Jurnal Teknik Informatika Vol. 17 No. 1, April 2024 (10-20)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v17i1.37422

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

SVM Optimization with Grid Search Cross Validation for Improving Accuracy of Schizophrenia Classification Based on EEG Signal

Masdar Desiawan[1*] and Achmad Solichin[2]

[1,2]Department of Computer Science, Faculty of Information Technology, Budi Luhur University.
[1,2]Ciledug Raya Road, North Petukangan, South Jakarta, DKI Jakarta, 12260, Indonesia

Article:
Accepted: March 23, 2024
Revised: January 12, 2024
Issued: April 25, 2024

***Correspondence Address:**
masdardesiawan@gmail.com

ABSTRACT

The advantage of the Support Vector Machine (SVM) is that it can solve classification and regression problems both linearly and non-linearly. SVM also has high accuracy and a relatively low error rate. However, SVM also has weaknesses, namely the difficulty of determining optimal parameter values, even though setting exact parameter values affects the accuracy of SVM classification. Therefore, to overcome the weaknesses of SVM, optimizing and finding optimal parameter values is necessary. The aim of this research is SVM optimization to find optimal parameter values using the Grid Search Cross-Validation method to increase accuracy in schizophrenia classification. Experiments show that optimization parameters always find a nearly optimal combination of parameters within a specific range. The results of this study show that the level of accuracy obtained by SVM with the grid search cross-validation method in the schizophrenia classification increased by 9.5% with the best parameters, namely C = 1000, gamma = scale, and kernel = RBF, the best parameters were applied to the SVM algorithm and obtained an accuracy of 99.75%, previously without optimizing the accuracy reached 90.25%. The optimal parameters of the SVM obtained by the grid search cross-validation method with a high degree of accuracy can be used as a model to overcome the classification of schizophrenia.

Keywords : *Optimization; support vector machin; grid search cross-validation; classification; schizophrenia;*

1. INTRODUCTION

Schizophrenia affects approximately 24 million people, i.e., one in 300 people (0.32%) world wide [1]. In addition, the underlying disease mechanisms are still poorly understood [2]. Misdiagnosis of schizophrenia occurs because the symptoms of schizophrenia are similar to other mental disorders, such as bipolar disorder. The many symptoms of schizophrenia complicate the diagnosis and classification of schizophrenia disorders and can delay treatment [3]. Traditional or manual classification methods, the diagnosis process is quite tricky because schizophrenic patients do not have specific features in appearance and behavior. Some of them can look and behave like ordinary people [4]. Therefore, classifying schizophrenia using machine learning is essential to obtain objective and accurate information [5]. Classification is the activity or process of grouping data into specific categories based on the characteristics and patterns of these categories. The classification process is applied when providing information as a reference in developing information classification rules [6]. Classification can be used to find a set of patterns or features that separate classes or classes of data, with the aim that these patterns can be used to predict the target class accurately [7]. Previous research conducted for schizophrenia classification using machine learning methods includes Artificial Neural Networks (ANNs) [8], Support Vector Machine (SVM) [9], Decision Tree C5.0 [3], Ensemble [10], and Deep Neural Networks [11]. Researchers in the world widely use these methods. Artificial Neural Networks (ANN) can model and understand very complex patterns in data, even when those patterns are challenging to define explicitly. Still, they also have the disadvantage of being prone to overfitting and requiring more excellent computational resources [12].

Mean while, the advantage of Support Vector Machine (SVM) is that it can solve both linear and non-linear classification and regression problems. SVM also has high accuracy and a relatively low error rate [13]. In theory, this algorithm examines the chromosome value of the sample and then finds the best hyperplane to classify the collected data [14]. In addition to having advantages, SVM has a disadvantage, namely the difficulty of determining the optimal parameter value [15].

Several studies have been done before related to optimization in SVM, including research on air quality prediction by optimizing SVM using grid search cross-validation [16]. Furthermore, research on SVM optimization using the firefly algorithm on film opinion data [17]. Research on optimization of support vector machine (SVM) parameters with particle swarm optimization (PSO) for blood donor classification with RFMTC dataset [18]. Research titled Optimizing Support Vector Machine Parameters Using Bat Optimization Algorithm [19]. Another study entitled Engineering Applications of Artificial Intelligence Optimization of support vector machine uses metaheuristic algorithms to forecast TBM advance rate [20]. Based on all the research on optimization above, the optimization of SVM is done because SVM has a weakness in finding optimal parameters. On that basis, hyperparameter tuning is carried out. After this is done, the optimal parameters will be obtained and used to produce a high level of accuracy. On average, previous research shows a high level of accuracy after optimization. Therefore, from various optimization techniques for SVM, this study proposes to apply optimization with the Grid Search Cross Validation method as a solution to the weaknesses of SVM with the object of research is the classification of schizophrenia.

