

Identification of Skin Disease Using K-Means Clustering, Discrete Wavelet Transform, Color Moments and Support Vector Machine

I Ketut Gede Darma Putra, Ni Putu Ayu Oka Wiastini, Kadek Suar Wibawa, and I Made Suwija Putra

Abstract—Skin disease is one of disease that is often found in tropical countries, such as Indonesia. People who suffered from skin disease in Indonesia were still relatively high, the prevalence could range between 20% - 80%. Therefore, the help of computer technology was expected to detect the disease earlier that attacked the skin in the human's body and it could reduce the possibility of the occurrence for other dangerous diseases. This study proposed the making of an application of identification image for skin disease by using one of the machines learning method, called Support Vector Machine (SVM) which was done by processing the image and machine learning processes that could perform early detection of skin diseases. This study aimed to determine the classification of skin diseases in humans into four classes, such as the class Benign Keratosis, Melanoma, Nevus, and Vascular. The segmentation method used was K-Means Clustering, while the feature extraction method that used was feature extraction of the Discrete Wavelet Transform (DWT) and Color Moments. Based on the results of the test that had conducted, the sensitivity was 95%, the specificity was 97.9% and the accuracy was 97.1% by using SVM parameters, that was kernel Radial Basis Function (RBF), Box Constraint = 1.5, RBF_Sigma (σ) = 1, and iterations = 1000.

Index Terms—Skin disease, K-Means clustering, discrete wavelet transform, color moments, SVM.

I. INTRODUCTION

Skin is the outer surface and the largest organ in the human's body, which limits the environment from the disease and its transmission to the skin, where the function is to show any damage occurs in the body with the change in pigmentation or color [1], [2]. Unmaintained skin can lead to various diseases of skin cancer, which able to impair the appearance and activity of the infected person.

Skin disease can be caused by infected bacteria, fungi, parasites or viruses [3]. Moreover, skin disease is one of disease which often found in tropical countries, such as Indonesia with the prevalence that can range between 20% - 80% [4]. This is due to less awareness of the society towards the surrounding environment along with the lack of knowledge about the types and the impacts of skin disease which caused people to ignore the fact of it. In addition, skin disease can infect anyone in any part of the body regardless

of age [2]. Generally, in the world of health, skin disease can be detected by using a biopsy method which is extremely painful. Because the action that is applied is to take partially the tissue in the skin for histopathological examination under a microscope [5]. Therefore, based on those problems, several automatic approaches were proposed assisted by the technology of computer to analyze dermoscopy images [6].

Recently, with the advanced development of technology, the studies about the classification of the skin disease combined with digital image processing have been performed. In the previous research, Revati Kadu and Dr. U. A. Belorkar [7], who conducted Wavelet extraction methods and Artificial Neural Network for classification methods to detect five types of skin diseases. The proposed method was successful in detecting various types of dermatological skin diseases. Other research also conducted by S. Reena Parvin and O.A. Mohamed Jafar [8], who discussed the prediction of skin diseases, such as melanoma and non-melanoma. The study applied 80 datasets with accuracy that was achieved at 97% to 98%. However, the diagnosis of skin diseases through images is very important in the medical world. When the skin disease is not properly treated, it will make the disease become more severe. Therefore, the appropriate method is necessary to detect human's skin disease, then society will be able to treat it immediately according to the type of skin disease they suffer.

The classification method proposed in the development of this study is Support Vector Machine (SVM). SVMs is known as the technique in finding a hyperplane that can separate the data sets from two or more different classes. This method used to classify four types of skin disease images. Previous studies have succeeded by using this method for the classification and the detection of biomedical imagery, such as identification of Circulating tumor DNA [9], the detection and the classification of brain tumors [10] through MRI images, the classification of skin lesions [11], segmentation and classification of cervical cancer [12] through MRI images, focal liver detection [13] and breast cancer classification [14] using Support Vector Machine (SVM). In the study [15], SVMs mentioned has the advantage, that is a regularisation parameter, which makes the user think about avoiding over-fitting.

The research related to the identification of developed skin diseases was proceeded by 43 test images, which consisted of Benign Keratosis, Melanoma, Nevus and Vascular. This study focuses on the results of K-Means Clustering segmentation method implementation, the Discrete Wavelet Transform (DWT) extraction method and the Support Vector Machine (SVM) classifications method

Manuscript received July 9, 2019; revised May 12, 2020. This research was supported by Information Technology, Engineering Faculty, Udayana University.

I Ketut Gede Darma Putra, Ni Putu Ayu Oka Wiastini, Kadek Suar Wibawa and I Made Suwija Putra are with the Department of Information Technology, Faculty of Engineering, Udayana University, Bali 80361, Indonesia (e-mail: ikgdarmaputra@unud.ac.id, ayuwiaast@gmail.com, suar_wibawa@unud.ac.id, putrasuwija@unud.ac.id).

which differentiate it from the previous similar studies.

