NEURAL NETWORK OPTIMIZATION BASED ON PARTICLE SWARM OPTIMIZATION AND BAGGING METHODS ON CLASSIFICATION OF SINGLE PAP SMEAR IMAGE CELLS

Robi Aziz Zuama[1]; Irwan Agus Sobari[2]

[1]Information System
Universitas Bina Sarana Informatika, Jakarta, Indonesia
www.bsi.ac.id
robi.rbz@bsi.ac.id

[2] Informatics Engineering
STMIK Nusa Mandiri, Jakarta, Indonesia
www.nusamandiri.ac.id
irwan.igb@nusamandiri.ac.id

***Abstract**— In this study, the analysis of automatic diagnosis of pap smear image extraction using neural network algorithm, the analysis includes a review of the results of pap smear Herlev image extraction with level 7 classes, 2 normal and abnormal classes, 3 normal level dysplasia classes and 4 abnormal level dysplasia classes. The problem is that the neural network is very difficult to designate optimal features in diagnosis and it is difficult to deal with class imbalances. This study proposes a combination of particle swarm optimization (PSO) to optimize bagging features and methods to deal with class imbalances, with the aim that the diagnosis results using a neural network can improve its accuracy. The results show that using PSO and bagging methods can improve the accuracy of the network neural algorithm. At level 7 classes could increase by 1.64%, 2 classes increased by 0.44%, 3 classes increased by 2.04%, and at level 4 classes increased by 5.47%.*

***Keywords:** Cervical Cancer, Pap Smear, Neural Network Algorithm, Particle Swarm Optimization, Bagging.*

Intisari—Dalam penelitian ini memaparkan analisis diagnosis otomatis hasil ektrasi citra pap smear menggunakan algoritma neural network, analisis meliputi tinjauan pada hasil ektraksi citra pap smear herlev dengan level 7 kelas, 2 kelas normal dan abnormal, 3 kelas dysplasia level normal dan 4 kelas level dysplasia abnormal. Permasalahanya neural network sangat susah menuntukan fitur yang optimal dalam pendiagnosisan dan susah dalam menangani ketidakseimbangan kelas. Penelitian ini mengusulkan kombinasi PSO untuk optimalkan fitur dan penanganan kelas yang tidak seimbang oleh metode bagging dengan tujuan agar hasil diagnosis menggunakan neural network dapat meningkat hasil akurasinya. Hasil menunjukan, bahwa dengan menggunakan PSO dan metode bagging dapat meningkatkan akurasi algoritma nerual network. Pada level 7 kelas dapar meningkat sebesar 1,64%, 2 kelas meningkat sebesar 0,44%, 3 kelas meningkat sebesar 2,04%, dan pada level 4 kelas meningkat sebesar 5,47%.

Kata Kunci: Kanker serviks, pap smear, Neural Network, Particle Swarm Optimization, Bagging.

INTRODUCTION

The World Health Organization (WHO) states that the most common cancer that is No. 2 exist in women is cervical cancer. This cancer usually exist on the territory of low- and middle-income countries (World Health Organization, 2013), (UNAIDS, 2016). More than 270,000 women in 2012 in the world died of cervical cancer, cervical cancer itself caused by the Human Papillomavirus (HPV) virus which is transmitted by sexual contact.

In 2018, WHO through the International Agency for Research on Cancer through GLOCOBAN 2018 data identified that there were around 32,469 new cases of the uterine cervix or about 9.3% so the number of cancer patients in Indonesia, that infects with an aged range 15 to 44 years since 2018 (Bruni et al., 2017). One of the controls of the spread of cervical cancer by doing screening tests to take cervical foot images of all women who are at risk of cervical cancer (WHO, 2013).

Cervical cells can be identified by analyzing the microscopic image of tissue cells of cervical

microorganisms with a retrieval procedure called pap smear, from the results of the analysis it is important to evaluate the gradation and prognosis of the disease. At present analysis is still mostly done by experts, which results in the detection process being slow and the results of accuracy are sometimes less accurate depending on the expertise of an expert (Ruberto & Putzu, 2016).

