

Cerebral infarction classification using multiple support vector machine with information gain feature selection

Zuherman Rustam¹, Arfiani², Jacob Pandelaki³

^{1,2}Department of Mathematics, University of Indonesia, Indonesia

³Department of Radiology, Cipto Mangunkusumo Hospital, Indonesia

Article Info

Article history:

Received Aug 10, 2019

Revised Oct 14, 2019

Accepted Dec 4, 2019

Keywords:

Cerebral infarction

Information gain

Support vector machine

ABSTRACT

Stroke ranks the third leading cause of death in the world after heart disease and cancer. It also occupies the first position as a disease that causes both mild and severe disability. The most common type of stroke is cerebral infarction, which increases every year in Indonesia. This disease does not only occur in the elderly, but in young and productive people which makes early detection very important. Although there are varied of medical methods used to classify cerebral infarction, this study uses a multiple support vector machine with information gain feature selection (MSVM-IG). MSVM-IG is a modification among IG Feature Selection and SVM, where SVM conducted doubly in the process of classification which utilizes the support vector as a new dataset. The data obtained from Cipto Mangunkusumo Hospital, Jakarta. Based on the results, the proposed method was able to achieve an accuracy value of 81%, therefore, this method can be considered to use for better classification result.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Zuherman Rustam,

Department of Mathematics,

University of Indonesia,

Depok 16424, Indonesia.

Email: rustam@ui.ac.id

1. INTRODUCTION

Stroke is a leading cause of mortality and disability throughout the world [1, 2]. This far, ischemic stroke is the most common type, which accounts for 70-90% of all stroke cases [3, 4]. Deaths that occur due to ischemic stroke are still of foremost concern [5]. This disease becomes an important global health problem, so that an effective way is needed to reduce mortality from this ischemic stroke. One way to diagnose whether a patient has cerebral infarction, an examination from the radiology agency is needed, and one diagnostic method often used to conduct these examinations is the computed tomography scanning (CT Scan). This method is used to obtain a picture of the patient's head area. When some firmly demarcated dark areas are visualized surrounding the brain tissue during the test, then that area is the chronic phase. As a result, a body function regulated by the area tends to be permanently disrupted when early treatment isn't provided.

Early medication helps to prevent diseases. Therefore, one important method used to prevent chronic cerebral infarction is early identification to enable the patient to obtain the right treatment and care immediately. One method used for this classification is machine learning such as the multiple support vector machines with information gain feature selection (MSVM-IG) as proposed in this study. The cerebral infarction data was obtained from RSCM hospital with as many as 206 patients who had undergone the examination. Each patient was informed of the feature used to determine the severity of cerebral infarction, and its data in this study consists of 10 features.

The previous researches on the classification of cerebral infarction had been carried out using the Support Vector Machine method [6, 7] with great results. Similarly, the information gain feature selection method has been used to detect Brain [8] and Lung Cancer [9]. In addition, the support vector machine method has been used for the classification of schizophrenia data [10], to construct process maps for additive manufacturing [11], often used for pattern recognition one of them in [12], for prediction of protein structural classes [13], hyperspectral imagery [14], traffic incident detection [15], for image retrieval and image process [16], fault interpretation, a study based on 3D seismic mapping of the Zhaozhuang coal mine in the Qinshui Basin, China [17], intrusion detection system [18], pattern recognition to AVO classification [19], for estimation of reservoir porosity and water saturation based on seismic attributes [20], elastic impedance based facies classification [21], and the application of svm for prediction of coal and gas outburst [22].

2. RESEARCH METHOD

This research proposes a Multiple Support Vector Machine with Information Gain Feature Selection (MSVM-IG) for early cerebral infarction classification. MSVM-IG is a method that uses support vector obtained from SVM as an input in feature selection. Therefore, the amount of data processed by the IG feature selection is not the same as the initial. The term multiple is used because after the feature selection process with IG, SVM is re-evaluated. Due to the decrease in the amount of input data, IG selection features are able to rank features more accurately with SVM producing better accuracy.

