

# Cervical cancer classification using convolutional neural network-support vector machine

Jane Eva Aurelia, Zuherman Rustam, Ihsya Wirasati

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia

## Article Info

### Article history:

Received Jul 22, 2020

Revised Jul 16, 2021

Accepted Aug 1, 2021

### Keywords:

Cervical cancer

Classification

Convolutional neural network

Machine learning

Support vector machine

## ABSTRACT

Cervical cancer is the second most common cancer in women worldwide, and occurs when there are presences of abnormal cells in the cervix, which continue to grow uncontrollably. In the early stages, cervical cancer indications are not perceptible; however, it is easily detected with different forms of machine learning methods, such as the convolutional neural network (CNN). This is a popular method with a wide range of applications and known for its high accuracy value. Moreover, there is a support vector machine (SVM) with several kernel functions that is commonly used in the classification of diseases, and also known for its high accuracy value. Therefore, the combination of CNN-SVM with several linear kernels functions as classifier for the categorization of cervical cancer.

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



## Corresponding Author:

Jane Eva Aurelia

Department of Mathematics

University of Indonesia

Prof. DR. Sudjono D. Puspongoro St., Pondok Cina, Depok, West Java 16424, Indonesia

Email: [corresp-author@mail.com](mailto:corresp-author@mail.com)

## 1. INTRODUCTION

Many countries ranked cancer as the second most common health issues, of which cervical cancer causes the most of deaths recorded [1]. Cervical cancer occurs when there are abnormal cells in the cervix, which continues to grow uncontrollably, and results in benign tumors, which later develops into cervical cancer cells that spread to other body parts [2].

This cancer is one of the most common disease in women throughout the world with nearly 500,000 women developing the disease each year, and ranked the fourth most communal malignant disease worldwide [3], [4]. In the initial stages, early cervical cancer and pre-cancer do not experience symptoms, since they do not show symptoms until the tumor is formed. Most cases are recorded in less developed countries with unavailability of effective screening systems [4].

Almost all cases are caused by human papillomavirus (HPV) and the risk factors include exposure to smoking, and immune-system dysfunction [4]. There are more than one hundred types of HPV; however, one of about 15 genotypes of carcinogenic HPV is very common among young women in their first sexual activity [5]. Besides carcinogenic risks that are linked to evolutionary species, each genotype acts as an independent infection [5].

In women's bodies, this virus produces 2 types of proteins, namely E6 and E7. Both of them are dangerous, since they deactivate certain genes that play crucial role in stopping tumor development. These two proteins also aggressively trigger the growth of uterine cell wall. This unnatural cell growth eventually causes gene changes or gene mutations, which then become the cause of cervical cancer that develops in the

body. The symptoms that characterized the disease are as follows, unusual bleeding from the vagina, irregular menstrual cycles, pain in the hip, low back pain, body weakness and tiredness, weight loss when not on a diet, loss of appetite, abnormal vaginal fluid, and leg inflammation.

Cervical cancers are mostly associated with the low and middle-income countries with approximately 90% HPV vaccination programs and unorganized screening [6]. The treatment depends on the level of the disease with respect to the available resources and diagnosis made in the early-stages. Fertility-preserving surgical procedures have been the care standard for women with low-risk. The overall prognosis remains poor for women with metastatic or recurrent disease. Yet, the period of survival is less than 12 months, however, the incorporation of the anti-vascular endothelial growth factor (VEGF) agent has been able to extend it.

## 2. RESEARCH METHOD

### 2.1. Convolutional neural network

Convolutional neural network (CNN) is a type of deep neural networks as a result of the multilayer perceptron (MLP) [7], [8]. The difference between CNNs and MLP is their ability of being used in the detection and recognition of objects in image forms. CNNs gives better results than neural networks (NNs), due to the addition of one layer to CNNs, which is known as the convolutional layer and consist of neurons with activation functions, bias, and weight [7]. CNNs is classified into two important parts which are, feature extraction and fully-connected layer [8], [9]. Illustration of CNNs is shown in Figure 1 [10].

#### 2.1.1. Feature extraction layer

Feature extraction layer "encodes" an image in the form of the object represented (feature extraction) [7]. Hence, CNNs is technically an architecture encompassing several stages, and each input and output process, features maps and numerous arrays, while the extraction layer individually comprises of two parts, as follows [11]-[13].