2. METHODS

This research uses a quantitative approach to solve a problem by performing mathematical calculations based on test and training data. According to [21], quantitative research tests objective theories by examining the relationship between variables. These variables can usually be measured instrumentally, so numbered data can be analyzed using statistical methods. This section describes the materials and methods used, including the schizophrenia dataset, data preprocessing, classification model, SVM, model evaluation, and optimization with grid search cross-validation.

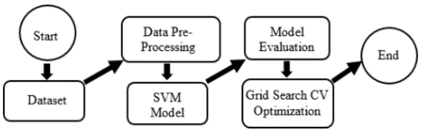


Figure 1. *Research method*

2.1. Dataset

This study used an EEG dataset of basic sensory tasks in schizophrenia from the last Kaggle update in 2019 and retrieved in 2022 [22]. The dataset consisted of 81 subjects, 32 normal subjects, and 49 schizophrenic subjects. The age group of the subjects was 16-63 years old and consisted of 67 males and 14 females. EEG signal data was obtained from each subject by attaching electrode wires to the subject's scalp. Electrodes were placed at 64 points (channels) including Fz N100, FCz N100, Cz N100, FC3 N100, FC4 N100, C3 N100, C4 N100, CP3 N100, CP4 N100, Fz P200, FCz P200, FCz. P200, C3 P200, C4 P200, CP3 P200, CP4 P200, Fz B0, FCz B0, Cz B0, FC3 B0, FC4 B0, C3 B0, C4 B0, CP3 B0, CP4 B0, Fz B1, Cz FC4,1 B1, C3B1, C4B1, CP3B1, CP4B1. In the EEG experiment, the subjects were in three states: the first subject was asked to press a button to produce a sound, the second subject passively listened to the same or repetitive tone, and the third subject was asked to press the button at the same time with a speed at a specific interval within a certain period. Each condition recorded EEG signal data consisting of 3072 attributes. Since there are three conditions and 81 objects, the size of the EEG signal data is 243 × 3072.

2.2. Preprocessing

Preprocessing is done because it is possible that the data still cannot be used in the classification model, so data processing, such as normalization and data cleaning, is needed. Tools used in data processing include Jupyter Notebook and Microsoft Excel. The following are the stages in preparing the data:

a. Data Selection: data selection is made by selecting attributes influencing the classification process.
b. Data Cleaning: data cleaning is done on data with no value (*null*),irrelevant data, and inconsistent data.
c. Data Transformation: data transformation is carried out on data that isnominal to numerical values.

2.3. Support Vector Machine

Support Vector Machine (SVM) is a machine learning method that works on the structural risk minimization principle and aims to find the best hyperplane separating two classes in the input space [23]. The simple idea behind SVM is to maximize the margin, which is the distance between data classes. Below is a graphical representation of the difference between two data classes in a linear SVM with a maximum margin. graphical representation of the difference between two data classes in a linear SVM with a maximum margin.

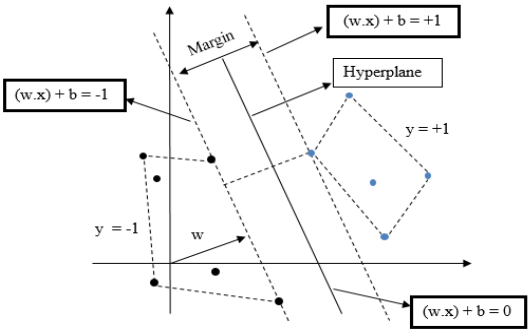


Figure 2. *Separation of two classes of data* [24]

In the above image, the hyperplane is found by measuring the edges and finding the maximum point. The margin is the distance between the hyperplane and the closest pattern in each class. The closest pattern is called the support vector. The solid line in the figure above, which lies between two classes, indicates the best hyperplane, while the black and blue dots are the support vectors. Finding the hyperplane is the core of the support vector machine (SVM) learning process [25]. Data is notated as $xi= \{x1, x2 , \dots , xn\} \in Rd$ while the label or class of the data is notated $yi \in \{-1, +1\}$ for i = 1, 2, … n, where n is the number of data. It is assumed that the two classes -1 and +1 can be perfectly separated by a d-dimensional hyperplane defined by Equation 1.

$$(w.x) + b = 0 \tag{1}$$

Description:

w = weight vectorx = data

b = bias value

The weight vector (w) is a vector line perpendicular to the coordinate center and the hyperplane line. Bias (b) is the coordinate of the line relative to the coordinate point. These two parameters are parameters whose values must be determined to get the best function to combine the input and output data. The pattern x_i belonging to class -1 can be formulated as a pattern that satisfies the inequality as in Section 2.

$$w.x_i + b \ \leq \ -1 \qquad (2)$$

Mean whileMeanwhileern xi, which belongs to class +1, satisfies inequality 3.

$$w.x_i + b \ \leq +1 \qquad (3)$$

Equation 4 is the equation used to calculate the value of b, while equation 5 calculates the value of w.

