Smooth Support Vector Machine for Suicide-Related Behaviours Prediction

G. Indrawan, I K. P. Sudiarsa, K. Agustini, Sariyasa
Department of Computer Science, Universitas Pendidikan Ganesha, Bali, Indonesia

Article Info	ABSTRACT
	Suicide-related behaviours need to be prevented on psychiatric patients. Prediction of those behaviours based on patient medical records would be very useful for the prevention by the psychiatric hospital. This research focused on developing this prediction at the only one psychiatric hospital of Bali Province by using Smooth Support Vector Machine method, as the further development of Support Vector Machine. The method used 30.660 patient medical records from the last five years. Data cleaning gave 2665 relevant data for this research, includes 111 patients that have suicide-related behaviours and under active treatment. Those cleaned data then were transformed into ten predictor variables and a response variable. Splitting training and testing data on those transformed data were done for building and accuracy evaluation of the method model. Based on the experiment, the best average accuracy at 63% can be obtained by using 30% of relevant data as data testing and by using training data which has one-to-one ratio in number between patients that have suicide-related behaviours and patients that have no such behaviours. In the future work, accuracy improvement need to be confirmed by using Reduced Support Vector Machine method, as the further development of Smooth Support Vector Machine.

Corresponding Author:

G. Indrawan,
Department of Computer Science,
Universitas Pendidikan Ganesha,
Jl. Udayana 11, Singaraja 81116, Bali, Indonesia.
Email: gindrawan@undiksha.ac.id

1.   INTRODUCTION

Suicide-related behaviours (SRBs) need to be prevented on psychiatric patients. SRBs includes suicide attempt or instrumental SRBs [1]. Suicide is the act of intentionally causing one's own death [2]. Risk factors include mental disorders such as depression, bipolar disorder, schizophrenia, personality disorders, alcoholism, or substance misuse [3], [4]. People have SRBs that do not result in death are at high risk for future self-injury and completed suicide [5], [6].

Prediction of those SRBs based on patient medical records would be very useful for the prevention by the psychiatric hospital. This research focused on developing this prediction at the only one psychiatric hospital of Bali Province by using Smooth Support Vector Machine (SSVM) method, as the further development of Support Vector Machine (SVM) [7]-[9]. According to [10], SVM utilizes quadratic programming optimization so that it is less efficient for high-dimensional and large data. Because of that, a developed smoothing technique is used to replace plus function of SVM by using integral of neural network sigmoid function. This smoothing technique is known as SSVM. When compared with SVM, SSVM has better running time and accuracy. The SSVM generated and solve an unconstrained smooth reformulation of the SVM for pattern classification using completely arbitrary kernel [8]. SSVM is solved by a Newton-Armijo algorithm and has been extended to nonlinear separation surfaces by using nonlinear kernel

techniques. The numerical results show that SSVM is faster than other methods and has better generalization ability [7].

2. SMOOTH SUPPORT VECTOR MACHINE

As a base of SSVM, SVM [11] is a method to find optimal hyperplane that separates two classes of input space. Separation of more than two classes have conducted previously by authors on fingerprint data [12]-[15]. Figure 1 shows several alternative hyperplanes (discrimination boundaries) and the best hyperplane of a data set consists of two classes, i.e. class {−1} and {+1}. The best hyperplane is the hyperplane which has a maximum margin obtained from alternative dividing lines (discriminant boundaries). Margin is the distance between the hyperplane to the nearest point of each class. This nearest point is so-called support vector [16].
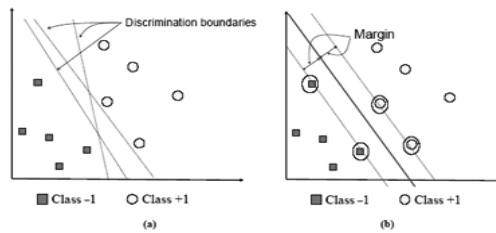

Figure 1. Several alternative hyperplanes (left), the best hyperplane (right)

