Combination of texture feature extraction and forward selection for one-class support vector machine improvement in self-portrait classification

Reina Alya Rahma, Radityo Adi Nugroho, Dwi Kartini, Mohammad Reza Faisal, Friska Abadi
Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Banjarbaru, Indonesia

Article Info

Article history:

Received Dec 24, 2021
Revised Aug 8, 2022
Accepted Sep 1, 2022

Keywords:

Feature selection
Haralick
Local binary pattern
One-class support vector machine

ABSTRACT

This study aims to validate self-portraits using one-class support vector machine (OCSVM). To validate accurately, we build a model by combining texture feature extraction methods, Haralick and local binary pattern (LBP). We also reduce irrelevant features using forward selection (FS). OCSVM was selected because it can solve the problem caused by the inadequate variation of the negative class population. In OCSVM, we only need to feed the algorithm using the true class data, and the data with pattern that does not match will be classified as false. However, combining the two feature extractions produces many features, leading to the curse of dimensionality. The FS method is used to overcome this problem by selecting the best features. From the experiments carried out, the Haralick+LBP+FS+OCSVM model outperformed other models with an accuracy of 95.25% on validation data and 91.75% on test data.

Corresponding Author:

Radityo Adi Nugroho
Department of Computer Science, Faculty of Mathematics and Natural Science
Lambung Mangkurat University
A. Yani St., KM 36, Banjarbaru, South Kalimantan, Indonesia
Email: radityo.adi@ulm.ac.id

1. INTRODUCTION

Currently, identity verification is performed on most types of digital transactions. Verification is done by uploading identity documents such as ID cards, passports, self-portraits, and others. One of the critical pieces of data in the identity verification process is a self-portrait. In Indonesia, self-portraits are still widely used as proof of a person's identity, such as registration for selecting prospective civil servants (CPNS), joint selection to enter state universities (SBMPTN), or e-commerce account verification. The accepted self-portrait generally has specific terms and conditions such as shooting position, background, and photo quality.

However, in reality, any form of self-portrait can be uploaded even if it does not meet the predetermined standards, thus reducing the efficiency of identity verification. This happens because there is no self-portrait validation process. Self-validation is needed to facilitate the face verification process from identity documents and faces in real-time. The literature on selfie portraits and photo documents have been carried out by [1]–[4]. Most of them use the convolutional neural network (CNN) approach. CNN is good at generating performance, but some works still adopt several other algorithms to get optimal results. However, CNN also requires significant computational resources. This process can be simplified by involving efficient machine learning methods.

This process can be simplified by involving machine learning algorithms. One of the algorithms is the one-class support vector machine (OCSVM). This algorithm can solve the problems faced because the negative class population cannot be appropriately represented due to the many variations of self-portrait errors. In OCSVM, these error variations are classified as anomalies, as has been done by [5]–[8]. OCSVM was proposed by [9] and is an adaptation of the support vector machine (SVM) methodology. Like the basic concept of SVM, OCSVM involves kernel functions to perform classification, including linear, polynomial, radial basis function (RBF), and sigmoid. OCSVM kernel comparison has been carried out by [10], the results show that RBF consistently provides the best performance.

In performing self-portrait classification, feature extraction steps are needed. Feature extraction generates features that are used to describe the content [11]. Image extraction is divided into several categories, namely color, texture, and shape feature extraction. The texture is a key element of human visual perception widely used in computer vision systems [12]. Some examples of texture feature extraction methods are Haralick and local binary patterns. In the study [12]–[14], Haralick feature extraction resulted in excellent classification performance, as well as the use of local binary pattern (LBP) carried out by [15]–[17]. In addition, Kaplan *et al.* [18] and Porebski *et al.* [19] states that efficient texture feature extraction in predicting sample variation is the Haralick feature and local binary pattern.

While they deliver good performance, Haralick and LBP produce a large number of features. The large number of features tends to reduce the prediction accuracy of the classification model [20]. This problem can be overcome by minimizing dimensions or irrelevant features, one of which is feature selection. Feature selection can improve machine learning model performance and reduce computation time [21], [22]. In addition, forward selection (FS) is utilized to perform feature selection and results in improved classification performance [22]–[25].