$$b = -\frac{1}{2}\ (w.x^{+} + w.x^{-}) \qquad (4)$$

$$w = \textstyle\sum_{i}^{n} \alpha_i y_i \, x_i \qquad (5)$$

Description:

b = bias value
w.x+ = weight value for the positive data class
w.x- = weight value for the negative data class
w = weight vector
αi = data weight value i
yi = data class i
xi = data i

The most significant margin can be found by maximizing the distance value between the hyperplane and its closest point. Equation 6 is the equation to calculate the margin.

$$\text{Margin} = \frac{2}{\|\text{w}\|} = \frac{2}{\sqrt{w_1^{\ 2} + w_2^{\ 2}}} \qquad (6)$$

Then, to determine the optimal hyperplane for both classes using the following equation:

$$\min \ J_1[w] = \frac{1}{2}\|\text{w}\|^2 \qquad (7)$$

subject to:

$$y_i(x_i.w + b) - 1 \geq 0, i = 1, \dots, n \qquad (8)$$

Support Vector Machine (SVM) is a supervised learning model with a learning algorithm that analyzes data for classification and regression analysis. It was developed at AT&T Bell Laboratories by Vladimir Vapnik and colleagues and was first presented in 1992 at the Annual Workshop on Computational Learning Theory [26]. The basic concept of SVM is one of the most potent prediction methods based on the statistical learning framework or VC theory proposed by Vapnik in 1982. Given a set of training examples, each characterized as belonging to one of two categories, the SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist for using SVMs in a probabilistic classification setting). The SVM maps the training examplesto points in the space to maximize the width of the gap between the two categories, then maps into the same spaceand predicts based on which side of the gap they fall on.

2.4. Model Evaluation

SVM model evaluation uses a confusion matrix to measure accuracy, precision, recall, and f1-score performance metrics. The confusion matrix is a method used to evaluate the performance of classification methods. The evaluation is done to determine the accuracy of the previous test results. The confusion matrix has four terms that are known as a representation of the classification results.

Table 1. *Confusion matrix*

Class	Positive Classificatifon	Negative Classification
+	True Positive (TP)	False Negative (FN)
-	False Positive (FP)	True Negative (TN)

True positive (TP) indicates the amount of positive data classified as valid. True negative (TN) indicates the amount of negative data classified as valid. False positive (FP) indicates the amount of negative data classified as positive. False negative (FN) indicates the number of positives but is classified as negative. The accuracy value is a value that describes how accurate the classification is and is the percentage of correctly classified data in the total data. The accuracy value is obtained from equation 9.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \qquad (9)$$

Precision is a positive ratio or degree of reliability where the number of correctly classified positive category data is divided by the total positive classified data. So equation 10 can be obtained.

$$Precision = \frac{TP}{TP+FP} * 100\% \qquad (10)$$

Recall is the percentage of positive category data correctly classified by the system. The recall value can be obtained with equation 11.

$$Recall = \frac{TP}{TP+FN} * 100\% \qquad (11)$$

F1-score summarizes all the results of precision and recall calculations. The f1-score value can be obtained with equation 12.

$$F1 - Score = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} \qquad (12)$$

2.5. Grid Search Cross Validation

Grid Search optimizes SVM parameters using cross-validation techniques such as performance metrics. The goal is to identify good hyperparameter combinations so that the *classifier* can accurately predict unknown data. The goal of *grid search* is to determine the combination that gives the best model performance that can be selected as the prediction model [27]. *Grid search is* usually combined with K-fold Cross-Validation (K-CV), creating an evaluation index for classification models [28].

The k-fold cross-validation procedure divides the trainingset into K equal-sized parts. Each part of the training set is considered a validation set, and the remaining part, K1, is considered a new one. Then, K models will be formed, trained, and validated with the K1 training and K validation sets.

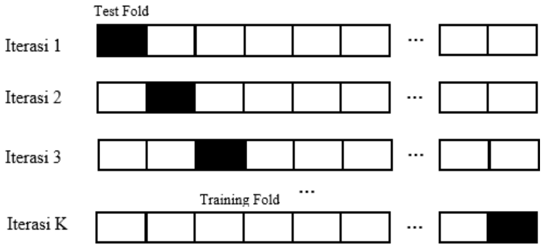


Figure 3. *K-CV split procedure of training and test data* [28]

K-fold cross-validation can repeat the training data and *testing* data for k repetitions and divide 1/k of the dataset usedas testing data [29]. The k-model accuracy can be obtained, and the performance of the k-fold cross-validation classification model is evaluated based on the average k-model accuracy. Then, based on the grid search, the classifier's parameters are changed, and the accuracy of the classifier is recalculated.

3. RESULTS AND DISCUSSION

3.1. Data Preprocessing

In this section, steps are taken using the methodology used to obtain results by the research objectives. The first is to determine the attributes used and the data, commonly called data selection. Several selection techniques are used for data selection, including correlation matrix and weighting values, using Rapid Miner tools with correlation feature operators. Based on the results of the two techniques, it was decided that the attributes used were as in Table 2.