##### – Convolutional layer

Convolutional layer is the main structure of a convolutional neural network (CNN). This layer is utilized in the transformation of inputs into a form that is easily processed by going through a filter or kernel of a fixed size without losing essential convoluted features [14]. In this layer, there are filters (kernels) that spread to the entire input, and each unit receives input from the previous layer. Therefore, through convolution, the input map is generated between each filter, then shifting the input and using the sum of dot products.

##### – Pooling

Pooling is a technique for reducing dimensions with the aid of two common approaches namely the average and maximum pooling [15]. This operation is called the max pooling when it uses the highest value, while the average pooling uses the medial value. After this, the flattening process takes place, which is the reshaping of a pooled structure into a one-dimensional vector, then placed into fully-connected neural networks or MLP for classification [14].

##### – MLP layers

MLP layers is a fully connected multi-layer perceptron that performs the classification operation. There are three layers in MLP namely, the hidden layers, input and output layers. The activation function uses the rectified linear unit (ReLU), which is quite popular in deep learning due to its simplicity.

#### 2.1.2. Fully-connected layer

Fully-connected layer functions based on the feature extraction layer, which is a multidimensional array, with flatten (reshape) in the vector feature map [16], [17]. In addition, all active neurons from previous layer are linked with the next layer as in neural networks. Therefore, in order to connect properly, individual activation (of the previous) ought to be converted into 1-D data. These usually use MLP term, which process data with proper classification [18]. Meanwhile, the contrast against convolution layers are the neurons, which are connected to a specific input area, while fully-connected occurs in almost all parts. However, both continue to perform "dot product" operations; therefore, their functions are not significantly different.

### 2.2. Support vector machine

Support vector machine (SVM) has received much attention in the classification aspect [19]. The main field of this study is used to develop SVM algorithm based on the statistical learning theory [20]. SVM is also known as one of the effective machine learning and has high classification efficiency [20]. Illustration of SVMs is shown in Figure 2 [21].

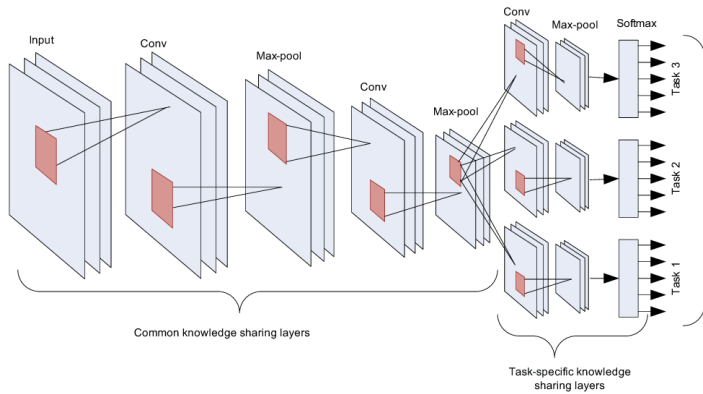


Figure 1. Illustration of convolutional neural network

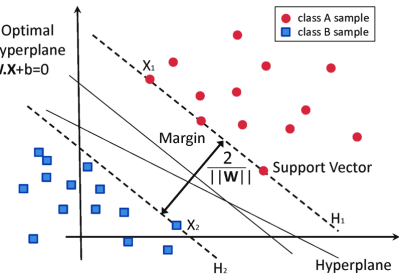


Figure 2. Illustration of support vector machine

SVM is a machine learning algorithm for classification and regression, which was introduced by Vapnik (1990) [22]. Then, Nello Cristianini researched about SVM based on Vapnik results [23]. Subsequently, Bernhard Scholkopf developed SVM theory and kernel function [24]. In addition, SVM is an initial form for binary classification; however, it is also for multiclass categorization.

SVM does mapping forms a higher dimensional space for supporting nonlinear classification, and constructing the maximal separating hyperplane. For instance, there is a set of firms represented by the value of their ratios  $\{x_i\}, i = 1, \dots,$  and a set of associated labels  $y_i \in \{-1, +1\}$  which describes results as failed or healthy.

The main purpose of SVM is to find the best hyperplane that is written as;

$$w \cdot x + b = 0 \tag{1}$$

The (1) above is able to maximize the margin.