Many studies on the method of automatic detection of pap smear images have been carried out such as research (Riana et al., 2017) that integrates the ensemble (bagging) method to handle imbalance classes and use the gray forward selection for feature selection to improve the results of the pap smear classification results (Kristiyanti & Wahyudi, 2017). by using the Naïve Bayes algorithm. Research (Mesquita et al., 2018) analyzed the pap smear image using LDA and SVM algorithms. The study (Herliana, 2016) compared the Naïve Bayes (NB) classification method with C4.5 optimized with CFS (Correlation Features Selection).

Neural networks (NN) have the advantage of strong fault tolerance and strong capabilities of dynamic nonlinear data processing, more appropriate in the case of many complex attributes and algorithms, but the practicality of NN is limited, due to the difficulty of choosing suitable network parameters, including the number hidden neurons (William et al., 2018) and the main problem in neural networks is overfitting (Srivastava et al., 2014).

Also, the image of pap smear cervical cancer has an unbalanced class problem (Riana et al., 2017) which causes a reduction in the performance of the classification algorithm. One technique used to deal with class imbalances is the ensemble technique. The most popular ensemble algorithm, namely boosting and bagging techniques, AdaBoost or adaptive boosting is one of the boosting techniques that can improve the performance of classifiers (Saifudin & Wahono, 2015).

Bagging can overcome class imbalances in the data to improve classification by a combination of classifications and the complexity of class imbalances by reducing variance by avoiding overfitting (Wahono, Herman, et al., 2014)(Herman & Giuliani, 2018).

Feature selection is generally used in machine learning (Wahono, Herman, et al., 2014) when classification algorithms involve high dimensions, Most feature selection algorithms try to find solutions in feature selection that range between suboptimal and almost optimal regions because they use local search in the whole process, not global search. The most widely used

metaheuristic optimization for feature selection (Wahono, Suryana, et al., 2014) includes genetic algorithm , particle swarm optimization and ant colony optimization (Gocic et al., 2016). Research by Bora et. al (Bora et al., 2016) developed an automation system for cervical cancer diagnosis using cervical cancer cytology images to propose neural network classification algorithms, the feature used for diagnosis is the ratio of nucleus to cytoplasm, shape and color intensity together with core, circumference and eccentricity areas. Research by Gomez et. al (Gomez et al., 2017) Developing an intelligent system for automatic categorization of pap smear images to detect cervical dysplasia levels, classification techniques based on shape (Taha et al., 2017) characteristics, texture and color features and classifying them into two levels namely normal/abnormal and three dysplasia class. The system was trained and tested on a dataset from Herlev University (Nirmal Jith et al., 2018). An ensemble classifier was designed using LSSVM, MLP, and Random Forest using Weighted Majority Voting (Bora et al., 2017), accuracy results obtained using 2 classes at 96.51%, and for 3 classes at 91.71%.

Research by Chatterjee (Chatterjee et al., 2017) uses assemble algorithms (Bagging and AdaBoost) to build Classifier, data used from Herlev University Hospital, the results of experiments carried out by combining bagging + Multilayer Perceptron and AdaBoost + LMT methods that produce 95.74 % for normal and abnormal classification using 20 attributes & formulated according to the data set from Herlev namely binary classification problem.

In previous studies, analysis of machine learning algorithms for automatic detection of pap smear images was proposed, such as neural networks, LSSVM, MLP, Random Forest, and others. The neural network is one of the proposed linear machine algorithms for automatic detection of pap smear images because of its superiority in handling high dimensions. In previous studies, machine learning algorithm analysis for automatic detection of pap smear images was widely proposed such as neural networks, LSSVM, MLP, Random Forest, etc. The neural network is one of the proposed machine alignment algorithms for automatic detection of pap smear images because of its superiority in handling high dimensions.

This study will propose a machine learning neural network algorithm by integrating Particle Swarm Optimization (PSO) for the selection of the best features of the 20 features from the feature extraction results and adding the Bootstrap Aggregating (Bagging) method that will deal with

class imbalances in the Pap smear image extraction results. The difference from this research is to compare some data mining methods to see the accuracy results that will be obtained from the results of feature selection and handling class imbalance in the Pap smear image dataset from Herlev Hospital University. The purpose of this study is to obtain a high accuracy value of neural network algorithms in predicting the diagnosis of pap smear microscopic image automation, with the help of ensemble particle swarm optimization models and bagging methods.

