

Improving Hierarchical Decision Approach for Single Image Classification of Pap Smear

Dwiza Riana¹, Yudi Ramdhani², Rizki Tri Prasetyo³, Achmad Nizar Hidayanto⁴

¹STMIK Nusa Mandiri, Indonesia

^{2,3}Universitas BSI, Indonesia

⁴Universitas Indonesia, Indonesia

Article Info

Article history:

Received Mar 28, 2018

Revised Jul 27, 2018

Accepted Aug 7, 2018

Keyword:

Cervical cancer
Genetic algorithm
Hierarchical Decision
Approach (HAD)
Neural Network (NN)
Pap smear

ABSTRACT

The single image classification of Pap smears is an important part of the early detection of cervical cancer through Pap smear tests. Unfortunately, most classification processes still require accuracy enhancement, especially to complete the classification in seven classes and to get a qualified classification process. In addition, attempts to improve the single image classification of Pap smears were performed to be able to distinguish normal and abnormal cells. This study proposes a better approach by providing different handling of the initial data preparation process in the form of the distribution for training data and testing data so that it resulted in a new model of Hierarchical Decision Approach (HDA) which has the higher learning rate and momentum values in the proposed new model. This study evaluated 20 different features in hierarchical decision approach model based on Neural Network (NN) and genetic algorithm method for single image classification of Pap smear which resulted in classification experiment using value learning rate of 0.3 and momentum of 0.2 and value of learning rate of 0.5 and momentum of 0.5 by generating classification of 7 classes (Normal Intermediate, Normal Columnar, Mild (Light) Dysplasia, Moderate Dysplasia, Severe Dysplasia and Carcinoma In Situ) better. The accuracy value enhancement were also influenced by the application of Genetic Algorithm to feature selection. Thus, from the results of model testing, it can be concluded that the Hierarchical Decision Approach (HDA) method for Pap Smear image classification can be used as a reference for initial screening process to analyze Pap Smear image classification.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Dwiza Riana,
STMIK Nusa Mandiri Jakarta,
Jalan Damai no 8 Jakarta Selatan, Indonesia.
Email: dwiza@nusamandiri.ac.id

1. INTRODUCTION

Research on the classification of single Pap smear image has been done. This attempt was intended to digitize the introduction of early detection of cervical cancer. As known that one type of malignant cancer that attacks women according to WHO body with the massive number of patients in Indonesia is cervical cancer. It is no wonder that Indonesia became one of the countries that have a lot of cervical cancer patients. Cervical cancer is generally caused by a virus called Human Papilloma Virus (HPV). Sexual intercourse became the largest case of HPV [1].

Pap smear is a method of early detection of cervical cancer. The process applied on Pap smear continuously and consistently in a country will help prevent early cervical cancer. This method was performed by a Pathologist in a clinical pathology laboratory, in which tests were performed on a woman's squamous epithelium. The results of pathologist's examination with a Pap smear will show whether the

woman has normal or abnormal cells [2]. There are various classifications in Pap Smear, but in this study, Pap smear images are classified up to 7 classes [3], in which the first three classes are normal cell class categories including Normal Superficial, Normal Intermediate, and Normal Columnar while the next four classes of abnormal cell categories are: Mild (Light) Dysplasia, Moderate Dysplasia, Dysplasia and Carcinoma In Situ [4].

General examination used to detect cervical cancer in Pap smear method is to prevent and detect the presence of pre-cancer and cancer situation in cervical cell samples. The problem of Pap smear image classification is caused by Pap smear image having unique characteristic so that the automatic identification of Pap smear image is a challenging problem for researchers. Different cell conditions and structures with high variations of image conditions make the identification and classification process of the Pap smear image need special handling. Particularly the process of Pap smear image classification until now is still experiencing difficulties and requires techniques and methods of classification that have a high accuracy.

The use of data mining so far is commonly used to obtain optimal information from a large group of large databases that have complexity [5]. In a study of single Pap smear image classification found in the Herlev dataset [4], data mining was used to get information from 20 features in the data to identify pathologic cases of cervical cancer. The previous researches which aimed to identify pathological cases with the same dataset include the study of classification methods on normal class images [6-8] and classification of abnormal classes [9]. Besides the classification of previous research forms, some researchers aim to segment the Pap smear image [6],[10]. Even the effort to identify the best features to solve the pathological case of cervical cancer has also been done. Feature [11] and texture analysis [6], [12] are some of the examples. The combination of several features (20 features) referring to 7 classes of diverse cases of pathological cancer, causing difficulties in the classification for 7 classes in this Pap smear image where it remains a challenge for researchers. Some algorithms aimed at selecting features such as genetic algorithms [13] perform a feature selection process by selecting some of the best individuals. Individual taking should be done randomly and proportionally including the proportion of its quality.

