Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 7, No. 3, August, Pp. 211-218

Feature selection based on multi-filters for classification of mammogram images to look for signs of breast cancer

Shofwatul Uyun
Universitas Islam Negeri Sunan Kalijaga, Indonesia

Article Info
Keywords:
Multi-Filter, Feature Selection, Classification, Mammogram

Article history:
Received: April 15, 2022
Accepted: June 14, 2022
Published: August 31, 2022

Cite:
S. 'Uyun, "Feature Selection Based on Multi-Filters for Classification of Mammogram Images to Look for Signs of Breast Cancer", *KINETIK*, vol. 7, no. 3, Aug. 2022. https://doi.org/10.22219/kinetik.v7i3.1437

*Corresponding author.
Shofwatul Uyun
E-mail address:
Shofwatul.uyun@uin-suka.ac.id

Abstract
The accuracy of classification results on mammogram images has a significant role in breast cancer diagnosis. Therefore, many stages consider finding the model has a high level of accuracy and minimizing the computing load, one of which is the accuracy in using the best feature. This needs to be prioritized considering that mammogram image has many features resulting from the mammogram extraction process. Our research has four stages: feature extraction, feature selection-multi filters, classification, and performance evaluation. Thus, in this research, we propose algorithms that can select the features by utilizing multiple filters simultaneously on the filter model for feature selection of mammogram images based on multi-filters/FSbMF. There are six feature selection algorithms with a filter approach (information gain, rule, relief, correlation, gini index, and chi-square) used in this research. Based on the testing result using 10-fold cross-validation, the features resulting from the FSbMF algorithm have the best performance based on the accuracy, recall, and precision from 72,63%, 70,38%, 75,01% to be 100%. Furthermore, the number of resulting features is the minimum because it results from intersection operation from the feature subsets resulting from the multi-filter.

1. Introduction

Mammogram image is a medical imaging technology widely used to detect and diagnose breast cancer. One of the advantages of mammogram images compared with other medical imaging technology is USG (ultrasound). In addition, the use of mammography technology does not depend on the operator's skill so that the resulting image is objective. However, analyzing an abnormality on a mammogram image is a challenging task because a mammogram image has deficient quality. Therefore, this research is trying to develop a method of Computer-Aided Design (CAD) to produce a better level of accuracy. Nevertheless, mammogram image classification can be used for screening or diagnosis [1], [2].

The feature extraction results on mammogram images produce many attributes or features. Nowadays, the data sample number or dimension development of data numbers is increasing rapidly on some learning machine apps such as pattern recognition, computer vision, or biomedical. This certainly will become a great challenge for some learning machines. The use of many irrelevant or less relevant features does not only make the learning process slower. It may result in lowering the performance of some learning tasks but will also complicate the model interoperability. Therefore, the feature selection process is an effective way to solve the problem by deleting data on irrelevant and redundant features. If the feature selection process is implemented, there are three advantages: faster computing time, increased accuracy level, and an easier way of analyzing and studying the learning method and data [3].

Feature selection finds relevant feature subsets from a set of source (original data) feature subsets. Feature selection plays the role of a data compressor on a small scale by deleting irrelevant and redundant features. Furthermore, it can also be used in preprocessing stages of a learning algorithm that is capable of producing good qualities so that it may increase the learning accuracy (the feature selection process will delete the insignificant features that may cause misleading due to overfitting; thus, the accuracy value will increase); reduce the learning time; minimize the overfitting (feature selection process may delete the redundant data and noise that may cause the overfitting on the following procedure) and simplify the learning results (the more precise the dimension of dataset is, the faster and more efficient the algorithm learning will be able to run). The feature selection process includes the combination of the search process, estimation of feature effects on data label determination, and evaluation using a machine learning algorithm. Feature selection involves many processes; therefore, it requires a heuristic search procedure to optimize the feature selection process combined with evaluator functioning to estimate the feature effect level [4].

