khazanah informatika
Jurnal Ilmu Komputer dan Informatika

Medical External Wound Image Classification Using Support Vector Machine Technique

Syifa'ah Setya Mawarni[1*], Murinto[2], Sunardi[3]
[1,2]Informatics Department
Universitas Ahmad Dahlan
Yogyakarta
[3]Electrical Engineering Department
Universitas Ahmad Dahlan
Yogyakarta
*syifa.mawarni19@gmail.com

Abstract-Diagnosis is an activity that refers to the examination of something. Diagnosis is often associated with medical activities as a determinant of a person's condition, in the health sector diagnosis means a procedure performed by a doctor to determine a patient's condition. Unfortunately, it is rare to diagnose disease using an object wound, whereas if the wound is not treated immediately it can lead to more serious illnesses such as ulcers and tetanus or in some cases it can cause infection which then becomes a complication, in the worst case amputation occurs. The skin protects the body from various threats, the skin is also the first fortress for the body. Before implementing a prototype external wound diagnosis, it is necessary to test the accuracy of the algorithm to be used so we know that the algorithm used is suitable for the system to be developed. The algorithm that can be used for diagnosis or classification is the Support Vector Machine or SVM which in the process goes through 3 stages, namely data collection, preprocessing, and classification. This research obtained the results of feature extraction on the wound image test data using GLCM with a contrast value of 0.0082, a correlation value of 0.9769, an energy value of 0.6391, and a homogeneity value of 0.9959 as well as the accuracy of using the SVM algorithm which was measured using a confusion matrix to get an accuracy value of 96.39%, 93.06% precision, recall 92.85%, and F1-score 92.58%. The results of the accuracy of the classification of external wound images using the SVM algorithm are 92.85%.

Keywords: Support Vector Machine, External Wound, Confusion Matrix.

Article info: *submitted May 19, 2023, revised June 19, 2023, accepted July19, 2023*

1. Introduction

Diagnosis is an activity that refers to the examination of something. Diagnosis is often associated with medical activities as a determinant of a person's condition, in the health sector diagnosis means a procedure performed by a doctor to determine a patient's condition. Diagnostic indicators are carried out in several ways by physical examination, laboratory tests, or the like and the use of computer technology in the form of specially designed programs in the assessment process [1].

The use of computer technology in the medical field from severe disease to mild disease has been developed as was done [2] by creating an expert system for diagnosing hypertension with forward chaining inference using the support vector machine (SVM) method to make it easier for medical staff and patients to identify symptoms early diagnosis of hypertension, so that treatment can be done quickly and accurately. The results of this research accuracy were 85% in the training process and 91.89% in the testing process with 247 patient data. Other diseases such as stroke have also been carried out by [3] using the SVM algorithm with influential attributes in predicting stroke using the implementation of the Relief-f Algorithm with the Ratio and Confusion Matrix methods using 3,426 rows and five-column variables. Data accuracy of 100% was obtained using 2398 training data and 1028 test data. Comparison between one algorithm and another is used to find the best results as in research conducted by [4] discussing the comparison of the Xgboost and SVM algorithms for diagnosing breast cancer with the results of the Xgboots algorithm having a better accuracy of 95.12% while for SVM it is 90.24%. Another comparison has also been carried out by [5] using the SVM algorithm by comparing two different accuracy calculations, ADASYN and SMOTE using 630 data with 9 features where 290 data are diabetic patient data and 340 non-diabetic patients. The dataset was obtained from Karya Medika

Laboratory, Indonesia. ADASYN obtained better results with an accuracy of 87.3% and SMOTE 85.4%. Research by [6] compared the performance of 3 feature extractions such as HOG, GLCM, and shape feature extraction for classification using SVM in breast cancer. the final result is feature extraction which can detect the location of the cancer and SVM identifies breast cancer better than Artificial Neural Networks.

Human organs not only consist of the breast and blood, the skin is also part of the organs that are located on the outside of the body. The body is protected by the skin from various threats, the skin is also the first fortress for the body. Diagnosis of disease using wound objects is rarely found, if the wound is not treated immediately it causes more serious diseases such as ulcers and tetanus or in some cases, it can cause infection which then becomes a complication, in the worst cases it causes amputation.

Implementation of an expert system for diagnosing skin diseases due to fungal infections has been carried out by [7] using the Naïve Bayes algorithm with an accuracy of 87.1% using 31 data, while research on the diagnosis of external wounds has never been discussed. Testing the level of accuracy of the algorithm is required before implementing an external wound diagnosis system. One of the algorithms that can be used for diagnosis or classification is the Support Vector Machine or SVM. SVM works by defining the boundary between two classes with the maximum distance from the closest data. The maximum distance is obtained by finding the best hyperplane (separation line) in the input space which is obtained by measuring the margins of the hyperplane [8], this makes SVM an uncomplicated algorithm [9].