2.1. Data

The numeric data used in this study obtained from the results of the CT scan of Cipto Mangunkusumo Hospital, Central Jakarta, which consists of 10 features, and they include: Gender, Age of patient, Cerebral infarction area, Air normal cavity, Minimum value of area, Maximum value of area, Sum of acute point, Length of area, Average of area, and Standar deviasi of area. The data include 206 observations with are 103 data labeled positive infarc and 103 data negative infarc.

2.2. Information gain feature selection

Information gain (IG) is one technique of filter type selection which works by sorting features based on each value. Measurements from IG itself are based on the basic concept of entropy by determining the difference between the entropy of all training data and the weighted sum of its subset of partition values on a feature [23]. IG is also one of the easiest and fastest methods of sorting features. For example, there is a training data set S with n -features and m -classes, $S(A_1, A_2, \dots, A_n, C)$ with C is an attribute consisting of different m -classes. The value for the entropy of all training data is calculated based on different m -classes, therefore:

$$Entropy(S) = -\sum_{i=1}^m p(c_i) \cdot \log_2 p(c_i) \quad (1)$$

with $p(c_i)$ is the probability (relative frequency) of the class (c_i) in the S training data, with l different values used to calculate the weighed total sum of the entropy subset or partitioned values. Each value contains an entropy value based on the class label in feature A such that call S_v acts as a subset, where $v = 1, 2, \dots, l$. Therefore, the weighted sum of the entropy subset of partition values on a feature is formulated as follows:

$$\sum_{v \in A} \frac{|S_v|}{|S|} * entropy(S_v) \quad (2)$$

As previously explained, IG is obtained by looking at the difference between the entropy of all training data and the weighted sum of the entropy subset of the partition values on a feature. Therefore, the difference from equations (1) and (2) is the IG of a feature [23]:

$$IG(S, A) = Entropy(S) - \sum_{v \in A} \frac{|S_v|}{|S|} * entropy(S_v) \quad (3)$$

$$= -\sum_{i=1}^m p(c_i) \cdot \log_2 p(c_i) - \sum_{v \in A} \frac{|S_v|}{|S|} * entropy(S_v) \quad (4)$$

2.3. Support vector machine

Support vector machine (SVM) which was introduced by Vapnik in the late 1990s, is a machine learning algorithm used for classification and regression. SVM is related to structural risk minimization (SRM) and was initially used for binary classification. It is currently used for multiclass classification

and takes the form of mapping input space into higher dimensional space to support nonlinear classification problem where the maximum separation of the hyperplane is constructed. The hyperplane is a linear pattern whose maximum margin provides separation between decision classes.

In the dataset $\{x_i, y_i\}_{i=1}^N$, N is the number of samples, $x_i \in R^D$ is a feature vectors from sample- i , with D is the number of features (dimension), and y_i is a class label. For the two-class classification problem $y_i \in \{-1, +1\}$, while in a multiclass $y_i \in \{1, 2, \dots, k\}$ with k is the number of class. The main goal of SVM is to determine the best hyperplane [24] and it illustrated in Figure 1:

$$w \cdot x + b = 0 \tag{5}$$

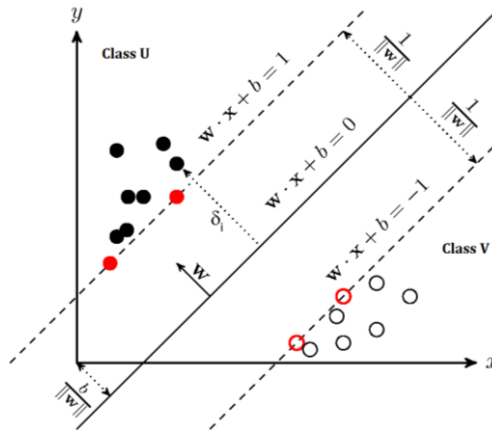


Figure 1. SVM is trying to determine the best hyperplane to separate two classes

The problem of SVM optimization is summarized as follows:

$$\min \frac{1}{2} \|w\|^2 \tag{6}$$

$$s. t. y_i(w^T \cdot x_i + b) \geq 1, \forall i = 1, \dots, N \tag{7}$$

Objective function (6) to determine $w \in R^n$ and $b \in R^n$ subject to (7), with w is the weights and b is bias. By completing the equation above, the formula w and b are obtained as follows:

$$w = \sum_{i=1}^N a_i y_i x_i \tag{8}$$

$$b = \frac{1}{N_s} \sum_{i \in S} (y_i - \sum_{m \in S} a_m y_m x_m) \tag{9}$$

and, the decision function as follows:

$$f(x) = \text{sign}(w \cdot x + b) \tag{10}$$

Below is the diagram flow of the proposed method, see Figure 2. First step is the data will be processed by SVM so that the support vector is generated. Then, the IG feature selection will select the selected features based on support vector. Lastly, SVM will be used again to get the measurement.

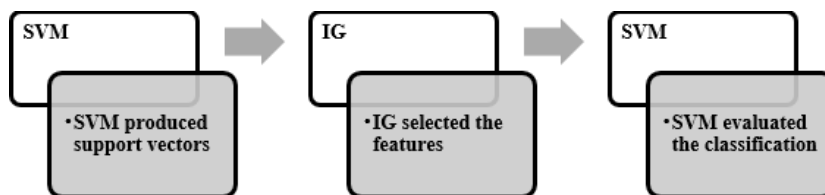


Figure 2. The flow diagram of MSVM-IG

2.4. Kernel function

This research utilizes two kernel functions, namely radial basis function and polynomial kernel functions with several parameters. The kernel function is given as follows:

$$K(\mathbf{x}_i, \mathbf{x}'_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}'_j) \rangle \quad (11)$$

with $\varphi(\mathbf{x})$ is a function that maps $\mathbf{x} \in \mathbb{R}^n$ to the feature space \mathbb{F} . Every time $\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}'_j) \rangle$ appears in the classification algorithm, it is replaced with $K(\mathbf{x}_i, \mathbf{x}'_j)$ [25]. By using kernel functions, it is expected that data is linearly separated linearly on higher dimensions. The formula of radial basis function (RBF) and polynomial are shown below [26].

– RBF Kernel Function:

$$k(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2) \quad (12)$$

– Polynomial kernel function:

$$k(\mathbf{x}, \mathbf{y}) = [(\mathbf{x} \cdot \mathbf{y}) + 1]^d \quad (13)$$

2.5. Model performance evaluation

In this study, a performance evaluation model was conducted by measuring accuracy, precision, sensitivity, specificity, and recall. Let TN, TP, FN, FP denote true negative, true positive, false negative, and false positive, respectively. The following formulas below are used [27]:

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (14)$$

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (16)$$

$$Specificity = \frac{TN}{TN+FP} \quad (17)$$

$$F1 - Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (18)$$

3. RESULTS AND ANALYSIS

The support vector machine with information gain (IG-SVM) feature selection conventional (without multiple SVM) is used to compare the proposed method. Two kernel functions were used, namely radial basis function (RBF) and polynomial. Approximately 10 values of σ and d are used in the RBF and polynomial kernels respectively with the same parameter values; $C = 1000$, k -fold = 3, and 5 main features.

3.1. Classification results with RBF kernel

For the RBF kernel we tried 10 different σ values that we determined randomly. The results are listed in Tables 1 and 2.