This study establishes a model based on OCSVM and applies it to simulate the self-portraits validation. In addition, this research is also intended to improve OCSVM performance by combining Haralick and local binary patterns, then reducing irrelevant features using forward selection. This paper is structured as: The second section describes the procedures and methodologies applied to our study. The third section covers the results achieved in our experiment, evaluation, and analysis. In the last section, the main findings of this study are highlighted and discussed.

2. METHOD

The following is the research procedure carried out. Figure 1 presents the block diagram of the various sub-stages in step by-step manner. The sub-stages are discussed in detail as shown in Figure 1.

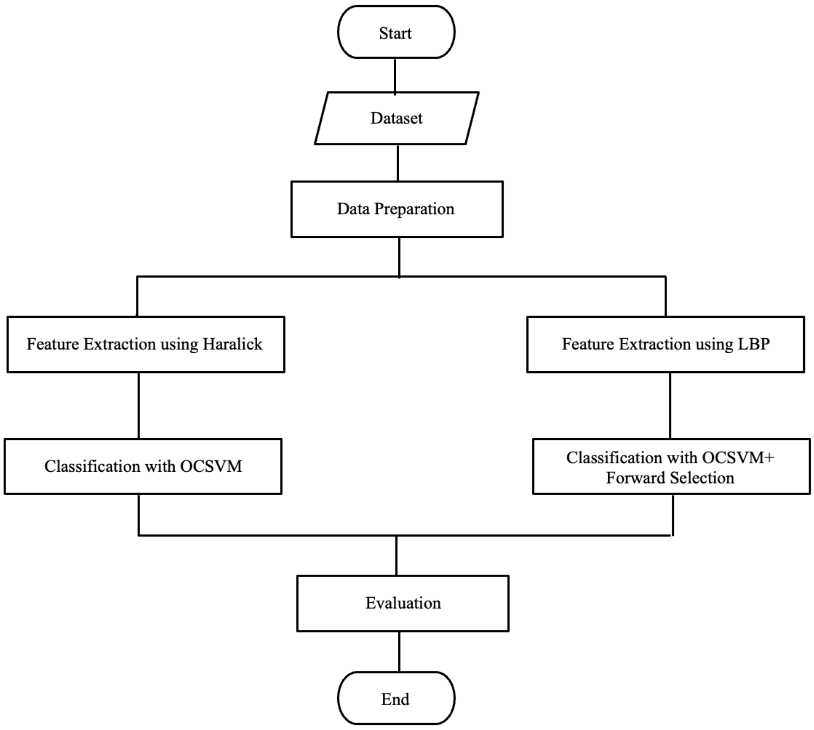


Figure 1. Proposed method abstraction

2.1. Dataset

This study uses student self-portrait datasets taken from the Lambung Mangkurat University Academic Portal application. Self-portraits are divided into 2 labels, namely true and false self-portraits. Self-portrait is considered correct if the picture is taken from the front, the position of the object is symmetrical, the background is monotonous, and the image quality is clear.

2.2. Feature extraction

Before classification, the dataset was extracted using texture feature extraction. The texture feature extraction used is Haralick and local binary pattern. In this study, feature extraction was carried out without pre-processing. The feature extraction process is implemented in Python and the Mahotas library.

2.2.1. Haralick

Haralick feature extraction was proposed by Haralick *et al.* [26]. The Haralick extraction result consists of 14 features calculated from the gray level co-occurrence matrix (GLCM). This method calculates the feature value of the 4 GLCM angles and takes the average value. GLCM is generated from the probability relationship of 2 pixels (Pi, j) with distance d, angle θ (0°, 45°, 90°, and 135°), and color level N. The formula can be seen in (1)-(14).