The proposed HDA classification model on single Pap smear image was started [14] from when the Pap smear classification model offered new process stages by utilizing both quantitative and qualitative features that was utilization of Importance Performance Analysis as the basis of the proposed multi-stage classification. The results of this study still have difficulties in classification for moderate dysplasia and severe dysplasia class [14].

The next attempt to classify the image of cervical cancer was to apply the Genetic Algorithm (GA) for feature selection. Furthermore, to classify healthy cells and cancer cells, we used SVM algorithm (Support Vector Machine) [15]. The results show that genetic algorithm is a better method for selection of features and optimization of parameters.

In this study, NN was selected as a tool of analysis on Papsmeared image dataset used. The use of this algorithm was to make data prediction and identify pathological cases of cervical cancer to be handled. The use of NN for medical data classification is commonly used such as classification to predict mortality prediction [16]. Optimization on NN algorithm can be done with the aim of improving the performance of NN [17]. The most commonly used optimization method is GA, Particle Swarm Optimization (PSO), and Ant Colony Optimization [18]. In this study, GA was selected as a feature selection algorithm. GA is one of algorithms that can select a relevant feature subset, learning rate, momentum, and initialization and weight optimization.

Based on the previous research [19], we focused this research to improve classification accuracy in the best model of the classification result based on the HDA model for single-cell Pap smear image classification. The comparison of classification results was done by using NN algorithm and feature optimization using GA to determine the increase of accuracy. The results show that there is a significant increase of accuracy from the proposed HDA model.

In this paper we propose methods for Pap smear cell image classification aimed at two specific objectives: a) selection of the best features on 20 features of pap smear and b) Pap smear image classification approach using hierarchical decision approach stage. Thus there are two main contributions in our paper. First, features of the Pap smear image that are not relevant in the classification process are not used like the longest diameter nucleus and nucleus roundness. Second, the uses of the hierarchical decision approach make the classification process more effective and increase the accuracy of classification results. In this way the automatic classification process to help pathologist allows to be realized.

This method is based on feature selection for less relevant features by using genetic algorithms and generates relevant features to be used in subsequent classification processes. This method combines the knowledge on the variations of classification stages between Pap smear and hierarchical decision approach class by optimizing the value of learning rate and momentum on NN algorithm. Based on this fact we propose a method that can classify Pap smear image into 7 classes which are 3 normal classes and 4

abnormal classes. This method exploits the features of the nucleus and cytoplasm through feature selection. Finally, this method is evaluated by using 917 sample dataset and has 20 features, divided into 90% training data and 10% testing data. The evaluation process uses applications built to support the proposed method. The remainder of this paper is organized as follows: section 2 about related work, section 3 about research method used in the study. Section 4 describes the results and analysis, then followed by conclusions and further research plans.

2. RELATED WORK

The automatic hybrid segmentation classification approach to select and enhance the segmentation of nucleus cells for Pap smear test images by using nested hierarchical portioning, segmentation level selection, and SVM classifier was already performed. The purpose of merging the end of the segmentation is to avoid over segmentation. The segmentation was done with morphological algorithm (watershed) and hierarchical merging (waterfall) algorithm based on spectral information and shape information as well as class information. SVM classifier is used to separate two classes of regions that are the nucleus and not the nucleus area (cytoplasm and background) by using a feature set (morphometric, edge-based, and convex hull-based). The results of segmentation and classification were compared with the segmentation provided by pathologist and showed improvement in the proposed method [20]. Unfortunately, this research has not yet reached the classification process of Pap smear image.

GA has been used in previous research and is considered as a better method for feature selection and parameter optimization in Pap smear images on the same dataset [15]. Support Vector Machine (SVM) Algorithm is used for classification. With this structure, new cells can be classified by observing the best feature values for cancer cell classification as cancer cells or benign cells. Unfortunately, the results show that the effectiveness of this method has not given the highest accuracy for the classification of 7 classes [15].

The hybrid ensemble technique is used for Pap smear image classification with the addition of new data [21] [22]. By comparing the methods of NN and SVM. The research stages are not thoroughly conducted in all class conditions, so the results obtained apply only to the class according to the simplified stages where the research does not produce a classification model of 7 classes but only presents class recall data [21].

This study compares Linear Discriminant Analysis (LDA) algorithm and Naïve Bayes algorithm to obtain the best classification results. The result of classification of LDA algorithm has poor accuracy on 7 classes whereas for Normal and Abnormal class classification, the result has good enough accuracy, and there is difficulty for abnormal classification with low accuracy value. The low accuracy of the abnormal class affects the classification into 7 classes [23].