Classification problem of m points in n-dimensional space (R^n) is represented as $m \times n$ -sized matrix **A.** The matrix element $\mathbf{A}_i^T$ to the class {−1} and {+1} is defined on $m \times m$ -sized diagonal matrix **D** with −1 and +1 at its diagonal. Linear SVM algorithm is shown by (1), with constrains $\mathbf{D}(\mathbf{Aw} - \mathbf{e}\gamma) + \mathbf{y} \geq \mathbf{e}$ and $\mathbf{y} \geq 0$; a positive value SVM parameter v; $m \times 1$ -sized slack variable vector **y** that measures classification error and has non-negative value; m-sized column vector **e** and has value of 1; $n \times 1$ -sized normal vector **w**; and bias value γ that determine hyperplane relative location to the original class.

$$\min_{(\mathbf{w},\, \gamma,\, \mathbf{y})\, \in\, R^{n+1+m}} v\mathbf{e}^T\mathbf{y} + \frac{1}{2}\mathbf{w}^T\mathbf{w} \tag{1}$$

The constrains equation above compares each vector element. When two classes can be separated perfectly by the defined hyperplane $\mathbf{x}^T\mathbf{w} + \gamma = 0$, there are two parallel hyperplane which are boundaries of those two classes, i.e. $\mathbf{x}^T\mathbf{w} + \gamma = -1$ of the class {−1} and $\mathbf{x}^T\mathbf{w} + \gamma = +1$ of the class {+1}. A non-linear hyperplane is obtained by transforming the standard SVM formulation (2), and by using "kernel trick" through a Gaussian kernel function (3), where μ is a kernel parameter and $i, j = 1, 2, ..., m$.

$$\mathbf{w} = \mathbf{A}^T\mathbf{Du} \tag{2}$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\mu\|\mathbf{x}_i - \mathbf{x}_j\|^2), \mu > 0 \tag{3}$$

By using (2) into (1), non-linear problem functions is obtained, as shown by (4), with constrains $\mathbf{D}(\mathbf{AA}^T\mathbf{Du} - \mathbf{e}\gamma) + \mathbf{y} \geq \mathbf{e}$ and $\mathbf{y} \geq 0$.

$$\min_{(\mathbf{w},\, \gamma,\, \mathbf{y})\, \in\, R^{n+1+m}} v\mathbf{e}^T\mathbf{y} + \frac{1}{2}\mathbf{u}^T\mathbf{DAA}^T\mathbf{Du} \tag{4}$$

The solution for the functions (4) is $K(\mathbf{x}^T, \mathbf{A}^T)\mathbf{Du} = \gamma$. By replacing $\mathbf{A}^T\mathbf{A}$ with non-linear kernel $K(\mathbf{A}, \mathbf{A}^T)$ and variable y is minimized by weighting $\frac{v}{2}$, generalized non-linear SVM is shown by (5) with constrains $\mathbf{D}(K(\mathbf{A}, \mathbf{A}^T)\mathbf{Du} - \mathbf{e}\gamma) + \mathbf{y} \geq \mathbf{e}$ and $\mathbf{y} \geq 0$.

$$\min_{(\mathbf{w},\, \gamma,\, \mathbf{y})\, \in\, R^{n+1+m}} \frac{v}{2}\mathbf{y}^T\mathbf{y} + \frac{1}{2}(\mathbf{u}^T\mathbf{u} + \gamma^2) \tag{5}$$

To solve (5), constraint function is defined by (6).

$$\mathbf{y} = (\mathbf{e} - \mathbf{D}(K(\mathbf{A}, \mathbf{A}^T)\mathbf{D}\mathbf{u} - \mathbf{e}\gamma))_+ \qquad (6)$$

By replacing (6) into (5), SVM problem equation is obtained which is equivalent to unconstrained SVM optimization, as shown by (7).

$$\min_{(\mathbf{u}, \gamma) \in R^{m+1}} \frac{v}{2}\|(\mathbf{e} - \mathbf{D}(K(\mathbf{A}, \mathbf{A}^T)\mathbf{D}\mathbf{u} - \mathbf{e}\gamma))_+\|_2^2 + \frac{1}{2}(\mathbf{u}^T\mathbf{u} + \gamma^2) \qquad (7)$$