There is a similarity between extraction and feature selection in reducing dimensions. However, both have different characters in their process. The studies on feature selection [3] have often discussed some fields such as image recognition [5], image retrieval [6], data mining [7], image mining [8], intrusion detection [9], malware classification [10], speaker identification [11], and bioinformatics [12]. Based on the use of data training, feature selection can be categorized to be three models: supervised [13][14][15], unsupervised [11] [16], and semi-supervised [17][10]. Whereas, based on the learning method [18], there are three models: filter [18], wrapper [19], and embedded [20] [21]. There are two types of output resulting from the feature selection process: ranking based on weighting on each feature and feature subset selection.

Compared with the wrapper model, the advantage of the filter model is the lower computing cost. The selection process on the filter model is only focused on the use of feature associations with their class labels. The critical point of the filter model lies in the determination of evaluation criteria. While the feature selection on the embedded model is conducted during the training process on the learning model, the best feature result will be automatically found when the training process is over. Model filter in conducting the feature selection is based on the score or correlation function based on an algorithm of the certain model independently evaluated and ranked on each feature. Some univariate filters are: information gain [22], correlation [23] and relief [24]. Therefore, in this paper, the researcher proposes the simultaneous use of a multi-filter on filter model for feature selection of mammogram images called feature selection based on multi-filters/FSbMF.

2. Research Merhod

This research uses two data sources of mammogram images: primary data (oncology clinic) and secondary data (MIAS public database). The image processing is then conducted from the data, usually called preprocessing, consisting of cropping and resizing with bilinear interpolation of 256 x 256 pixels. In addition, reducing noise with median filtering, increasing image quality using CLAHE with block size 127, histogram bins 256 and slope maximum three, and histogram equalization with saturated pixel 0.4%. This research proposes a feature selection method using some filters simultaneously (feature selection based on multi-filters) that later is called FSbMF. There are four stages: feature extraction, feature selection-multi filters, classification, and performance evaluation. The detail is shown in Figure 1.