– Angular second moment

$$\sum_i \sum_j p(i,j)^2 \tag{1}$$

– Contrast

$$\sum_{n=0}^{N_g-1} n^2 \left\{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)\right\}, |i-j| = \tag{2}$$

– Correlation

$$\frac{\sum_i \sum_j (i,j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{3}$$

– Sum of squares: variance

$$\sum_i \sum_j (i-\mu)^2 p(i,j) \tag{4}$$

– Inverse difference moment

$$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j) \tag{5}$$

– Sum average

$$\sum_{i=2}^{2N_g} i p_{x+y}(i) \tag{6}$$

– Sum variance

$$\sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i) \tag{7}$$

– Sum entropy

$$\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} = f_8 \tag{8}$$

– Entropy

$$\sum_i \sum_j p(i-j)\ \log(p(i,j)) \tag{9}$$

– Difference variance

$$\sum_{n=0}^{N_g-1} i^2 p_{x+y}\,(i) \tag{10}$$

– Difference entropy

$$\sum_{n=0}^{N_g-1} p_{x+y}(i)\ \log\{p_{x+y}(i)\} \tag{11}$$

– Information measure of collection I

$$\frac{HXY-HXY1}{max\{HX,HY\}} \tag{12}$$

– Information measure of collection 2

$$1-exp\,[-2(HXY2-HXY])^{\frac{1}{2}} \tag{13}$$

– Maximum correlation coefficient, the square root of Q, where:

$$Q_{(i,j)}=\sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)} \tag{14}$$

2.2.2. Local binary pattern

LBP feature extraction was first proposed by Ojala *et al.* [27]. LBP is calculated from the grayscale image. The LBP calculation begins with the localization of pixels determined from the sampling point p on a circle of radius r as shown in Figure 2. Meanwhile, the principle of calculating LBP is shown in Figure 3. This method compares the value of the center pixel with the values of the pixels around it. If the intensity of the center pixel is greater than the center pixel, the value is set to 1, if it is smaller than 0. The value after thresholding is multiplied by the weight of each pixel and the additive result is the LBP value.

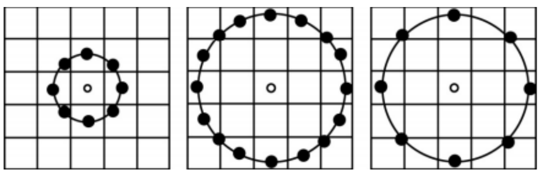


Figure 2. Sampling point P on a circle of radius R

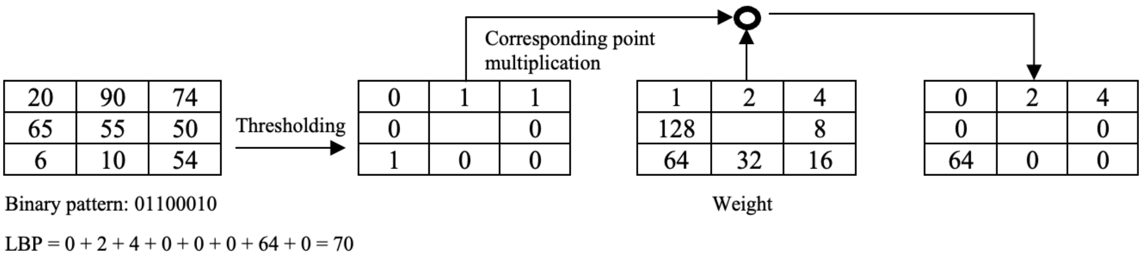


Figure 3. LBP principle

2.3. Combining extraction results

After feature extraction, the next step is to combine the results of Haralick and LBP. The new dataset consists of all Haralick features plus all LBP features. Now, there are 3 datasets formed, namely Haralick, LBP, and Haralick+LBP.

2.4. Data mining

Data mining modeling is done using Rapidminer. Classification is performed on the dataset generated in the previous step. Each is classified with OCSVM, with or without forward selection.