Table 1. Results of cerebral infarction classification using MSVM-IG with RBF kernel

σ	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
0.0001	81.863	81.553	83.333	81.372	81.951
0.001	81.127	79.812	83.333	78.922	81.535
0.05	80.882	79.439	83.333	78.431	81.34
0.1	80.76	79.254	83.333	78.186	81.243
1	80.686	79.143	83.333	78.039	81.184
10	80.637	79.07	83.333	77.941	81.146
50	80.602	79.017	83.333	77.871	81.118
100	80.576	78.978	83.333	77.819	81.097
1000	80.556	78.947	83.333	77.778	81.081
10000	80.539	78.923	82.353	77.745	81.068

Table 2. Results of cerebral infarction classification using IG-SVM with RBF kernel

σ	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
0.0001	80.763	80.490	82.493	80.872	80.661
0.001	80.137	79.612	82.343	79.722	80.524
0.05	80.782	79.339	81.433	79.431	80.245
0.1	79.751	78.154	81.433	79.187	80.143
1	79.586	78.133	81.333	79.029	80.104
10	79.507	78.071	81.333	78.831	79.144
50	78.492	78.017	81.333	78.771	79.137
100	78.466	77.968	80.443	78.519	79.035
1000	78.446	77.847	80.443	77.769	79.033
10000	78.429	77.623	80.443	77.750	79.021

According to Tables 1 and 2, the smaller the value of σ the greater the classification results with the highest accuracy, precision, sensitivity, specificity, and f1-score values obtained when the value of $\sigma = 0.0001$ for both methods. This is because the smaller the value of σ the faster the classification method to learn data patterns and produce better results. The MSVM-IG produces better results than IG-SVM with the highest accuracy, precision, sensitivity, specificity, and f1-score obtained by 81.863%, 81.553%, 83.333%, 81.372%, and 81.951% respectively. There was an approximate total difference of 1% between the two methods, however, MSVM-IG is the method of choice for the classification of cerebral infarction.

3.2. Classification results with polynomial kernel

Also, for the polynomial kernel we tried 10 different d values that we determined randomly. The results are listed in Tables 3 and 4. The result shows that for experiment d values from 1 to 10 produced the same accuracy, precision, sensitivity, specificity, and F1-score.

Table 3. Results of cerebral infarction classification using MSVM-IG with polynomial kernel

d	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
1	80.392	78.704	83.333	77.451	80.952
2	80.392	78.704	83.333	77.451	80.952
3	80.392	78.704	83.333	77.451	80.952
4	80.392	78.704	83.333	77.451	80.952
5	80.392	78.704	83.333	77.451	80.952
6	80.392	78.704	83.333	77.451	80.952
7	80.392	78.704	83.333	77.451	80.952
8	80.392	78.704	83.333	77.451	80.952
9	80.392	78.704	83.333	77.451	80.952
10	80.392	78.704	83.333	77.451	80.952

Table 4. Results of cerebral infarction classification using IG-SVM with polynomial kernel

d	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
1	79.882	78.145	82.954	77.211	79.534
2	79.792	78.145	82.833	77.211	79.534
3	79.592	78.144	82.573	77.211	79.534
4	78.456	77.765	82.573	76.352	78.726
5	78.455	77.765	82.573	76.352	78.726
6	78.444	77.765	82.573	76.352	78.726
7	78.340	77.765	82.573	76.352	78.726
8	78.340	77.765	81.997	76.352	77.942
9	78.340	77.765	81.997	75.451	77.942
10	78.340	77.765	81.997	75.441	77.942

According to Tables 3 and 4, the smaller the value of d the greater the classification results, the higher the accuracy, precision, sensitivity, and specificity, with f1-score values are obtained when $d = 1$ for both methods. The smaller the value of d the faster the classification method to quickly learn data patterns and produce better results. The results of MSVM-IG is better than IG-SVM with the highest accuracy, precision, sensitivity, specificity, and f1-score obtained by 80.392%, 78.704%, 83.333%, 77.451%, and 80.952% respectively. The difference between the two methods is approximately 1%, however, the MSVM-IG tends to be the method of choice for the classification of cerebral infarction.