At $(.)_+$, negative components are replaced by zeros. Equation (7) has a unique solution but its objective function is not twice differentiable which precludes the use of a fast Newton method. For that, a smoothing technique was proposed [10] that replaces plus function $(.)_+$ by using integral of sigmoid function $(1 + \exp(-\alpha\mathbf{x}))^{-1}$ of neural network. Equation (8) shows the SSVM where α is the smoothing parameter.

$$\min_{(\mathbf{u}, \gamma) \in R^{m+1}} \Phi_\alpha(\mathbf{w}, \gamma) = \min_{(\mathbf{u}, \gamma) \in R^{m+1}} \frac{v}{2}\| p(\mathbf{e} - \mathbf{D}(K(\mathbf{A}, \mathbf{A}^T)\mathbf{D}\mathbf{u} - \mathbf{e}\gamma), \alpha) \|_2^2 + \frac{1}{2}(\mathbf{u}^T\mathbf{u} + \gamma^2) \qquad (8)$$

Equation (8) can be optimized by using numerical approach through the Newton-Armijo method. The first step is to initiate $(\mathbf{w}^{(0)}, \gamma^{(0)}) \in R^{m+1}$ where $\mathbf{w}^{(i)}$ indicates ith iteration of $\mathbf{w}$. The second step is to repeat the iteration until the gradient of the objective function at (8) is equal to zero or $\nabla\Phi_\alpha(\mathbf{w}^{(i)}, \gamma^{(i)}) = 0$. The third step is to calculate $(\mathbf{w}^{(i+1)}, \gamma^{(i+1)}) = 0$ as follows:

a. Newton Direction: determine the direction of $d^{(i)} \in R^{n+1}$, as shown by (9).

$$\nabla^2\Phi_\alpha(\mathbf{w}^{(i)}, \gamma^{(i)})\, d^{(i)} = -\nabla\Phi_\alpha(\mathbf{w}^{(i)}, \gamma^{(i)})^T \qquad (9)$$

b. Armijo Stepsize: choose the stepsize $\lambda_i \in R$, such that

$$(\mathbf{w}^{(i+1)}, \gamma^{(i+1)}) = (\mathbf{w}^{(i)}, \gamma^{(i)}) + \lambda_i\, d^{(i)} \qquad (10)$$

where $\lambda_i = \max\{1, \frac{1}{2}, \frac{1}{4}, \dots\}$, such that

$$\Phi_\alpha(\mathbf{w}^{(i)}, \gamma^{(i)}) - \Phi_\alpha((\mathbf{w}^{(i)}, \gamma^{(i)}) + \lambda_i\, d^{(i)}) \geq -\delta\lambda_i\nabla\Phi_\alpha(\mathbf{w}^{(i)}, \gamma^{(i)})\, d^{(i)} \qquad (11)$$

where $\delta = (0, \frac{1}{2})$

When $\nabla\Phi_\alpha(\mathbf{w}^{(i)}, \gamma^{(i)}) = 0$, the Newton-Armijo algorithm iteration stopped and convergent value of $\mathbf{w}$ and γ were obtained for hyperplane function, as shown by (12).

$$f(\mathbf{x}) = sign\,(\mathbf{x}^T\mathbf{w} - \gamma) \qquad (12)$$

3. RESEARCH METHOD

For SRBs prediction using SSVM, five stages of research method were conducted, i.e.: 1) data preparation; 2) data transformation; 3) training and testing data selection; 4) SSVM model development; and 5) SSVM model evaluation.

Data preparation is related to the data collection from electronic and non-electronic medical record of the only one psychiatric hospital in Bali Province. There are 30.660 inpatient and outpatient medical record from the last five years up to April 2016. Data were collected through database query on the hospital information system and then they were exported to CSV format. Data cleaning gave 2665 relevant data for this research, includes 111 patients that have SRBs and under active treatment. Removed data have one or more than one of this three condition, i.e. 1) not psychiatric-disorder patient data (drug-free or psychiatric-disorder-free certificate applicant, dental patient, drug patient, neurology patient, or physiotherapy patient); 2) incomplete data (manual data that was migrated into information system and related patient has not been

inpatient or outpatient since data migration time); and 3) inactive patient data (pass away patient or not an outpatient).