2.4.1. One class support vector machine

Based on the research proposed by [9], OCSVM aims to find the best hyperplane to separate the target data from the origin/second class. The hyperplane is affected by the $v(nu)$ parameter. OCSVM uses a kernel trick to map data into a high-dimensional space. Kernels used in OCSVM include linear, polynomial, RBF, and sigmoid. This research uses the RBF kernel. The OCSVM is formulated in (15):

$$\min_{\omega,\xi_i,\rho} \frac{1}{2}\left|\left|\omega\right|\right|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho \tag{15}$$

where n is number of data points, ν is regularization parameter, ξi is slack variable for point xi that allows it to be placed outside the decision boundary, and ω and ρ is parameters that determine the decision boundary.

2.4.2. Forward selection

Classification is performed on the dataset generated in the previous step. Each is classified as OCSVM, with or without forward selection. Forward selection will select the most influential attribute and remove irrelevant attributes. How the forward selection works starts from the empty model, then one by one the attributes are entered until certain criteria are met.

2.5. Evaluation

The performance of the model is evaluated based on its accuracy. After the performance of each model is obtained, the next step is to make a comparison. Comparison is made by comparing the results of the accuracy of the proposed model (Haralick+LBP→FS+OCSVM) with other models to prove that the proposed model provides increased accuracy.

3. RESULTS AND DISCUSSION

3.1. Dataset

Student self-portraits are obtained from the ULM Student Academic Portal application. The dataset contains 59,000 photos from the student intake between the years 2013 and 2020. Of these, we took 2,400 photos labeled true and false. An example of data labeling can be seen in Table 1. After being labeled, the data is then divided into 3 categories, namely training, validation, and testing data. Details of the amount of data can be seen in Table 2.

Table 1. Labeling photo

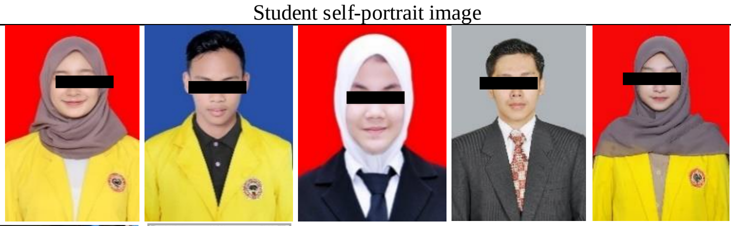
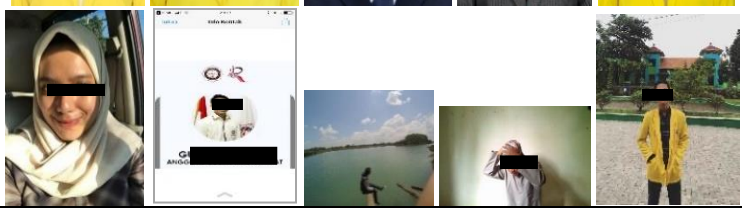
Student self-portrait image	Label
	True
	False

Table 2. Details of the amount of data

No	Data	Label		Total
		True	False	
1	Data Training	1,600	-	1,600
2	Data Validation	200	200	400
3	Data Testing	200	200	400
	Total Data			2,400

3.2. Feature extraction

After the photo data is obtained, the next step is to perform feature extraction with Haralick and LBP. To perform this extraction, we use the Python programming language with the Mahotas library. In Haralick feature extraction, we take all its features, which are 14 features. The results of Haralick's extraction can be seen in Table 3. While the LBP feature extraction used parameters p=8 and distance r=1 which resulted in 36 features. The results of LBP feature extraction can be seen in Table 4. These tables show example results from some of the selected images.

Table 3. Haralick extraction example results

Attributes	File			Attributes	File		
	1.jpg	2.jpg	3.jpg		1.jpg	2.jpg	3.jpg
ASM	0.027705	0.02869	0.000126	SE	6.495217	6.777125	8.813428
Cnt	2992.924	5291.095	6871.323	Etrp	10.15162	9.950032	14.47072
Corr	0.51061587	0.608174	0.393398	DV	0.000358	0.000427	0.000101
SSV	3260.136	7079.089	5624.2	DE	4.74229	4.600722	6.524528
IDM	0.173995	0.198313	0.081346	IMC 1	0.31729	0.42064	0.17505
SA	360.7437	246.5232	249.17	IMC 2	0.986353	0.996576	0.919458
SV	10047.62	23025.26	15625.48	MCC	5,717714	5.248164	5.509102