Table 2. *Attributes*

#	Column Name	Description
1	Subject	Subject identifiers can be used to join with other data files
2	Trial	Numeric trial identifiers (in order of occurrence for a given subject and condition)
3	Condition	Numeric code for the condition, where 1 =press button + tone, 2 = playback tone, 3 = Press button
4	Group	Grouping ofThe groupingpants value 0 if valuesarticipant is control andin one of the participants is schizophrenic as *class* l, which is al attribute.
5	Gender	Gender is symbolized as M (Male) and F (Female), M is Male, and F is Female.
6	Age	Age in years
7	Education	Education in years
8	Fz_N100	Amplitude N100 for electrode Fz, where N100 is the average amplitude between 80-100 ms (inclusive)
9	FCz_N100	N100 amplitude for FCz electrode
10	Cz_N100	N100 amplitude for Cz electrode
11	FC3_N100	N100 amplitude for FC3 electrode
12	FC4_N100	N100 amplitude for FC4 electrode
13	C3_N100	Amplitude N100 for electrode C3
14	C4_N100	Amplitude N100 for electrode C4
15	CP3_N100	N100 amplitude for CP3 electrode
16	CP4_N100	N100 amplitude for CP4 electrode

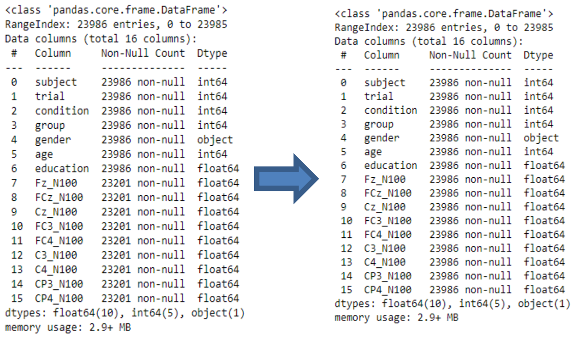
After data selection, data cleaning was carried out because there were data containing missing values. To solve this problem, several approaches are used, one of which is to replace the missing data with the average value of the available historical data [30].

In Table 3, the electrode channels used in this study all have the same missing values, from channel Fz_N100 to CP4_N100, each with a total number of missing values of 785 and a percentage of 3.3%.

Table 3. *Missing value*

Attributes	Total Zero MissingValues	% Total Zero MissingValues
Fz_N100	785	3,3
FCz_N100	785	3,3
Cz_N100	785	3,3
FC3_N100	785	3,3
FC4_N100	785	3,3
C3_N100	785	3,3
C4_N100	785	3,3
CP3_N100	785	3,3
CP4_N100	785	3,3

After knowing the number and percentage of missing values, the data was cleaned by replacing the value with the average value. The results can be seen in Figure 4.

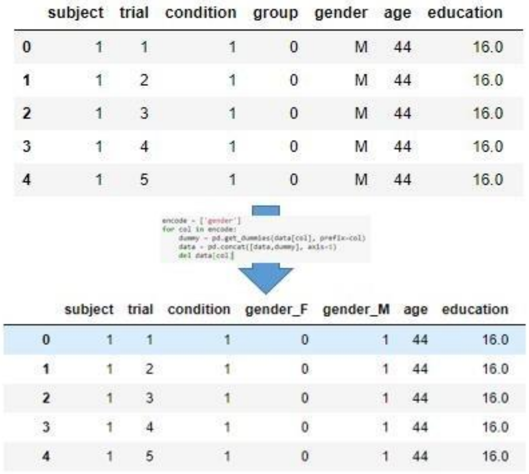


```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23986 entries, 0 to 23985
Data columns (total 16 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   subject    23986 non-null  int64  
 1   trial      23986 non-null  int64  
 2   condition  23986 non-null  int64  
 3   group      23986 non-null  int64  
 4   gender     23986 non-null  object 
 5   age        23986 non-null  int64  
 6   education  23986 non-null  float64
 7   Fz_N100    23201 non-null  float64
 8   FCz_N100   23201 non-null  float64
 9   Cz_N100    23201 non-null  float64
 10  FC3_N100   23201 non-null  float64
 11  FC4_N100   23201 non-null  float64
 12  C3_N100    23201 non-null  float64
 13  C4_N100    23201 non-null  float64
 14  CP3_N100   23201 non-null  float64
 15  CP4_N100   23201 non-null  float64
dtypes: float64(10), int64(5), object(1)
memory usage: 2.9+ MB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23986 entries, 0 to 23985
Data columns (total 16 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   subject    23986 non-null  int64  
 1   trial      23986 non-null  int64  
 2   condition  23986 non-null  int64  
 3   group      23986 non-null  int64  
 4   gender     23986 non-null  object 
 5   age        23986 non-null  int64  
 6   education  23986 non-null  float64
 7   Fz_N100    23986 non-null  float64
 8   FCz_N100   23986 non-null  float64
 9   Cz_N100    23986 non-null  float64
 10  FC3_N100   23986 non-null  float64
 11  FC4_N100   23986 non-null  float64
 12  C3_N100    23986 non-null  float64
 13  C4_N100    23986 non-null  float64
 14  CP3_N100   23986 non-null  float64
 15  CP4_N100   23986 non-null  float64
dtypes: float64(10), int64(5), object(1)
memory usage: 2.9+ MB
```