II. METHODOLOGY

This section gives a general view of the proposed algorithms for diagnosis of skin diseases. The system structure proposed to identify skin diseases in humans is shown in Fig. 1.

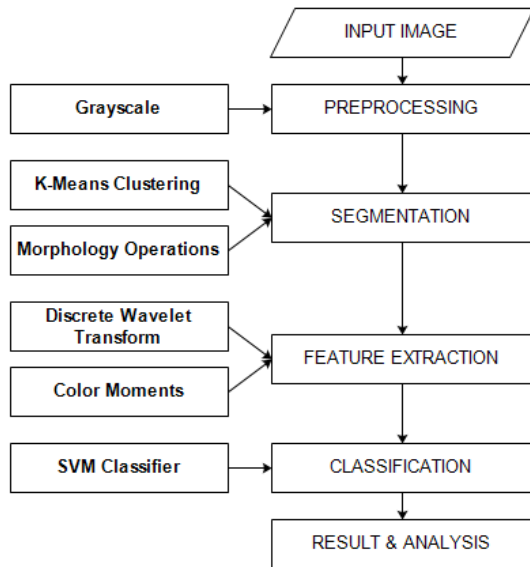


Fig. 1. Architecture of the proposed system.

A. Input Image

The input image applied is a dermatoscopic image of skin diseases about 131 images which are divided into 4 classes, such as Benign Keratosis, Melanoma, Nevus, and Vascular. Thus skin disease images were taken and downloaded from The International Skin Imaging Collaboration (ISIC) [16].

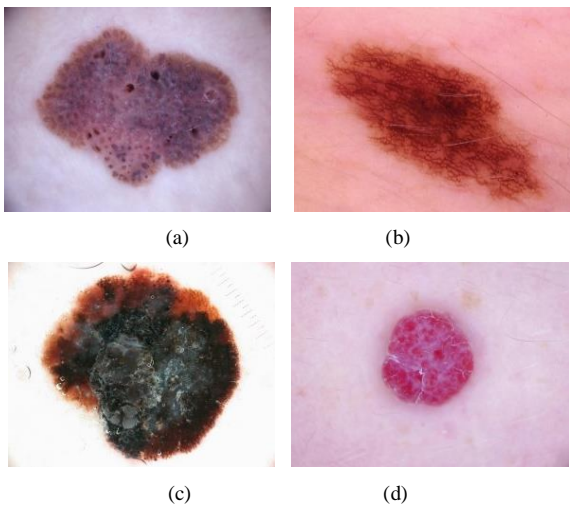


Fig. 2. Skin Disease images, (a) Benign keratosis, (b) Melanoma, (c) Nevus, (d) Vascular.

B. Preprocessing

The pre-processing stage is the first process that will transform the input data into a suitable data form with the format and ready to be processed. The pre-processing process was proposed to convert the RGB image into a

grayscale image, this is used to simplify the model of the image and simplify it during the segmentation process, which can also disguise the noises exist in the image.

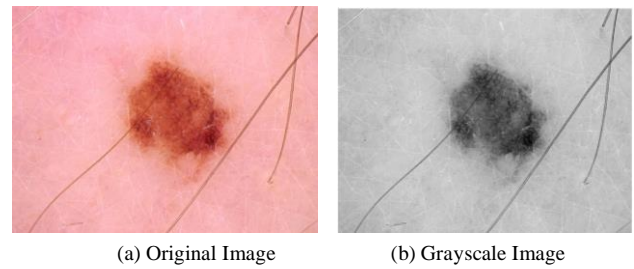


Fig. 3. Preprocessing.

C. Input Image

Image segmentation is an important operation in image recognition. The purpose of image segmentation is to divide the image into different parts, dividing which areas need the main focus compared to the background [17].

The K-Means Clustering method is used to segment the image of skin disease, where this method will split the data into several cluster regions based on the closest distance between the data and the centroid of each cluster. The output of the image segmentation is in a binary image where the foreground is white (1), while the background that wants to remove is black (0).

