. The dataset of classification accuracies was constructed for each selection criterion and used in the analysis. Each problem was validated via LOSOCV in the Results section; the data of one of the subjects were used for test data, and the results of all subjects were averaged for each problem. The classification accuracies of all selection criteria were divided by the optimal value for the same tested subject in each problem to normalize the range of the accuracies from zero to one. The normalized classification accuracies from all nine problems were included in the dataset of each selection criterion. The dataset of the LSSC was compared with others using the *t*-test, and *p*-values were calculated (Table 5). When the calculated *p*-value is lower than 0.01, it can be considered that the null hypothesis of two groups is rejected at the confidence level of 99%. The *p*-value was lower than 0.01 in all cases in Table 5, and statistically significant differences could be verified between the classification accuracies by the LSSC and others. This means that the LSSC was clearly more effective than others in the nine problems on physiological data.

Table 5. *T*-test results for classification accuracies obtained by the LSSC and others.

	KSC	KSC2	DBTC	ESDR	XAB	GACV	
<i>p</i> value	$1.2  imes 10^{-5}$	$9.6 \times 10^{-6}$	$6.5 \times 10^{-4}$	$1.9 \times 10^{-4}$	$9.0 \times 10^{-5}$	$2.7 \times 10^{-5}$	

#### 4.2. Relationship between ISV and LSSC

An additional analysis was performed to understand the relationship between the effectiveness of LSSC and the degree of ISV in the data. Hence, two indices, namely *Index*1 and *Index*2, were defined for each classification problem. *Index*1 was used to reflect the

degree of ISV in each problem; the ISV of the data are disclosed when the data are evaluated by LOSOCV because it is a subject-independent test which separates the data of a tested subject from those of the others. Therefore, all data from each classification problem were evaluated using the LOSOCV and five-fold cross validation as the subject-independent and subject-dependent tests, respectively. The classification accuracy for the best SVM model was found for each test. Let us define the accuracy for the subject-independent test as  $ACC_{SIT}$  and that for the subject-dependent test as  $ACC_{SDT}$ ; *Index1* can then be obtained as

$$Index1 = \frac{ACC_{SDT} - ACC_{SIT}}{ACC_{SDT}}$$
(6)

to represent the relative deterioration of the classification accuracy by ISV in the data. *Index2* shows the relative effectiveness of the LSSC compared with the conventional selection criterion, namely KSC. When the classification accuracies by KSC and LSSC (in Table 2) are defined as  $ACC_{KSC}$  and  $ACC_{LSSC}$ , *Index2* can be calculated as

$$Index2 = \frac{ACC_{LSSC} - ACC_{KSC}}{ACC_{KSC}}.$$
(7)

Table 6 lists *Index*1 and *Index*2 values for each of the problems. *Index*1 was observed to be low in problems 3, 5, and 6 because of the small ISV in the data. *Index*2 was also low in the cases of problems 3, 5, and 6. However, *Index*1 was relatively high for problems 2, 7, and 8, and *Index*2 was also high for the same problems. The relationships between the indices can be observed more obviously in the scatter plot of Figure 1. A problem that had a high *Index*1 had a higher *Index*2 than others overall, and the correlation coefficient between *Index*1 and *Index*2 was 0.9862. This indicates that the effectiveness of the LSSC was generally greater for classification problems having higher ISV in the data. Therefore, the LSSC could be considered and used preferentially compared with the others for classification problems having high *Index*1 values that quantified the degree of ISV in the data.

**Table 6.** Index1 and Index2 values for each classification problem.

Problem	1	2	3	4	5	6	7	8	9
Index1	0.103	0.187	0.037	0.063	0.040	0.043	0.248	0.236	0.075
Index2	0.030	0.072	0.017	0.018	0.011	0.008	0.093	0.104	0.019



Figure 1. Scatter plot of *Index*1 vs. *Index*2 for the nine studied problems.

