Classification of Schizophrenia Data Using Support Vector Machine (SVM)

T V Rampisela¹ and Z Rustam²

¹,²Mathematics Department, University of Indonesia, Indonesia

Abstract. Schizophrenia is a severe and chronic mental disorder. This disorder is marked with disturbances in thoughts, perceptions, and behaviours. Due to these disturbances that can trigger Schizophrenics to commit suicide or attempt to do so, Schizophrenics have a lower life expectancy than the general population. Schizophrenia is also difficult to diagnose as there is no physical test to diagnose it yet and its symptoms are very similar to several other mental disorders. Using Northwestern University Schizophrenia Data, this research aims to distinguish people who are Schizophrenics and people who are not. The data consists of 392 observations and 65 variables that are demographic data and clinician-filled Scale for the Assessment of Positive and Negative Symptoms questionnaires. Classification method used is machine learning with Support Vector Machines (SVM). Simulations are done with different data and percentage of training data. In each simulation, accuracy is measured. Model performance validation and evaluation are done by averaging ten times Hold-Out Validations that were done. In conclusion, SVM successfully classified Schizophrenia data with final accuracy of 90.1%. Furthermore, SVM with linear kernel and Gaussian kernel reached an accuracy of 95.0% in at least one simulation in classifying Schizophrenia data.

1. Introduction
Schizophrenia is a severe and chronic mental illness characterized by disturbance in thinking, perception, and behaviour. In general, Schizophrenia symptoms are divided into positive, negative, and cognitive symptoms. Positive symptoms are additional brain activities that are not supposed to exist, for example hallucination and delusion while negative symptoms are the opposite: things that are supposed to exist but are not present, namely apathy and lack of emotion. Cognitive symptoms are symptoms related to disturbances in memory and difficulty in concentrating [1]. All these symptoms of Schizophrenia can affect the quality of life and productivity significantly [2].

Schizophrenics’ life expectancy are 2-3 times lower than the general population [2], as 10% committed suicide. Furthermore, 20-40% Schizophrenics attempted suicide at least once [3]. Anyone could suffer from Schizophrenia even though the cause of it has not been determined for sure [2]. This is exacerbated by the difficulty of detecting Schizophrenia as there is no radiology, laboratory, or psychometric test that can be used to diagnose this illness yet and the similarities of Schizophrenia symptoms with several mental illnesses, e.g. Major Depression and Bipolar [1]. Hence, an easy, objective, and accurate method to diagnose Schizophrenics is needed so that they can get the suitable treatment. One of the methods is through machine learning.

Detecting Schizophrenia by making use of psychological evaluation results is considered as a classification problem in machine learning. Machine learning learns the patterns produced from data training and the results from the learning process would then be used to predict data testing [4]. A few past researches have attempted to apply machine learning in diagnosing Schizophrenia or
Schizophrenia-related problems. There are several machine learning methods used in those studies, namely Fisher Linear Discriminant Analysis [5], Linear Discriminant Analysis and k-Nearest Neighbour [6], Random Forest, Support Vector Classification, Elastic Net, as well as Least Absolute Shrinkage and Selection Operator [7]. However, there are still not many studies on the application machine learning in Schizophrenia-related problems.

The Support Vector Machines (SVM) method is a binary classification method that aims to create a model with good generalization ability [8] with an optimum global solution [9]. SVM creates a hyperplane that separates two target values such that margin, the nearest distance between data and hyperplane, is maximized. In this study, SVM is chosen as it can generalize well and has good accuracy [8]. SVM itself has been used to solve classification problems in numerous studies, for example in intrusion detection system [10], policyholders satisfactory [11], insolvency prediction in insurance companies [12], and brain cancer classification [13].

2. Data
Northwestern University Schizophrenia Data [14] is used in this research. This research uses clinical data from the database, i.e. the group data, demographic data (gender, dominant hand, ethnic, race, and age), as well as Scale for the Assessment of Negative Symptoms (SANS) dan Scale for the Assessment of Negative Symptoms (SAPS) questionnaires data. The data consists of 392 observations and 65 variables. There are 171 data with the label “Schizophrenics” and 221 with “non-Schizophrenics”.