The research that tried to overcome the difficulties of single Pap smear image classification in 7 classes was done by [24]. This study observed a number of classes that has different amounts of data, ie, the dataset has a class with a number of different and unbalanced classes. Another condition is that the data has features that are suspected to be irrelevant, so it is still difficult to classify especially abnormal classes. To handle the class imbalance, this study used ensemble method (Bagging). For handling data that HDA features and HDA no contribution, we made feature selection of Greedy Forward Selection. Furthermore, Naïve Bayes was used as learning algorithms. Although this method can handle imbalance classes, but the classification of 7 classes has not achieved the maximum results [24].

We have implemented Pap smear classification algorithms by using NN classification algorithm and feature selection by using GA. The best model of the classification result became the Hierarchical HDA model, a new classification approach for Pap Smear image. The comparison of classification results by using NN algorithm and feature optimization by using GA to determine the increase of accuracy was conducted. Pap smear image classification into 7 classes using HDA method has good classification value while classification using NN algorithm and feature optimization using GA have lower value compared to HDA algorithm [19]. However, the present study is an improvement of the research by giving special attention to the more proportional initial data-sharing process by using split validation method that improves the process of previous research methods. This resulted in accuracy values for both normal and abnormal classification, and the classification of 7 classes experienced a significant increase.

3. RESEARCH METHOD

3.1. Data Collection

At this stage, we determined the data to be processed, searched for available data, obtained the additional data required, and integrated all data into data sets including variables required in the process. The data used for training and testing is secondary data classified carefully by cyto-technicians and doctors. To

improve the classification of Pap smear cell images in this experiment we used Herlev 917 data [4]. In Table 1, it can be seen that the 20 features found in the dataset feature was optimized by using GA.

Table 1. The Feature of Herlev Dataset [4]

Name Of Feature	Name Of Feature	Name Of Feature	Name Of Feature
Nucleus Area or Kerne_A	Nucleus Shortest Diameter or KerneShort	Cytoplasm Longest Diameter or CytoLong	Nucleus Realtive Position or KernePos
Cytoplasm Area or Cyto_A	Nucleus Longest Diameter or KerneLong	Cytoplasm Elongation or CytoElong	Nucleus Maximum or KerneMax
N/C ratio or K/C	Nucleus Elongation or KerneElong	Cytoplasm Roundness or CytoRund	Nucleus Minimum or KerneMin
Nucleus Brightness or Kerne_Ycol	Nucleus Roundness or KerneRund	Nucleus Perimeter or KernePeri	Cytoplasm Maximum or CytoMax
Cytoplasm Brightness or Cyto_Ycol	Cytoplasm Shortest Diameter or CytoShort	Cytoplasm Perimeter or CytoPeri	Cytoplasm Minimum or CytoMin

3.2. Proposed Method

At this stage the data was analyzed and grouped into variables that are related to each other. After the data was analyzed, the models according to the data type were applied. Data sharing into training data and test data was also required for modeling. This study will select and apply appropriate techniques for Pap smear image classification. The first stage in this study was to divide the Pap smear cell dataset into two parts ie, traning data and testing data. The next step was to perform the best feature selection in the Pap smear image dataset by using GA, and then the selected feature was classified by using NN algorithm. The best model from the classification result was used as HDA model, so a new classification method approach was proposed for Pap smear image. The results of the model classification will be measured with an accuracy value. The research design can be seen in Figure 1.