#### 4.3. Comparison with the Results of Previous Studies

Some previous studies compared the effectiveness of selection criteria for various datasets as mentioned in the Introduction [26,27]. The compared selection criteria were not the same with this study, and the MS situation on physiological data was not considered in the previous studies. However, the results of previous studies were analyzed together with the results of this research to identify the difference in the effectiveness of each selection criterion for physiological data and data without ISV. The relative ranking of selection criteria for classification accuracy was investigated in this analysis; the results of ranking for all problems were averaged, and the relative ranking was obtained based on comparisons with the averaged values (Table 7). KSC was the most effective in previous studies, but it is less advantageous than DBTC and GACV for physiological data. DBTC was relatively superior in the previous studies, and it was still ranked high in this study. The effectiveness of DBTC was more robust for ISV, and yielded consistent performance compared with KSC. XAB and GACV were not as good as KSC and DBTC as reported in previous studies, but the ranking of GACV was higher than KSC, and it was the same with DBTC for physiological data. Nevertheless, the variation of its effectiveness was severe according to a classification problem, and the averaged classification accuracy was lower than KSC as indicated in the results in Table 2. Therefore, it is difficult to conclude that GACV is superior to KSC for physiological data, and a more rigorous comparison using big-sized datasets will be needed.

 Table 7. Relative ranking of selection criteria in each study.

Study	KSC	KSC2	DBTC	ESDR	XAB	GACV	LSSC
[26]	1	-	-	-	2	3	-
[27]	1	-	2	-	4	3	-
This paper	4	6	2	5	7	2	1

## 4.4. Contributions of This Study

This study compared and analyzed seven well-known selection criteria for the MS of SVM on physiological data. Readers can utilize the results of this study by considering its contributions outlined below:

- Comparison outcomes of selection criteria on classification accuracy and computational time were presented and analyzed. The results can be used when a selection criterion is selected for a classification problem on physiological data. The LSSC yielded the highest classification accuracy, but its computational time was the longest. DBTC was lower than the LSSC in classification accuracy, but it was associated with a relatively short computational time.
- The effectiveness of LSSC in classification accuracy was verified experimentally. It
  outperformed the classification accuracies of others, and the difference was statistically
  significant. In addition, the advantages were more pronounced for data that had
  larger ISV.
- The change in relative superiority was investigated when selection criteria were used for the data that contained ISV. The relative effectiveness of DBTC was increased more than KSC when the selection criterion was utilized for physiological data compared with data without ISV.

## 5. Conclusions

The ISV is an important constraint that induces difficulties when analyzing physiological data. This study focused on the influence of the ISV in the MS for SVM and compared several well-known selection criteria to identify a selection criterion suitable for physiological data with ISV. Seven selection criteria were simulated for nine classification problems. The results showed that LSSC was the most effective for the classification problems and that the effect was greater according to the degree of ISV in the data. Therefore, the LSSC could be tested preferentially for MS related to physiological data with large ISV. MS is one of the factors that determines the performance of an SVM classifier, and the aforementioned comparison results are expected to be helpful for the construction of a better SVM classifier for classification problems based on physiological data. The advantage of the proposed approach can be utilized in various physiological-data-based applications, including health-care, human–computer interactions, and medical decision systems [44,45]. In the future, we plan to conduct more rigorous verification of these results using large-sized datasets. In addition, the design of a new selection criterion for physiological data is one of our future research objectives.

Author Contributions: Conceptualization, M.C. and J.J.J.; methodology, M.C. and J.J.J.; software, M.C.; validation, M.C. and J.J.J.; writing—original draft preparation, M.C.; writing—review and editing, M.C. and J.J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

#### Informed Consent Statement: Not applicable.

Data Availability Statement: This study constructed nine problems for simulations. Among them, data for problems 1, 2, and 6–9 are available online as follows : [1: https://physionet.org/content/drivedb/1.0.0/] (accessed on 3 December 2021), [2: https://www.kaggle.com/qiriro/swell-heart-rate-variability-hrv] (accessed on 3 December 2021), [6: http://bbci.de/competition/iv/] (accessed on 3 December 2021), [7–8: https://www.kaggle.com/qiriro/comfort] (accessed on 3 December 2021), [9: https://www.kaggle.com/mondejar/mitbih-database] (accessed on 3 December 2021).

