Gender Classification for Anime Character Face Image Using Random Forest Classifier Method and GLCM Feature Extraction

Dadang Iskandar Mulyana[1] , Vika Vitaloka Pramansah[2]

[1]*Department of Informatics Engineering , STIKOM Cipta Karya Informatika, Indonesia,*
[2]*Department of Information Systems , STIKOM Cipta Karya Informatika, Indonesia*

[1]mahvin2012@gmail.com, [2]vitalokavika@gmail.com

Abstract - Japan has many entertaining and unique artworks, especially its signature animation, called anime. Anime is an animation art that is unique in that the characterizations, characters, and storylines are made to resemble human life. The characters have 2 genders called male and female with unique visuals and are the characteristics of each anime character to entertain the audience. Training large-scale data and complex textures because not all of the anime images owned are of high quality, making classification by Machine Learning Algorithms low in accuracy. This study will describe an experiment using an anime face image dataset to classify the gender, namely male or female. From this problem, this research implements feature extraction to produce unique features of anime images with Gray-Level Cooccurrence Matrix (GLCM) and uses the Random Forest Classifier which is a classification algorithm in Machine Learning to classify gender. The results of this study get a good accuracy value of 95%, using 3,612 images where the test data used is 723 images and Homogeneity5 feature being the most relevant feature in increasing the accuracy value with a value of 0.06378389.

Keywords: Random Forest Classifier, GLCM, anime, classification, gender.

I. INTRODUCTION

Classifying/identifying an object from an image or video has been widely studied using the Machine Learning method. For example with fuzzy or Artificial Neural Networks which is a classic supervised classifier. However, there is an unsolved problem where in complex remote sensing data does not always apply the Gaussian distribution assumed by Machine Learning. A large number of training samples is required by a large number of parameters in the Neural Network Classifier in order to optimize its capabilities, especially when the input dimensions are increasing. In the Fuzzy Classifier, pay attention to how to present the last output to the user so that good problem handling is needed. Then in the classic classifier, if the size of the data set is new and complex it will make the classifier difficult. In addition, the lower the accuracy and the complex texture of the image due to the more datasets [1].

Until now, gender recognition by direct visual observation has been emphasized in many jobs, but in performing this task it must be done by computers [2]. The criteria used by humans are difficult to define or translate into a format usable by computers, especially for classical machine learning (ML) techniques [3]. Much research is working on facial feature extraction and facial recognition for still images and videos. By using static images and videos, many studies have been carried out in terms of facial feature extraction and facial recognition. Various applications including identity authentication, human and computer interaction [4], a number of studies using facial images, and access control can be enhanced by gender classification games. [5]. Human and computer interaction interfaces, video surveillance systems, collection and evaluation systems for demographic indicators, robotics, intellectual digital information representation systems, and so on have widely used gender classification systems [6]. Machine learning techniques used in gender classification have a universal character that allows to application solutions and knowledge in other image understanding gender recognition or object classification tasks [7].

In this study, I propose the Random Forest Classifier method for classifiers with GLCM (Gray Level Co-occurrence Matrix) as feature extractor to analyze features from the provided data to simplify images that have complex textures. The dataset used is a male and female anime face image dataset. After loading the training data, the evaluated features, and the analysis after the experimental training data is loaded, the image feature data set will be generated by feature extraction. This process provides various features such as Mean and texture-based features such as GLCM and Entropy [8].

Feature extraction is the extraction of characteristics in an image [9]. In this process, high-level information such as shape, texture, and color can be obtained from an image [10]. GLCM is statistically similar in the technique of analyzing textures in images [11]. GLCM counts the occurrence of the same image matrix in the image pixels to obtain feature values and ave distance, grayscale angle and intensity in relation between 2 neighboring pixels are representative of GLCM [12]. GLCM examined through arithmetic methods by considering the association of pixel dimensions [13]. There are 8 angles that can be used in GLCM, including 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315° [14]. The features of dissimilarity, angular second moment, energy, homogeneity, and contrast are statistical features of GLCM and a solitary feature vector is formed from all of these features [15]. GLCM has been shown to be more intense in the world than in the world played by images with a linear line space or two pixels [16].