Table 4. LBP extraction example results

Attributes	File			Attributes	File		
	1.jpg	2.jpg	3.jpg		1.jpg	2.jpg	3.jpg
LBP0	49913	168299	28983	LBP18	2	4	8
LBP1	7701	20721	6716	LBP19	3	3	0
LBP2	8225	22094	13909	LBP20	0	0	0
LBP3	1367	3550	953	LBP21	5	11	7
LBP4	4418	12081	5198	LBP22	4	2	1
LBP5	1	5	1	LBP23	0	0	2
LBP6	1141	2732	879	LBP24	894	2367	180
LBP7	1062	2799	445	LBP25	1092	2376	289
LBP8	3062	8230	4553	LBP26	0	0	12
LBP9	0	0	0	LBP27	16	28	116
LBP10	745	2825	336	LBP28	0	0	0
LBP11	0	0	0	LBP29	0	0	0
LBP12	991	2750	501	LBP30	1146	2747	525
LBP13	1063	2534	1933	LBP31	0	0	0
LBP14	3050	8797	3705	LBP32	3848	10143	3305
LBP15	1237	2571	641	LBP33	2610	5017	184
LBP16	6554	19788	21480	LBP34	12758	30526	20050
LBP17	0	0	0	LBP35	0	0	0

3.3. Combining extraction results

This step is to form a new dataset by combining the data extracted by Haralick with the results of the LBP extraction. The number of features in the combined dataset is 50 features, consisting of 14 Haralick features and 36 LBP features. From here, there are 3 datasets that are ready to be entered into the next step, namely: Haralick, LBP, and Haralick+LBP.

3.4. Data mining

We created the schema model using Rapidminer, as shown in Figure 4. Based on this scheme, the best accuracy is obtained in each model based on the relevant attribute FS and optimal hyper-parameter OCSVM. Details can be seen in Table 5. At this stage, there are 6 models built, namely:

Haralick → OCSVM
LBP → OCSVM

Haralick+LBP → OCSVM
Haralick → FS+OCSVM
LBP → FS+OCSVM
Haralick+LBP → FS+OCSVM

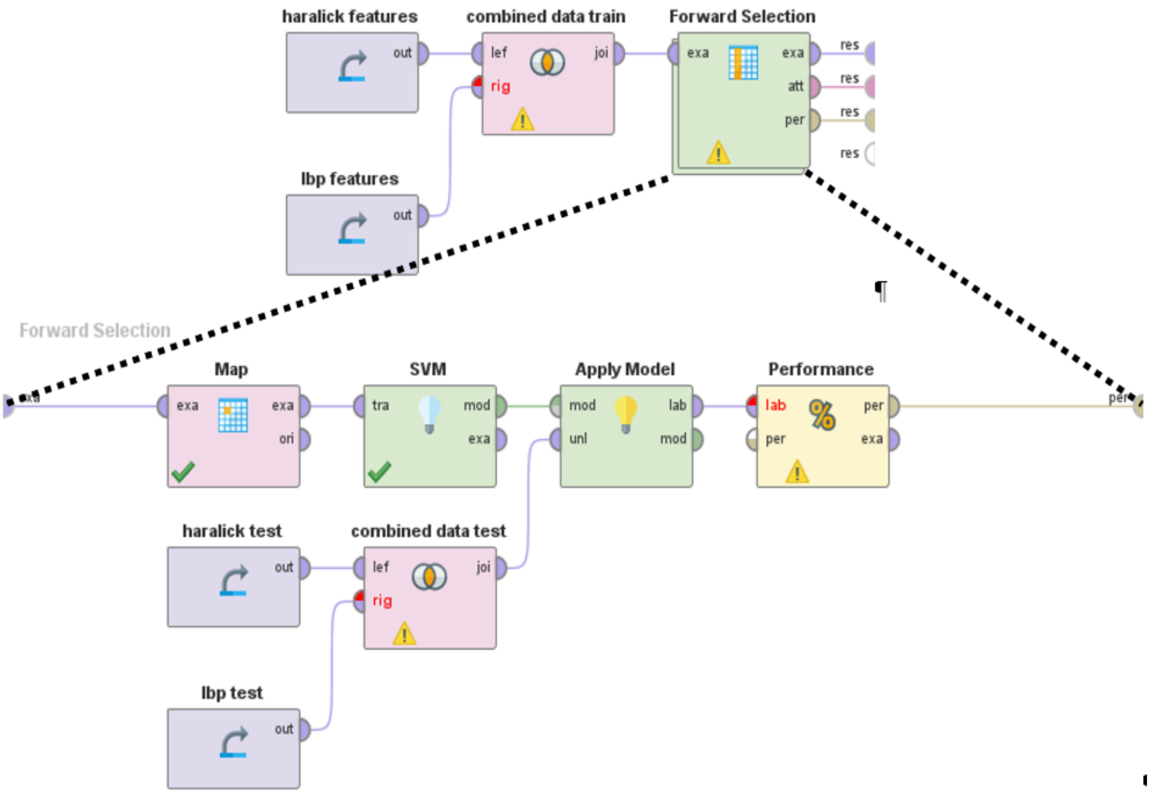


Figure 4. Proposed data mining schema in Rapidminer

Table 5. The best hyper-parameter values and accuracy

No	Model	Best Hyper-parameter nu	gamma	Data validation accuracy
1	Haralick+OCSVM	0.12637	5.0E-9	85.75%
2	LBP+OCSVM	0.199005	5.0E-10	85.25%
3	Haralick+LBP+OCSVM	0.197015	5.0E-9	80.50%
4	Haralick+FS+OCSVM	0.07264	0.005	94.25%
5	LBP+FS+OCSVM	0.0209	0.0005	89.50%
6	Haralick+LBP+FS+OCSVM	0.06468	0.5	95.25%

3.5. Evaluation

After obtaining the optimum hyper-parameter data, our proposed model is re-examined on the test data that has been prepared. The results can be seen in Table 6. The table shows that the proposed model consistently produces the best performance.

Table 6. Data test accuracy

No	Model	Accuracy
1	Haralick+OCSVM	86.25%
2	LBP+OCSVM	80.00%
3	Haralick+LBP+OCSVM	80.25%
4	Haralick+FS+OCSVM	86.75%
5	LBP+FS+OCSVM	78.50%
6	Haralick+LBP+FS+OCSVM	91.75%

Several misclassifications were obtained from the evaluation results because the test images had a pattern similar to the training data. However, we grouped the data into the false class because they had photo variables that did not match the training data, such as wrong dimensions, incorrect color, and object position. This misclassification is reasonable because these three things do not affect the extraction of Haralick and LBP texture features. On the other hand, the true class was misclassified as false due to the hyperplane (boundary) in OCSVM being affected by the Nu parameter. The Nu parameter controls how many outliers we want to allow. In this study, some true class data were not classified in the correct class due to a strict hyperplane because the optimum Nu value is too small. Table 7 shows the examples of misclassification results.

Table 7. Misclassification example results

Student self-portrait	True	Prediction	Student self-portrait	True	Prediction
	False	True		True	False
	False	True		True	False

4. CONCLUSION

The experiments have proven that the proposed model, which is a combination of Haralick and LBP feature extraction, is classified with the OCSVM algorithm and forward selection feature selection (Haralick+LBP+FS+OCSVM), outperforms other models with an accuracy of 95.25% on data validation, and 91.75% on testing data. Thus, we conclude that combining the results of Haralick and LBP feature extraction selected with forward selection can improve the performance of OCSVM in classifying self-portrait. For further research, because profile photos require precise dimensions, color, and object position, it is necessary to use a method that is not invariant to these three factors.

REFERENCES

[1] F. Costa, M. V. L. Melo, I. Gadelha, G. Fôlego, L. Gambaro, and A. Rodrigues, “Self-portrait to ID Document face matching: CNN-Based face verification in cross-domain scenario,” in *Anais do XVII Workshop de Visão Computacional (WVC 2021)*, Nov. 2021, pp. 31–36, doi: 10.5753/wvc.2021.18885.