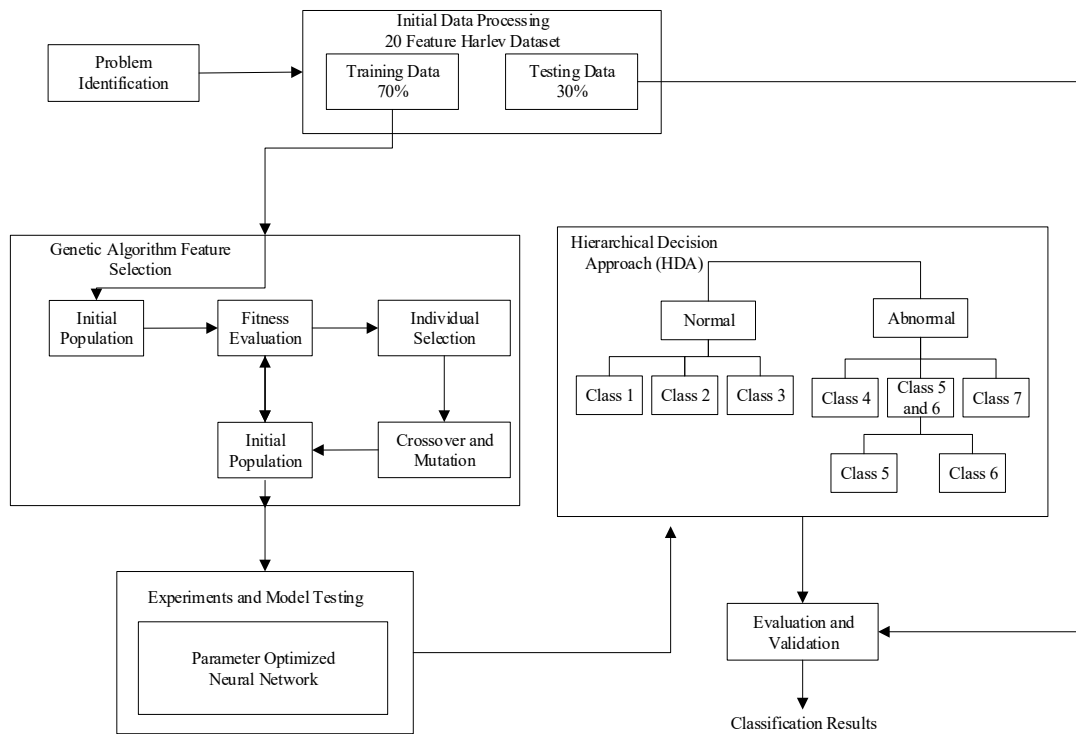


Figure 1. Research Design

a) Initial Data Processing

In this stage, data selection was conducted. The data was cleaned and transformed into the desired shape so that it can be done in preparation of model making. At this stage, exploration of the datasets

provided is required. First of all, it is known that the main goal to be achieved is to know the best classification result of Pap smear cell image. This study used Herlev dataset with the records of 917. To test the model developed, the data would be divided into two parts, namely training data and data testing. The data training was used for model development while data testing was used for model testing. It is known that the amount of data is 917 with a division of 70% (642) used for training data and 30% (275) used for data testing. The next stage was to select data that would be used as training data and data testing by using split validation. Furthermore, the feature selection method was performed in this research which is GA method. GA create a population composed of many individuals that evolve according to certain selection rules that have optimization determination and value.

b) Experiments and Model Testing

At this stage the proposed model will be tested to see the results of a rule that will be utilized in decision making. This research will conduct experiments on the classification of data mining using NN algorithm. The modeling will be done by using Rapidminer software. The models that have been obtained are transformed into the programming language of Visual Basic .Net 2017, and modeling translation of research design that has been done before are performed because the model of HDA cannot be done on software Rapidminer programming.

c) Evaluation and Validation

At this stage an evaluation of the model determined to find out the level of model accuracy was done. The evaluation was performed by using the confusion matrix table to determine the algorithm performance measurement on the classification algorithm model. The measured performance is Accuracy. The validation performed used the data that had been divided manually into testing data and training data. The model performance will be compared with NN algorithm by performing feature optimization by using GA and compared with Neural Network algorithm without doing optimization. Accuracy was used to compare the results so that the results obtained are more accurate.

4. RESULTS AND ANALYSIS

In this research, we will perform feature selection experiments by using GA and Pap smear classification by using NN algorithm. The experiments were conducted by using Herlev dataset where the initial data processing had been done with the distribution of training and testing data. In this section we will show the experimental results by using the NN algorithm and feature selection using GA by using 20 attributes shown in Table 2 in the Herlev dataset.

In the early stages of this research, the process of separation of training data and testing data was conducted, and the feature selection using Genetic Algorithm was then performed. The best attribute will be used as the Pap smear classification model using NN method. The classification process using NN algorithm was done by optimizing the best value of NN algorithm parameter with the value of Learning Rate and Momentum into 2 models. The first model used the learning rate (lr) value of 0.3 and momentum (m) of 0.2 while the second model uses the learning rate (lr) value of 0.5 and momentum (m) of 0.5. Furthermore, the highest accuracy value analysis was used for the HDA model. From the results, it is known that the value of learning rate and momentum greatly affects the accuracy of the classification.