Classification using the SVM algorithm is not only used for medical images, as in research [10] using plastic bottle images with the highest accuracy results of 84% with a total of 70% training data and 30% test data, while the Multiclass SVM classification on textile images produces accuracy 78% [11]. SVM is not only used for image classification, SVM can also be used to determine whether a programmer's performance is good or bad based on social media postings with an accuracy of 85.1% [12].

Research [8] compared the SVM and CNN methods for classifying flowers and obtained better CNN results with an accuracy of 91.6% while SVM was 78.3%. Algorithmic comparison is also discussed [13] by comparing KNN, SVM, Decision Tree, and AdaBoost in the classification of facial images based on ethnicity in Indonesia with the accuracy of KNN 31.95%, SVM 38.05%, Decision Trees 29.4%, and AdaBoost 31.55%. Contrary to [8] and [13], [14] optimized the SVM algorithm using the Grid Search method on Robusta coffee beans with the highest accuracy of 94% on the RBF and Polynomial kernels.

2. Methods

The flowchart in this research includes image dataset collection, preprocessing, and classification. Details of the flowchart are shown in Figure 1.

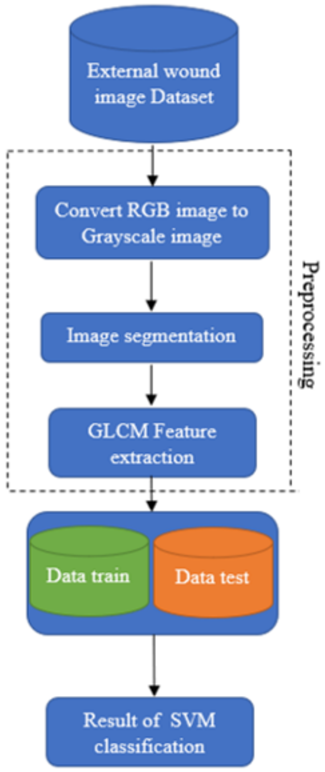


Figure 1. SVM classification flowchart

a. Image Dataset

External wound image dataset was obtained from the Regional General Hospital (RSUD) KRT Setjonegoro Wonosobo, Central Java and using a dataset from Kaggle (www.kaggle.com). Figure 2 shows the original image which is 640x640 pixels in size. The image dataset consists of 4 classes, which are punctures, lacerations, abrasions and burns with a total of 280 data and each class of wounds has 70 image data.

a) Laseration Wound b) Burn Wound

c) Punctum Wound d) Abration Wound

Figure 2. Wound Dataset

The image dataset is then divided into 70% train data and 30% test data. Table 1 shows the results of dividing the wound image dataset.

Table 1. Dataset Distribution

Category	Total Datasets	Train Data	Test Data
Punctum	70	49	21
Laseration	70	49	21
Abrastion	70	49	21
Burn	70	49	21

The train data and test data get the same preprocessing stage, namely converting color images (RGB) into gray images which then the images will be segmented using the Active Contour algorithm and get feature extraction using the Gray Level Co-occurrence Matrix (GLCM) technique with the final result is accuracy from the classification of external wound image data using Support Vector Machine.

b. Feature Extraction

The features used in this research are extracted from the results of image segmentation. This feature is normalized within a normal range [0.1] and then used as SVM classifier input data. These features are average, variance, standard deviation, and skewness. Feature extraction is used to convert image data from matrix to vector.

Feature extraction used in this research is feature extraction GLCM or Gray Level Co-occurrence Matrix where GLCM is an image texture analysis technique that represents the relationship between two adjacent pixels with intensity, distance, and grayscale angle [15]. GLCM has eight angles 0º, 45º, 90º, 135º, 180º, 225º, 270º, 315º. GLCM has a distance parameter calculation based on the number of pixels between reference pixels and neighboring pixels [16]. Four GLCM features use is homogeneity, contrast, correlation, and energy.