Acknowledgments: This work was supported by the National Research Foudation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1G1A1003578).

Conflicts of Interest: The authors declare no conflict of interest.

#### Appendix A. Description on Datasets and Extracted Features

#### Appendix A.1. Problem 1

Healey measured ECG, EDA, EMG, and respiration of drivers and extracted 22 features to detect stress (Table A1) [33,34]. ECG is the signal caused by heart activity, and RR intervals can be obtained from the ECG. The RR intervals contain information on our autonomic nervous system, and the features extracted from RR intervals were used. EDA represents the activity of sweat glands in the skin measured on the hand and foot, and electrical activity from skeletal muscles was recorded as EMG in the research. Problem 1 uses the feature data.

Signal	Features
ECG	Normalized mean and variance of RR intervals, ratio of the low (0–0.08 Hz) to the high-frequency (0.15–0.5 Hz) spectral power for RR intervals
EDA (hand)	Normalized mean and variance, the number of orienting re- sponses, sum of the startle magnitudes, sum of the response duration, sum of the areas under responses
EDA (foot)	Same as EDA (hand)
EMG	Normalized mean
Respiration	Normalized mean and variance, spectral power (0–0.1, 0.1–0.2, 0.2–0.3, and 0.3–0.4 Hz)

Table A1. Extracted features for problem 1.

#### Appendix A.2. Problem 2

Nkurikiyeyezu analyzed 5-minute ECG in the SWELL database and extracted HRV-related features [36]. Table A2 shows the features, and the extracted feature data are used for problem 2.

Table A2. HRV-related features for problem 2.

Classification	Features
Time domain	Mean and median of RR intervals, root mean square of the successive RR interval differences (RMSSD), standard deviation of RR interval differences (SDSD), previous four features for relative RR intervals, standard deviation, skew- ness, and kurtosis of RR intervals, ratio of SDRR over RMSSD, percentage of RR interval differences more than 25 and 50 ms
Spectral feature	Spectral power of RR intervals for very low ( $\leq 0.04$ Hz), low (0.04–0.15 Hz), and high (0.15–0.4 Hz) frequency bands, ratio of the low to the high-frequency spectral power for RR intervals
Nonlinear measure	Two Poincaré plot descriptors

#### Appendix A.3. Problems 3–5

PPG and EDA were obtained using a wearable device in our previous research [9]. PPG is to measure changes in the volume of blood vessels using light sources, and it contains cardiovascular information of our body. Obtained signals were processed every 10 s, and features were extracted to detect stress, fatigue, and drowsiness. Problems 3–5 are based on the feature data (Table A3).

Table A3. Extracted features for problems 3–5.

Problem	Signal	Features
3	PPG	Standard deviation of pulse amplitudes, DC amplitude, standard deviation of amplitude differences between pulses, signal energy, bandwidth
4	PPG	Mean of pulse intervals, normalized power in low- frequency range (0.04–0.15 Hz) of HRV, DC amplitude, mean rise time (time from valley to peak), mean ratio of rise to fall time (time from peak to valley), bandwidth
	EDA	Mean
5	PPG EDA	Mean of pulse intervals, normalized power in low- frequency range of HRV, mean amplitude of pulse ampli- tudes, DC amplitude, mean rise time, signal energy, band- width Mean

#### Appendix A.4. Problem 6

EEG is a physiological signal that is related to brain activity, and it is possible to find out the intention of a person to move one's body by analyzing the EEG. Brain–computer interface competition IV-2b database contains EEG with label information indicating whether a subject imagined movement of the right or left hand when the signal was measured. Problem 6 is based on the database, and a common spatial pattern algorithm was used to extract features from EEG [41]. The algorithm divides EEG into several frequency bands including 4–8, 8–12,..., and 36–40 Hz and uses spatial filters to increase the discriminability of each class.