Figure 4. *Data changes after filling in the average value*

Next is data transformation, which is changing object-type data to numeric data. In the dataset used, the gender attribute is changed to numeric data because the gender attribute is still object data. The purpose of changing object data to numeric is to improve performance during modeling. The results can be seen in Figure 5.



	subject	trial	condition	group	gender	age	education
0	1	1	1	0	M	44	16.0
1	1	2	1	0	M	44	16.0
2	1	3	1	0	M	44	16.0
3	1	4	1	0	M	44	16.0
4	1	5	1	0	M	44	16.0

```
encode = ['gender']
for col in encode:
    dummy = pd.get_dummies(data[col], prefix=col)
    data = pd.concat([data,dummy], axis=1)
    del data[col]
```

	subject	trial	condition	gender_F	gender_M	age	education
0	1	1	1	0	1	44	16.0
1	1	2	1	0	1	44	16.0
2	1	3	1	0	1	44	16.0
3	1	4	1	0	1	44	16.0
4	1	5	1	0	1	44	16.0

Figure 5. *Data transformation process*

3.2. SVM Model

After data preprocessing, the next step is to experiment using the SVM algorithm without optimization, dividing the data into training data and testing datausing the holdout validation method. Data division comprises as many as four ratios: 90%:10%, 75%:25%, 70%:30%, and 65%:35%. The total data is 23986. As in Table 4.

Table 4. *Data sharing ratio (hold out validation)*

Ratio	Training Data	Data Testing
1	90% = 21587	10% = 2399
2	75% = 17989	25% = 5997
3	70% = 16790	30% = 7196
4	65% = 15590	35% = 8396

Next, test each ratio that has been made. The first ratio is 90%: 10%, the implementation results of SVM testing on the schizophrenia classification model show an accuracy of 90%. As in Table 5.

Table 5. *SVM classification results with a ratio of 90%:10%*

	precision	recall	f1-score	support
0	0.74	1.00	0.85	677
1	1.00	0.86	0.93	1722
Accuracy			0.90	2399
avg macros	0.87	0.93	0.89	2399
weighted avg	0.93	0.90	0.91	2399

The implementation of the SVM algorithm with a ratio of 75%:25% shows the accuracy in rounding is 90%. As in Table 6.

Table 6. *SVM classification results with a ratio of 75%:25%*

	precision	recall	f1-score	support
0	0.74	1.00	0.85	1767
1	1.00	0.85	0.92	4230
Accuracy			0.90	5997
avg macros	0.87	0.93	0.88	5997
weighted avg	0.92	0.90	0.90	5997

The implementation of the SVM algorithm with a ratio of 70%:30% shows the accuracy in rounding is 90%. As in Table 7.

Table 7. *SVM classification results with rasio 70% : 30%*

	precision	recall	f1-score	support
0	0.74	1.00	0.85	2111
1	1.00	0.86	0.92	5085
Accuracy			0.90	7196
avg macros	0.87	0.93	0.89	7196
weighted avg	0.92	0.90	0.90	7196

The implementation of the SVM algorithm with a ratio of 65%:35% shows an accuracy of 89%. As in Table 8.

Table 8. *SVM classification results with ratio 65%: 35%*

	precision	recall	f1-score	support
0	0.73	1.00	0.85	2438
1	1.00	0.85	0.92	5958
Accuracy			0.89	8396
avg macros	0.87	0.93	0.88	8396
weighted avg	0.92	0.90	0.90	8396

So, it can be concluded that the overall experiment above shows the accuracy value in rounding from the first to the third ratio reaches 90% accuracy, and the fourth ratio is 89%. Therefore, model evaluation is needed to see the difference in accuracy value in detail.

3.3. Evaluation Model

The model evaluation results obtained accuracy, precision, recall, and F1-score values based on experiments with four *training* and *testing* data-sharing ratios. The results are as in Table 9.

Table 9. *SVM results before optimization*

Ratio	Accuracy	Precision	Recall	F1-Score
90:10	90.25%	100%	86.41%	92.71%
75:25	89.51%	99.97%	85.20%	91.97%
70:30	89.72%	99.93%	85.50%	92.16%
65:35	89.50%	99.94%	85.21%	91.99%

The accuracy results of the SVM algorithm without optimization of each data division ratio obtained the highest accuracy value at the first data division ratio (90:10), whichis 90.25%.