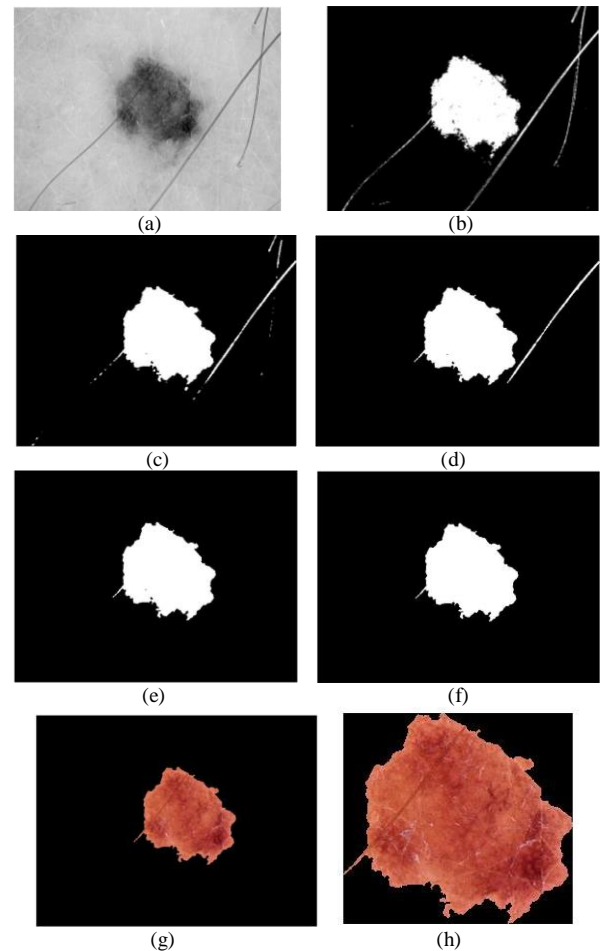


Fig. 4. Steps of an image segmentation process, (a) Grayscale image, (b) K-Means clustering, (c) Median filter, (d) Noise removal, (e) Border cleaning, (f) Fill the hole, (g) Masking and (h) Cropping.

The post-preprocessing stage is done on the results of

image segmentation with K-Means Clustering in Fig. 4(b) because these results are considered to be less than perfect and there are some noise/ small objects that go inside. The technique used is operations on binary images, such as the median filter, noise removal, and border cleaning, holes filling in objects, the masking process and cropping images of skin diseases.

D. Feature Extraction

There are several unique features that distinguish four classes of the human's skin disease, such as, the texture features and color features. These features are selected for the classification. There are two types of extraction methods that are proposed, such as the Discrete Wavelet Transform (DWT) and Color Feature.

1) Discrete wavelet transform (DWT)

The feature extraction method that is proposed to extract the texture features in skin disease images is known as Discrete Wavelet Transform (DWT). DWT can be used for decomposition of images. Meanwhile, the wavelet transformation process is done in a simple concept. The implementation of DWT is passing high-frequency signals through the highpass filter and the low-frequency filters through lowpass filters, which transformed the original image that is decomposed into four sub-new images to replace it. Each sub-images of ¼ times the size of the original image.

Sub-image at the top right, bottom left and bottom right will seem like a crude version of the original image. As for the first sub-image on the left looks like the original image and looks smoother (smooth) because it contains a low-frequency component [18].

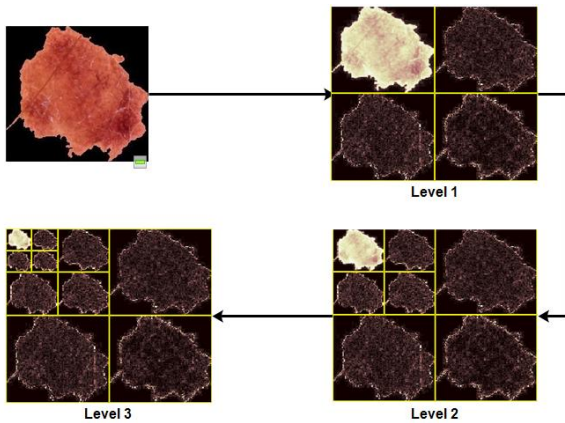


Fig. 5. Wavelet 2D decomposition.

In 2-D images, the decomposition process is done on rows and columns in two-dimensional arrays, which each of it corresponds to the horizontal direction and vertical direction in the image. The decomposition process will be done based on the level of the Wavelet. The decomposition process at the next level is done at the approximation coefficient (LL) which is the lowest frequency because it contains most of the image information. Based on the decomposition value, the energy which is the characteristics of the image will be calculated. Energy represents the resemblance of the image texture. Eventually, there are energy values which can be taken from the decomposition results, such as the approximation coefficient (LL), the horizontal detail coefficient (HL), the vertical detail

coefficient (LH) and the diagonal detail coefficient (HH).

2) Color moments

Skin disease images also can be divided based on the color variation of each class. The color feature extraction method proposed is the color moments. Color moments are a solid representation of color features in characterizing the image colors. The calculation of moments is used to get the color similarity of an image, where the value of similarity itself is used to compare the images contained in the image database [19].

Color moments assume the color distribution of an image as a probability distribution. This study will use two main moments from the color distribution of the image, such as mean and standard deviation. Therefore, this method produces two values for each color component. Mean is known as the average of color value in the image, while the standard deviation is the root of the variance in the distribution or the range of the spreading data from Mean. The mean and standard deviation can be calculated by using the following resemblance [19].

$$E = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (1)$$

$$\sigma_i = \sqrt{\left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^2 \right)} \quad (2)$$

In this research [20], the color moment method that tested had the best accuracy to recognize the color features of the skin diseases. The basis of this method is the assumption that the color distribution in an image can be expressed as a probability distribution. Therefore, the resulting accuracy is constant even though the image size changes.