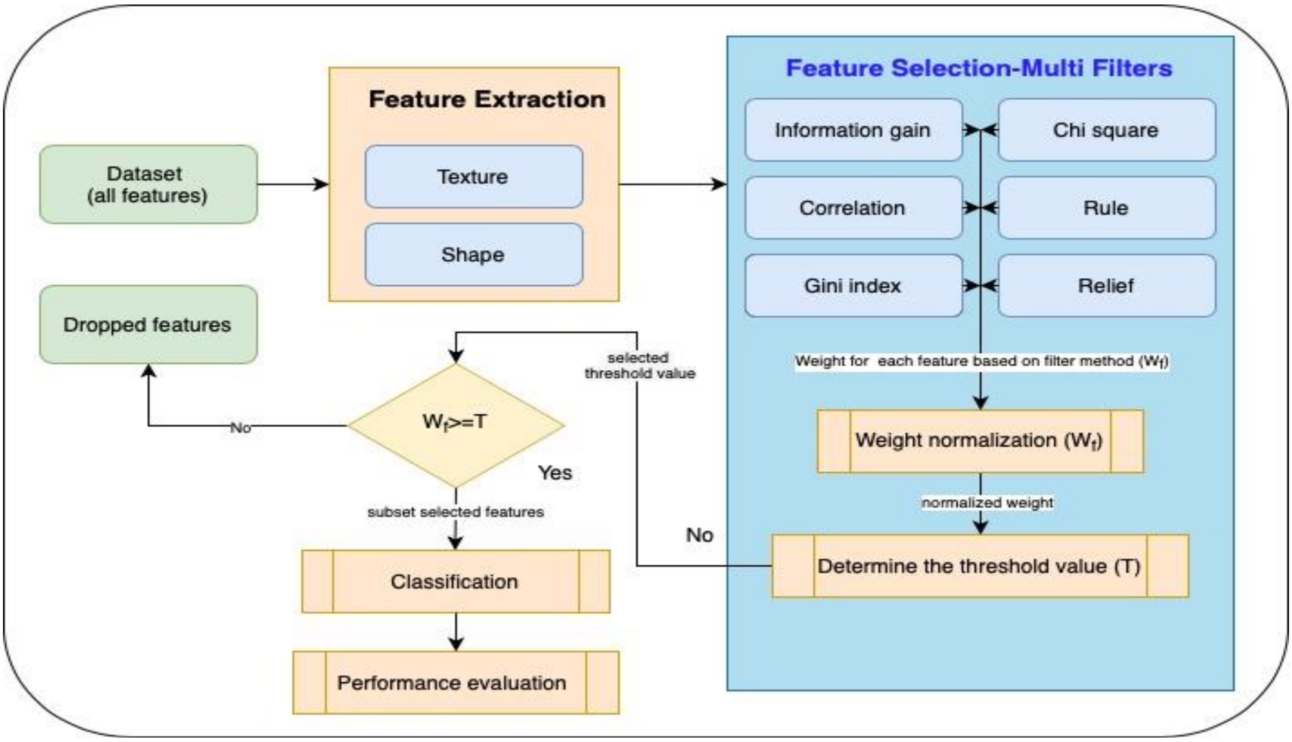

Figure 1. A Method Proposed for Feature Selection Based on Multi-filter

2.1 Preprocessing

After preprocessing has been conducted on the mammogram image, the feature extraction process is undertaken. This research uses some features involved in either texture or shape features that also have been used [8]. Texture features save information about the surface structure of an image. In the first-order texture measurement process, statistical calculation is used only based on the pixel value of the original image. In contrast, in the second order, a pair of two original image pixels is calculated as the gray level retrieval matrix, commonly known as the Gray Level Co-occurrence Matrix (GLCM). Co-occurrence can be interpreted as two or more things occurring together or simultaneously, which means the number of occurrences at the one-pixel level that is adjacent to another pixel value based on a certain distance and angle orientation (Ɵ). Distance is represented as pixels, while the orientation is defined in degrees [25]. The orientation is formed from four corner directions with an interval of 45^0 those are 0^0, 45^0, 90^0, and 135^0 and the distance between pixels is determined by one pixel. At the same time, the shape features can show the character of an object based on the line and contour configuration categories used. There are two categories: based on the boundary (boundary-based) to describe the region's shape using external characteristics and based on the region (region-based). The total of features used is forty-six features. The shape features include area, a center of mass, modal gray value, centroid, perimeter, integrated density, median, area fraction, stack position, circularity, aspect ratio, roundness, and solidity. At the same time, the texture feature includes the meaning, standard deviation, skewness, and kurtosis. Four features have four directions of (contrast, correlation, energy, and homogeneity) [8]. At this stage, the n x m matrix is formed; n indicates the number of data, and m denotes the data feature.

2.2 Feature Selection – Multi Filters

At this stage, the filtering process is conducted to all data sets with all features using six filter methods of feature selection. The six filter methods include information gain, chi-square, correlation, rule, gini-index, and relief. Each method produces a subset of the feature set with its weight value, and then the feature selection process is conducted using more than one filter method called FSbMF.

A. Weight normalization
   The normalization process is conducted to the weight owned by each feature produced from the feature selection method with a value range [0 1].
B. Determination of threshold value and feature
   The weight normalization result found the feature produced by each feature extraction method with the normalized weight. To determine what features are the best requires a threshold value. The algorithm to find the threshold value is indicated by the following algorithm 1.

Algorithm 1: Determination of threshold value and feature
Input: feature and its weight
$Threshold_{[0.1-0.9]}$ value
Make $_{mx1}$ matrix (m=feature produced by each filter method based on its weight and threshold value)
Select $threshold_{(t)}$ value for each filter method
if $feature_{(t)}$ *set* = $feature_{(t+1)} set$ **then**
    Determine $threshold_{(t)} value$ = sum($threshold_{\in(metode\ filter)}$)/$\sum filter\ method$
Final of features set = intersection(set features$_{[1..n]}$)
End