Data transformation is related to the transformation of previous data into predictor variables and response variable. Ten predictor variables were obtained from database query above, i.e. disease diagnosis (x_1), profession (x_2), education (x_3), payment type / health insurance type (x_4), domicile (x_5), age (x_6), age range (x_7), sex (x_8), marital status (x_9), and family history (x_{10}). A response variable (y) is a variable with value −1 and +1 that represents class of patients that have no SRBs and class of patients that have SRBs, respectively. Response variable data was obtained from non-electronic medical record related to data of suicide attempt or instrumental SRBs [1]. Figure 2 shows research sample of predictor (instance) and response (label) matrix by using SSVM toolbox library [17]. Row and column of instance matrix represent a patient data and his/her related ten predictor variables, respectively.

Figure 2. Predictor and response matrix

Training and testing data selection is related to the next stage of SSVM model development and evaluation. Two data selection mechanisms were used to get the best hyperplane in Figure 1, i.e. by using:

a. Ten-folds cross validation (10-fcv) selection [18] and data ratio selection of 2665 relevant data. k-fold cross validation splits the data into k sections at random. Each section has the same class proportion to the initial class proportion. Each section will be used as training data and the rest is used as testing data, so there will be k accuracy. Final accuracy is the average of those k accuracy. On data ratio, data of patients that have SRBs and patients that have no SRBs were randomly selected in certain ratio for training and testing data, respectively. For an example, ratio 90:10 means 90% of relevant data as training data and 10% of relevant data as testing data. Training data consist of 90% of data of patients that have SRBs and 90% of data of patients that have no SRBs, while testing data consist of 10% of data of patients that have SRBs and 10% of data of patients that have no SRBs;

b. Data ratio selection based on 111 data of patients that have SRBs from 2665 relevant data. For an example, data ratio 1:2 means training data (also used as testing data) consist of data of patients that have SRBs and 222 data of patients that have no SRBs. Several best results of obtained SSVM models than were tested by using randomly selected data in number of 10%, 20% … and 100% of 2665 relevant data.

SSVM model development is related to parameter **w** and γ (12) that was computed by using SSVM toolbox library [17]. Several parameter **w** and γ was computed based on various training and testing data selection above. SSVM model evaluation is related to the classification accuracy that can be determined by using contingency table [19], as shown by Table 1. Based on that table, the classification accuracy can be measured by using (13).

Table 1. Classification accuracy contingency

Actual	Prediction	
	I (Negative)	II (Positive)
Negative	True Negative (TN)	False Positive (FP)
Positive	False Negative (FN)	True Positive (TP)

$$accuracy\ (\%) = \frac{TN + TP}{TP + FP + FN + TN} \qquad (13)$$

where TN is the number of prediction of patients that have no SRBs and in fact that patients have no such behaviours; TP is the number of prediction of patients that have SRBs and in fact that patients have such behaviours; FP is the number of prediction of patients that have SRBs and in fact that patients have no such behaviours; and FN is the number of prediction of patients that have no SRBs and in fact that patients have such behaviours.

4. RESULTS AND ANALYSIS

Experiment was based on previous two selection mechanisms of training and testing data, running on Intel Core™ i5-4460T CPU @1.90GHz with 4GB RAM and Windows 10 64-bit Operating System.

4.1. Data selection mechanism 1

Table 2 shows high accuracy of SSVM model on every data selection type but all of their *TP* were zero that make all of those model cannot be used for prediction of patients that have SRBs. High accuracy came from high number of *TN* (13) since many patients that have no SRBs are on the data in this research.

4.2. Data selection mechanism 2

Table 3 shows non-zero *TP*s by six SSVM models that were obtained by using six data ratio selections from 1:05 up to 1:1. Each of those SSVM models then were tested by using ten portions of 2665 data. So, each SSVM model will give ten accuracy result where its average accuracy is shown by Figure 3.