[2] A. Chinapas, P. Polpinit, N. Intiruk, and K. R. Saikaew, “Personal verification system using ID card and face photo,” *International Journal of Machine Learning and Computing*, vol. 9, no. 4, pp. 407–412, Aug. 2019, doi: 10.18178/ijmlc.2019.9.4.818.

[3] D. Praveenbalaji, R. Srinivas, S. Roopa, M. Suresh, A. Gayathri, “ID photo verification by face recognition,” in *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2020, pp. 1449–1453, doi: 10.1109/ICACCS48705.2020.9074246.

[4] A. Castelblanco, J. Solano, C. Lopez, E. Rivera, L. Tengana, and M. Ochoa, “Machine learning techniques for identity document verification in uncontrolled environments: a case study,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12088, Springer International Publishing, 2020, pp. 271–281.

[5] D. Devi, S. K. Biswas, and B. Purkayastha, “Learning in presence of class imbalance and class overlapping by using one-class SVM and undersampling technique,” *Connection Science*, vol. 31, no. 2, pp. 105–142, Apr. 2019, doi: 10.1080/09540091.2018.1560394.

[6] A. Barbado, Ó. Corcho, and R. Benjamins, “Rule extraction in unsupervised anomaly detection for model explainability: application to OneClass SVM,” *Expert Systems with Applications*, vol. 189, Mar. 2022, doi: 10.1016/j.eswa.2021.116100.

[7] E. H. Budiarto, A. Erna Permanasari, and S. Fauziati, “Unsupervised anomaly detection using K-means, local outlier factor and one class SVM,” in *5th International Conference on Science and Technology (ICST)*, Jul. 2019, pp. 1–5, doi: 10.1109/ICST47872.2019.9166366.

[8] J. Sun, J. Shao, and C. He, "Abnormal event detection for video surveillance using deep one-class learning," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 3633–3647, Feb. 2019, doi: 10.1007/s11042-017-5244-2.

[9] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *NIPS*, 1999, vol. 12, pp. 582–588.

[10] K.-K. Seo, "An application of one-class support vector machines in content-based image retrieval," *Expert Systems with Applications*, vol. 33, no. 2, pp. 491–498, Aug. 2007, doi: 10.1016/j.eswa.2006.05.030.

[11] S. Choras, Ryszard, "Image feature extraction techniques and their applications for CBIR and biometrics systems," *International Journal of Biology and Biomedical Engineering*, vol. 1, no. 1, pp. 6–16, 2007.

[12] F. Tomita and S. Tsuji, *Computer analysis of visual textures*. Boston, MA: Kluwer Academic Publishers, 1990.

[13] S. Samanta, S. S. Ahmed, M. A. M. M. Salem, S. S. Nath, N. Dey, and S. S. Chowdhury, "Haralick features based automated glaucoma classification using back propagation neural network," in *Advances in Intelligent Systems and Computing*, vol. 327, Springer International Publishing, 2014, pp. 351–358.

[14] V. Bhateja *et al.*, "Haralick features-based classification of mammograms using SVM," in *Advances in Intelligent Systems and Computing*, vol. 672, Springer Singapore, 2018, pp. 787–795.

[15] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and one class classifiers," *Computers and Electronics in Agriculture*, vol. 156, pp. 96–104, Jan. 2019, doi: 10.1016/j.compag.2018.11.005.

[16] Y. Luo, C. M. Wu, and Y. Zhang, "Facial expression feature extraction using hybrid PCA and LBP," *Journal of China Universities of Posts and Telecommunications*, vol. 20, no. 2, pp. 120–124, Apr. 2013, doi: 10.1016/S1005-8885(13)60038-2.

[17] K.-C. Song, Y.-H. Yan, W.-H. Chen, and X. Zhang, "Research and perspective on local binary pattern," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 730–744, Jun. 2013, doi: 10.1016/s1874-1029(13)60051-8.