2.3 Classification

At this stage, the threshold value used for selecting the subset of the feature set is the threshold value resulting from the previous process. Afterward, the training process and testing are conducted to the subset of the feature set resulting from each filtering method using different classification algorithms. This research uses five classification algorithms consisting of decision trees [26], Bayes [27], K-Nearest Neighbor (KNN), random forest [27] [28], and random trees [29]. The k-Nearest Neighbor (K-NN) Algorithm is one of the classification algorithms by searching the object K group on the data training nearest (similar) to the object on new data or data testing. The number of nearest neighbors is determined based on the K (neighborhood) value [30]. The K value cannot be greater than the number of training data, and it should be odd and more than one. K-NN algorithm can model a predictive case with high accuracy. K-NN is included in the supervised learning algorithm with the results of the newly classified query instance (test sample) based on the majority of the categories in K-NN. The class that appears the most will be the class resulting from the classification [31].

2.4 Performance Evaluation

The evaluation of the proposed method uses k-fold cross validation by dividing the dataset to be data training also data testing by dividing into ten parts. The technique that can be used to evaluate the performance of the classification model is k-fold cross-validation. K-fold is one of the cross-validation methods. K-fold cross-validation concept does not only make some test data samples repeatedly but divides the dataset to be separated parts with the same measure. The model is trained by the training data subset and validated by the validation (test data) subset amounted to k. The k-fold cross validation may reduce the computing time by maintaining the model estimation's accuracy. Afterward, the performance assessment is conducted using a subset of feature sets produced from the feature selection process using some classification algorithms. Three parameters used to measure the performance are accuracy, precision, and recall, part of the confusion matrix. The confusion matrix consists of information on the classification between facts and predictions produced by the classification system. The illustration of how to get the values of three parameters is shown in Table 1.

Table 1. Representation of confusion matrix

Predicted	Actual	
	True	False
True	True Positive-TP (Correct result)	False Positive-FP (Unexpected result)
False	False Negative-FN (Missing result)	True Negative-TN (Correct absence of result)

Accuracy is the amount of data predicted correctly by the classification system, either negative or positive. The calculation of accuracy value can be calculated using Equation 1.

$$Accuray = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

By definition, precision compares True Positive (TP) with the amount of data with positive prediction. The calculation of precision is calculated using Equation 2.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

While for recall, by definition, it is a comparison between True Positive (TP) with the amount of positive data that can be calculated using Equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

3. Results and Discussion

The process of feature extraction on mammogram images is conducted using two features those are texture and shape features. The total of features produced from the extraction process of the two features is forty-two, then a coding process is conducted for each feature value produced, which detail can be seen in Table 2.

Table 2. Detailed Data of Features Extracted from Mammogram Images

No	Feature	No	Feature
f_1	Area	f_{22}	energy (0^0)
f_2	center of massa	f_{23}	energy (90^0)
f_3	modal gray value	f_{24}	energy (180^0)
f_4	Centroid	f_{25}	energy (270^0)
f_5	Perimeter	f_{26}	homogeneity (0^0)
f_6	integrated density	f_{27}	homogeneity (90^0)
f_7	Median	f_{28}	homogeneity (180^0)
f_8	area fraction	f_{29}	homogeneity (270^0)
f_9	stack position	f_{30}	inverse difference moment (0^0)
f_{10}	Mean	f_{31}	inverse difference moment (90^0)
f_{11}	standard deviation	f_{32}	inverse difference moment (180^0)
f_{12}	Skewness	f_{33}	inverse difference moment (270^0)
f_{13}	Kurtosis	f_{34}	entropy (0^0)
f_{14}	contrast(0^0)	f_{35}	entropy (90^0)
f_{15}	contrast (90^0)	f_{36}	entropy (180^0)
f_{16}	contrast (180^0)	f_{37}	entropy (270^0)
f_{17}	contrast (270^0)	f_{38}	angular second moment (0^0)
f_{18}	correlation(0^0)	f_{39}	angular second moment (90^0),
f_{19}	correlation(90^0)	f_{40}	angular second moment (180^0),
f_{20}	correlation(180^0)	f_{41}	angular second moment (270^0),
f_{21}	correlation(270^0)	f_{42}	Slice