3.4. Grid Search Cross Validation Optimization.

Based on Table 9, the highest accuracy value is 90.25%. This accuracy can still be optimized to get the best accuracy value. The grid search cv method finds parameter values that will give the best performance model. The first optimization process is to define the parameter values of C, gamma, degree, and kernel (RBF, poly, linear). Therange of C value is 0.1, 1, 10, 100, 1000 for all kernels, gamma (auto, scale) is used in the RBF kernel, and degree is the valueof degree 2 and 3 in poly kernel. The selection of values for each parameter is based on theory and previous research. TheCV value is set by default to 5, meaning that five times, the combination of models and parameters will be validated by randomly dividing the data by five equal parts. Then, from this process, the best estimator can be seen based on the mean_test_score, as shown in Table 10.

Table 10. *Best estimator based on mean test score*

Kernel	Params	Rank	Mean Test Score	Std Test Score
1000_scale_rbf	{'C': 1000, 'gamma': 'scale', 'kernel': 'RBF'}	1	0.999573	0.000154
100_scale_rbf	{'C': 100, 'gamma': 'scale', 'kernel': 'RBF'}	2	0.997258	0.000537
10_scale_rbf	{'C': 10, 'gamma': 'scale', 'kernel': 'RBF'}	3	0.994851	0.000794
0.1_auto_rbf	{'C': 0.1, 'gamma': 'auto', 'kernel': 'RBF'}	4	0.966075	0.001846
10_auto_rbf	{'C': 10, 'gamma': 'auto', 'kernel': 'RBF'}	5	0.966041	0.001901
100_auto_rbf	{'C': 100, 'gamma': 'auto', 'kernel': 'RBF'}	5	0.966041	0.001901
1000_auto_rbf	{'C': 1000, 'gamma': 'auto', 'kernel': 'RBF'}	5	0.966041	0.001901
1_auto_rbf	{'C': 1, 'gamma': 'auto', 'kernel': 'rbf'}	8	0.65946	0.001833
1000_3_poly	{'C': 1000, 'degree': 3, 'kernel': 'poly'}	9	0.961339	0.003444
1_scale_rbf	{'C': 1, 'gamma': 'scale', 'kernel': 'rbf'}	10	0.956110	0.003330
100_3_poly	{'C': 100, 'degree': 3, 'kernel': 'poly'}	11	0.945291	0.004304
1000_2_poly	{'C': 1000, 'degree': 2, 'kernel': 'poly'}	12	0.933742	0.003508

Table 10 continued…

Kernel	Params	Rank	Mean Test Score	Std Test Score
10_3_poly	{'C': 10, 'degree': 3, 'kernel': 'poly'}	13	0.923341	0.005253
100_2_poly	{'C': 100, 'degree': 2, 'kernel': 'poly'}	14	0.92203	0.003658
10_2_poly	{'C': 10, 'degree': 2, 'kernel': 'poly'}	15	0.909599	0.005205
1_2_poly	{'C': 1, 'degree': 2, 'kernel': 'poly'}	16	0.904910	0.006115
1000_linear	{'C': 1000, 'kernel': 'linear'}	17	0.899169	0.006309
100_linear	{'C': 100, 'kernel': 'linear'}	18	0.899032	0.006223
1_3_poly	{'C': 1, 'degree': 3, 'kernel': 'poly'}	19	0.894257	0.006035
10_linear	{'C': 10, 'kernel': 'linear'}	20	0.892982	0.007286
1_linear	{'C': 1, 'kernel': 'linear'}	21	0.891750	0.006999
0.1_linear	{'C': 0.1, 'kernel': 'linear'}	22	0.891689	0.006906
0.1_scale_rbf	{'C': 0.1, 'gamma': 'scale', 'kernel': 'RBF'}	23	0.886418	0.006372
0.1_2_poly	{'C': 0.1, 'degree': 2, 'kernel': 'poly'}	24	0.880029	0.007634
0.1_3_poly	{'C': 0.1, 'degree': 3, 'kernel': 'poly'}	25	0.879869	0.007701

From the Table 10, it can be seen that ,after tuning, thebest kernel is RBF and the best parameters {'C': 1000, 'gamma':'scale', 'kernel': 'RBF'} with *mean test score* 0.999573 and *std test score* 0.000154. The *poly* kernel with degree 3 and C=0.1 shows the lowest performance of all models. Specifically, to distinguish the performance of the models, the AUC score can be compared statistically. All models were evaluated on the same partition with 25 candidates and five folds of 125 partitions, thus increasing the correlation between modelperformance. Furthermore, to see the effect of partitioning, we plot the performance of all models at each fold and calculatethe correlation between models across folds.