Table 2. Classification Result of NN Algorithm and GA

No	Type Of Classification	NN	GA + NN (0.3 lr and 0.2 m)	GA +NN (0.5 lr and 0.5 m)
1	7 classes	64.00%	70.18%	66.91%
2	Normal & Abnormal	93.12%	96.01%	97.10%
3	Normal 1,2,3	97.22%	98.61%	100%
4	Abnormal 4, 5&6, 7	57.14%	74.88%	73.40%
5	Abnormal 5&6	74.76%	85.44%	84.47%

In Table 2, the classification comparison result shows that the classification with 7 classes using NN algorithm with the accuracy value of 64.00% after using feature selection by using GA and classification by using NN algorithm with the learning rate of 0.3 and momentum of 0.2 experiences the improvement of accuracy with a value of 70.18% and with the value of learning rate of 0.5 and momentum of 0.5 but produces an accuracy value of 66.91% where the accuracy results still look less. Thus, the process of classification using the HDA model by taking the best model in each classification was done. From the best classification result of each class, the best model was taken for the formation of HDA model. From the model

whose highest accuracy value had been known, the best feature separation using GA was performed with the distribution of data shown in Table 3.

This stage will perform an evaluation that aims to determine the level of accuracy of the classification testing results using NN algorithm and feature selection using GA by counting the amount of testing data that can be classified correctly. The test was done by using rapidminer software to get the best model and get the result of accuracy value. After obtaining the best model of the results obtained, then the process of classification using HDA was conducted to obtain classification results with 7 classes by using Visual Studio program 2017. Based on the research that has been classified with 7 classes, it has the highest accuracy of 70.18% in which the accuracy produced has not been optimal, so then we proposed classification process using HDA model. It was performed by separating the classification model into some of the best models including: Normal and Abnormal Classification with the accuracy of 97.10%, Normal Classification 1,2,3 with the accuracy of 100%, Abnormal Classification 4,5+6,7 with the accuracy of 74, 88%, by referring to Table 3. Class 5+6 was made into one because there were classification difficulties for the Moderate Dysplasia class and Severe Dysplasia [14]. The final step was to classify class 5 and 6 with the accuracy of 85.443%.

The HDA model highly depends on the model that has been derived from the classification of each class to be the reference model for making the HDA algorithm. Therefore, each of the best features that have been selected by using GA is presented in Table 3 as a representation of HDA model formation. Each class has a different Hidden layer depending on the number of features selected and the most relevant feature to the accuracy value.

HDA algorithm model not only affects the accuracy of each class but also affects the weight value of each node where nodes are obtained from each attribute that has been selected. Each weight has different values.

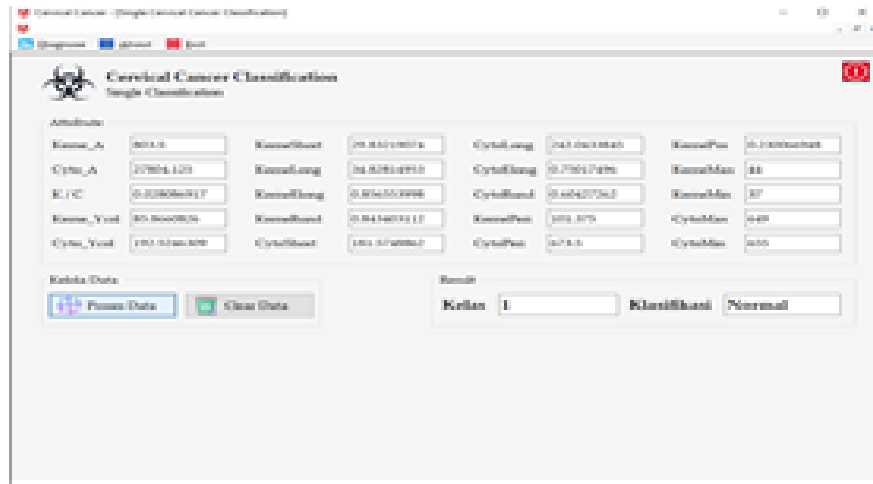
Table 3. Selected Attributes Using GA

No	Normal and abnormal Classification	Classification 1,2,3	Classification of Class 4, 5, 6, 7	Class 5, 6
1	Kerne_Ycol	Kerne_A	Cyto_A	Kerne_A
2	Cyto_Ycol	Cyto_A	K/C	Kerne_Ycol
3	KerneShort	Kerne_Ycol	Cyto_Ycol	Cyto_Ycol
4	KerneLong	Cyto_Ycol	KerneLong	KerneShort
5	CytoLong	CytoElong	KerneMax	KerneLong
6	CytoRund	CytoRund	KerneMin	CytoShort
7	CytoPeri	CytoPeri	CytoMax	CytoLong
8	KernePos	KernePos		CytoRund
9	KerneMax	CytoMin		KernePeri
10	KerneMin			CytoPeri
11				KerneMax
12				CytoMax
13				CytoMin