The color space of YCbCr that is used consists of three color components (Y, Cb, and Cr). YCbCr is a color space that is often applied in the video and digital photography systems [21]. Y represents the luminance component, while Cb and C represent the chrominance components of red and blue.

E. Image Identification

The final step divided into two, such as the identification and classification of skin diseases images based on the four types of classes, such as Benign Keratosis, Melanoma, Nevus, and Vascular. This stage is divided into two processes, such as the training process and the testing process. The training process is conducted to train the data, therefore it produces a model which can be used to classify the new data (test data). Meanwhile, the testing process is used to test the results of classification by using the best models that are obtained from the training process. The method that is proposed in this stage is Support Vector Machine (SVM).

The concept of classification of SVM is the attempt in searching for the hyperplane that has functioned as a separator between the two data classes in the input space. Hyperplane (decision boundary) is the separator between two classes that can be found by measuring the hyperplane margin and looking for the maximum point. The margin is the distance between the hyperplane and the closest data from each class. Therefore, the closest data is called a support vector. The attempt to find a hyperplane location is in the essence of the training process for SVM.

SVMs uses a kernel trick that aims to classify the data

that cannot be classified linearly. The idea of this kernel is to map the input vectors from the original input space to high-dimensional feature space with several non-linear mapping functions [22]. There are types of kernel that can be used in SVM, which are Linear, Polynomial, and RBF.

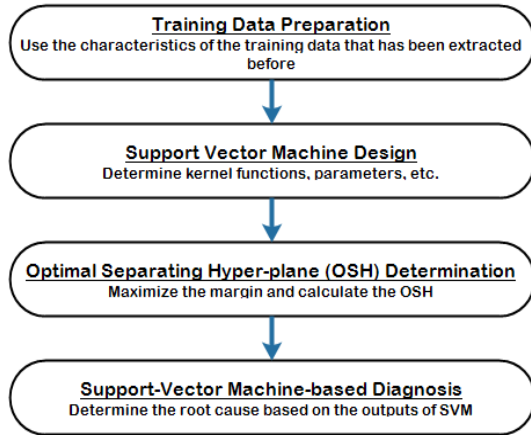


Fig. 6. Identification flow using SVM.

The flow of identification of skin disease images from the proposed method called Support Vector Machine (SVM) is shown in Fig. 6 [23].

The first step is the preparation of the training data which is taken based on the characteristics of all training images had extracted. This feature is used as the input and output for SVM training.

The second step, which is to determine the model or design of SVM. Proper kernel function and penalty parameter are chosen to determine the SVMs training. The choice of these two parameters affects the performance of SVMs. However, there are no generic rules to select the best kernel and other parameters of SVMs for a specific problem. According to [24] and [25], choosing the SVM with a Radial Basis Function (RBF) kernel and followed by relatively large penalty parameter are chosen to provide the highest diagnostic accuracy.

The third step is the determination of OSH or the optimal separation of hyperplane on the input data based on the kernel and the parameters used. In this step, SVM training has begun. The optimal determination of hyperplane is specified in Equations (3) and (4).

$$\omega = \sum_{i=1}^S \alpha_i \gamma_i x_i \tag{3}$$

$$b = \frac{1}{NS} \sum_{i=1}^S (\gamma_i - w \cdot x_i) \tag{4}$$

The fourth step, which is calculating SVM output using the decision function in Equation (5). After OSH is determined in step three, the SVM-based identification system is ready to identify a new data.

$$f(x) = \text{sgn}(\omega^T x + b) \tag{5}$$

III. RESULT AND ANALYSIS

The skin disease identification system became the better solution in the early prevention of the disease. Also, it was considered to be more efficient than the biopsy method in the medical world. This could minimize the costs and the time for the treatment that was needed by the society. This

research combined artificial intelligence and digital image processing to detect skin disease. The Support Vector Machine (SVM) classification method proved to be very efficient in the decision making and the pattern recognition. The test was done by using 40 test images. The steps of the test were to find out the level of accuracy that was produced based on the method that had proposed. Preprocessing techniques followed by image segmentation using K-Means Clustering were considered to be able to separate the foreground and background from the tested images.