MATERIALS AND METHODS

A.   Materials

The dataset of single pap smear cell images was collected at Herlev Hospital University, data taken by skilled pantologists using a microscope with a resolution of 0.201μm pixels to take digital images of a single cell (Norup, 2005 p.15). Each cell image is then classified manually into 7 different cell types which are described in Figure 1 for normal data and Figure 2 for abnormal data. Table 1 shows the distribution of a pap smear data set with normal classes and table 2 shows the distribution of pap smear data with abnormal classes.

In this study, the research object was extracted based on quantitative characteristics, which were obtained from public Helev Hospital University data developed by Norup (2005) taken from the website http://mde-lab.aegean.gr/index.php/downloads. The nucleus is large, dark in color, and irregular in shape (Dewi et al., 2018). The cytoplasm is dark in color. Comparison of the area of the nucleus with the area of the cytoplasm is very large, for more details see figure 1 (normal) and picture 2 (abnormal).

Table 1 Distribution of normal class smear data

Normal – 242 Sel	
Superficial squamous epithelial	74 Sel
Intermediate squamous epithelial	70 Sel
Columnar epithelial	98 Sel

Source: (Zuama & Sobari, 2020)

Table 2 Distribution of abnormal class smear data

Abnormal – 675 Sel	
Mild squamous non-keratinizing dysplasia	182 Sel
Moderate squamous non-keratinizing dysplasia	146 Sel
Severe squamous non-keratinizing dysplasia	197 Sel
Squamous cell carcinoma in situ intermediate	150 Sel

Source: (Zuama & Sobari, 2020)
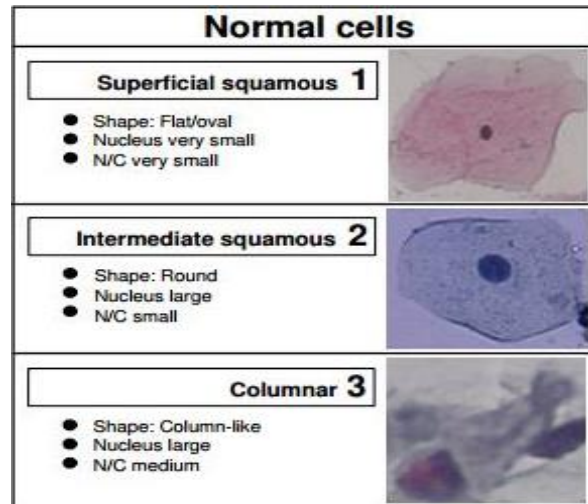
Source: (Zuama & Sobari, 2020)
Figure 1. Cell type characteristics for normal single Pap smear cells
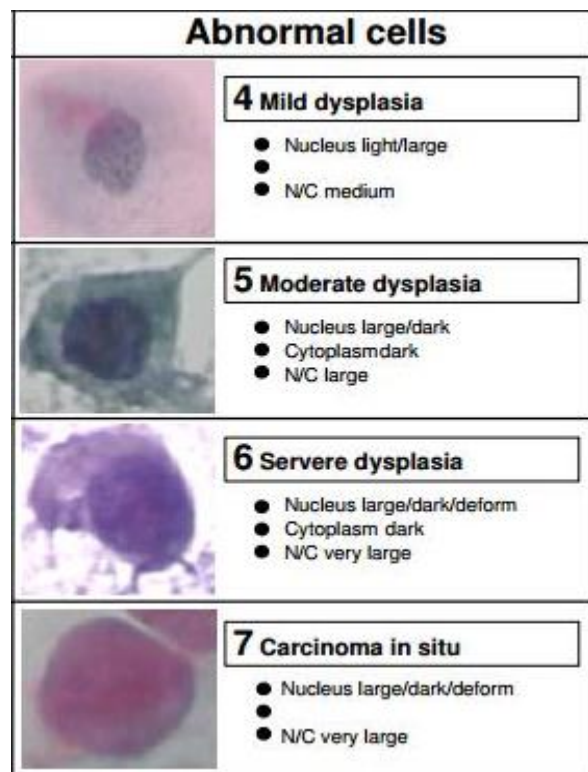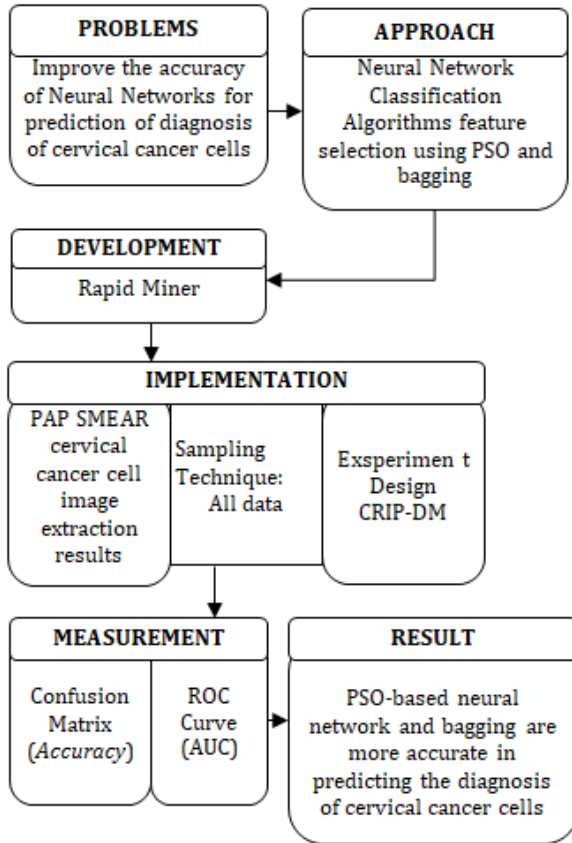