Based on the data shown in Table 2, there are 42 features extracted from mammogram images. The number of features makes the data of mammogram image features have a very high dimension. Therefore, selecting the best feature needs to be conducted using six feature selection methods: information gain, chi-square, gini index, relief, rule, and correlation. Each process of feature selection produces weight for each feature. Feature with more weight compared with other features means that the feature has a strong level of correlation compared with other features towards the determination of classification results on mammogram images. So each method of feature selection produces a set of feature sets with its weight. Afterward, the weight normalization process is conducted as the limited value to select the relevant feature. The threshold value on each feature selection method is determined based on the feature set produced with the range of threshold 0.1 to 0.9. So each process of feature selection has a bunch of feature sets based on its threshold value. The feature results of six feature selection methods based on threshold values are shown in Table 3. The multi-filter method requires a threshold value applied to all feature selection methods to collaborate. This research proposes the FSbMF algorithm for selecting the feature set produced by many feature selection methods (weight normalization, determination of threshold value and feature), which detail of the process can be seen in Figure 1.

Table 3. The Feature Set Produced by Each Method of Feature Selection is Based on its Threshold Value

Feature selection algorithm	Features	Threshold	Feature ranking
Information gain	12	0.1	f_{39}, f_{41}, f_7, f_8, f_{13}, f_{42}, f_{10}, f_6, f_9, f_{12}, f_1, f_5
	2	0.2	f_1, f_5
Chi square	20	0.1	f_{21}, f_{31}, f_{14}, f_{18}, f_{31}, f_{20}, f_{15}, f_{33}, f_{39}, f_3, f_{19}, f_{13}, f_7, f_6, f_9, f_{12}, f_{10}, f_1, f_5
	7	0.2	f_7, f_6, f_9, f_{12}, f_{10}, f_1, f_5
	2	0.3	f_1, f_5
Gini index	23	0.1	f_3, f_{11}, f_{18}, f_{19}, f_{21}, f_{31}, f_{33}, f_{20}, f_{35}, f_{36}, f_{37}, f_{38}, f_{39}, f_7, f_8, f_{13}, f_{10}, f_6, f_9, f_{12}, f_1, f_5
	9	0.2	f_7, f_8, f_{13}, f_{10}, f_6, f_9, f_{12}, f_1, f_5
	2	0.3	f_1, f_5
Relief	2	0.1	f_1, f_5
Rule	14	0.1	f_3, f_{33}, f_{11}, f_7, f_{31}, f_{10}, f_6, f_9, f_{16}, f_{34}, f_8, f_{15}, f_1, f_5
	9	0.4	f_{10}, f_6, f_9, f_{16}, f_{34}, f_8, f_{15}, f_1, f_5
	6	0.5	f_{16}, f_{34}, f_8, f_{15}, f_1, f_5
	5	0.8	f_{34}, f_8, f_{15}, f_1, f_5
	4	0.9	f_8, f_{15}, f_1, f_5
Correlation	28	0.1	f_8, f_{36}, f_{38}, f_{37}, f_{35}, f_{14}, f_{16}, f_{30}, f_{32}, f_{32}, f_{34}, f_3, f_{19}, f_{11}, f_{21}, f_{18}, f_{20}, f_{33}, f_{31}, f_{39}, f_{13}, f_{12}, f_7, f_6, f_9, f_{10}, f_1, f_5
	19	0.2	f_{32}, f_{34}, f_3, f_{19}, f_{11}, f_{21}, f_{18}, f_{20}, f_{33}, f_{31}, f_{39}, f_{13}, f_{12}, f_7, f_6, f_9, f_{10}, f_1, f_5
	16	0.3	f_{19}, f_{11}, f_{21}, f_{18}, f_{20}, f_{33}, f_{31}, f_{39}, f_{13}, f_{12}, f_7, f_6, f_9, f_{10}, f_1, f_5
	8	0.4	f_{13}, f_{12}, f_7, f_6, f_9, f_{10}, f_1, f_5
	2	0.5	f_1, f_5

The output of the FSbMF algorithm is the selected threshold value implemented on all feature selection methods. This research recommends that the selected feature be divided into three groups of feature sets. The selected feature set is then classified using five classification algorithms (decision tree, Bayes, KNN, random forest, and random tree) shown in Table 4.