Table 2. SSVM Model Performance by using Data Selection Mechanism 1

Selection	Testing Data	w	γ	TN	FN	FP	TP	Acuracy	Time (s)
10-fcv	variable	0.01	0.0017	2554	111	0	0	0.9583	907.954
90:10	266	0.01	0.0018	255	11	0	0	0.9586	722.597
80:20	533	0.01	0.0019	511	22	0	0	0.9587	555.397
70:30	799	0.01	0.0016	766	33	0	0	0.9587	361.729
60:40	1065	0.01	0.0023	1021	22	0	0	0.9587	250.639
50:50	1282	0.01	0.0018	1227	55	0	0	0.9571	165.321

Table 3. SSVM Model Performance by using Data Selection Mechanism 2

Selection	Testing Data	w	γ	TN	FN	FP	TP	Acuracy	Time (s)
1:0.5	167	0.01	7.88E-05	8	5	48	106	0.6826	1.312
1:0.6	178	0.01	8.83E-05	10	5	57	106	0.6517	1.317
1:0.7	189	1.78E+03	5.79E-04	77	52	1	59	0.7196	1.627
1:0.8	200	1.78E+03	5.60E-04	86	53	3	58	0.72	1.978
1:0.9	211	1.00E+04	7.78E-05	96	49	4	62	0.7488	2.634
1:1	222	1.7783	2.67E-05	101	57	10	54	0.6982	2.094
1:2	333	0.0562	0.0323	222	111	0	0	0.6667	4.155
1:3	444	0.0562	0.0268	333	111	0	0	0.75	10.11
1:4	555	0.0562	0.0245	444	111	0	0	0.8	18.337
1:5	666	0.0562	0.0216	555	111	0	0	0.8333	32.273
1:6	777	0.01	0.002	666	111	0	0	0.8571	40.845
1:7	888	0.01	0.003	777	111	0	0	0.875	69.532
1:8	999	0.01	0.0022	888	111	0	0	0.8889	80.796
1:9	1110	0.01	0.0021	999	111	0	0	0.9	127.757
1:10	1221	0.01	0.0022	1110	111	0	0	0.9091	155.998
1:11	1332	0.01	0.0021	1221	111	0	0	0.9167	197.339
1:12	1443	0.01	0.0028	1332	111	0	0	0.9231	233.046
1:13	1554	0.01	0.0015	1443	111	0	0	0.9286	224.112
1:14	1665	0.01	0.0021	1554	111	0	0	0.9333	278.644
1:15	1776	0.01	0.0022	1665	111	0	0	0.9375	327.802
1:16	1887	0.01	0.0023	1776	111	0	0	0.9412	396.622
1:17	1998	0.01	0.0021	1887	111	0	0	0.9444	465.335
1:18	2109	0.01	0.0022	1998	111	0	0	0.9474	560.466
1:19	2220	0.01	0.0017	2109	111	0	0	0.95	576.861
1:20	2331	0.01	0.0023	2220	111	0	0	0.9524	637.395
1:21	2442	0.01	0.0024	2331	111	0	0	0.9545	768.092
1:22	2553	0.01	0.0023	2442	111	0	0	0.9565	860.427
1:23	2665	0.01	0.0017	2554	111	0	0	0.9583	907.954

Based on Figure 3, SSVM model generated by data ratio 1:1 (training data consist of 111 data of patients that have SRBs and 111 data of patients that have no SRBs) gave the best average accuracy at about 63%. On running time, most of the time was used for SSVM model development related to parameter **w** and γ (12). No significant increase on running time for different portion of relevant data, as shown by Table 4.
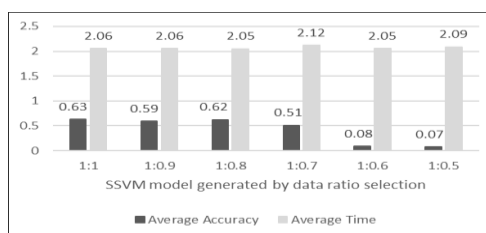