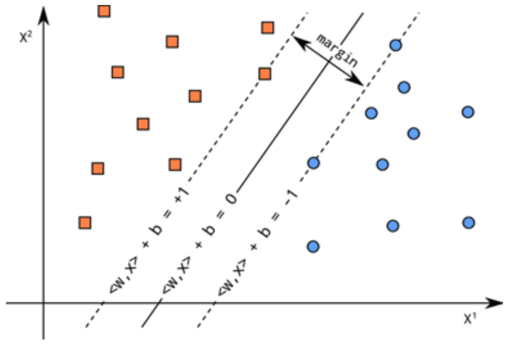
Entropy gives information about the random size of the image. Homogeneity is the value of measuring the distance of GLCM distribution elements with values 0 and 1. Value 1 means GLCM diagonal homogeneity and value 0 means GLCM vertical homogeneity. The contrast will adjust the high and low colors of the image. All of these values provide information as analytical material in determining the type of tumor or cancer [17].

c. Classification Using Support Vector Machine

Datasets consisting of two classes that can be separated linearly with a straight line are called linear data, while data that cannot be separated linearly is called non-linear data. Non-linear data are usually separated using irregular curves. SVM algorithm is difficult to separate non-linear data so it requires a trick kernel to map non-linear data with low dimensions to higher dimensions so that it can be separated into linear data. There are 3 kinds of kernel tricks.

1. Linear Kernel
   Linear kernel is used to separate linear data or data that only consists of two classes with a straight line as shown in Figure 3. linear kernel formula is in equation 1.

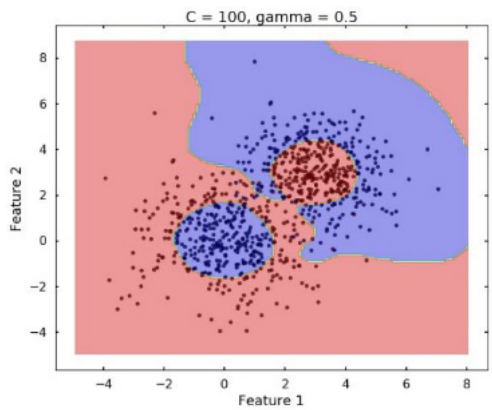
$$K(x,xi) = sum(x*xi) \quad (1)$$



Linear Kernel

2. RBF Kernel
   The RBF kernel is used for non-linear data which is separated using an irregular curve as shown in Figure 4. RBF kernel training results have fewer errors than other kernels using the gamma parameter. Equation 2 shows the RBF kernel formula.

$$K(x,xi) = exp(-gamma*sum(x-xi^2)) \quad (2)$$



RBF Kernel

3. Polynomial Kernel
   Polynomial kernels are also used for non-linear data where the degree in the polynomial kernel controls the flexibility of the classification results, more higher degree value in the polynomial kernel allows for more flexible decision limits [18]. The polynomial kernel formula is in equation 3.

$$K(x,xi) = exp(-gamma*sum(x-xi^2)) \quad (3)$$

SVM includes supervised learning algorithms where machine learning is trained to learn labeled data. Classification divides the data into train data and test data.

Train data is labeled prediction, test data is labeled target. SVM classification uses test data.

d. Model Evaluation Classification

Calculation of accuracy in classification is the most important final stage, this stage determines whether or not the use of the algorithm is suitable for the problem at hand. This stage will show how suitable the use of the SVM algorithm is for the case of external wound image classification where the measurement of accuracy uses the confusion matrix method as shown in Table 2. The confusion matrix has four terms that represent the results of the classification process including:

1. True Positive (TP), is true data and predicted correctly.
2. True Negative (TN), is false data and predicted wrong.
3. False Positive (FP), is false data but predicted to be correct.
4. False Negative (FN), is true data but predicted to be wrong.

Table 2. Confusion Matrix

		Prediction	
		True	False
True data	Positive	TP	FP
	Negative	TN	FN

Confusion matrix using accuracy, precision, recall, and F1 score [8]. Accuracy is used to see how much the percentage results of the dataset are correctly classified by the system [19]. Equation 4 calculates accuracy.

$$Accuracy = \frac{TP+TN}{Total\ Data}\ x\ 100\% \quad (4)$$

The degree of suitability of the information provided by the system and the information required by the user is called precision, while the success rate of the system in finding information is called recall [20]. Equation 5 and equation 6 are used to calculate precision and recall.

$$precision = \frac{TP}{TP+FP}\ x\ 100 \quad (5)$$

$$recall = \frac{TP}{TP+FN}\ x\ 100\% \quad (6)$$

F1 Score, is the average formula of the correct recall value and precision value [21]. Equation 7 calculates the F1 score.

$$F1 = 2\,\frac{precision*recall}{precision+recall} \quad (7)$$

3. Result and Discussion

Classification of data through three stages, namely data collection, preprocessing, and classification.