Source: (Zuama & Sobari, 2020)
Figure 2. Cell type characteristics for abnormal single Pap smear cells

B.   Methods

The study began by identifying the existing problem, which is not yet known whether using the PSO and Bagging feature selection to deal with class imbalances can improve the accuracy of the classification algorithm and which algorithm produces the best accuracy.

CRISP-DM is a standard data mining process is used to get the rules of each algorithm, the rules are then validated using the confusion matrix and the ROC curve.

Source: (Zuama & Sobari, 2020)
Figure 3. Research Framework

The results of the validation of each algorithm are then compared to find the best rule that has the highest accuracy. The selected rule will be used to apply to new data. Figure 3 shows the skeleton model used in this study. The study was conducted using a dataset obtained from Herlev Hospital University and can be downloaded on the page:
http://mde-lab.aegean.gr/index.php/downloads.

RESULTS AND DISCUSSION

A. Level Seven Class

Comparative results of the comparison of the use of neural network and neural network + PSO & Bagging algorithms are presented in a confusion matrix table. The results show the performance of the neural network algorithm optimized by PSO as feature selection and bagging method use in recall (see. Table 3) has increased by an average of 2.12% shown in the table and the precision table (see table 4) has increased in average 2.56%.

From table 5 also shows the results of the accuracy of neural network algorithms optimized by the PSO algorithm and the bagging method can increase accuracy by a difference of 1.64%, Figure 4 visualizes the difference in accuracy between neural networks and neural networks + PSO and Bagging.

Table 1. Recall level 7 class

Algoritma	Recall						
	Superficial	*Intermediate*	*Columnar*	*Mild Squamous*	*Moderate Squamous*	*Severe squamous*	*Squamous cell*
Neural Network	86,49%	91,43%	69,39%	73,08%	42,47%	50,25%	58,00%
NN + PSO & Bagging	90,54%	92,86%	76,53%	75,82%	42,47%	49,75%	58,00%

Source:(Zuama & Sobari, 2020)

Table 2. Precision level 7 class

| Algoritma	Precision						
	Superficial	Intermediate	Columnar	Mild Squamous	Moderate Squamous	*Severe squamous*	Squamous cell
Neural Network	94,12%	82,05%	66,02%	69,63%	47,33%	51,03%	57,24%
NN + PSO & Bagging	94,37%	89,04%	75,76%	68,66%	49,60%	48,76%	59,18%

Source: (Zuama & Sobari, 2020)