3. The use of kernel function
The kernel function is defined as follows:

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$$

where $\varphi(x)$ is a function that maps $x \in R^n$ to feature space. Each time $\varphi(x_i) \cdot \varphi(x_j)$ appears in classification algorithm, it can be replaced by $K(x_i, x_j)$ [15]. By using kernel function, it is hoped that the data can be separated linearly in a higher dimension.

In this study, three different kernels [16] are used:

- Linear kernel, $K(x_i, x_j) = x_i \cdot x_j$
- Polynomial kernel, $K(x_i, x_j) = (x_i \cdot x_j + 1)^p$
- Gaussian kernel, $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2), \gamma = 1$ [17].

4. Support Vector Machines (SVM)
Let the pattern classified be a set of $m$ row vectors, $A_i (i = 1, 2, ..., m) \in R^n$, where $A_i = (A_{i1}, A_{i2}, ..., A_{in})^T$. Let $y_i \in \{1, -1\}$ be the class of the $i$-th pattern. On the case with data that can be separated linearly, $w \in R^n$ and $b \in R$ are determined such that: $A_i w \geq 1 - b$ for $y_i = 1$ and $A_i w \leq -1 - b$ for $y_i = -1$ [18].

The plane defined by:

$$w^T x + b = 0$$

is located between two planes defined by $w^T x + b = 1$ and $w^T x + b = -1$. The plane in equation (1) separates each class from another by a margin of $\frac{1}{\|w\|}$ on each side. The SVM model is obtained by maximizing the margin and is equivalent to the following problem [18]:

(SVM 1) $\min_{w,b} \frac{1}{2} w^T w$

subject to:

$$A_i w \geq 1 - b \text{ for } y_i = 1 \text{ and } A_i w \leq -1 - b \text{ for } y_i = -1$$

(2)
However, when the two classes cannot be fully separated linearly, there will be error in satisfying the inequality in equation (2) on several patterns. Hence, equation (2) can be modified to:

$$A_iw + q_i \geq 1 - b \text{ for } y_i = 1 \text{ and } A_iw - q_i \leq -1 - b \text{ for } y_i = -1$$

$$q_i \geq 0, i = 1, 2, ..., m$$  \hspace{0.5cm} (3)

where $q_i$ denotes the error variable of the $i$-th data sample. This classification method is considered as one with soft margin as it classifies data with some errors. Soft margin depends on the value of $q_i$. Classification of the sample for testing is obtained by determining the sign of $w^T x + b$ [18].

Formulation of SVM with soft margin and linear kernel is given by:

$$(SVM \ 2) \ \min_{w,b,q} C e^T q + \frac{1}{2} w^T w$$

subject to:

$$A_iw + q_i \geq 1 - b \text{ for } y_i = 1$$

$$A_iw - q_i \leq -1 - b \text{ for } y_i = -1$$

$$q_i \geq 0, i = 1, 2, ..., m$$

where $e$ is a vector which all its entries are 1 and the scalar $C$ is a trade-off parameter. A larger value of $C$ means that the model would prioritize the minimization of $q$, while a smaller one would prioritize the minimization of $w$ [18].

5. Parameter Optimization
In this study, several parameters are optimized by using the grid search method. The grid search method finds one by one the combination of parameters that produces the optimum model [19]. The parameters that are optimized are:

- $p = 2, 3, 4$ on the polynomial kernel, and
- $C$ on the interval $[10^{-3}, 10^3]$ by logarithmic scale.

6. Model Performance Validation
To validate the model performance, Hold-Out Validation is done. In Hold-Out Validation, the data is randomly separated into two parts: training and testing data. The model evaluation will be used on the testing data. Computationally, this validation method is easy and fast [20].

Hold-Out Validation is done on different percentages of data in this study. To overcome the weakness of Hold-Out Validation, which is very dependent on the data used for training and testing [20], simulations are done 10 times with a different set of data for each percentage of data used.

7. Model Performance Evaluation
Model performance evaluation in this study is done by measuring accuracy, which is how often the classification model makes a correct prediction. Accuracy is the ratio of the number of correct predictions and the total number of predictions [20]. A classification model performance is said to be good if the accuracy of the model is high [21].