Table 4. Feature Set Produced Based on the Selected Threshold Use

Nama	#Features	Feature	Feature selection algorithm
Set features [1]	2	f1, f5	information gain, chi square, gini index, relief
Set features [2]	14	f3, f36, f11, f7, f34, f10, f6, f9, f16, f36, f8, f15, f1, f5	Rule
Set features [3]	16	f19, f11, f21, f18, f20, f36, f34, f42, f13, f12, f7, f6, f9, f10, f1, f5	Correlation

The highest accuracy value, recall, and precision have resulted from the classification results using the 1st feature set-valued 100%, and the results of the 2nd and the 3rd decrease. The detail can be seen in Figure 2. Therefore, based on the FSbMF algorithm that we propose, the selected feature is the one as the intersection result of three feature sets.

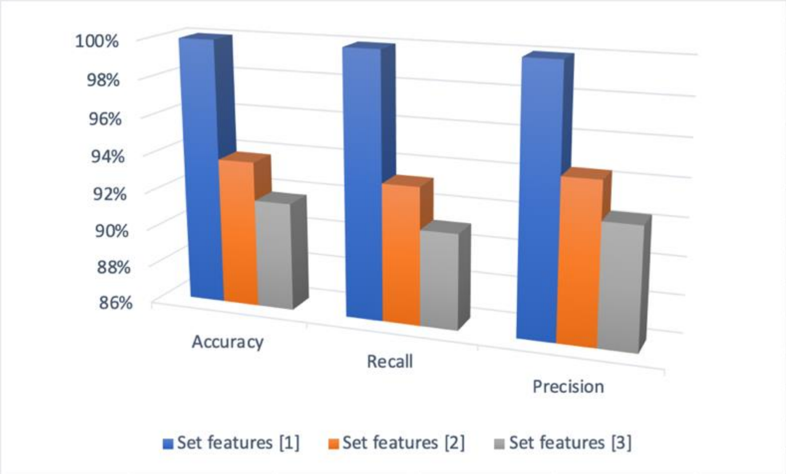

Figure 2. Accuracy Value, Recall, and Precision for Three Candidates of Selected Feature Set

In general, the use of feature sets resulting from the feature selection algorithm can increase the accuracy value, recall, and precision and decrease the root mean square error value when involving all features or without the feature selection process. For example, based on the FSbMF algorithm that we propose, the recommendation of selected feature (the result of intersection from three feature subsets of one or some feature selection methods) from forty-two features to become two features. This certainly dramatically helps to reduce the dimension. The comparison of the performance of the use of each feature subset is shown in Table 5.

Table 5. Evaluation of pPerformance on the Use of Feature Subset

Subset features	Performance evaluation using parameters		
	Accuracy	Recall	Precision
f_1, f_5 (selected)	100.00%	100.00%	100.00%
f_3, f_{36}, f_{11}, f_7, f_{34}, f_{10}, f_6, f_9, f_{16}, f_{36}, f_8, f_{15}, f_1, f_5	93.76%	93.27%	94.43%
f_{19}, f_{11}, f_{21}, f_{18}, f_{20}, f_{36}, f_{34}, f_{42}, f_{13}, f_{12}, f_7, f_6, f_9, f_{10}, f_1, f_5	91.73%	91.06%	92.41%
All features	72.63%	70.38%	75.01%

4. Conclusion

The proper feature selection as the base to conduct the classification process on mammogram images has a significant role in reducing data dimension. We propose the FSbMF algorithm that can produce the best feature from the feature extraction result on mammogram images. Previously, the number of features from the feature extraction process, the shape and texture ones, was forty-two. The selection process is conducted to the features using multi-filters. The feature selection process is undertaken using six methods of feature selection that produce three feature subsets as the result of the grouping of six feature selection methods. The selected final features are certainly based on the intersection operation result of three feature subsets and have the best performance. Based on the testing result using 10-fold cross-validation, the elements resulting from the FSbMF algorithm have the best performance based on the accuracy, recall, and precision from 72,63%, 70,38%, and 75,01% to be 100%.