Random Forest (RF) is a new and promising classifier by Breiman in 2001 [17]. Used as handling input variables that number in the thousands without having to delete variables and in their classification can estimate which variables are important [18]. Random Forest is known as an efficient algorithm because it can classify large amounts of data well [19]. Merging trees in training the sample data provided is the way of the Random Forest classifier. Random Forest works by randomly sampling the training set in turns [20]. The goal is to generate a new training set using the bootstrap resampling method. Each tree that forms a random forest consists of different subsets and is followed by selecting a terminal node. The results of the classification are generated from the most votes in the last process [21]. Random Forest has a bagging principle and its ability to avoid overfitting is new to the classification [22]. Small decision tree input variables can be assigned based on the physiological characteristics of the face, both the distance between the feature points and the depth gradient of the points of various facial features in gender recognition [23].

Random Forest algorithm is used as a classifier in large amounts of data. This is done by combining a number of decision trees that conduct training on the training data they have. Random retrieval of training data and random retrieval of training features are two aspects of randomness in Random Forest. This can avoid overfitting problems, different image features and trees grown using subsets of data is the reason [24]. The advantages of the RandomForest method are that it runs efficiently on the use of large data, the method is effective in estimating missing data, and variables in important classifications can be estimated [25]. The wide acceptance of the Random Forest Classifier is also due to its easy implementation [26].

Anime is entertainment in the form of animation that can be made by hand or with computer applications originating from Japan. Anime is very popular in its own country and other countries, based on the report " The Report on Japanese Animation Industry " issued by The Association of Japanese Animation (AJA) in 2020 reported that sales from 2010 - 2019 recorded a significant figure of 2.511 Trillion Yen . Table 1 shows the growth of the anime market from 2002 – 2019 submitted by The Associaction Japanese Animation and other publicly available statistics as in Table I.

It can be seen that the anime market growth is very good from the anime market table above. Anime has characters, visual characters, and storylines like humans, similar to television content such as soap operas, FTV, dramas and the like. Anime images are unique so they can be used for computer classification, but have complex textures because not all images are of good quality, sometimes the images obtained are not clear/blurred so we need a good method that produces good classification accuracy.

Previously, some studies had been conducted regarding the use of classification algorithms in Machine Learning with GLCM feature extraction. The first research is conducted by [19]. For his experiments using handwritten form sections from 130 authors with 5 samples per author. The results obtained show a promising ability of the CSCA-Based Classifier in Writer identification and the SVM classification method produces high accuracy of 84.23% and Random Forest with an accuracy of 79.23% [19].

The second study is conducted by [18]. This study evaluates the potential of mono- and multi-seasonal geostatistical texture measurements together with a relatively new technique, the RF classifier, to classify various categories of highly heterogeneous mediterranean land cover.

TABLE I
ANIME MARKET (SOURCE: THE ASSOCIATION OF JAPANESE ANIMATION (AJA))

Year	Total Anime Market (in Billion Yen)
2002	1,097
2003	1,118
2004	1,223
2005	1.304
2006	1.350
2007	1.314
2008	1.389
2009	1,266
2010	1,324
2011	1,338
2012	1,339
2013	1,476
2014	1,636
2015	1,829
2016	2,001
2017	2,162
2018	2,181
2019	2,511

Random Forest yielded an accuracy of 0.85 for MKT data and 0.92 for MKT + GT31 data with 972 potential input variables [18].

The third study is conducted by [23]. In this paper, the authors propose to combine the physiological characteristic point depth gradient feature of the face with the physiological characteristic point distance feature. Then build a recognition model based on the above two features using a random forest algorithm, and realize gender recognition. The database used is Texas 3D Facial Recognition, which contains 1149 pairs of high-resolution facial color images resulting in an average accuracy of 89.3% with the Random Forest algorithm for gender classification [23].

The fourth study is conducted by [25]. In this paper, we investigate the set level of active contour segmentation of mass-based classification in digital mammograms. For classification, both SVM and RF (random forest) were investigated for classification. Experiments were tested using a database of 236 clinical mammograms. Accuracy results obtained with SVM and RF were 86% and 83%, respectively [25].

That's research from other researchers and has a good accuracy value. This study aims to classify the gender of anime characters which consists of 2 classes, namely male and female through the provided images of 3,612 images using GLCM (Gray Level Co-occurrence Matrix) as feature extractor and the Random Forest Classifier classification method.

II. METHOD

The method used can be seen in flowchart Fig. 1.