8. Simulation Results and Analysis
In this section, results and analysis of classification of Schizophrenia data with SVM will be covered. The software MATLAB R2017a was used to form the SVM with soft margin model.

8.1. Classification of Schizophrenia data using SVM with linear kernel
The result of Schizophrenia data classification using SVM with linear kernel can be seen below.

| Table 1. Accuracy of Schizophrenia data classification using SVM with linear kernel. |
|---------------------------------------------------------------|
| Accuracy of the $i$-th simulation (%)                        |

3
Table 2. Final accuracy of Schizophrenia data classification using SVM with linear kernel.

| Percentage of Training Data (%) | Accuracy (%) |
|---------------------------------|--------------|
| 10                              | 88.1 ± 1.47  |
| 20                              | 89.0 ± 1.22  |
| 30                              | 88.8 ± 1.09  |
| 40                              | 88.8 ± 1.27  |
| 50                              | 89.0 ± 1.39  |
| 60                              | 89.9 ± 1.68  |
| 70                              | **90.1 ± 2.30** |
| 80                              | 90.1 ± 2.79  |
| 90                              | 89.8 ± 4.48  |

It can be seen from Table 1 that the SVM model with linear kernel that produces the best accuracy for classification of Schizophrenia data is the one with 90% training data with an accuracy of 95.0%. On the other hand, the worst-performing model is also obtained from the one with 90% training data with an accuracy of 82.5%. This can happen as the Hold-Out Validation method is very dependent on the data used for training and testing.

Based on the final accuracy of Schizophrenia data classification using SVM with linear kernel (Table 2), in general, the more percentage of training data used in building the model, the more accurate is the final accuracy. The SVM model with linear kernel that gives the best final accuracy for the classification of Schizophrenia data is the one that uses 70% training data with a mean of 90.1%, whereas the worst one is the model that uses 10% training data with a mean of 88.1%.

8.2. Classification of Schizophrenia data using SVM with polynomial kernel

It can be seen from Table 3 that the SVM model with polynomial kernel that produces the best accuracy for classification of Schizophrenia data is the one with 90% training data with an accuracy of 92.5%. However, the model that has the worst performance is also obtained from the one with 90% training data with an accuracy of 68.0%.

Based on the final accuracy of Schizophrenia data classification using SVM with polynomial kernel (Table 4), in general, the more percentage of training data used in building the model, the more accurate is the final accuracy. The SVM model with polynomial kernel that gives the best final accuracy for the classification of Schizophrenia data is the one that uses 90% training data with a mean of 87.0%, whereas the model that gives the worst performance uses 30% training data with a mean of 79.1%.
On the other hand, the worst for classification of Schizophrenia data is the one with 90%.

It can be seen from Table 5 that the SVM model with gaussian kernel that produces the best accuracy is 8.3.

Based on the final accuracy of Schizophrenia data classification using SVM with Gaussian kernel (Table 6), in general, the more percentage of training data used in building the model, the better is the

### Table 3. Accuracy of Schizophrenia data classification using SVM polynomial kernel

| Percentage of Training Data (%) | Accuracy of the i-th simulation (%) |
|-------------------------------|-----------------------------------|
| 10                            | 90.1   68.0  76.2  83.0  82.2  85.3  84.7  76.8  69.1  79.3 |
| 20                            | 77.7   75.8  77.4  75.2  84.7  78.7  82.2  82.5  76.1  81.2 |
| 30                            | 70.9   79.3  77.8  73.5  82.9  80.0  81.5  80.7  81.8  82.2 |
| 40                            | 75.0   78.4  79.2  74.6  86.4  80.1  84.3  81.8  79.7  79.2 |
| 50                            | 75.0   79.6  78.6  81.6  83.2  83.7  83.1  75.0  86.2  80.1 |
| 60                            | 85.4   89.2  78.3  79.0  82.8  84.1  82.2  77.1  82.8  84.1 |
| 70                            | 87.3   80.5  91.5  84.7  82.2  87.3  83.1  78.0  91.5  90.7 |
| 80                            | 91.1   82.3  87.3  83.5  82.3  87.3  82.3  86.1  87.3  78.5 |
| 90                            | 92.5   87.5  92.5  77.5  80.0  87.5  90.0  82.5  90.0  90.0 |