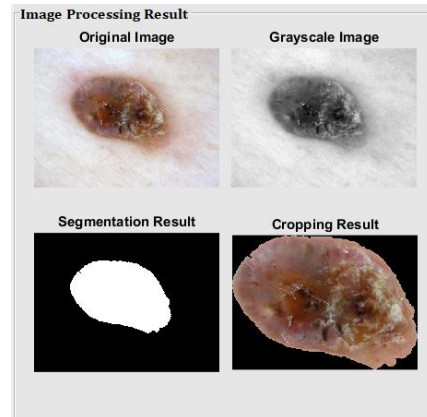
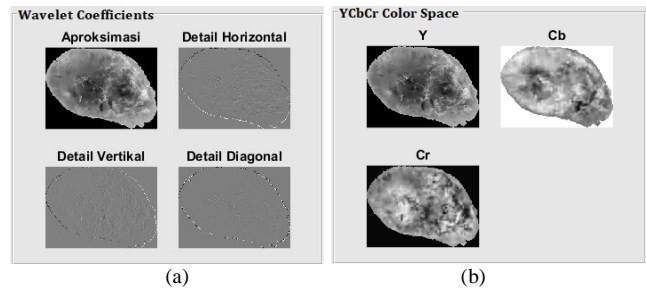


Fig. 7. Test Image segmentation.

Feature extraction that was tested use Discrete Wavelet Transform (DWT) in Level 1 decomposition which would produce four subband pieces, such as, the approximation coefficient, horizontal detail coefficient, vertical detail coefficient and diagonal detail coefficient. Meanwhile, the color moments proposed for color feature extraction would extract the moment mean and standard deviation in the image with the color spaces Y, Cb and Cr.



(a)

(b)

Feature Extraction		
	Property	Nilai
1	Energi CA	152.8194
2	Energi CH	-0.0639
3	Energi CV	0.0738
4	Energi CD	-0.0439
5	Mean Y	81.6156
6	Mean Cb	118.8445
7	Mean Cr	140.9586
8	Standar Deviasi Y	48.0834
9	Standar Deviasi Cb	7.1541
10	Standar Deviasi Cr	9.1740

(c)

Fig. 8. Feature extraction, (a) Wavelet transform, (b) YCbCr Color dan, (c) Feature extraction result.

The value of the features that were extracted from each images, found that there were ten features and this features used for the process of training and testing used the Support Vector Machine (SVM) method. The parameters with the

best hyperplane separator model in this study used Radial Basis Function (RBF) kernel.

On non-linear SVM separation, the data were mapped by functions of kernel with the high dimensional vector spaces. Thus, every data that exists in the input space were transformed into a new higher dimensional vector space, which was third dimension. In this new vector space, each class could be easily separated linearly by a hyperplane.

1) Evaluation metrics

In this study, the evaluation metrics were used to evaluate the performance of classification methods, such as, sensitivity, specificity and accuracy. The criteria for this metric were defined in the following equation.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (8)$$

2) Performance analysis

The determination of the better SVM model would be found out based on the parameters that influenced the training and testing process. These parameters were known as testing on the data ratio, type of kernel, C parameters, RBF Sigma parameters and the type of feature extraction that were used. The test was conducted to find out which scenario gets the best results in the process of classification.

TABLE I: KERNEL FUNCTION

Kernel	Sensitivity	Specificity	Accuracy
Linear	92.5%	97.3%	96.1%
Polynomial	90%	96.4%	94.7%
RBF	95%	97.9%	97.1%

The accuracy of classification used the SVM method depended on the function of the kernel and also the parameters. In SVM, learning about the kernel tricks models could help when overcoming the problems. Therefore, the choice of the kernel influenced the accuracy that was produced. The experiments used three SVM kernels resulted in the maximum classification accuracy that was achieved by using the RBF kernel.

The parameters were considered to be used when the RBF kernel, which were RBF_Sigma and BoxConstraint. RBF_Sigma was used to determine the width of the distribution of the RBF kernel function, while BoxConstraint used to determine the support vector of each class. These parameters were determinable to prevent overfitting in the training dataset.

TABLE II: BOXCONSTRAINT PARAMETERS

BoxConstraint	Sensitivity	Specificity	Accuracy
0.05	80%	92.3%	88.8%
0.08	80%	92.3%	88.8%
0.2	85%	94.4%	91.8%
0.5	95%	97.9%	97.1%
0.8	95%	97.9%	97.1%
1	95%	97.9%	97.1%

In the RBF kernel function, the influence of the

BoxConstraint parameter for SVM classification is shown in Table II. When RBF_Sigma is 1, the test is carried out for some BoxConstraint values which are 0.05, 0.08, 0.2, 0.5, 0.8 and 1, where the largest percentage classification accuracy is in the parameter value 1, 1.2 and 1.5. So, when the BoxConstraint value increases, it will produce maximum classification accuracy.