Appendix A.5. Problems 7 and 8

A dataset containing HRV-based features extracted from ECG is utilized for problem 7 [42]. The features are listed in Table A4, and the features are also used in problem 8.

Table A4. Extracted features for problems 7 and 8.

Classification	Features
Time domain	Mean, median, standard deviation, skewness, and kurtosis of RR intervals, root mean square of the successive RR interval differences (RMSSD), standard deviation of RR interval differences (SDSD), ratio of SDRR over RMSSD, previous eight features for relative RR intervals, percentage of RR interval differences more than 25 and 50 ms, HR, sample entropy of RR intervals
Spectral features	Spectral power of RR intervals for very low ( $\leq 0.04$ Hz), low (0.04–0.15 Hz), high (0.15–0.4 Hz), and whole frequency bands, ratio of the low to the high-frequency spectral power for RR intervals, ratio of the high to the low-frequency spectral power for RR intervals
Nonlinear measure	Higuchi fractal dimension, two Poincaré plot descriptors

#### Appendix A.6. Problem 9

ECG can be used to identify an abnormality of heart. VEB is one of the abnormal cases and contains premature ventricular contraction and ventricular escape beats. Each heartbeat of ECG is tested, and VEBs are detected in problem 9 using 45 features (Table A5) [40].

Table A5. Extracted features for problem 9.

Classification	Features
RR interval	Distance between current and previous beats, distance be- tween current and next beats, average of previous ten-RR in- tervals, average of previous RR intervals in the last 20 min, normalized version of previous four features
Higher order statistics	Kurtosis and skewness for five segments of each beat
Wavelet transform	23-dimensional descriptors using Daubechies wavelet func- tion with three levels of decomposition
Morphological descriptor	Maximum values for the first and last segments and mini- mum values for the second and third segments when each beat is divided into four segments

#### References

- 1. Lee, J.S.; Lee, S.J.; Choi, M.; Seo, M.; Kim, S.W. QRS detection method based on fully convolutional networks for capacitive electrocardiogram. *Expert Syst. Appl.* **2019**, *134*, 66–78. [CrossRef]
- Lynn, H.M.; Kim, P.; Pan, S.B. Data Independent Acquisition Based Bi-Directional Deep Networks for Biometric ECG Authentication. *Appl. Sci.* 2021, 11, 1125. [CrossRef]
- Tirado-Martin, P.; Sanchez-Reillo, R. BioECG: Improving ECG Biometrics with Deep Learning and Enhanced Datasets. *Appl. Sci.* 2021, 11, 5880. [CrossRef]
- Zhang, P.; Li, F.; Zhao, R.; Zhou, R.; Du, L.; Zhao, Z.; Chen, X.; Fang, Z. Real-Time Psychological Stress Detection According to ECG Using Deep Learning. *Appl. Sci.* 2021, *11*, 3838. [CrossRef]