[18] K. Kaplan, Y. Kaya, M. Kuncan, and H. M. Ertunç, "Brain tumor classification using modified local binary patterns (LBP) feature extraction methods," *Medical Hypotheses*, vol. 139, Jun. 2020, doi: 10.1016/j.mehy.2020.109696.

[19] A. Porebski, N. Vandenbroucke, and L. Macaire, "Haralick feature extraction from LBP images for color texture classification," in *First Workshops on Image Processing Theory, Tools and Applications*, Nov. 2008, pp. 1–8, doi: 10.1109/IPTA.2008.4743780.

[20] S. B. Yengec Tasdemir, K. Tasdemir, and Z. Aydin, "ROI detection in mammogram images using wavelet-based Haralick and HOG features," in *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2018, pp. 105–109, doi: 10.1109/ICMLA.2018.00023.

[21] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018, doi: 10.1016/j.neucom.2017.11.077.

[22] J. Zeniarja, A. Ukhifahdhina, and A. Salam, "Diagnosis of heart disease using k-nearest neighbor method based on forward selection," *Journal of Applied Intelligent System*, vol. 4, no. 2, pp. 39–47, Mar. 2020, doi: 10.33633/jais.v4i2.2749.

[23] T. Löfstedt, P. Brynolfsson, T. Asklund, T. Nyholm, and A. Garpebring, "Gray-level invariant Haralick texture features," *PLOS ONE*, vol. 14, no. 2, Feb. 2019, doi: 10.1371/journal.pone.0212110.

[24] D. Riana, A. N. Hidayanto, and Fitriyani, "Integration of bagging and greedy forward selection on image pap smear classification using Naïve Bayes," Aug. 2017, doi: 10.1109/CITSM.2017.8089320.

[25] M. Reif and F. Shafait, "Efficient feature size reduction via predictive forward selection," *Pattern Recognition*, vol. 47, no. 4, pp. 1664–1673, Apr. 2014, doi: 10.1016/j.patcog.2013.10.009.

[26] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[27] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996, doi: 10.1016/0031-3203(95)00067-4.

BIOGRAPHIES OF AUTHORS

Reina Alya Rahma received her bachelor's degree in computer science from Lambung Mangkurat University. Reina has an interest in the field of data mining. She can be contacted at email: reina.alya@ulm.ac.id.

Radityo Adi Nugroho received his bachelor's degree in Informatics from the Islamic University of Indonesia and a master's degree in Computer Science from Gadjah Mada University. Currently, he is an assistant professor in the Department of Computer Science at Lambung Mangkurat University. His research interests include software defect prediction and computer vision. He can be contacted at email: radityo.adi@ulm.ac.id.

Dwi Kartini received her bachelor's and master's degrees in computer science from the Faculty of Computer Science, Putra Indonesia University "YPTK" Padang, Indonesia. Her research interests include the applications of artificial intelligence and data mining. She is an assistant professor in the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University in Banjarbaru, Indonesia. She can be contacted at email: dwikartini@ulm.ac.id.

Mohammad Reza Faisal received the B.Sc. and M.Eng. degrees in physics and informatics from Bandung Institute of Technology, Bandung, Indonesia, in 2004 and 2013. He also received a B.Eng. degree in informatics from Pasundan University, Bandung, Indonesia, in 2002 and a Ph.D. in computer science from Kanazawa University, Ishikawa, Japan, in 2018. He is currently the lecturer in the Computer Science Department, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University in Banjarbaru, Indonesia. His research interests include artificial intelligence applications, text mining, and software engineering. He can be contacted at email: reza.faisal@ulm.ac.id.

Friska Abadi received his bachelor's degree in computer science from Lambung Mangkurat University, Banjarbaru, Indonesia, in 2011. He also received a master's degree in informatics from STMIK Amikom, Yogyakarta, in 2016. He is currently the lecturer in the Computer Science Department, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarbaru, Indonesia. His research interests include data mining and software engineering. He can be contacted at email: friska.abadi@ulm.ac.id.