A. Data Collection

The dataset used in this study was downloaded from the public dataset on the Kaggle website uploaded by the Spencer Churchill account entitled "Anime Face Dataset", which can be accessed via the link. The dataset is separated by folders with male and female labels where the data used are 3,612 images with details of 2,505 images labeled as female for example as in Fig. 3 and 1,107 images labeled as male for example as in Fig. 2.

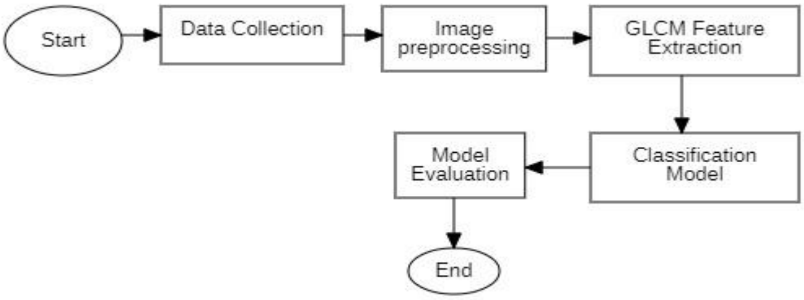

Fig. 1 Research flowchart

Fig. 2 An example of male anime

Fig. 3 An example of female anime

B. Image Preprocessing

Image preprocessing is the dataset processing stage, either by cleaning the data or converting it to the required data format before being trained or tested. This process makes it easier for the model to study the data because the level of complexity of the data is reduced. There are several techniques in image preprocessing such as image resizing, image augmentation, and changing color images to grayscale. An example of the difference before and after preprocessing as in Fig. 4.

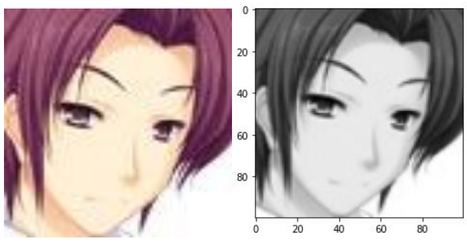

Fig. 4 Real image (left) and image after preprocessing (right)

C. GLCM Feature Extraction

GLCM is an image texture analysis technique. GLCM counts the occurrence of the same image matrix in the image pixels to get the feature value and where the grayscale intensity is owned between 2 neighboring pixels to represent relationships, distances, and angles. There are 8 angles that can be used in GLCM, namely angles 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. In GLCM, the features in this study consist of contrast, correlation, energy, homogeneity, and dissimilarity.

Energy is the total element squared in GLCM and can be calculated using (1).

$$E = \sum_{x,y=0}^{level-1} P_{x,y}^{2} \tag{1}$$

Contrast is a measure of local intensity variation in the GLCM and can be calculated using (2).

$$Con = \sum_{x,y=0}^{level-1} P_{x,y}\ (x-y)^2 \tag{2}$$

Homogeneity is a measure of the uniformity of the gray level in the image and can be calculated using (3).

$$H = \sum_{x,y=0}^{level-1} \frac{P_{x,y}}{1+(x,y)^2} \tag{3}$$

Dissimilarity is a measure of the distance between pairs of an object in pixels in the expected region and can be calculated using (4).

$$D = \sum_{x,y=0}^{level-1} P_{x,y}\ |x-y| \tag{4}$$

Correlation is used to measure the probability of occurrence of a composite pair of pixels and can be calculated using (5).

$$Cor = \sum_{x,y=0}^{level-1} P_{x,y} \left[\frac{(x-\mu_x)(y-\mu_y)}{\sqrt{(\sigma_x^2)(\sigma_y^2)}} \right] \tag{5}$$

Information :
E : Energy
Con : Contrast
H : Homogenity
D : Dissimilarity
Cor : Correlation
x,y : Coordinates of pixels in the GLCM
level : Gray tone range, in digital images, is 0–255 (level=256)
$P_{x,y}$: P pixel values at coordinates x,y matrix GLCM

D. Classification Model

Classification is estimating an image whose label is not known. The classification model in Machine Learning used in this research is the Random Forest, which works by randomly sampling the training sets in turn. The goal is that new training sets can be generated using the bootstrap resampling method. Each tree that makes up a Random Forest which consists of different subsets is followed by selecting a terminal node. The classification results are generated from the most votes in the last process. The workflow can be seen as in Fig. 5.