- Ayoobi, N.; Sharifrazi, D.; Alizadehsani, R.; Shoeibi, A.; Gorriz, J.M.; Moosaei, H.; Khosravi, A.; Nahavandi, S.; Chofreh, A.G.; Goni, F.A.; et al. Time Series Forecasting of New Cases and New Deaths Rate for COVID-19 using Deep Learning Methods. *arXiv* 2021, arXiv:2104.15007.
- Moosaei, H.; Ketabchi, S.; Razzaghi, M.; Tanveer, M. Generalized Twin Support Vector Machines. *Neural Process. Lett.* 2021, 53, 1545–1564. [CrossRef]
- Mangasarian, O.L. Data mining via support vector machines. In Proceedings of the IFIP Conference on System Modeling and Optimization, Trier, Germany, 23–27 July 2001; pp. 91–112.
- Lee, Y.J.; Mangasarian, O.L. SSVM: A smooth support vector machine for classification. *Comput. Optim. Appl.* 2001, 20, 5–22. [CrossRef]
- 9. Choi, M.; Koo, G.; Seo, M.; Kim, S.W. Wearable Device-Based System to Monitor a Driver's Stress, Fatigue, and Drowsiness. *IEEE Trans. Instrum. Meas.* 2018, 67, 634–645. [CrossRef]
- 10. Ortega, S.; Fabelo, H.; Halicek, M.; Camacho, R.; Plaza, M.d.I.L.; Callicó, G.M.; Fei, B. Hyperspectral superpixel-wise glioblastoma tumor detection in histological samples. *Appl. Sci.* 2020, *10*, 4448. [CrossRef]
- Setiowati, S.; Franita, E.L.; Ardiyanto, I. A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods. In Proceedings of the 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Phuket, Thailand, 12–13 October 2017; pp. 1–6.
- 12. Pandit, R.; Kolios, A. SCADA data-based support vector machine wind turbine power curve uncertainty estimation and its comparative studies. *Appl. Sci.* **2020**, *10*, 8685. [CrossRef]
- 13. Rizwan, A.; Iqbal, N.; Ahmad, R.; Kim, D.H. WR-SVM Model Based on the Margin Radius Approach for Solving the Minimum Enclosing Ball Problem in Support Vector Machine Classification. *Appl. Sci.* **2021**, *11*, 4657. [CrossRef]
- 14. Ayat, N.E.; Cheriet, M.; Suen, C.Y. Automatic model selection for the optimization of SVM kernels. *Pattern Recogn.* 2005, 38, 1733–1745. [CrossRef]
- 15. Adankon, M.M.; Cheriet, M. Optimizing resources in model selection for support vector machine. *Pattern Recogn.* 2007, 40, 953–963. [CrossRef]
- 16. Zhang, X.; Qiu, D.; Chen, F. Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis. *Neurocomputing* **2015**, *149*, 641–651. [CrossRef]
- 17. Kapp, M.N.; Sabourin, R.; Maupin, P. A dynamic model selection strategy for support vector machine classifiers. *Appl. Soft Comput.* 2012, 12, 2550–2565. [CrossRef]
- Li, W.; Liu, L.; Gong, W. Multi-objective uniform design as a SVM model selection tool for face recognition. *Expert Syst. Appl.* 2011, *38*, 6689–6695. [CrossRef]
- 19. Huang, C.M.; Lee, Y.J.; Lin, D.K.; Huang, S.Y. Model selection for support vector machines via uniform design. *Comput. Stat. Data Anal.* **2007**, *52*, 335–346. [CrossRef]
- Wu, C.H.; Tzeng, G.H.; Goo, Y.J.; Fang, W.C. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Syst. Appl.* 2007, 32, 397–408. [CrossRef]
- Namdeo, A.; Singh, D. Challenges in evolutionary algorithm to find optimal parameters of SVM: A review. *Mater. Today-Proc.* 2021. [CrossRef]
- 22. Vapnik, V.; Chapelle, O. Bounds on error expectation for support vector machines. Neural Comput. 2000, 12, 2013–2036. [CrossRef]
- Anguita, D.; Ridella, S.; Rivieccio, F.; Zunino, R. Hyperparameter design criteria for support vector classifiers. *Neurocomputing* 2003, 55, 109–134. [CrossRef]
- 24. Sun, J.; Zheng, C.; Li, X.; Zhou, Y. Analysis of the distance between two classes for tuning SVM hyperparameters. *IEEE Trans. Neural Netw.* **2010**, *21*, 305–318. [PubMed]
- Yin, S.; Yin, J. Tuning kernel parameters for SVM based on expected square distance ratio. *Inform. Sci.* 2016, 370, 92–102. [CrossRef]
- Duan, K.; Keerthi, S.S.; Poo, A.N. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing* 2003, *51*, 41–59. [CrossRef]
- 27. Duarte, E.; Wainer, J. Empirical comparison of cross-validation and internal metrics for tuning SVM hyperparameters. *Pattern Recogn. Lett.* **2017**, *88*, 6–11. [CrossRef]
- Choi, M.; Seo, M.; Lee, J.S.; Kim, S.W. Fuzzy support vector machine-based personalizing method to address the inter-subject variance problem of physiological signals in a driver monitoring system. *Artif. Intell. Med.* 2020, 105, 101843. [CrossRef]
- 29. Gholamiangonabadi, D.; Kiselov, N.; Grolinger, K. Deep Neural Networks for Human Activity Recognition with Wearable Sensors: Leave-one-subject-out Cross-validation for Model Selection. *IEEE Access* **2020**, *8*, 133982–133994. [CrossRef]
- Rojas-Domínguez, A.; Padierna, L.C.; Valadez, J.M.C.; Puga-Soberanes, H.J.; Fraire, H.J. Optimal hyper-parameter tuning of SVM classifiers with application to medical diagnosis. *IEEE Access* 2017, *6*, 7164–7176. [CrossRef]
- 31. Kumar, S. Neural Networks: A Classroom Approach; Tata McGraw-Hill Education: New York, NY, USA, 2004.
- Diosan, L.; Rogozan, A.; Pecuchet, J.P. Improving classification performance of support vector machine by genetically optimising kernel shape and hyper-parameters. *Appl. Intell.* 2012, *36*, 280–294. [CrossRef]
- Healey, J.A.; Picard, R.W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* 2005, 6, 156–166. [CrossRef]
- Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. Physiobank, physiotoolkit, and physionet. *Circulation* 2000, 101, e215–e220. [CrossRef] [PubMed]