External wound image dataset was obtained from the Regional General Hospital (RSUD) KRT Setjonegoro Wonosobo, Central Java and using a dataset from Kaggle (www.kaggle.com). The original image in this dataset has 640x640 pixels in size. The image dataset consists of 4 classes, which are punctures, lacerations, abrasions and burns with a total of 280 data and each class of wounds has 70 image data. Train data and test data are divided into 70% and 30% with a total of 49 train images and 21 test images for each class.

Preprocessing is the stage of processing the dataset that has been obtained. The preprocessing stage includes the process of changing the color image to a gray image, image segmentation using the Active Contour algorithm and feature extraction using the segmented image whose features are extracted using GLCM with average features. GLCM uses 4 features, namely homogeneity, contrast, correlation, and energy where the results of each feature can be seen in Table 3.

After obtaining the features of each image data, then dividing the train data and test data. SVM is an algorithm that only applies to 2 classes in training, because this research uses 4 classes, the training data used is cross-grouped while the data tested are all classes grouped together.

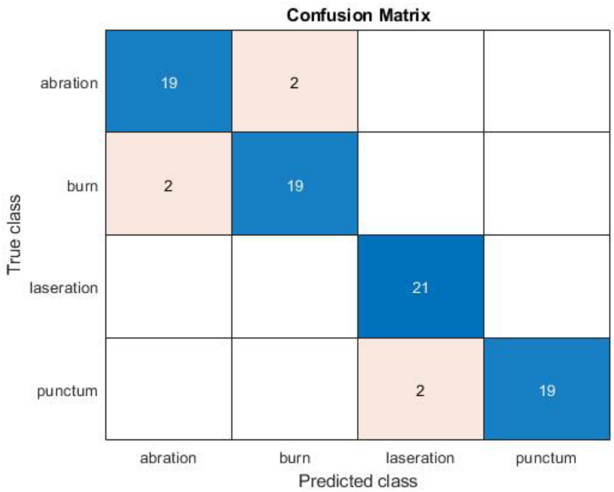
Measurement of the accuracy of the SVM classification using the confusion matrix method is in Figure 5 with the results of accuracy, precision, recall, and F1 scores in Table 4 and the accuracy reaches 92.85%.

Table 3. GLCM Feature Extraction Results

	Train Data				Test Data
	Burn	Laseration	Abration	Punctum	
Contrast	0.0035	0.0061	0.0029	8.0380	0.0082
Corre lation	0.9825	0.9742	0.9879	0.9101	0.9769
Energy	0.7953	0.7573	0.7552	0.9903	0.6391
Homogenity	0.9982	0.9970	0.9985	0.9996	0.9959

Table 4. SVM Classification Result

	Accuracy	Precision	Recall	F1-score
Burn	95.23%	90.47%	90.47%	90%
Laseration	97.47%	91.30%	100%	95.28%
Abration	95.23%	90.47%	90.47%	90%
Punctum	97.61%	100%	90.47%	95.03%



Confusion Matrix SVM

4. Conclusion

Based on the research, after testing the segmentation of the external wound image using the SVM algorithm to determine the suitability of the algorithm with the image, the results of feature extraction on the wound image test data using GLCM obtained a contrast value of 0.0082, a correlation value of 0.9769, an energy value of 0.6391, and a homogeneity value of 0.9959. The accuracy of the SVM algorithm is measured using the confusion matrix to get an accuracy of 96.39%, precision of 93.06%, recall of 92.85%, and F1-score of 92.58%. The accuracy of the SVM algorithm is measured using the confusion matrix to get an accuracy of 96.39%, precision of 93.06%, recall of 92.85%, and F1-score of 92.58%.

The results of the classification accuracy of external wound image segmentation using the SVM algorithm are 92.85%. Based on the result of classification accuracy, SVM algorithm is considered capable of classifying the result of external wound image segmentation.

References

[1] D. Mauli, "Tanggung Jawab Hukum Dokter Terhadap Kesalahan Diagnosis Penyakit Kepada Pasien," *Cepalo*, vol. 2, no. 1, pp. 33, Sep 2019, doi: 10.25041/cepalo.v2no1.1760.

[2] Supriyono, Erida, and Fadila, "Sistem Pakar Penegakan Diagnosa Penyakit Hipertensi Dengan Inferensi Forward Chaining Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Kesehatan Medika Udayana,* vol. 8, no. 2, pp. 207–221, 2022.

[3] U. Amelia, J. Indra, and A. F. N. Masruriyah, "Implementasi Algoritma Support Vector Machine (SVM) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh" *Scientific Student Journal for Information, Technology and Science,* vol. 3, no. 2, pp. 254–259, 2022.