4. CONCLUSION

Stroke holds the second place of leading cause of death and the third the leading cause of disability. Ischemic stroke is the most common type so we have to find the way to label stroke efficiently. This study proposed a multiple support vector machine using the information gain feature selection (MSVM-IG) for the classification of cerebral infarction. Additionally, the RBF and polynomial kernel functions are used and based on the results as well as discussion, it was found that MSVM-IG tends to produce good accuracy, sensitivity, specificity, and F1-score when using the RBF kernel ($\sigma = 0.0001$) with a high enough accuracy of 81.863%. When compared with the conventional method, namely support vector machine with information gain feature selection (IG-SVM), the difference was approximately 1% with MSVM-IG results greater than IG-SVM. This indicated that MSVM-IG has a better result than the conventional method. For future work, this modification could be improved again and the other kernel functions and techniques can be used for comparison.

ACKNOWLEDGEMENTS

This research was financially supported by The Ministry of Research and Higher Education, Republic of Indonesia (KEMENRISTEKDIKTI), with a PDUPT 2020 research grant scheme.

REFERENCES

- [1] V. L. Feigin et al., "Global and regional burden of stroke during 1990-2010: Findings from the global burden of disease study 2010," *Lancet*, vol. 838, pp. 245-255, 2014.
- [2] D. Mukherjee and C. G. Patil, "Epidemiology and global burden of stroke," *World Neurosurgery*, vol. 76; no. 6, pp. 585-590, 2011.
- [3] W. Wang et al., "Prevalence, incidence, and mortality of stroke in China: Results from a nationwide population-based survey of 480 678 adults," *Circulation*, vol. 135, no. 8, pp. 759-771, 2017.
- [4] T. Xu et al., "Smoking, heart rate, and ischemic stroke: A population-based prospective cohort study among inner Mongolians in China," *Stroke*, vol. 44, no. 9, pp. 2457-2461, 2013.
- [5] K. B. Slot et al., "Impact of functional status at six months on long term survival in patients with ischaemic stroke: Prospective cohort studies," *BMJ*, vol. 336, no. 7640, pp. 376-379 2008.
- [6] P. Bentley et al., "Prediction of stroke thrombolysis outcome using CT brain machine learning," *NeuroImage Clinical*, vol. 4, pp. 635-640, 2014.
- [7] T. Kim et al., "Machine learning for detecting moyamoya disease in plain skull radiography using a convolutional neural network," *EBioMedicine*, vol. 40, pp. 636-642, 2018.
- [8] C. M. Lai, W. C. Yeh, and C. Y. Chang, "Gene selection using information gain and improved simplified swarm optimization," *Neurocomputing*, vol. 218, pp. 331-338, Dec 2016.
- [9] L. Gao, M. Ye, X. Lu, and D. Huang., "Hybrid method based on information gain and support vector machine for gene selection in cancer classification," *Genomics Proteomics Bioinformatics*, vol. 15, no. 6, pp. 389-395, 2017.
- [10] T. Rampisela and Z. Rustam, "Classification of schizophrenia data using support vector machine," *Journal of Physics: Conference Series*, vol. 1108, no. 1, pp. 1-7, 2018.
- [11] K. Aoyagi, H. Wang, H. Sudo, and A. Chiba, "Simple method to construct process maps for additive manufacturing using a support vector machine," *Additive Manufacturing*, vol. 27, pp. 353-362, 2019.
- [12] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.
- [13] Y-D. Cai, X-J. Liu, X-B. Xu, and K-C. Chou, "Prediction of protein structural classes by support vector machines," *Computing & Chemistry*, vol. 26, no. 3, pp. 293-296, 2002.
- [14] Z. Chunhui, G. Bing, Z. Lejun, and W. Xiaoqing, "Classification of hyperspectral imagery based on spectral gradient, SVM and spatial random forest," *Infrared Physics and Technology*, vol. 95, pp. 61-69, 2018.
- [15] J. Xiao, "SVM and KNN ensemble learning for traffic incident detection," *Physica A: Statistical Mechanics and its Applications*, vol. 517, pp. 29-35, 2019.
- [16] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 2, pp. 721-724, 2001.
- [17] G. Zhou, K. Ren, Z. Sun, S. Peng, and Y. Tang, "Fault interpretation using a support vectormachine: A study based on 3D seismic mapping of the Zhaozhuang coal mine in the Qinshui Basin, China," *Journal of Applied Geophysics*, vol. 171, pp. 1-11, 2019.
- [18] J. Maharani and Z. Rustam, "The application of multi-class support vector machines on intrusion detection system with the feature selection using information gain," *1st Annual International Conference on Mathematics, Science, and Education (ICoMSE)*, pp. 1-4, 2017.
- [19] J. Li and J. Castagna, "Support vector machine (SVM) pattern recognition to AVO classification," *Geophysical Research Letter*, vol. 31, no.2, pp. 475-496, 2004.
- [20] S. R. Na'imi, S. R. Shadizadeh, M. A. Riahi, and M. Mirzakhani, "Estimation of reservoir porosity and water saturation based in seismic attributes using support vector regression approach," *Journal of Applied Geophysics*, vol. 107, pp. 93-101, 2014.