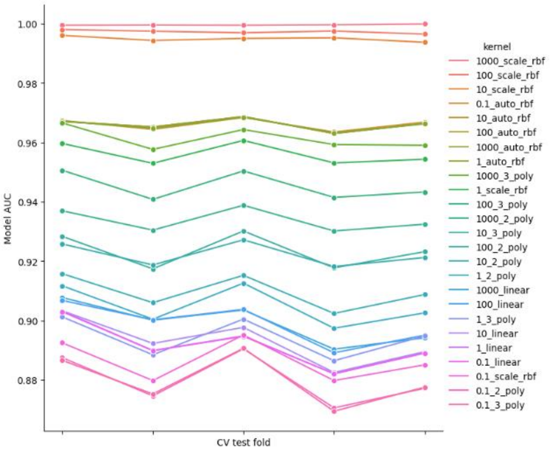


Figure 6. *Model performance graph*

From all the above folds, it can be observed that the model's performance depends on the folds. Furthermore,the best parameters and estimators are applied for classification, and the report is displayed. So you can see the classification accuracy results after optimization as in Table 11.

Table 11. *Classification result after optimization*

	precision	recall	f1-score	Support
0	1.00	1.00	1.00	911
1	1.00	1.00	1.00	1488
Accuracy			1.00	2399
avg macros	1.00	1.00	1.00	2399
weighted avg	1.00	1.00	1.00	2399

From Table 11, it can be see that the accuracy has increased significantly. The value above is the rounding result of 99.75%. In detail, the mathematical calculation below can be seen:

Actual Classification	0	908	3
	1	3	1485
		0	1
		Prediction Classification	

If :
TP = 1:1 = 1485
FN = 1:0 = 3
TN = 0:0 = 908
FP = 0:1 = 3

1. Accuracy

$$Accuracy \;=\; \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

$$= \frac{1485+908}{1485+908+3+3} * 100\% \quad (13)$$

$$= 99.75\%$$

2. Precision

$$Precision \;=\; \frac{TP}{TP+FP} \times 100\%$$

$$= \frac{1485}{1485+3} \times 100\% \quad (14)$$

$$= 99.80\%$$

3. Recall

$$Recall \;=\; \frac{TP}{TP+FP} \times 100$$

$$= \frac{1485}{1485+3} \times 100\% \quad (15)$$

$$= 99.80\%$$

$$F1-Score \;=\; \frac{2 \times Precision \;\times Recall}{Precision + Recall}$$

$$= \frac{2 \times 99.80\% \times 99.80\%}{99.80\% + 99.80\%} \quad (16)$$

$$= 99.80\%$$

The above calculation obtains the value details after applying the best estimator.

a. Accuracy = 99.75%
b. Precision = 99.80%
c. Recall = 99.80%
d. F1-Score = 99.80%

Next, a comparison of the results of the SVM algorithm before and after optimization is carried out at a ratio of 90:10, as in Table 12.

Table 12. *Comparison of svm algorithm results before and after optimization*

BEFORE OPTIMIZATION	Precision	100%
	Recall	86.41%
	F1-Score	92.71%
	Accuracy	90.25%
AFTER OPTIMIZATION	Precision	99.80%
	Recall	99.80%
	F1-Score	99.80%
	Accuracy	99.75%

Table 12, shows that the accuracy increased by 9.5% from 90.25% to 99.75% before optimization.

3.5. SVM Optimization Results in Other Research

As a comparison, SVM optimization carried out by previous research with different methods shows good results. However, at least in this study, optimization of SVM with Grid Search Cross-Validation was slightly better than in previous studies, this can be seen in table 13.

Table 13. *SVM optimization results in other research*

Author	Purpose	Optimization Methode	Result
Setyawati et al., 2021	Finding optimal SVM parameter values based on accuracy on film opinion data	Firefly	87.84%
Zhou et al., 2020	Forcasting tunnel boring machine advance rate	GWO, WOA, MFO	MFO-SVM: R2(0,9623 and 0,9724), RMSE (0,1269 and 0,1155), VAP (96,24% and 97,34%
Tharwat and Hassanien, 2019	SVM optimization for classification with Iris, Wine, Liver datasets	Bat	Achieve competitive results with optimal parameter values
Susena et al., 2018	SVM optimization for blood donor behavior classification	PSO	90%

CONCLUSION

Based on the results of research that has been carried out related to support vector machine optimization with the grid search cross-validation method for schizophrenia classification, it can be concluded that the optimal parameters in SVM obtained by the grid search cross-validation method, namely C = 1000, gamma = scale, kernel = RBF, affect the accuracy rate and can be used as a model to overcome schizophrenia classification. The accuracy rate obtained by Support Vector Machine (SVM) by applying the grid search cross-validation method for schizophrenia classification increased by 9.5%, which resulted in an accuracy of 99.75%, without optimization accuracy reaching 90.25%.

REFERENCES

[1] WHO, "Schizophrenia," 2022. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/schizophrenia.

[2] M.-H. Hiesh, Y.-Y. Andy Lam, C.-P. Shen, W. Chen, F.-S. Lin, H.-Y. Sung, J.-W. Lin, M.-J. Chiu, and F. Lai, *Classification of Schizophrenia Using Genetic Algorithm-Support Vector Machine (GA-SVM)*. 2013.