Figure 3. Six SSVM models performance by using six data ratio selections

Table 4. SSVM Model Performance by using Data Ratio 1:1

Data Portion	Testing Data	TN	FN	FP	TP	Acuracy	Time (s)
10%	266	170	4	85	7	0.6654	2.042
20%	533	394	12	117	10	0.75797	2.044
30%	799	672	19	94	14	0.8586	2.042
40%	1065	531	27	490	17	0.51455	2.202
50%	1383	860	28	467	28	0.6421	2.042
60%	1600	959	30	574	37	0.6225	2.034
70%	1866	818	38	970	40	0.4598	2.02
80%	2243	1207	45	947	44	0.5577	2.038
90%	2399	1320	53	979	47	0.5698	2.064
100%	2665	1490	57	1064	54	0.57936	2.058

Results above were based on ten predictor variables, as described previously. Table 5 gave the result about the influence of each of those variables to the SSVM model performance by using data ratio 1:1 and testing data at 30% of 2665 relevant data as shown in Table 4.

Table 5. SSVM Model Performance on Reduced Predictor Variables

No	Reduced Predictor Variables	TN	FN	FP	TP	Acuracy
1	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$	629	18	137	15	0.80601
2	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_{10}$	629	18	137	15	0.80601
3	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9, x_{10}$	629	18	137	15	0.80601
4	$x_1, x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{10}$	629	18	137	15	0.80601
5	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9, x_{10}$	619	18	147	15	0.79349
6	$x_1, x_2, x_3, x_4, x_6, x_7, x_8, x_9, x_{10}$	667	30	99	3	0.83855
7	$x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}$	759	32	7	1	0.95129
8	$x_1, x_2, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$	722	30	44	3	0.9074
9	$x_1, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$	737	33	29	0	0.9224
10	$x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$	716	32	50	1	0.8974

Table 5 shows that six predictor variables, i.e. disease diagnosis (x_1), profession (x_2), education (x_3), payment type/health insurance type (x_4), domicile (x_5), and age (x_6), have much influence to the SSVM model result because of the decreasing value of its *TP* and/or the increasing value of its *FP* without each of those variables. Hypothetically, age range (x_7), sex (x_8), marital status (x_9), and family history (x_{10}) have influence on the prediction. Age range between 19−45 years old (that was used as reference by the psychiatric hospital) apparently has relatively small influence in this SSVM model, neither do sex, marital status, nor family history, even though female or unmarried status or social network was considered to be the risk factors of SRBs [1].

5. CONCLUSION AND FUTURE WORK

Suicide-related behaviours (SRBs) prediction with SSVM gave the best average accuracy at 63%. This accuracy can be obtained by using 30% of 2665 relevant data as data testing and by using training data which have one-to-one ratio in number between patients that have SRBs and patients that have no SRBs. In the future work, accuracy improvement need to be confirmed by using Reduced Support Vector Machine (RSVM) method, as the further development of SSVM [10].

REFERENCES

[1]   J. L. Skeem, E. Silver, P. S. Aippelbaum, and J. Tiemann, "Suicide-Related Behavior after Psychiatric Hospital Discharge: Implications for Risk Assessment and Management", *Behavioral Sciences and the Law*, vol. 24, pp. 731-746, 2006.

[2]   T. L. Stedman, Stedman's Medical Dictionary, 28th ed. Philadelphia: Lippincott Williams & Wilkins, 2006.

[3]   K. Hawton and K. van Heeringen, "Suicide", *Lancet*, vol. 373, no. 9672, pp. 1372-1381, 2009.

[4]   WHO, "Suicide Fact Sheet," 2016.

[5]   G. K. Brown, A. T. Beck, R. Steer, and J. R. Grisham, "Risk Factors for Suicide in Psychiatric Outpatients: A 20-year Prospective Study", *Journal of Consulting and Clinical Psychology*, vol. 68, no. 5, pp. 371-377, 2000.

[6]   J. Cooper, *et al.*, "Suicide after Deliberate Self-Harm: A 4-year Cohort Study", *American Journal of Psychiatry*, vol. 162, no. 2, pp. 297-303, 2005.

[7]   Y. J. Lee and O. L. Mangasarian, "A Smooth Support Vector Machine for Cassification", *Journal of Computational Optimization and Applications*, pp. 5-22, 2001.