Table 3. Comparison of accuracy results at level 7 class

Algoritma	Accuracy

Neural Network	62,92%
NN + PSO & Bagging	64,56%

Source:(Zuama & Sobari, 2020)
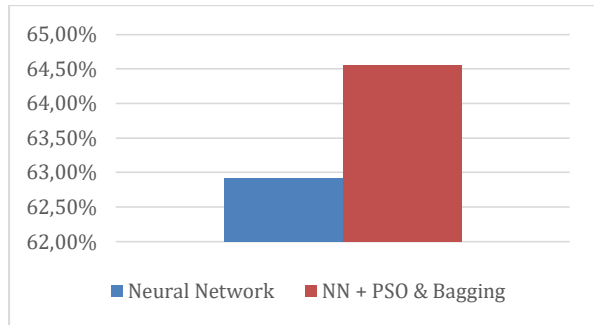
Source: (Zuama & Sobari, 2020)
Figure 4. Visualization of accuracy results at level 7 class
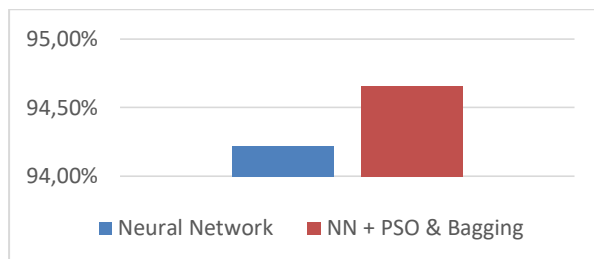
B.   Level 2 class
From the results of the analysis it is proven that the NN + PSO and Bagging algorithm can improve the accuracy of the results of the Neural Network classification on a prediction of normal and abnormal pap smear levels. The accuracy value for Neural itself is 94.22% and the accuracy value of NN + PSO and Bagging is 94.66% with a difference of 0.44% accuracy can be seen in Table 6 and the visualization results can be seen in Figure 5.

Table 4. Comparison of accuracy results at level 2 class

| Accuracy	
Neural Network	94,22%
NN + PSO & Bagging	94,66%

Source: (Zuama & Sobari, 2020)

Source: (Zuama & Sobari, 2020)
Figure 5. Visualization of accuracy results at level 2 class

To evaluate using the ROC curve so that the AUC (Area Under Curve) value for the Neural Network classification algorithm model produces a value of 0.9422 with an Excellent Classification diagnostic value, whereas for the Neural Network

+ PSO & Bagging classification algorithm produces a value of 0.9466 with a diagnosis value of Excellent Classification, and the difference the value of both is 0.044, Roc curve can be seen in Figures 6 & 7.
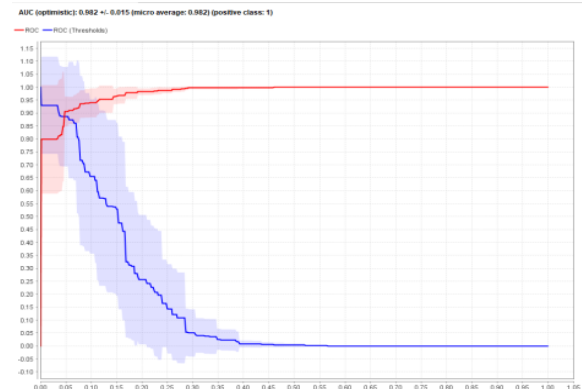
Source: (Zuama & Sobari, 2020)
Figure 6. Roc Curve Neural Network
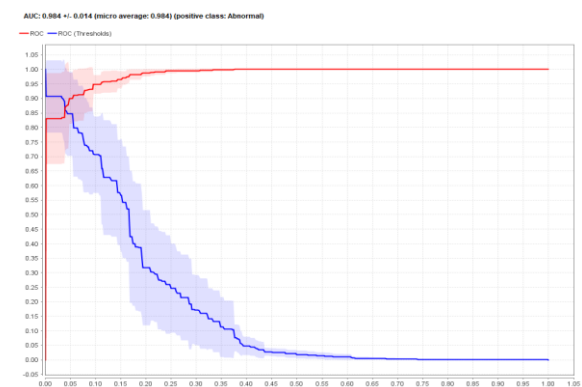
Source: (Zuama & Sobari, 2020)
Figure 7. Roc Curve Neural Network + PSO & Bangging

CONCLUSION

From the results of the analysis presented, it can be concluded that by using PSO to optimize parameters and bagging methods to deal with the imbalance classes of the Pap Smear dataset the neural network algorithm can be improved its accuracy. This means that a neural network can improve its accuracy by using the PSO and Bagging algorithm at the level of seven classes, two classes, three classes, and four classes. At the level of seven classes the increase in accuracy of the neural network by a difference of 1.64%, at the level of two classes the increase is not too significant only increase accuracy by a difference of 0.44%, at the level of three classes the increase is very significant with a difference of 2.04% and at level four class also gained a significant increase of 5.47%. In the optimization of the class classification of two classes, three classes have

been said to be excellent classification because each accuracy result obtained is 94.66% for two classes using neural network + PSO and bagging and 94.60% at level 3 class using the NN + PSO algorithm & bagging. Whereas at level 7 classes and 4 classes are still said to be poor classification because each NN algorithm gets 64.56% and 59.54% accuracy even though it has been optimized using PSO and bagging methods.