- Koldijk, S.; Sappelli, M.; Verberne, S.; Neerincx, M.A.; Kraaij, W. The swell knowledge work dataset for stress and user modeling research. In Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, 12–16 November 2014; pp. 291–298.
- Nkurikiyeyezu, K.; Yokokubo, A.; Lopez, G. The Effect of Person-Specific Biometrics in Improving Generic Stress Predictive Models. arXiv 2019, arXiv:1910.01770.
- Leeb, R.; Brunner, C.; Müller-Putz, G.; Schlögl, A.; Pfurtscheller, G. BCI Competition 2008–Graz Data Set B; Graz University of Technology: Graz, Austria, 2008.
- 38. Lopez, G.; Kawahara, Y.; Suzuki, Y.; Takahashi, M.; Takahashi, H.; Wada, M. Effect of direct neck cooling on psychological and physiological state in summer heat environment. *Mech. Eng. J.* **2016**, *3*, 15-00537. [CrossRef]
- 39. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH arrhythmia database. IEEE Eng. Med. Biol. 2001, 20, 45–50. [CrossRef]
- Mondéjar-Guerra, V.; Novo, J.; Rouco, J.; Penedo, M.G.; Ortega, M. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomed. Signal Proces.* 2019, 47, 41–48. [CrossRef]
- Ang, K.K.; Chin, Z.Y.; Wang, C.; Guan, C.; Zhang, H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* 2012, 6, 39. [CrossRef]
- Nkurikiyeyezu, K.; Yokokubo, A.; Lopez, G. Affect-aware thermal comfort provision in intelligent buildings. In Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, UK, 3–6 September 2019; pp. 331–336.
- 43. Son, Y.; Kim, W. Missing Value Imputation in Stature Estimation by Learning Algorithms Using Anthropometric Data: A Comparative Study. *Appl. Sci.* **2020**, *10*, 5020. [CrossRef]
- Le, N.Q.K.; Hung, T.N.K.; Do, D.T.; Lam, L.H.T.; Dang, L.H.; Huynh, T.T. Radiomics-based machine learning model for efficiently classifying transcriptome subtypes in glioblastoma patients from MRI. *Comput. Biol. Med.* 2021, 132, 104320. [CrossRef]
- 45. Do, D.T.; Le, N.Q.K. Using extreme gradient boosting to identify origin of replication in Saccharomyces cerevisiae via hybrid features. *Genomics* **2020**, *112*, 2445–2451. [CrossRef]