[3] F. Riyanda, I. Cholissodin, and Sutrisno, “Klasifikasi Gangguan Jiwa Skizofrenia Menggunakan Algoritme Decision Tree C5.0,” 2019.

[4] W. F. Maramis, “Buku Ajar Ilmu Kedokteran Jiwa.” p. 783, 2009.

[5] T. V. Rampisela and Z. Rustam, "Classification of Schizophrenia Data Using Support Vector Machine (SVM)," in *Journal of Physics: Conference Series*, 2018, vol. 1108, no. 1.

[6] D. Kurniawaty, I. Cholissodin, and P. P. Adikara, “Klasifikasi Gangguan Jiwa Skizofrenia Menggunakan Algoritme Support Vector Machine (SVM),” 2018.

[7] P. Sinha and P. Sinha, "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM," *Int. J. Eng. Res.*, vol. V4, no. 12, pp. 608–612, 2015.

[8] I. M. Elgohary, A. M. M Eissa, H. Mohamed, M. I. Elgohary, T. A. Alzohairy, A. M. Eissa, and sally Eldeghaidy, "An intelligent System for Diagnosing Schizophrenia and Bipolar Disorder based on MLNN and RBF," vol. 4, no. 2, pp. 117–123, 2016.

[9] I. K. Laksono and M. E. Imah, “Klasifikasi Schizophrenia Berdasarkan Sinyal EEG Menggunakan Algoritma Support Vector Machine,” *J. Ilm. Mat.*, vol. 7, no. 2, 2019.

[10] M. Ahmad, N. Wahid, R. A. Hamid, S. Sadiq, A. Mehmood, and G. S. Choi, "Decision Level Fusion using Ensemble Classifier for Mental Disease Classification," pp. 1–23, 2020.

[11] J. Sun, R. Cao, M. Zhou, W. Hussain, B. Wang, J. Xue, and J. Xiang, "A hybrid deep neural network for classification of schizophrenia using EEG Data," *Sci. Rep.*, vol. 11, no. 1, pp. 1–17, 2021.

[12] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, 1996.

[13] H. Harafani and R. S. Wahono, “Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan,” *J. Intell. Syst.*, vol. 1, no. 2, 2015.

[14] W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.

[15] I. Ilhan and G. Tezel, "A genetic algorithm-support vector machine method with parameter optimization for selecting the tag SNPs," *J. Biomed. Inform.*, vol. 46, no. 2, pp. 328–340, 2013.

[16] A. Toha, P. Purwono, and W. Gata, “Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter Grid Search CV,” *Bul. Ilm. Sarj. Tek. Elektro*, vol. 4, no. 1, pp. 12–21, 2022.

[17] Styawati, A. Nurkholis, Z. Abidin, and H. Sulistiani, “Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 904–910, 2021.

[18] I. G. N. E. Susena, M. T. Furqon, and R. C. Wihandika, “Optimasi Parameter Support Vector Machine (SVM) dengan Particle Swarm Optimization (PSO) Untuk Klasifikasi Pendonor Darah Dengan Dataset RFMTC,” *S1 Univ. Brawijaya*, vol. 2, no. 12, pp. 7278–7284, 2018.

[19] A. Tharwat and A. E. Hassanien, "Optimizing support vector machine parameters using bat optimization algorithm," in *Studies in Computational Intelligence*, vol. 801, A. E. Hassanien, Ed. Cham: Springer International Publishing, 2019, pp. 351–374.

[20] J. Zhou, Y. Qiu, S. Zhu, D. Jahed, and C. Li, "Engineering Applications of Artificial Intelligence Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate," *Eng. Appl. Artif. Intell.*, vol. 97, no.

September 2020, p. 104015, 2021.

[21] J. W. Creswell, *John W. Creswell's Research Design 3rd Ed*. 2009.

[22] B. Roach, "EEG data from basic sensory task in Schizophrenia," 2019. [Online]. Available: https://www.kaggle.com/datasets/broach/button-tone-sz.

[23] R. Feldman and J. Sanger, "The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data," 2006.

[24] W. K. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*. 2019.

[25] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support vector machine," *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 1–5, 2003.

[26] B. E. Boser, V. N. Vapnik, and I. M. Guyon, "Training Algorithm Margin for Optimal Classifiers," *Perception*, pp. 144–152, 1992.

[27] G. S. K. Ranjan, A. Kumar Verma, and S. Radhika, "K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries," *2019 IEEE 5th Int. Conf. Converg. Technol. I2CT 2019*, pp. 9–13, 2019.

[28] T. Yan, S. L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm," *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1292–1303, 2022.

[29] K. R. Singh, K. P. Neethu, K. Madhurekaa, A. Harita, and P. Mohan, "Parallel SVM model for forest fire prediction," *Soft Comput. Lett.*, vol. 3, no. June, p. 100014, 2021.

[30] E. Hartini, "Classification of Missing Values Handling Method During Data Mining: Rview," vol. 21, no. 2, 2017.