Acknowledgement

The author would like to express appreciation and gratitude to Universitas Islam Negeri Sunan Kalijaga for funding this research.

References

[1] D. Carvalho, P. R. Pinheiro, and M. C. D. Pinheiro, "A Hybrid Model to Support the Early Diagnosis of Breast Cancer," in *Procedia Computer Science*, 2016, vol. 91, pp. 927–934. https://doi.org/10.1016/j.procs.2016.07.112

[2] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolution Neural Network Improvemenr for Breast Cancer Classification," *Expert Systems With Applications*, vol. 120, pp. 103–115, 2018. https://doi.org/10.1016/j.eswa.2018.11.008

[3] V. Bolón-Canedo and A. Alonso-Betanzos, "Feature selection," *Intelligent Systems Reference Library*, vol. 147, pp. 13–37, 2018. https://doi.org/10.1007/978-3-319-90080-3_2

[4] S. Dash, M. R. Senapati, and U. R. Jena, "K-NN based automated reasoning using bilateral filter based texture descriptor for computing texture classification," *Egyptian Informatics Journal*, vol. 19, no. 2, pp. 133–144, Jul. 2018. https://doi.org/10.1016/j.eij.2018.01.003

[5] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. https://doi.org/10.1007/s11263-015-0816-y

[6] X. S. Zhou, I. Cohen, Q. Tian, and T. S. Huang, "Feature Extraction and Selection for Image Retrieval," ACM Multimedia, pp. 1–7, 2000.