[8]   S. W. Purnami and A. Embong, "Smooth Support Vector Machine for Breast Cancer Classification", in *The 4th IMT-GT 2008 Conference of Mathematics, Statistics and Its Application (ICMSA 2008)*, 2008.

[9]   M. Furqan, A. Embong, A. Suryanti, S. W. Purnami, and S. Sajadin, "Smooth Support Vector Machine for Face Recognition using Principal Componen Analysis", in *2nd International Conference On Green Technology and Engineering (ICGTE)*, 2009.

[10] Y. J. Lee and O. L. Mangasarian, "RSVM:Reduced Support Vector Machine", in *The First SIAM International Conference on Data Mining*, 2001.

[11] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2006.

[12] G. Indrawan, S. Akbar, and B. Sitohang, "Review of Sequential Access Method for Fingerprint Identification", *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 10, no. 2, pp. 199-206, Jun. 2012.

[13] G. Indrawan, A. S. Nugroho, S. Akbar, and B. Sitohang, "A Multi-Threaded Fingerprint Direct-Access Strategy Using Local-Star-Structure-based Discriminator Features", *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 12, no. 5, pp. 4079-4090, 2014.

[14] G. Indrawan, S. Akbar, and B. Sitohang, "Fingerprint Direct-Access Strategy Using Local-Star-Structure-based Discriminator Features: A Comparison Study", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, no. 5, pp. 817-829, 2014.

[15] G. Indrawan, S. Akbar, and B. Sitohang, "On Analyzing of Fingerprint Direct-Access Strategies", in *International Conference on Data and Software Engineering (ICoDSE)*, 2014.

[16] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Application of Support Vector Machine in Bioinformatics", in *Indonesian Scientific Meeting in Central Japan*, 2003.

[17] DSMI, "Smooth Support Vector Machine Toolbox | Data Science and Machine Intelligence Lab", 2014. [Online]. Available: http://dmlab8.csie.ntust.edu.tw/#toolbox. [Accessed: 06-Apr-2017].

[18] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross Validation", *Encyclopedia of Database Systems*. Springer, 2009.

[19] S. W. Purnami, J. M. Zain, and T. Heriawan, "An Alternative Algorithm for Classification Large Categorical Dataset: k-mode Clustering Reduced Support Vector Machine", *International Journal of Database Theory and Application*, vol. 4, no. 1, 2011.

BIOGRAPHIES OF AUTHORS

Dr. G. Indrawan. Head of Computer Science Department, Graduate Program, Universitas Pendidikan Ganesha, Bali, Indonesia. He received his doctoral degree in Electrical Engineering and Informatics from Bandung Institute of Technology, Indonesia. His research interests include biometrics, pattern recognition, and robotics. He can be reached at gindrawan@undiksha.ac.id.

I K. P. Sudiarsa, M.Kom. IT Administrator of Psychiatric Hospital of Bali, Indonesia. He received his master degree in Computer Science from Universitas Pendidikan Ganesha, Bali, Indonesia. His research interests include machine learning, and data mining. He can be reached at kompassudiarsa@gmail.com.

Dr. K. Agustini. Vice Head of Instructional Technology Department, Graduate Program, Universitas Pendidikan Ganesha, Bali, Indonesia. She is also lecturer at Computer Science Department, Graduate Program, Universitas Pendidikan Ganesha, Bali, Indonesia. She received her doctoral degree in Education Technology from Jakarta State University and her master degree in Computer Science from Bogor Agricultural Institute. His research interests include pattern recognition, and learning media. She can be reached at ketutagustini@undiksha.ac.id.

Prof. Dr. Sariyasa. Head of Mathematics Education Department, Graduate Program, Universitas Pendidikan Ganesha, Bali, Indonesia. He is also lecturer at Computer Science Department, Graduate Program, Universitas Pendidikan Ganesha, Bali, Indonesia. He received his master and doctoral degree in Mathematics from Flinders University, Australia. His research interests include complex analysis, and numerical method. He can be reached at sariyasa@undiksha.ac.id.