[4] M. R. Andryan, M. Fajri, & N. Sulistyowati, "Komparasi Kinerja Algoritma Xgboost dan Algoritma Support Vector Machine (SVM) untuk Diagnosis Penyakit Kanker Payudara," *JIKO (Jurnal Informatika dan Komputer),* vol. 6, no. 1, pp. 1-5, 2022.

[5] N. G. Ramadhan, "Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus," *Scientific Journal of Informatics,* vol. 8, no. 2, pp. 276–282, Nov 2021, doi: 10.15294/sji.v8i2.32484.

[6] H. Mu'jizah and D. C. R. Novitasari, "Comparison of the Histogram of Oriented Gradient, GLCM, and Shape Feature Extraction Methods for Breast Cancer Classification Using SVM," *Jurnal Teknologi dan Sistem Komputer,* vol. 9, no. 3, pp. 150–156, Jul 2021, doi: 10.14710/jtsiskom.2021.14104.

[7] F. Rahmadsyah and M. Akbar, "Sistem Pakar Diagnosa Penyakit Kulit Akibat Infeksi Jamur Menggunakan Teorema Bayes," *Jurnal Komputer dan Informatika,* vol. 9, no. 1, pp. 38–44, Mar 2021, doi: 10.35508/jicon.v9i1.3170.

[8] A. Peryanto, A. Yudhana, and R. Umar,

"Convolutional neural network and support vector machine in classification of flower images," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika,* vol. 8, no. 1, pp. 1-7, 2022.

[9] M. Yang and G. Thung, "Classification of trash for recyclability status," *CS229 project report 2016.1*, pp. 3, 2016.

[10] A. Setyo Nugroho, R. Umar, and F. Abdul, "Klasifikasi Botol Plastik Menggunakan Multiclass Support Vector Machine," *JURNAL KHATULISTIWA INFORMATIKA,* vol. 9, no. 2, pp. 79–85, 2021.

[11] F. Arnia Ramadhani, R. Muharar, "Klasifikasi Otomatis Motif Tekstil Menggunakan Support Vector Machine Multi Kelas," *Jurnal Teknologi Informasi dan Ilmu Komputer,* vol. 7, no. 1, pp. 99-108, 2020, doi: 10.25126/jtiik.202071428.

[12] R. Umar, I. Riadi, and P. Purwono, "Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial dengan Metode Support Vector Machines," *Cybernetics*, vol. 4, no. 1, pp. 32-40, 2020.

[13] S. Hartono, H. Sujaini, and A. Perwitasari, "Komparasi Algoritma Nonparametrik untuk Klasifikasi Citra Wajah Berdasarkan Suku di Indonesia," *Jurnal Edukasi dan Penelitian Informatika,* vol. 6, no. 3, pp. 337–343, Des 2020.

[14] H. Apriani, J. H. Jaman, and R. I. Adam, "Optimasi SVM menggunakan algoritme grid search untuk identifikasi citra biji kopi robusta berdasarkan circularity dan eccentricity," *Jurnal Teknologi dan Sistem Komputer,* vol. 10, no. 1, pp. 12-19, 2022.

[15] V. Punithavathi and D. Devakumari, "Detection of breast lesion using improved GLCM feature based extraction in mammogram images," published by *SSRN eLibrary*, 2019.. [Online]. Available in: https://ssrn.com/abstract=3663875

[16] A. H. Farhan and M. Y. Kamil, "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques," in *IOP Conference Series: Materials Science and Engineering,* vol. 928, no. 7, p. 072098, November 2020, IOP Publishing. doi: 10.1088/1757-899X/928/7/072098.

[17] A. R. W. Putri, "Classification of Breast Cancer Using the Digital Mammogram Method," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi),* vol. 9, no. 4, pp. 2752-2761, 2022.

[18] A. Ben-Hur and J. Weston, "A User's Guide to Support Vector Machines," in O. Carugo and F. Eisenhaber (Eds.), *Data Mining Techniques for the Life Sciences,* vol. 609, pp. 223-239, Humana Press, 2010. doi: 10.1007/978-1-60327-241-4_13.

[19] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi untuk diagnosa diabetes menggunakan metode bayesian regularization neural network (RBNN)," *Jurnal Informatika,* vol. 11, no. 1, p. 36, 2017.

[20] I. Saputra and D. Rosiyadi, "Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp," *Faktor Exacta,* vol. 12, no. 2, pp. 101, Jul 2019, doi: 10.30998/faktorexacta.v12i2.4181.

[21] G. Hackeling, *Mastering Machine Learning with scikit-learn.* Birmingham: Packt Publishing Ltd., 2014.