[7] M. N. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, “Systematic ensemble model selection approach for educational data mining,” *Knowledge-Based Systems*, vol. 200, pp. 1–16, Jul. 2020. https://doi.org/10.1016/j.knosys.2020.105992
[8] S. Uyun and L. Choridah, “Feature selection mammogram based on breast cancer mining,” *International Journal of Electrical and Computer Engineering*, vol. 8, no. 1, 2018 http://doi.org/10.11591/ijece.v8i1.pp60-69.
[9] M. S. Abirami, U. Yash, and S. Singh, “Building an Ensemble Learning Based Algorithm for Improving Intrusion Detection System,” in *Advances in Intelligent Systems and Computing*, 2020, pp. 635–649. https://doi.org/10.1007/978-981-15-0199-9_55
[10] X. Gao, C. Hu, C. Shan, B. Liu, Z. Niu, and H. Xie, “Malware classification for the cloud via semi-supervised transfer learning,” *Journal of Information Security and Applications*, vol. 55, Dec. 2020. https://doi.org/10.1016/j.jisa.2020.102661
[11] V. de A. Campos and D. C. G. Pedronette, “A framework for speaker retrieval and identification through unsupervised learning,” *Computer Speech and Language*, vol. 58, pp. 153–174, Nov. 2019. https://doi.org/10.1016/j.csl.2019.04.004
[12] Y. Jiao and P. Du, “Performance measures in evaluating machine learning based bioinformatics predictors for classifications,” *Quantitative Biology*, vol. 4, no. 4, pp. 320–330, 2016. https://doi.org/10.1007/s40484-016-0081-2
[13] F. Fabris, J. P. de Magalhães, and A. A. Freitas, “A review of supervised machine learning applied to ageing research,” *Biogerontology*, vol. 18, no. 2, pp. 171–188, 2017. https://doi.org/10.1007/s10522-017-9683-y
[14] H. Li, L. Zhang, M. Jiang, and Y. Li, “Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network,” *Pattern Recognition Letters*, vol. 141, pp. 45–53, Dec. 2020. https://doi.org/10.1016/j.patrec.2020.11.014
[15] T. Jiang, J. L. Gradus, and A. J. Rosellini, “Supervised Machine Learning: A Brief Primer,” 2020. [Online]. Available: www.elsevier.com/locate/bt
[16] L. Vivona, D. Cascio, F. Fauci, and G. Raso, “Fuzzy technique for microcalcifications clustering in digital mammograms,” *BMC Medical Imaging*, vol. 14, no. 1, pp. 1–18, 2014. https://doi.org/10.1186/1471-2342-14-10
[17] X. Chang, Z. Ma, X. Wei, X. Hong, and Y. Gong, “Transductive semi-supervised metric learning for person re-identification,” *Pattern Recognition*, vol. 108, pp. 1–12, Dec. 2020. https://doi.org/10.1016/j.patcog.2020.107569
[18] L. H. N. Lorena, A. C. P. L. F. Carvalho, and A. C. Lorena, “Filter Feature Selection for One-Class Classification,” *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 80, no. Icmc, pp. 227–243, 2015. https://doi.org/10.1007/s10846-014-0101-2
[19] R. Panthong and A. Srivihok, “Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm,” in *Procedia Computer Science*, 2015, vol. 72, pp. 162–169. https://doi.org/10.1016/j.procs.2015.12.117
[20] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, “Hybrid filter-wrapper feature selection for short-term load forecasting,” *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 17–27, Apr. 2015.https://doi.org/10.1016/j.engappai.2014.12.014
[21] J. Apolloni, G. Leguizamón, and E. Alba, “Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments,” *Applied Soft Computing Journal*, vol. 38, pp. 922–932, Jan. 2016. https://doi.org/10.1016/j.asoc.2015.10.037
[22] T. A. Alhaj, M. M. Siraj, A. Zainal, H. T. Elshoush, and F. Elhaj, “Feature selection using information gain for improved structural-based alert correlation,” *PLoS ONE*, vol. 11, no. 11, pp. 1–18, 2016. https://doi.org/10.1371/journal.pone.0166017
[23] I. Koprinska, M. Rana, and V. G. Agelidis, “Correlation and instance based feature selection for electricity load forecasting,” *Knowledge-Based Systems*, vol. 82, pp. 29–40, 2015. https://doi.org/10.1016/j.knosys.2015.02.017
[24] R. J. Urbanowicz, M. Meeker, W. la Cava, R. S. Olson, and J. H. Moore, “Relief-based feature selection: Introduction and review,” *Journal of Biomedical Informatics*, vol. 85. Academic Press Inc., pp. 189–203, Sep. 01, 2018. https://doi.org/10.1016/j.jbi.2018.07.014
[25] M. A. Berbar, “Hybrid methods for feature extraction for breast masses classification,” *Egyptian Informatics Journal*, vol. 19, no. 1, pp. 63–73, Mar. 2018. https://doi.org/10.1016/j.eij.2017.08.001
[26] J. Singh, K. Singh, and J. Singh, “Reengineering framework for open source software using decision tree approach,” *International Journal of Electrical and Computer Engineering*, vol. 9, no. 3, pp. 2041–2048, 2019. http://doi.org/10.11591/ijece.v9i3.pp2041-2048
[27] K. S. Reddy and E. S. Reddy, “Integrated approach to detect spam in social media networks using hybrid features,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, p. 562, 2019. http://doi.org/10.11591/ijece.v9i1.pp562-569
[28] M. Ouzzani, I. Ilyas, H. Hammady, A. Elmagarmid, and M. Khabsa, “Learning to identify relevant studies for systematic reviews using random forest and external information,” *Machine Learning*, vol. 102, no. 3, pp. 465–482, 2015. https://doi.org/10.1007/s10994-015-5535-7
[29] L. Ma and S. Fan, “CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests,” *BMC Bioinformatics*, vol. 18, no. 1, pp. 1–18, 2017. https://doi.org/10.1186/s12859-017-1578-z
[30] M. Dong, Z. Wang, C. Dong, X. Mu, and Y. Ma, “Classification of Region of Interest in Mammograms Using Dual Contourlet Transform and Improved KNN,” vol. 2017, 2017.
[31] K. Huang, S. Li, X. Kang, and L. Fang, “Spectral–Spatial Hyperspectral Image Classification Based on KNN,” *Sensing and Imaging*, vol. 17, no. 1, pp. 1–13, 2016. https://doi.org/10.1007/s11220-015-0126-z