### Table 4. Final accuracy of Schizophrenia data classification using SVM with polynomial kernel

| Percentage of Training Data (%) | Accuracy (%) |
|-------------------------------|--------------|
| 10                            | 79.5 ± 7.07  |
| 20                            | 79.1 ± 3.29  |
| 30                            | 79.1 ± 3.96  |
| 40                            | 79.9 ± 3.67  |
| 50                            | 80.7 ± 3.75  |
| 60                            | 82.5 ± 3.61  |
| 70                            | 85.7 ± 4.77  |
| 80                            | 84.8 ± 3.68  |
| 90                            | 87.0 ± 5.24  |

8.3. Classification of Schizophrenia data using SVM with Gaussian kernel

It can be seen from Table 5 that the SVM model with gaussian kernel that produces the best accuracy for classification of Schizophrenia data is the one with 90% training data with an accuracy of 95.0%. On the other hand, the worst-performing model is also obtained from the one with 90% training data with an accuracy of 82.5%. Similar to the case with linear kernel, this can happen because the Hold-Out Validation method relies heavily on the data used for training and testing.

### Table 5. Accuracy of Schizophrenia data classification using SVM with Gaussian kernel

| Percentage of Training Data (%) | Accuracy of the i-th simulation (%) |
|-------------------------------|-----------------------------------|
| 10                            | 88.1   89.2  87.5  88.7  88.7  88.7  88.7  88.4  88.4  88.4 |
| 20                            | 89.2   89.5  86.9  89.5  88.5  89.5  88.5  88.2  87.9  87.9 |
| 30                            | 90.5   88.7  88.0  88.7  88.7  88.7  87.6  88.0  89.5  87.6 |
| 40                            | 90.7   91.1  87.7  88.6  89.0  88.6  88.1  88.1  89.0  87.3 |
| 50                            | 90.3   91.8  89.3  89.3  88.8  86.7  88.3  88.8  93.2  87.8 |
| 60                            | 90.4   93.6  91.1  88.5  88.5  90.4  88.5  88.5  90.4  88.5 |
| 70                            | 89.0   94.9  91.5  89.0  87.3  90.7  87.3  89.0  91.5  90.7 |
| 80                            | 91.1   94.9  92.4  89.9  86.1  88.6  89.9  86.1  92.4  89.9 |
| 90                            | 92.5   95.0  92.5  85.0  82.5  87.5  90.0  85.0  95.0  92.5 |

Based on the final accuracy of Schizophrenia data classification using SVM with Gaussian kernel (Table 6), in general, the more percentage of training data used in building the model, the better is the
final accuracy. The SVM model with Gaussian kernel that performs for the classification of Schizophrenia data is the one that uses 70% training data with a mean of 90.1%, whereas the worst one is the model that uses 10% training data with a mean of 88.5%.

Table 6. Final accuracy of Schizophrenia data classification using SVM with Gaussian kernel

| Percentage of Training Data (%) | Accuracy (%)     |
|--------------------------------|------------------|
| 10                             | 88.5 ± 0.444     |
| 20                             | 88.6 ± 0.855     |
| 30                             | 88.6 ± 0.892     |
| 40                             | 88.8 ± 1.22      |
| 50                             | 89.0 ± 1.39      |
| 60                             | 89.9 ± 1.68      |
| 70                             | 90.1 ± 2.30      |
| 80                             | 90.1 ± 2.79      |
| 90                             | 89.8 ± 4.48      |

9. Conclusion
Machine learning can solve the classification problem of Schizophrenia data using the Northwestern University Schizophrenia Data. There are three types of simulations done in this study: SVM with linear, polynomial, and Gaussian kernel. In each simulation, the parameters of the models are optimized with grid search method, while the validation and evaluation of model performance are done by Hold-Out Validation and calculating the accuracy.