TABLE III: RBF_SIGMA PARAMETERS

RBF_Sigma	Sensitivity	Specificity	Accuracy
0.4	45%	71%	62%
0.6	60%	81.8%	75%
0.8	77.5%	91.1%	87.3%
1	95%	97.9%	97.1%
1.2	95%	97.9%	97.1%

The test for the RBF_Sigma parameter value is conducted when the value of BoxConstraint is 1.2 with the value of the RBF_Sigma parameter is 0.4, 0.6, 0.8, 1, and 1.2. The maximum percentage accuracy is obtained when the value of RBF_Sigma gets bigger, that is at the value of 1 and then the accuracy continues until the next value.

The influence illustration of BoxConstraint and RBF_Sigma parameters is shown by the graphs in Fig. 9 and Fig. 10.

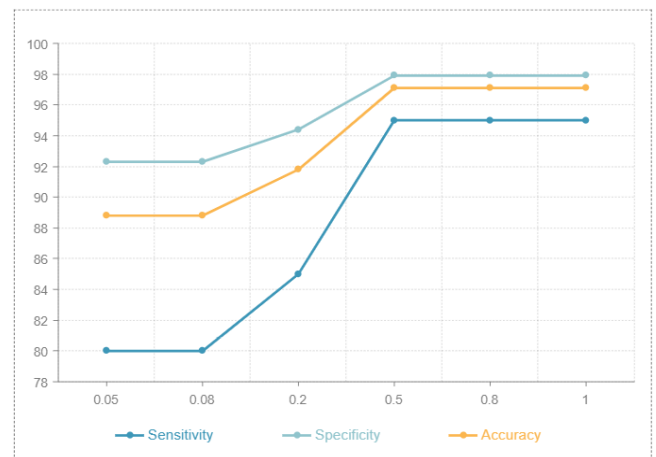


Fig. 9. Influence chart of BoxConstraint parameter.

The graph shown in Fig. 9 and Fig. 10 illustrate that, the larger the BoxConstraint and RBF_Sigma parameters are used, the hyperplane can separate data on input space properly.

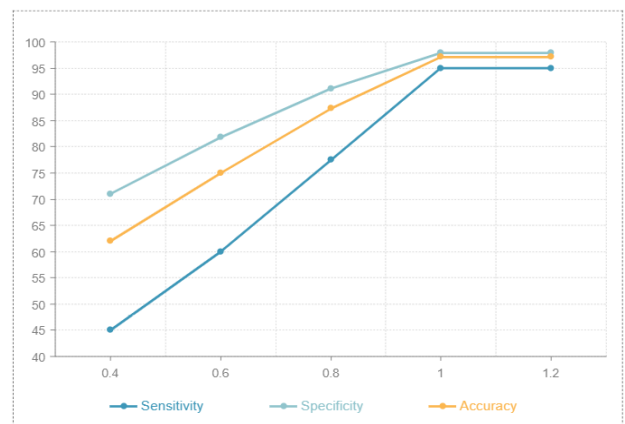


Fig. 10. Influence chart of RBF_Sigma parameter.

The next test, is the ratio of data to find out the

comparison of the accuracy of the used dataset. The dataset will be used both training data and testing data totaling 131 and will be divided according to a predetermined ratio of ratios to see the best accuracy results. The ratio determined, that is the ratio between the number of percentages of training data compared to the percentage number of test data. Tests were also carried out using the RBF kernel and the parameter values for the sequential training used in this test were BoxConstraint = 1 and RBF_Sigma = 1 and iterations = 1000.

TABLE IV: DATA RATIO

Ratio	Sensitivity	Specificity	Accuracy
90%:10%	100%	100%	100%
80%:20%	96.1%	98.6%	98%
70%:30%	95%	97.9%	97.1%
60%:40%	92.4%	97.3%	96%
50%:50%	92.4%	97.3%	96%
40%:60%	88.4%	95.8%	93.8%
30%:70%	74.4%	94.2%	91.5%
20%:80%	68.6%	85.5%	80.2%
10%:90%	48.2%	73.6%	65%

Based on Table IV, it was concluded that the more training data used would be the more precise the recognition made by SVM. Graphical display that illustrates the effect of dataset ratios on test accuracy is shown in Fig. 11.

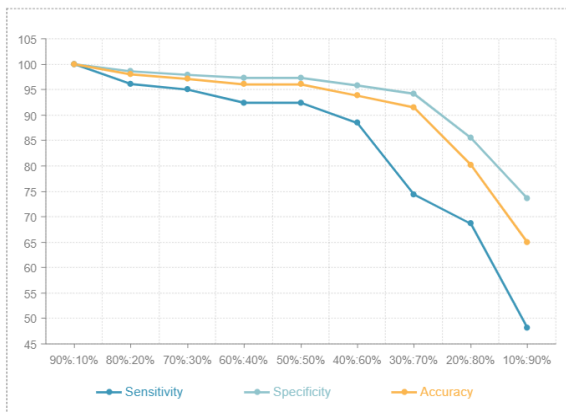


Fig. 11. Influence chart of dataset ratios.