The optimization problem of SVM is summarized as follow;

$$\text{Minimize } \frac{1}{2} ||w||^2 \tag{2}$$

Subject to;

$$y_i(w^T \cdot x_i + b) \geq 1, \forall i = 1, \dots, N \tag{3}$$

The (2) finds  $w \in R^n$  and  $b \in R^n$  with constrains to (3), along  $w$  (weights) and  $b$  (bias). Problem in (2) is quadratic optimization.

Therefore, the Lagrange multipliers  $\alpha_i$  for each of the constraints in (2) is shown by giving the function as;

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^N \alpha_i \{y_i(w \cdot x_i + b) - 1\} \tag{4}$$

where  $a = (a_1, a_2, \dots, a_n)^T$ .

When  $w$  and  $b$  equal to zero, setting the derivatives of  $L(w, b, \alpha)$  the equations obtained are,

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i x_i = 0 \rightarrow w = \sum_{i=1}^N \alpha_i y_i x_i \tag{5}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^N \alpha_i y_i = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \tag{6}$$

Then, eliminating  $w$  and  $b$  from  $L(w, b, \alpha)$  using (5) and (6), obtained the dual form as;

$$L(\alpha) = \max \left\{ -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{i=1}^N \alpha_i \right\} \tag{7}$$

$$\sum_{i=1}^N y_i \alpha_i = 0, \alpha_i \geq 0 \tag{8}$$

From (1) which is  $f(x) = w \cdot x + b$ , the  $w$  and  $b$  of regression function is finally obtained as follows;

$$\mathbf{w} = \sum_{i=1}^N a_i y_i x_i \quad (9)$$

$$b = \frac{1}{N_s} \sum_{i \in S} (y_i - \sum_{m \in S} a_m y_m x_m) \quad (10)$$

In this study, linear kernels are used for support vector machines (SVM) [25]. Kernel function resolves linear dimension problems and also for algorithms expression in the inner product between two vectors [25]. There are several kernel functions with their parameters in Table 1.

Table 1. Kernel function

Name	Kernel Function
Linier	$K(x_i, x_j) = \llbracket x_i \rrbracket^T x_j$
Polynomial	$K(x_i, x_j) = \llbracket (t + \llbracket x_i \rrbracket^T x_j) \rrbracket^d$
Gaussian Radial Basis Function (RBF)	$K(x_i, x_j) = \exp(-\ x_i - x_j\ ^2 / \sigma^2)$

### 2.3. Confusion matrix

Accuracy is one of main parameter that used to observe a classification's success. Refers to the percentage of correct answers at testing stage, confusion matrix used to measures the accuracy. The confusion matrix used is shown in Table 2 [25].

The formula of accuracy is written as:

$$accuracy = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (11)$$

$T_p$ : Number of samples having cervical cancer and classified correctly.

$F_p$ : Number of healthy individuals that are incorrectly classified to cervical cancer.

$F_N$ : Number of samples with cervical cancer that are incorrectly classified as healthy.

$T_N$ : Number of healthy individuals correctly spotted.

Table 2. Confusion matrix

Actual	Prediction	
	Positive	Negative
Positive	$T_p$	$F_p$
Negative	$F_N$	$T_N$

## 3. RESULTS AND ANALYSIS

### 3.1. Data

This paper received database of cervical cancer sufferers, which consisted of 652 informations with actual amounts of 607 major and 45 minor data. The minor represented the classes that indicated the presence of cervical cancer with label '1', while the major represented the classes that do not indicate the presence of cervical cancer with label '0'. There were 25 features used in this study, namely age, number of sexual partners, first sexual intercourse, number of pregnancies, smokes (years, packs/year), hormonal contraceptives (years), intrauterine device (years), sexually transmitted diseases (STD) (number, condylomatosis, vulvo-perineal condylomatosis, syphilis, human immunodeficiency virus (HIV), number of diagnosis), diagnosis (cancer, human papillomavirus (HPV)), hinselmann, schiller, and cytology.

### 3.2. Results

For the classification method, this research used 20% data for training and 80% data for testing. In this study, 1,000 amount of epochs were used for the convolutional neural network with the combination of several kernel functions used for the support vector machine. The results were shown in Figures 3, 4, and 5. In Figure 3 (a), there was a rise in the accuracy level of the model as many epochs increase. The blue line which stands for training data gave higher accuracy of 100%, while the orange line for testing data gave an accuracy value of 93.67%. Figure 3 (b) showed that the number of loss (error) decreases as the number of epochs decrease. The error found on training data was 0, while error on test data was 0.06.