- [21] Y. Nishitsuji and R. Exley, "Elastic impedance based facies classification using support vector machine and deep learning," *Geophysics Prospect*, vol. 67 no. 4, pp. 1040-1054, 2019.
- [22] H. Zhang, L. U. Guangli, X. U. Lu, and Q. Yang, "Application of support vector machine in prediction of coal and gas outburst," *Mining Safety & Environmental Protection*, vol. 8, no. 2, pp. 2670-2681, 2013.
- [23] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data," *Proceedings 5th European Conference PKDD*, vol. 2168, pp. 42-53, 2001.
- [24] N. Cristianini and J. S. Taylor, "An introduction to support vector machines and other kernel-based learning methods," *Cambridge University Press*, 2000.
- [25] I. H. Witten, E. Frank, and M. A. Hall, "Data mining: Practical machine learning tools and techniques-third edition," *Elsevier Inc.*, Burlington, 2011.
- [26] L. Liu, B. Shen, and X. Wang, "Research on kernel function of support vector machine," *In: Huang YM., Chao HC., Deng DJ., Park J. (eds) Advanced Technologies, Embedded and Multimedia for Human-centric Computing. Lecture Notes in Electrical Engineering*, vol. 260, pp 827-834, 2014.
- [27] M. Kohl, "Performance Measures in binary classification," *International Journal of Statistics in Medical Research* vol. 1, pp. 79-81, 2012.

BIOGRAPHIES OF AUTHORS



Zuherman Rustam is an Associate Professor and a lecturer of the intelligence computation at the Department of Mathematics, University of Indonesia. He obtained his Master of Science in 1989 in informatics, Paris Diderot University, French, and completed his Ph.D. in 2006 from computer science, University of Indonesia.

Assoc. Prof. Dr. Rustam is a member of IEEE who is actively researching machine learning, pattern recognition, neural network, artificial intelligence.



Arfiani is a Bachelor of Science from the Department of Mathematics, University of Indonesia. Her current research is machine learning and deep learning in various fields.



Jacub Pandelaki is an academic senate and lecturer at the Faculty of Medicine, University of Indonesia.

He graduated from the medical doctor at the faculty of medicine, the University of Indonesia in 1989, and obtained his Ph.D. degree at the same university in 2010.

Dr. dr. Jacub Pandelaki, Sp. Rad works in the Department of Radiology, Cipto Mangunkusumo Hospital, Indonesia as the doctor and the chair of the Interventional.