The next experiment was testing the proposed feature extraction method, namely color and texture features. The method is tested separately and combined. The classification results show that maximum accuracy is generated when the two methods are combined.

TABLE V: FEATURE TYPE

Feature Type	Accuracy	
	Train Data	Test Data
Color Features	99.4%	94.7%
Texture Features	86.2%	75%
Color + Texture Features	100%	97.1%

Next, a trial using color feature extraction was carried out using the Color Moments method with the use of different color spaces. The color space of the image to be tested is RGB, HSV and YCbCr color spaces. Test accuracy results from the three color spaces, as follows.

Table VI shows the accuracy obtained, after the feature extraction trials using the Color Moments method in

different color spaces, they are RGB, HSV and YCbCr. The test results show that the YCbCr color space has maximum accuracy in recognition of skin diseases. That is because, the YCbCr color space does not contain the effects of light which can change the characteristics of the skin color, so that many feature information can be obtained.

TABLE VI: COLOR FEATURE TYPE

Color Feature Type	Accuracy	
	Train Data	Test Data
RGB	100%	87.3%
HSV	100%	90.4%
YCbCr	100%	97.1%

BLE VII: TEXTURE FEATURE TYPE

Texture Feature Type	Accuracy	
	Train Data	Test Data
Haar	100%	97.1%
Daubechies	86.2%	75%
Coiflet	99.4%	94.7%

Besides to experiments with color feature, trials with DWT feature extraction were also carried out by trying out the types of families from Wavelet, namely Haar, Daubechies and Coiflet. Based on the tests performed, the type of Haar Wavelet gets better accuracy than Daubechies and Coiflet.

IV. CONCLUSION

The early diagnosis of skin disease system by using computer-based technique, it was proved to be more efficient than the conventional Biopsy methods. In addition, in this proposed methodology, whether the cost was less and the time spent efficiently. This method combined Artificial Intelligence and digital image processing for skin disease to identify skin disease. Support Vector Machine (SVM) had proved to be more succeeded in doing image recognition. In future work, we will multiply the training data to improve testing accuracy and extend our method that is available for the real time recognition.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

I Ketut Gede Darma Putra as the first author is in charge of corrected the errors of the program and concuted the research. Ni Putu Ayu Oka Wiasitini as the second author is in charge of analyzed the data, made skin disease classification programs and testing it. Kadek Suar Wibawa and I Made Suwija Putra as the third and fourth author are in charge of collected the data and wrote the paper.

REFERENCES

- [1] M. S. Manerkar, U. Snehalatha, S. Harsh, J. Saxena, S. P. Sarma, and M. Anburajan, "Automated skin disease segmentation and classification using multi-class SVM classifier," *Inst. Electr. Electron. Eng.*, pp. 1–6, 2016.
- [2] P. K. W. R. S. Gound, P. S. Gadre, and J. B. Gaikwad, "Skin disease diagnosis system using image processing and data mining," *Int. J. Comput. Appl.*, vol. 179, no. 16, pp. 38–40, 2018.

[3] D. S. Mason *et al.*, "The prevalence of scabies and impetigo in the Solomon Islands: A population-based survey," *PLoS Negl. Trop. Dis.*, vol. 10, no. 6, p. e0004803, 2016.

[4] M. A. Sahala, S. Soedarman, L. A. Rizky, A. P. Natanegara, M. S. Advani, and S. Sungkar, "The prevalence of skin diseases and its association with hygiene behavior and level of education in a Pesantren, Jakarta Selatan 2013," *eJournal Kedokt. Indones.*, vol. 2, no. 4, pp. 119–124, 2016.

[5] A. Mohd, G. K. Ram, and A. Shafeeq, "Skin Cancer classification using K-means clustering," *Int. J. Tech. Res. Appl.*, vol. 5, no. 1, pp. 62–65, 2017.

[6] A. Pal, S. Ray, and U. Garain, "Skin disease identification from dermoscopy images using deep convolutional neural network," *arXiv Prepr. arXiv1807.09163*, 2018.

[7] D. U. A. B. Revati Kadu, "Review: Dermatological skin disease detection and classification based on wavelet transform and artificial neural network," *Int. Journals Adv. Res. Comput. Sci. Softw. Eng.*, vol. 8, no. 5, pp. 1–6, 2018.

[8] S. R. Parvin and O. A. M. Jafar, *Prediction of Skin Diseases using Data Mining Techniques*, vol. 6, no. 7, pp. 313–318, 2017.

[9] L. Fernandez-Cuesta *et al.*, "Identification of circulating tumor DNA for the early detection of small-cell lung cancer," *EBioMedicine*, vol. 10, pp. 117–123, 2016.

[10] S. Suhag and L. M. Saini, "Automatic brain tumor detection and classification using SVM classifier," *Int. J. Adv. Sci. Eng. Technol.*, vol. 3, no. 4, pp. 119–123, 2015.