Random Forest Classifier workflow image as in Fig. 5, it can be seen that the existing dataset will be trained and then generate several decision trees from the Decision Tree model then each tree has a decision the image being trained will be categorized into what label, and then the most votes from all the decision trees generated will determine the label of the image class or final class in Random Forest Classifier.

Previously, in this study, classification was carried out using another classification method, namely SVM (Support Vector Machine) and Light GBM (Light Gradient Boosting Machine) to determine the method that has the best accuracy. SVM is used to find the best hyperplane by maximizing the distance between classes/labels. Hyperplane in SVM is a function that can be used for separators between classes. In a 2-dimensional dataset, the function used for classification between classes is called line whereas, the function used for classification between classes in 3 dimensions is called plane similarly, while hyperplane is a classification function in higher-dimensional class space. Light GBM is a high performance framework, based on decision tree algorithms, gradient enhancement, classification, used for ranking, and various other Machine Learning tasks.

E. Model Evaluation

Evaluation of the model used for anime gender classification was carried out to find out which model had the highest accuracy. Before deciding to use the Random Forest Classifier, this research has experimented with other methods, namely SVM and Light GBM. Of these three methods using the Confusion matrix to determine the results of accuracy, f-score, recall, and precision.

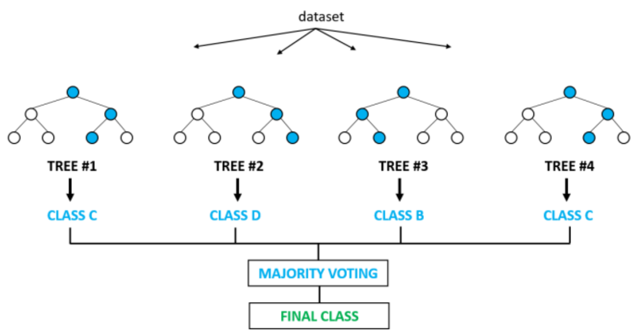

Fig. 5 Random forest classifier workflow

III. RESULT AND DISCUSSION

A. Anime Character Face Dataset

In this study, the dataset used was 3,612 images with details of 2,505 images labeled as female and 1,107 images labeled as male. The dataset will be divided into two parts, namely training data as much as 80% and test data as much as 20% of the entire dataset.

B. Image Preprocessing

The dataset used will be processed first to fit the desired format which is intended to facilitate the model in studying image data. In this research, the image is converted from RGB to grayscale to simplify the image model used, besides that the image we have will be converted to a width and height of 100 pixels for its size. An example of the difference before and after converted to grayscale and the size is 100 x 100 pixels as in Fig. 7.

After the image is changed in color and size, the image intensity is then normalized with a value between 0 and 1 using the Image Processing Python library, namely Scikit-image.

Image files are usually stored as integers for storage space efficiency, but transformations and other mathematical operations often result in conversions to fractions. To work around this, the image uses the img as_ubyte method to convert it back to an integer before being saved again. After that, the distribution of training data and test data is 80:20.

C. GLCM Feature Extraction

In this research, GLCM is used for texture analysis of anime imagery. Obtaining/extracting features prior to classification, can make it easier for images to be classified according to their textures because the image quality is sometimes not always good and anime images are artificial images that have unique characteristics. The GLCM matrix and the overall texture matrix of anime images are searched using the grey comatrix and grey coprops functions in the Scikit-Image library.

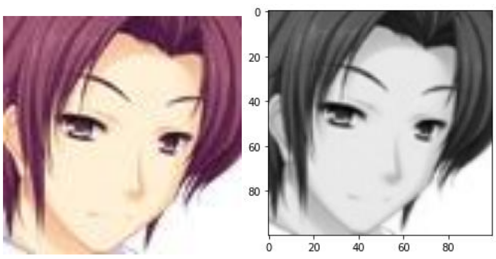

Fig. 7 The image before (left) and after converted to grayscale and the size is 100 x 100 pixels (right)

The parameters of the grey comatrix function used in this study are image, distances, and angles. The distances parameter used to construct the GLCM determines the scale at which textures are analyzed. The optimal choice of this distance depends on the scale of the texture being analyzed. Many implementations only consider distance 1. The distances used are 1 and 5. While the angles used are 0, 45, 90 and 135. GLCM will produce 25 data. An example of the resulting data matrix is as in Table II.