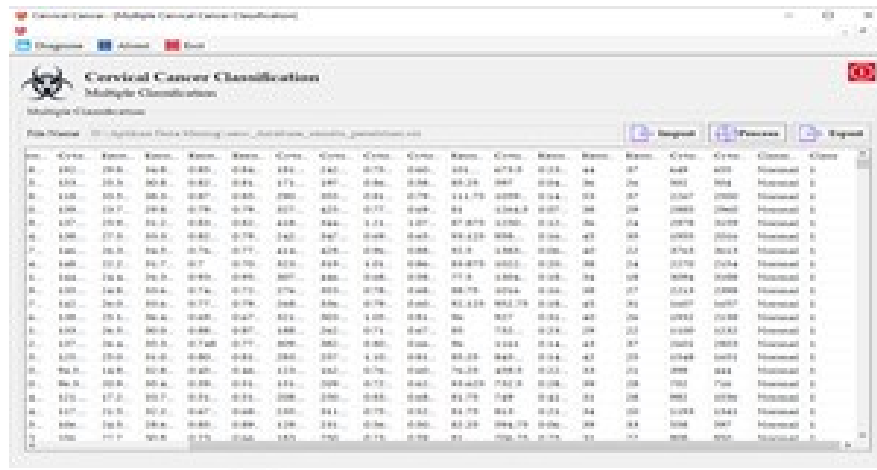
4.1. Application Development of Hierarchy Model

From the results obtained, then the best model was implemented in Visual Studio .Net 2017 application for the classification of 7 classes. The modeling stage used the Visual Studio .Net 2017 application with interface display in Figure 2(a) using each attribute input and 2(b) interface views for classification using datasets with multiple inputs.

The next step was the classification modeling implementation of 7 classes with the following stages: normal and abnormal classification model, normal classification model 1, 2, 3, abnormal classification model 4, 5 and 6, 7, and classification of class 5 and 6 with the following explanation: At this stage, the modeling for the normal and abnormal classification was performed by using the procedure described in the following stages of the program.



(a)



(b)

Figure 2. Application interface of the classification of 7 classes

Normal and abnormal classification algorithm

Input : Hidden layer Weight of each attribute, Max and Min Weight of each attribute, Output Weight of each attribute.

Output : Classification dataset of Normal and Abnormal Pap smear image.

Process :

- a. *Start.* Attribute normalization. $Normalization = ((data - min) / (max - min)) * (1 - (-1)) + (-1)$; Perform normalization on each attribute * Minimum and maximum value on training attribute.
- b. Calculate the weight of each hidden layer/node with as much weight as the hidden layer in the normal and abnormal classification model. Begin by calculating each hidden layer obtained from the multiplication of attributes that have been normalized with each weight that has been determined in selected attributes using GA. Furthermore, calculate the weight of each attribute on normal and abnormal class from the calculation of the initial weight. Calculate the output weights of each Normal and Abnormal Class output value. $Node\ 1 = (normaisasi_attribute * attribute\ weight) + bias$
 $Node\ Weight\ 1 = 1 / (1 + Exp(-node1))$
- c. Calculate the weights of each normal and abnormal classification. Calculate each classification weight obtained from the multiplication of each weight of the hidden layer with the weight of the nodes specified in calculation weight of hidden layer. Hidden layers are obtained

from the multiplication of attributes that have been normalized with each weight that has been determined in selected attributes using GA.

Classification=(hidden_layer_weights * node weight)+threshold

Classification weight=1/(1+Exp (-classification))

- d. Compare the classification weight that has been calculated with the normal and abnormal weight. If the normal weight is greater than the abnormal weight, the classification results are normal, but otherwise the classification becomes abnormal.

Classification=if normal weight>abnormal weight

Result=normal

If not

Result=abnormal

- e. In the next stage, perform the same process from stage a-d by performing calculations in each classification including: normal class classification 1,2,3, Abnormal class 4, 5 and 6, 7 and abnormal class 5 and 6.

4.2. Comparison Results of Accuracy Values

Table 4 shows that the HDA model has a superior accuracy value compared to the classification algorithm result shown in Table 4. The results obtained from the research shows that the classification model of HDA and NN algorithm has superior accuracy compared to other classification algorithms. After doing the research, the classification results of 4 classes that became the main goal were compared to see which algorithm and which method is best for the classification into 7 classes.

Based on the test that has been obtained on the Pap smear image dataset, it is known that the NN and HDA algorithms have the highest accuracy with the value of 87.02% when compared with other classification algorithms.