[11] F. Thompson and M. K. Jeyakumar, "Vector based classification of dermoscopic images using SURF," *Int. J. Appl. Eng. Res.*, vol. 12, no. 8, pp. 1758–1764, 2017.

[12] S. S. Dhumal and S. S. Agrawal, *MRI Classification and Segmentation of Cervical Cancer to Find the Area of Tumor*, vol. 3, no. vii, pp. 21–26, 2015.

[13] R. D. Mittal, A. Rani, and Ritambhara, "Detection and classification of focal liver lesions using support vector machine classifiers," *Journal of Biomedical Engineering and Medical Imaging*, 2016.

[14] E. E. E. Ali and W. Z. Feng, "Breast cancer classification using support vector machine and neural network," *Int. J. Sci. Res.*, vol. 5, no. 3, pp. 1–6, 2016.

[15] S. Karamizadeh, S. M. Abdullah, M. Halimi, J. Shayan, and M. J. Rajabi, "Advantage and drawback of support vector machine functionality," in *Proc. 2014 International Conference on Computer, Communications, and Control Technology (I4CT)*, 2014, pp. 63–65.

[16] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions," *Scientific Data*, Harvard Dataverse, 2018.

[17] X. Zheng, Q. Lei, R. Yao, Y. Gong, and Q. Yin, "Image segmentation based on adaptive K-Means algorithm," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, p. 68, 2018.

[18] J. A. Putra, "Mammogram classification scheme using 2D-discrete wavelet and local binary pattern for detection of breast cancer," *Journal of Physics: Conference Serie*, vol. 1008, no. 1, p. 12004 s, 2018.

[19] S. R. Singh and S. Kohli, "Enhanced CBIR using color moments, HSV histogram, color auto correlogram and gabor texture," *Int. J. Comput. Syst.*, vol. 2, no. 5, pp. 161–165, 2015.

[20] I. G. A. Triwayuni, I. K. G. D. Putra, and I. P. A. E. Pratama, "Content based image retrieval using lacunarity and color moments of skin diseases," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 1, pp. 243–248, 2018.

[21] P. Ganesan, V. Rajini, B. S. Sathish, V. Kalist, and S. K. K. Basha, "Satellite image segmentation based on YCbCr color space," *Indian J. Sci. Technol.*, vol. 8, no. 1, p. 35, 2015.

[22] M. Awad and R. Khanna, "Support vector machines for classification," *Efficient Learning Machines*, Springer, pp. 39–66, 2015.

[23] F. Ye, Z. Zhang, K. Chakrabarty, and X. Gu, "Diagnosis using support vector machines (SVM) BT - knowledge-driven board-level functional fault diagnosis," *Cham: Springer International Publishing*, pp. 23–42, 2017.

[24] S. V. G. Reddy, K. T. Reddy, V. V. Kumari, and K. V. Varma, "An SVM based approach to breast cancer classification using RBF and polynomial kernel functions with varying arguments," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5901–5904, 2014.

[25] S. Vachiravel, *Diagnosis of Heart Disease Using Decision Tree and SVM Classifier*, vol. 10, 2015.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



I Ketut Gede Darma Putra was born in Mengwitani on April 24, 1974. He is a lecturer in the Department of Electrical Engineering and Information Technology, Udayana University Bali, Indonesia. He received his S.Kom degree in informatics engineering from the Institute of Sepuluh November Technology Surabaya, Indonesia in 1997. He received his master degree in informatics and computer engineering from Electrical Engineering Department, Gadjah Mada University, Indonesia in 2000 and achieved his doctorate degree in informatics and computer engineering from Electrical Engineering Department, Gadjah Mada University, Indonesia in 2006. His research interests are biometrics, image processing, data mining, and soft computing.



Ni Putu Ayu Oka Wiastini was born in Keramas on June 23, 1997. She is a student at the Department of Information Technology, Udayana University. Her research interests are digital image processing, machine learning and data mining.



developer.

Kadek Suar Wibawa was born in Sangsit on August 16, 1983. He is a lecturer in the Department of Information Technology, Udayana University Bali, Indonesia. He received his bachelor degree from Electrical Engineering Department, Bandung Institute of Technology, Indonesia in 2009. He received his master degree from Informatics Technology Department, Bandung Institute of Technology, Indonesia in 2013. His research interests are internet of things, computer science, computer network and android



I Made Suwija Putra was born in Bali on August 7, 1988. He is one of the lecturers at the Department of Information Technology, Udayana University Bali, Indonesia since 2018. He received his S.T degree in the Department of Electrical Engineering from Udayana University, Indonesia on 2010. He earned a master's degree at Udayana University in 2014 and is currently active in researching data mining, database, and information technology.