From the data matrix table, the energy value will be of high value when the pixel values are similar to each other and vice versa will be of small value indicating the value of the normalized GLCM matrix is heterogeneous. The dissimilarity value will be high if it is random and vice versa it will be small if it is uniform, the contrast value will be 0 if the neighboring pixels have the same value, and the homogeneity value will be high when all pixels have the same/uniform value. GLCM can analyze quite well from the image of the trained anime.

TABLE II
DATA MATRIX OF 3 TRAINING DATA

	 **Image 1**	 **Image 2**	 **Image 3**
Energy	0.416	0.327	0.305
Correlation	0.924	0.936	0.935
Dissimilarity	0.321	0.387	0.442
Homogeneit	0.847	0.831	0.800
Contrast	0.397	0.638	0.660
Energy2	0.400	0.316	0.297
Correlation2	0.891	0.898	0.914
Dissimilarity	0.404	0.509	0.507
Homogeneit	0.814	0.793	0.781
Contrast2	0.571	1.015	0.870
Energy3	0.411	0.319	0.300
Correlation3	0.899	0.915	0.921
Dissimilarity	0.364	0.464	0.497
Homogeneit	0.834	0.805	0.781
Contrast3	0.526	0.850	0.796
Energy4	0.434	0.342	0.345
Correlation4	0.946	0.954	0.967
Dissimilarity	0.229	0.299	0.259
Homogeneit	0.891	0.867	0.878
Contrast4	0.284	0.460	0.334
Energy5	0.325	0.264	0.230
Correlation5	0.588	0.606	0.653
Dissimilarity	0.934	1.153	1,207
Homogeneit	0.644	0.640	0.590
Contrast5	2,149	3.858	5,272

D. Classification Model

In the anime gender classification, the Confusion matrix is used as a testing method. This method uses a matrix to represent the results of the classification consisting of True Positive (TP), which is the number of positive data that has been successfully classified as positive. False Positive (FP) is positive data that is wrongly classified as negative. False Negative (FN) is a negative record that is incorrectly classified as positive. True Negative (TN) is a negative record that has been successfully classified as a negative record. The formula in the Confusion matrix is as follows:

Precision is the ratio of TP to the total image that has been predicted to be positive and can be calculated using (6).

$$P = \frac{TP}{(FP+TP)} \text{ x } 100\% \quad (6)$$

The recall is the ratio of TP to the total image that should be positive and can be calculated using (7).

$$R = \frac{TP}{(FN+TP)} \text{ x } 100\% \quad (7)$$

Accuracy is the comparison between the correct total image and all existing images and can be calculated (8).

$$A = \frac{(TP+TN)}{(TP+TN+FP+FN)} \text{ x } 100\% \quad (8)$$

F1 Score is the average harmony of recall and precision and can be calculated using (9).

$$F1 = \frac{2TP}{(2TP+FN+FP)} \text{ x } 100\% \quad (9)$$

Information :

P : Precision
R : Recall
A : Accuracy
F1 : F1 Score
TP : True Positive
FP : False Positive
FN : False Negative
TN : True Negative

In this anime gender classification, it was tested with Machine Learning models, namely Random Forest Classifier, SVM, and LightGBM to find the highest accuracy. Here are the results obtained as in Table III. From the training data accuracy table and the confusion matrix of each of the above models, it can be seen that the highest accuracy value is the Random Forest Classifier model with an accuracy value of 95%. That is why the author uses the Random Forest Classifier method as the method proposed for gender classification in the image of anime characters. This is because the Random Forest Classifier method can train image data and test anime images better than the other tested methods.

E. Model Evaluation

The final result of the test shows that the value of gender classification accuracy of anime character facial dataset images has a fairly good result, namely 95% as in Table IV, but the Random Forest Classifier classifying datasets takes a long time in this study, which is 1,565 seconds or approximately 26 minutes.

Based on the feature importance graph as in Fig. 8, it is found that the Homogeneity5 feature gets the highest value, which is 0.06378389. That means the Homogeneity5 feature which uses an angle of 0 degrees and a distance of 5 pixels is the most relevant feature in this model.