Figure 4 (a), showed that the accuracy of the model increases as many epochs increase, since the blue line (training data) gave higher accuracy than the orange line (testing data). The accuracy of the training data was 100%, while for testing data it was 92.72%. Figure 4 (b) showed that the number of loss

(error) decreases as the number of epochs decrease. The error in training data was 0, while in testing data it was 0.07.

Figure 5 (a) showed that the accuracy of the model increases as many epochs increase. Then, the training data (blue line) gave higher accuracy than the testing data (orange line). The accuracy of training data was 100%, while the testing data was 92.91%. Besides this, Figure 5 (b) showed that the number of loss (error) decreases as the number of epoch decrease, which was 0 for training data and 0.07 on test data. Table 3 showed the accuracy of each method.

The comparison of convolutional neural network–support vector machine with some kernels for the classification of cervical cancer, were found to properly and correctly predict data. Result showed that the convolutional neural network–support vector machine with linear kernel had the best accuracy value of 93.67% on the test data. While on the training data, all methods gave best accuracy for database categorization. Therefore, the best method for the classification of cervical cancer is the convolutional neural network–support vector machine with linear kernel.

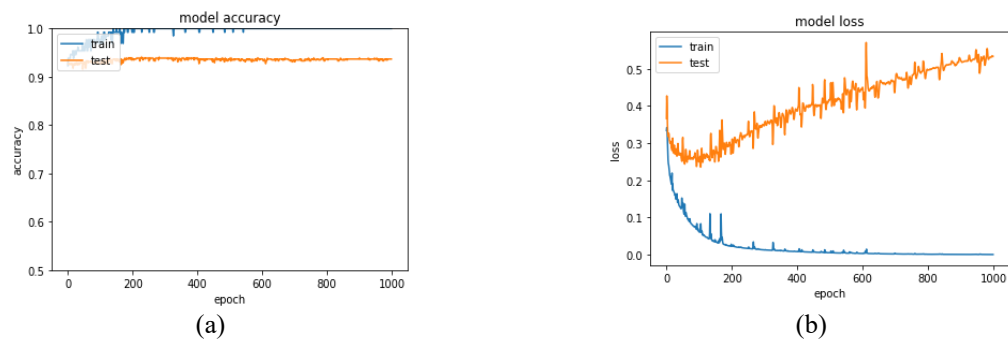


Figure 3. Convolutional neural network–support vector machine with linear kernel; (a) model accuracy of CNN–SVM linear kernel, and (b) model loss of CNN–SVM linear kernel

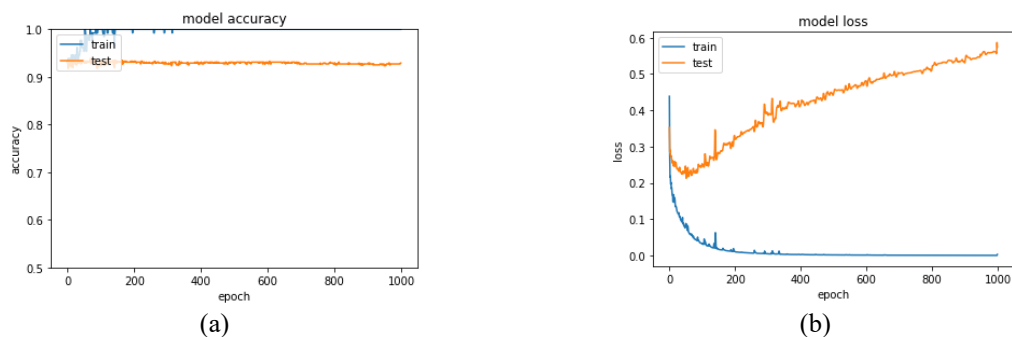


Figure 4. Convolutional neural network–support vector machine with polynomial kernel; (a) model accuracy of CNN–SVM polynomial kernel and (b) model loss of CNN–SVM polynomial kernel