Table 4. Comparison of Accuracy Values

Algorithm	Accuracy
Proposed Method	87,02%
Decision Tree + HDA [14]	83,26
GA + HDA Non Optimized NN [19]	79,78%
Hybrid Ensemble Learning [22]	78,00%
GF + Bagging + Naïve Bayes [24]	63,25%
GA + LDA [23]	62,92%
GA + Naïve Bayes [23]	62,16

5. CONCLUSION

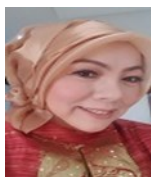
Pap smear image classification by using HDA method with the classification test into 7 classes (normal superficial, normal intermediate, normal columnar, mild (light) dysplasia, moderate dysplasia, severe dysplasia and carcinoma in situ) has the highest accuracy value of 87.02%. The results obtained from the HDA model for Pap smear image classification into 7 classes were compared to the classification results using the NN algorithm and feature optimization using GA to improve accuracy. In this work we propose a classification methodology in a single cell Pap smear image. This task is particularly useful for normal and abnormal cell image classification in each class. We can come out with the fact that the proposed method has not reached a very high level of accuracy. However, we need a more practical, practical alternative method to classify Pap smear images more accurately. As future work, we intend to expand our method using hybrid modeling classification. In hopes it can further improve the accuracy achieved. Thus, from the results of our model testing, it can be concluded that the HDA method for Pap smear image classification can be used as a reference for initial screening process to analyze Pap smear image classification. Further research will be done by making web-based applications, and the performance measurement of web-based applications will be conducted by users who are pathologists and researchers in the field of cervical cancer.

ACKNOWLEDGEMENTS

Authors would like to thank RISTEKDIKTI. This research was supported by The Ministry of Research, Technology, and Higher Education, Indonesia, for supporting this research through The Pasca Doctoral Research Grant (2017). This work is using the data from: Pap smear Benchmark Data for Pattern Classification J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, University Dept. of Pathology Herlev Ringvej 75, DK-2730 Herlev, Denmark.