The SVM method with linear and Gaussian kernel has a similar performance, but better than the one that uses polynomial kernel. With linear and Gaussian kernel, SVM managed to classify Schizophrenia data with a final accuracy of 90.1%. Additionally, in classifying Schizophrenia data, SVM with linear kernel and Gaussian kernel successfully reached an accuracy of 95.0% in at least one simulation.

For future researches, it is recommended to use other dataset, kernels, or evaluation and validation methods. By doing this research, it is hoped that the results could help the medical field in predicting whether someone suffers from Schizophrenia or not, so that they can receive the right treatment for their illness.

10. References
[1] American Psychiatric Association 2013 Diagnostic and statistical manual of mental disorders, fifth edition (Arlington, VA: American Psychiatric Association)
[2] World Health Organization 2018(online) Schizophrenia Key Facts [online] http://www.who.int/en/news-room/fact-sheets/detail/schizophrenia
[3] Schizophrenia.com 2010(online) Schizophrenia symptoms and diagnosis of Schizophrenia [online] http://schizophrenia.com/szfacts.htm
[4] Witten I H and Frank E 2005 Data mining: practical machine learning tools and techniques, second edition (Burlington: Elsevier Inc)
[5] Neuhaus A H, Popescu F C, Bates J A, Goldberg, T E, and Malhotra, A K 2013 Single-subject classification of schizophrenia using event-related potentials obtained during auditory and visual oddball paradigms European Archives of Psychiatry and Clinical Neuroscience 263(241)
[6] Ahn M, Hong J H, and Jun S C 2012 Feasibility of approaches combining sensor and source features in brain-computer interface Journal of neuroscience methods 204(1) 168–78
[7] Hettige N C, Nguyen T B, Yuan C, Rajakulendran T, Baddour J, Bhagwat N, Bani-Fatemi A, Voineskos A N, Mallar Chakravarty M, De Luca V 2017 Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach Gen Hosp Psychiatry 47 20-8
[8] Cortes C and Vapnik V N 1995 Support-vector networks Machine Learning 20(3) 273–97
[9] Burges C J C and Crisp D J 1999 Uniqueness of the SVM solution Neural Information Processing Systems 12 223-9
[10] Rustam Z and Zahras D 2018 Comparison between Support Vector Machine and Fuzzy C-Means as classifier for intrusion detection system Journal of Physics: Conference Series 1028(1)
[11] Rustam Z and Arians C P A 2018 Support Vector Machines for classifying policyholders satisfactorily in automobile insurance Journal of Physics: Conference Series 1028(1) 012005
[12] Rustam Z and Yaurita F 2018 Insolvency prediction in insurance companies using Support Vector Machines and Fuzzy Kernel C-Means Journal of Physics: Conference Series 1028(1), 012118
[13] Panca V and Rustam Z 2017 Application of machine learning on brain cancer multiclass classification AIP Conference Proceedings 1862(1) 030133
[14] Wang L, Kogan A, Cobia D, Alpert K, Kolansy A, Miller M I, and Marcus, D 2013 Northwestern University Schizophrenia Data and Software Tool (NUSDAST) Frontiers in Neuroinformatics 7 25
[15] Witten I H, Frank E, and Hall M A 2011 Data mining: practical machine learning tools and techniques, third edition (Burlington: Elsevier Inc)
[16] Han J and Kamber M 2006 Data mining concepts and techniques, second edition (San Francisco: Elsevier Inc)
[17] Xue H, Yang Q, and Chen S 2009 The top ten algorithms in data mining (New York: Chapman and Hall/CRC)
[18] Jayadeva, Khemchandani R, and Chandra S 2007 Twin Support Vector Machines for pattern classification IEEE Transactions on Pattern Analysis and Machine Intelligence 29(5) 905–10
[19] Lin C-J, Hsu C-W, and Chang C-C 2003 A Practical Guide to Support Vector Classification
[20] Zheng A 2015 Evaluating machine learning models: a beginner’s guide to key concepts and pitfalls (Sebastopol, CA: O’Reilly Media, Inc)
[21] Tan P N, Steinbach M, and Kumar V 2006 Introduction to Data Mining (Boston: Pearson Education)

Acknowledgments
The authors would like to express appreciation for the support of data collection and sharing for this project, which were funded by NIMH grant 1R01 MH084803.