TABLE III
ACCURACY OF TRAINING AND TESTING DATA

Data	Random Forest Classifier	Support Vector Machine	LightGBM
Training	99%	77%	70%
Testing	95%	80%	71%

TABLE IV
DETAILS OF THE CONFUSION MATRIX OF THE PROPOSED METHOD

Class	Precision	Recall	F1 Score	Amount
Male	94%	88%	91%	208
Female	95%	98%	97%	515
Average	95%	93%	94%	723
Accuracy		95%		

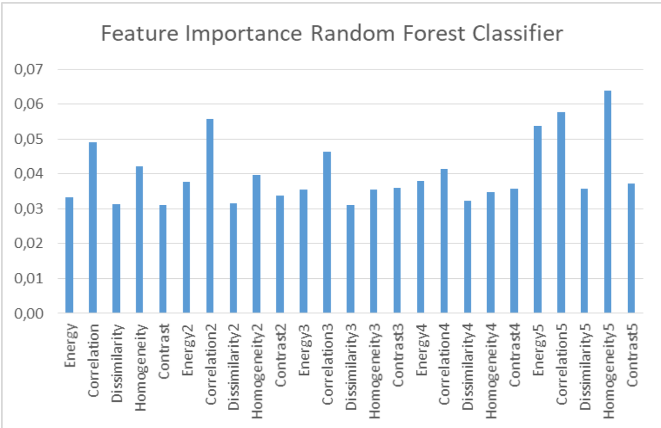

Fig. 8 Graph of feature importance

IV. CONCLUSION

With quite a lot of datasets and textures of anime images that have complex or complicated textures and images of varying intensity quality, the Random Forest Classifier produces a classification accuracy value of 95% which indicates it is good for classifying the owned dataset which uses test data as many as 723 images and trained data. as many as 2,889 images. This is also supported by the use of Gray-Level Cooccurrence Matrix (GLCM) as texture analysis and GridSearchCV which makes it easier to find the best Random Forest Classifier parameters where before the texture of the image is analyzed, the image is preprocessed first to make it easier for the model to study the dataset. And based on this research using the Random Forest Classifier, the most important feature or also called feature importance is the Homogeneity5 feature because it gets the highest value, which is 0.06378389. Therefore, the Homogeneity5 feature is the most relevant in increasing the accuracy value.

REFERENCES

[1] A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier, “Evaluating the performance of a Random Forest Kernel for land cover classification,” *Remote Sens.*, vol. 11, no. 5, pp. 1–20, 2019, doi: 10.3390/rs11050575.

[2] E. O. Abdulali, A. S. Huwedi, and K. A. Bozed, “Gender detection using random forest,” *ACM Int. Conf. Proceeding Ser.*, 2020, doi: 10.1145/3410352.3410799.

[3] Y. Nie, W. Ren, B. Liang, J. Dai, and P. Huang, “A study on image based gender classification using convolutional neural network,” *ACM Int. Conf. Proceeding Ser.*, pp. 81–84, 2019, doi: 10.1145/3342999.3343011.

[4] G. S. Lin, M. K. Chang, Y. J. Chang, and C. H. Yeh, “A gender classification scheme based on multi-region feature extraction and information fusion for unconstrained images,” *Multimed. Tools Appl.*, vol. 75, no. 16, pp. 9775–9795, 2016, doi: 10.1007/s11042-015-2797-9.

[5] A. Hasnat and S. Halder, “a Proposed System for Gender Classification,” pp. 581–585, 2015.

[6] V. V. Khryashchev, L. A. Shmaglit, A. L. Priorov, and A. M. Shemyakov, “Extracting adaptive features for gender classification of human face images,” *Program. Comput. Softw.*, vol. 40, no. 4, pp. 215–221, 2014, doi: 10.1134/S0361768814040057.

[7] L. Shmaglit and V. Khryashchev, “Gender classification of human face images based on adaptive features and support vector machines,” *Opt. Mem. Neural Networks (Information Opt.*, vol. 22, no. 4, pp. 228–235, 2013, doi: 10.3103/S1060992X13040036.

[8] P. Kathale and S. Thorat, “Breast Cancer Detection and Classification,” *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, vol. ৬, pp. 1–5, 2020, doi: 10.1109/ic-ETITE47903.2020.367.

[9] P. Panchal, V. C. Raman, and S. Mantri, “Plant Diseases Detection and Classification using Machine Learning Models,” *CSITSS 2019 - 2019 4th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. Proc.*, vol. 4, pp. 1–6, 2019, doi: 10.1109/CSITSS47250.2019.9031029.

[10] M. V. Suhas and A. Mishra, “Classification of benign and malignant bone lesions on CT images using random forest,” *2016 IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2016 - Proc.*, pp. 1807–1810, 2017, doi: 10.1109/RTEICT.2016.7808146.