REFERENCES

- [1] "World's health ministers renew commitment to cancer prevention and control.," May 2017. [Online]. Available: <http://www.who.int/cancer/media/news/cancerprevention-resolution/en/>.
- [2] D. Riana, W. H. Dwi, D. E. Dewi and T. L. R. Mengko, "Segmentasi Luas Nukleus Sel Normal Superfisial Pap Smear Menggunakan Operasi Kanal Warna Dan Deteksi Tepi," *Seminar Nasional Inovasi dan Teknologi (SNIT)*, 2012.
- [3] J. Jantzen, J. Norup, G. Dounias and B. Bjerregaard, "Pap-smear Benchmark Data For Pattern Classification," 2005.
- [4] E. Martin, "Pap-Smear Classification," *Technical University of Denmark - DTU*, 2003.
- [5] Pruengkarn, Ratchakoon, Kok Wai Wong and Chun Che Fung, "A review of data mining techniques and applications," *Journal of Advanced Computational Intelligence and Intelligent Informatics 21*, vol. 1, pp. 31-48, 2017.
- [6] D. Riana, D. H. Widyantoro and T. L. Mengko, "Extraction and classification texture of inflammatory cells and nuclei in normal Pap smear images," in *ICI-BME*, Bandung, 2015.
- [7] D. Riana, D. E. O. Dewi, D. H. Widyantoro and T. L. r. Mengko, "Color canals modification with canny edge detection and morphological reconstruction for cell nucleus segmentation and area measurement in normal Pap smear images," in *AIP*, Bandung, 2014.
- [8] D. Riana, D. H. Widyantoro and T. L. R. Mengko, "Inflammatory cell extraction and nuclei detection in Pap smear images," *Int. J. e-Health Med. Commun.*, vol. 6, pp. 27-43, 2015.
- [9] D. Riana, D. E. O. Dewi, D. H. Widyantoro and T. L. R. Mengko, "Segmentation and Area Measurement in Abnormal Pap smear Images Using Color Canals Modification with Canny Edge Detection," in *In International Conference on Women's Health in Science & Engineering*, Bandung, 2012.
- [10] R. Kurniawan, A. Kurniawardhani and I. Muhimmah, "Inflammatory Cell Extraction in Pap smear Images: A Combination of Distance Criterion and Image Transformation Approach," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, Vol 16 (5): 2048-2056; 2018.
- [11] J. Hyeon, H.-J. Choi and B. D. Lee, "Diagnosing Cervical Cell Images Using Pre-trained Convolutional Neural Network as Feature Extractor," in *in Big Data and Smart Computing (BigComp)*, 2017.
- [12] D. Kashyap, A. Somani and J. Shekhar, "Cervical Cancer Detection And Classification Using Independent Level Sets And Multi SVMs," in *9th Int. Conf. Telecommun. Signal Process*, 2016.
- [13] N. Lassouaoui, L. Hamami and N. Nouali, "Morphological description of cervical cell images for the pathological recognition," *Int. J. Med. Health*, Vols. 1, No 5, pp. 313-316, 2007.
- [14] D. Riana, "Hierarchical Decision Approach Berdasarkan Importance Performance Analysis Untuk Klasifikasi Citra Tunggal Pap Smear Menggunakan Fitur Kuantitatif dan Kualitatif," 2010. [Online]. Available: <http://www.digilib.ui.ac.id/detail?id=31702&lokasi=12>. [Accessed 20 March 2018].
- [15] T. K. Mansoori, A. Suman and S. K. Mishra, "Feature Selection by Genetic Algorithm and SVM Classification for Cancer Detection," *International Journal of Advanced Research in Computer Science and Software Engineering* Vol 4, pp. 357-365, 2014.
- [16] M. H. A. Yazid, S. Talib, M. H. Satria and A. A. Ghazi, "Neural Network on Mortality Prediction for the Patient Admitted with ADHF (Acute Decompensated Heart Failure)," in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics*, Yogyakarta, 2017.
- [17] M. F. Mohammed and T. H. Rassem, "An Ensemble of Enhanced Fuzzy Min Max Neural Networks for Data Classification," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, Vol 15 (2); 2017.
- [18] A. G. Karegowda, A. Manjunath and M. Jayaram, "Application of Genetic Algorithm Optimized Neural Network Connection Weights for Medical Diagnosis Of Pima Indians Diabetes," *International Journal on Soft Computing (IJSC)*, pp. 15-23, 2011.
- [19] Y. Ramdhani and D. Riana, "Hierarchical Decision Approach Based on Neural Network and Genetic Algorithm Method for Single Image Classification of Pap Smear," in *Informatik and Computing (ICIC)*, Jayapura, 2017.
- [20] M. Orozco-Montegudo, C. Mihai, H. Sahli and A. Taboada-Crispi, "Combined Hierarchical Watershed Segmentation and SVM Classification for Pap Smear Cell Nucleus Extraction," 2012.
- [21] E. J. Mariarputham and A. Stephen, "Nominated Texture Based Cervical Cancer Classification," *Computational and Mathematical Methods in Medicine*, pp. 1-10, 2015.
- [22] A. Sarwar, V. Sharma and R. Gupta, "Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis," *Personalized Medicine Universe*, pp. 1-9, 2015.
- [23] Y. Ramdhani, "Komparasi Algoritma LDA DAN Naive Bayes Dengan Optimasi Fitur Untuk Klasifikasi Citra Tunggal Pap Smear," *Informatika*, Vols. III, No, 2, pp. 434-441, 2015.
- [24] D. Riana, A. N. Hidayanto and Fitriyani, "Integration of Bagging and greedy forward selection on image pap smear classification using Naïve Bayes," in *Cyber and IT Service Management (CITSM)*, Bali, 2017.

BIOGRAPHIES OF AUTHORS

Dwiza Riana was born in Indonesia in 1970. She is an associate professor in STMIK Nusa Mandiri Jakarta and principal of the Magister Ilmu Komputer at STMIK Nusa Mandiri. She did her BA in Mathematics at the University of Sriwijaya, Indonesia, Magister of Management at University of Budi Luhur, Indonesia, Magister of Computer Science at University of Indonesia and PhD in Electrical Engineering in Informatics at Institut Teknologi Bandung, Indonesia. Her research is in the area of Computer Science, Biomedical Engineering, Data Mining, and Information System.



Yudi Ramdhani was born in Indonesia in 1990. He did his BA in Information Technical at the University BSI, Indonesia, Magister of Computer Science at STMIK Nusa Mandiri Jakarta, Indonesia. His research is in the area of Computer Science, Biomedical Engineering, Information System, and Big Data Analysis.



Rizki Tri Prasetyo was born in Indonesia in 1989. He is a Lecturer in Universitas BSI. He did his Bachelor of Science in Information System at Universitas BSI, Indonesia and Magister of Computer Science in STMIK Nusa Mandiri Jakarta, Indonesia. His research is in the area of Software Engineering, Big Data and Data Mining, and Mobile Computing.



Achmad Nizar Hidayanto is the Vice Dean for Resources, Ventures, and General Administration, Faculty of Computer Science, Universitas Indonesia. He received his PhD in Computer Science from Universitas Indonesia. His research interests are related to information management, IT diffusion and adoption, e-commerce, e-government, information systems security, change management, knowledge management and information retrieval.