REFERENSI

Bora, K., Chowdhury, M., Mahanta, L. B., Kundu, M. K., & Das, A. K. (2016). Pap smear image classification using convolutional neural network. *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing - ICVGIP '16*, 1–8. https://doi.org/10.1145/3009977.3010068

Bruni, L., Albero, G., Serrano, B., Mena, M., Gómez, D., Muñoz, J., … Information, I. C. on H. and C. (HPV. (2019). *Human Papillomavirus and Related Diseases in the World*. (October). Retrieved from www.hpvcentre.com

Bruni, L., L, B.-R., Albero, G., Serrano, B., Mena, M., Gómez, D., … De, S. S. (2017). *Human Papillomavirus and Related Diseases Report in Indonesia*. (July), 72. Retrieved from http://www.hpvcentre.net/statistics/reports/MYS.pdf

Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A. S., & Balas, V. E. (2017). Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. *Neural Computing and Applications*, *28*(8), 2005–2016. https://doi.org/10.1007/s00521-016-2190-2

Di Ruberto, C., & Putzu, L. (2016). A feature learning framework for histology images classification. In *High Temperature*.

Gomez, O. H., DelaCruz, E. S., & Mata, A. P. de la. (2017). Classification of Cervical Cancer Using Assembled Algorithms in Microscopic Images of Papanicolaou. *Research in Computing Science*, *139*(2017), 125–134.

Herliana, A. (2016). *Optimasi Klasifikasi Sel Tunggal Pap Smear Menggunakan Correlation Based Features Selection (Cfs) Berbasis C4 . 5 Dan Naive Bayes*. *3*(September), 148–155.

Mesquita, J. joaci de, Backes, A. R., & Bruno, O. M. (2018). *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 10657*, 677–684. https://doi.org/10.1007/978-3-319-

75193-1

Norup, J. (2005). *Classification of Pap-smear data by transductive neuro-fuzzy methods.*

Riana, D., Hidayanto, A. N., & Fitriyani. (2017). Integration of Bagging and greedy forward selection on image pap smear classification using Naïve Bayes. *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*. https://doi.org/10.1109/CITSM.2017.8089320

Saifudin, A., & Wahono, R. S. (2015). Penerapan Teknik Ensemble untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, *1*(1), 28–37. https://doi.org/10.1016/S1896-1126(14)00030-3

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958. https://doi.org/10.1214/12-AOS1000

Wahono, R. S., Herman, N. S., & Ahmad, S. (2014). Neural network parameter optimization based on genetic algorithm for software defect prediction. *Advanced Science Letters*, *20*(10–12), 1951–1955. https://doi.org/10.1166/asl.2014.5641

WHO. (2013). Guidelines for screening and treatment of precancerous lesions for cervical cancer prevention. *WHO Guidelines*, 60. Retrieved from http://www.who.int/reproductivehealth/publications/cancers/screening_and_treatment_of_precancerous_lesions/en/index.html

William, W., Ware, A., Basaza-Ejiri, A. H., & Obungoloch, J. (2018). A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Computer Methods and Programs in Biomedicine*, *164*, 15–22. https://doi.org/10.1016/j.cmpb.2018.05.034

World Health Organization. (2014). Comprehensive Cervical Cancer Control: a healthier future of girls and women. *WHO Library Cataloguing-in-Publication Data*, 364. Retrieved from http://www.who.int/reproductivehealth/publications/cancers/cervical-cancer-guide/en/

Zuama, R. A., & Sobari, I. A. (2020). *Optimasi Neural Network Dengan Particle Swarm Optimization Dan Metode Bagging Pada Klasifikasi Sel Tunggal Citra Pap Smear*. STMIK Nusa Mandiri.