[11] H. Zhang, Q. Li, J. Liu, X. Du, T. Dong, H. McNairn, C. Champagne, M. Liu & J. Shang, “Object-based crop classification using multi-temporal SPOT-5 imagery and textural features with a Random Forest classifier,” *Geocarto Int.*, vol. 33, no. 10, pp. 1017–1035, 2018, doi: 10.1080/10106049.2017.1333533.

[12] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, “Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection,” *Biomed. Signal Process. Control*, vol. 33, pp. 272–280, 2017, doi: 10.1016/j.bspc.2016.11.021.

[13] A. Murugan, S. A. H. Nair, and K. P. S. Kumar, “Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers,” *J. Med. Syst.*, vol. 43, no. 8, 2019, doi: 10.1007/s10916-019-1400-8.

[14] Y. Wang, H. Xia, X. Yuan, L. Li, and B. Sun, “Distributed defect recognition on steel surfaces using an improved random forest algorithm with optimal multi-feature-set fusion,” *Multimed. Tools Appl.*, vol. 77, no. 13, pp. 16741–16770, 2018, doi: 10.1007/s11042-017-5238-0.

[15] S. D. Thepade, S. V. Bang, P. R. Chaudhari, and M. R. Dindorkar, “Covid19 Identification from Chest X-ray Images Using Machine Learning Classifiers with GLCM Features,” *Electron. Lett. Comput. Vis. Image Anal.*, vol. 19, no. 3, pp. 85–97, 2020, doi: 10.5565/REV/ELCVIA.1277.

[16] M. Muhathir, M. H. Santoso, and D. A. Larasati, “Wayang Image Classification Using SVM Method and GLCM Feature Extraction,” *J. Informatics Telecommun. Eng.*, vol. 4, no. 2, pp. 373–382, 2021, doi: 10.31289/jite.v4i2.4524.

[17] J. Liang, Q. Liu, N. Nie, B. Zeng, and Z. Zhang, “An improved algorithm based on KNN and random forest,” *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3331453.3360963.

[18] V. F. Rodriguez-Galiano, M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan, “Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture,” *Remote Sens. Environ.*, vol. 121, pp. 93–107, 2012, doi: 10.1016/j.rse.2011.12.003.

[19] A. Bennour, "Clonal selection classification algorithm applied to Arabic writer identification," *ACM Int. Conf. Proceeding Ser.*, pp. 1–5, 2018, doi: 10.1145/3200842.3208087.

[20] S. Ayas and M. Ekinci, "Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples," *Signal, Image Video Process.*, vol. 8, no. 1, pp. 49–61, 2014, doi: 10.1007/s11760-014-0708-6.

[21] S. Nadgeri and A. Kumar, "An Image Texture based approach in understanding and classifying Baby Sign Language," *2019 2nd Int. Conf. Intell. Comput. Instrum. Control Technol. ICICICT 2019*, pp. 854–858, 2019, doi: 10.1109/ICICICT46008.2019.8993381.

[22] D. Chutia, N. Borah, D. Baruah, D. K. Bhattacharyya, P. L. N. Raju, and K. K. Sarma, "An effective approach for improving the accuracy of a random forest classifier in the classification of Hyperion data," *Appl. Geomatics*, vol. 12, no. 1, pp. 95–105, 2020, doi: 10.1007/s12518-019-00281-8.

[23] G. Liao, W. Chen, and Y. Wu, "Facial features for gender recognition," *Chinese Control Conf. CCC*, vol. 2016-Augus, pp. 4161–4165, 2016, doi: 10.1109/ChiCC.2016.7554002.

[24] Y. Dong, Y. Zhang, J. Yue, and Z. Hu, "Comparison of random forest, random ferns and support vector machine for eye state classification," *Multimed. Tools Appl.*, vol. 75, no. 19, pp. 11763–11783, 2016, doi: 10.1007/s11042-015-2635-0.

[25] J. Liu, J. Chen, X. Liu, and J. Tang, "An investigate of mass diagnosis in mammogram with random forest," *Proc. 4th Int. Work. Adv. Comput. Intell. IWACI 2011*, pp. 638–641, 2011, doi: 10.1109/IWACI.2011.6160086.

[26] A. Paul and D. P. Mukherjee, "Reinforced random forest," *ACM Int. Conf. Proceeding Ser.*, 2016, doi: 10.1145/3009977.3010003.