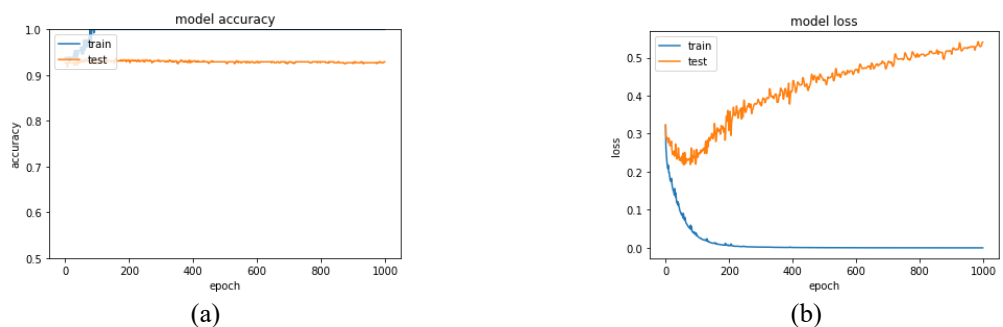


Figure 5. Convolutional neural network–support vector machine with Gaussian RBF kernel; (a) model accuracy of CNN–SVM Gaussian RBF kernel, and (b) model loss of CNN–SVM Gaussian RBF kernel

Table 3. The accuracy of each method

No	Model	Training Accuracy	Testing Accuracy
1.	CNN-SVM Linear	100%	93.67%
2.	CNN-SVM Polynomial	100%	92.72%
3.	CNN-SVM RBF	100%	92.91%

#### 4. CONCLUSION

Predicting the presence of disease by diagnosing with machine learning method help medical staff to classify ailments. An early detection of disease is important, since it makes the patient to receive a prompt right treatment, which helps to increase the chance of survival and reduce the health risk. Therefore, this research focuses on cervical cancer which is a common health problem with 652 data collected and 25 features observed. The method used was the combination of convolutional neural network-support vector machine with several kernel functions as classifier. The experimental results showed that the methods used, properly and correctly predicted the data. Based on findings, convolutional neural network-support vector machine with linear kernel is the best model for the classification of cervical cancer data as shown in Table 3. Therefore, in the future research, this method develops to give higher accuracy and uses a larger database, in order to give better results for predicting and classifying different diseases.

#### ACKNOWLEDGEMENTS

This research was supported financially by the Indonesian Ministry of Research Technology; with a KEMENRISTEK/BRIN PDUPT 2021 research grant scheme, ID number NKB-2830/UN2.RST/HKP.05.00/2020.

#### REFERENCES

- [1] N. Salmi and Z. Rustam, "Naïve Bayes Classifier Models for Predicting the Colon Cancer," 2019 *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052068.
- [2] A. A. Rachman and Z. Rustam, "Cancer Classification using Fuzzy C-Means with Feature Selection," 2016 *12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA)*, 2016, pp. 31-34, doi: 10.1109/ICMSA.2016.7954302.
- [3] R. Geetha, S. Sivasubramanian, and M. Kaliappan, "Cervical Cancer Identification with Synthetic Minority Oversampling Technique and PCA Analysis using Random Forest Classifier," 2019 *J Med Syst.*, vol. 43, no. 9, pp. 1-19, doi: 10.1007/s10916-019-1402-6.
- [4] S. E. Waggoner, "Cervical Cancer," 2003 *The Lancet*, vol. 361, no. 9376, pp. 2217-2225, 2003, doi: 10.1016/S0140-6736(03)13778-6.
- [5] M. Schiffman, P. E. Castle, J. Jeronimo, A. C. Rodriguez, and S. Wacholder, "Human Papillomavirus and Cervical Cancer," *The Lancet*, vol. 370, no. 9590, pp. 890-907, 2007, doi: 10.1016/S0140-6736(07)61416-0.
- [6] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, "Cervical Cancer," *The Lancet*, vol. 393, no. 10167, pp. 169-182, 2019, doi: 10.1016/S0140-6736(18)32470-X.
- [7] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," 2015 *arXiv preprint arXiv:1511.08458*.
- [8] S. C. Turaga, "Convolutional Networks can learn to Generate Affinity Graphs for Image Segmentation," *Neural computation*, vol. 22, no. 2, pp. 511-538, 2010, doi: 10.1162/neco.2009.10-08-881.
- [9] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-Column Deep Neural Networks for Image Classification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642-3649, doi: 10.1109/CVPR.2012.6248110.
- [10] T. L. Nwe, T. H. Dat, and B. Ma, "Convolutional Neural Network with Multi-Task Learning Scheme for Acoustic Scene Classification," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017, pp. 1347-1350, doi: 10.1109/APSIPA.2017.8282241.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image Net Classification with Deep Convolutional Neural Networks," *Advances Neural Information Processing Systems*, vol. 25, no. 2, pp. 1907-1105, 2012, doi: 10.1145/3065386.
- [12] D. C. Cireşan, "High Performance Convolutional Neural Networks for Image Classification," *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, 2011, pp. 1237-1242.
- [13] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional Neural Network Committees for Handwritten Character Classification," *IEEE Conference on Document Analysis and Recognition*, 2011, pp. 1135-1139, doi: 10.1109/ICDAR.2011.229.
- [14] X. S. Yang, "Introduction to Algorithms for Data Mining and Machine Learning," *Academic press*, 2019.
- [15] B. Navaneeth and M. Suchetha, "PSO Optimized 1-D CNN-SVM Architecture for Real-Time Detection and Classification Applications," *Computers in Biology and Medicine*, vol. 108, pp. 85-92, 2019, doi: 10.1016/j.compbiomed.2019.03.017.
- [16] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, "Pedestrian Detection with Convolutional Neural Networks," *Intelligent Vehicles Symposium IEEE*, 2005, pp. 224-229, doi: 10.1109/IVS.2005.1505106.

- [17] C. Szegedy, A. Toshev, and D. Erhan, "Deep Neural Networks for Object Detection," *Advances in Neural Information Processing Systems*, pp. 2553-2561, 2013.
- [18] F. H. C. Tivive and A. Bouzerdoum, "A New Class of Convolutional Neural Networks (Siconnets) and their Application of Face Detection," *Proceedings of The International Joint Conference*, vol. 3, 2003, pp. 2157-2162, doi: 10.1109/IJCNN.2003.1223742.
- [19] Z. Rustam and F. Yaurita, "Insolvency Prediction in Insurance Companies Using Support Vector Machines and Fuzzy Kernel C-Means," *J. Phys.: Conf. Ser.*, vol. 1028, no. 1, 2018, doi: 10.1088/1742-6596/1028/1/012118.
- [20] Z. Rustam and N. P. A. A. Ariantari, "Support Vector Machines for Classifying Policyholders Satisfactorily in Automobile Insurance," 2018 *J. Phys.: Conf. Ser.*, vol. 1028, no. 1, doi: 10.1088/1742-6596/1028/1/012005.
- [21] E. G. Gonzalo, Z. F. Muñoz, P. J. G. Nieto, A. S. Bernardo, and M. M. Fernandes, "Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers," *Materials*, vol. 9, no. 7, 2016, doi: 10.3390/ma9070531.
- [22] Z. Rustam and D. Zahras, "Comparison between Support Vector Machine and Fuzzy C-Means as Classifier for Intrusion Detection," *J. Phys.: Conf. Ser.*, vol. 1028, no. 1, 2018, doi: 10.1088/1742-6596/1028/1/012227.
- [23] H. K. Penawar and Z. Rustam, "A Fuzzy Logic Model to Forecast Stock Market Momentum in Indonesia's Property and Real Estate Sector," 2016 *AIP Conference Proceedings*, vol. 1862, no. 1, 2016, doi: 10.1063/1.4991229.
- [24] N. Christianini and J. S. Taylor, "An Introduction to Support Vector Machines and Other Kernel based Learning Methods," *Cambridge university press*, 2000.
- [25] Z. Rustam, D. A. Utami, R. Hidayat, J. Pandelaki and W. A. Nugroho, "Hybrid Preprocessing Method for Support Vector Machine for Classification of Imbalanced Cerebral Infarction Datasets," 2019 *International Journal on Advanced Science Engineering Information Technology*, vol. 9, no. 2, doi: 10.18517/ijaseit.9.2.8615.

## BIOGRAPHIES OF AUTHORS



**Jane Eva Aurelia** was born in Jakarta, 19 June 1998. She is a final year student in the Department of Mathematics, University of Indonesia. She is currently working on her thesis, which is firmly about applied mathematics using machine learning. Also, Ms. Jane's specialties in research are mostly about machine learning, mathematical modeling, and data mining.



**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.



**Ilsya Wirasati** is a final year student in the Department of Mathematics, University of Indonesia, who is currently working on her thesis. Her research is firmly about applied mathematics using machine learning in medical field. Ms. Ilsya's specialties in research are mostly about machine learning, mathematical modeling, and data mining