panunen_fix_080225.docx

by Turnitin Student

Submission date: 08-Feb-2025 10:40AM (UTC-0500)
Submission ID: 2582893155
File name: panunen_fix_080225.docx (236.05K)
Word count: 3139
Character count: 16758

Journal of Advanced Health Informatics Research (JAHIR)
Vol. 1, No. 3, November 2023, pp. xx-xx
ISSN: 2985-6124, DOI: 10.26555/jahir.v8i2.xxxx

Classification Of Skin Disease Images Using K-Nearest Neighbour (KNN)

Ari Peryanto [1], Dwi Susanto [2], Bagus Hayatul Jihad [3]
[1,2] *Department of Informatics, Universitas Madani, Yogyakarta, Indonesia*
[3] *Curriculum and Instruction, International Islamic University Malaysia, Malaysia*

ARTICLE INFO

Article history:

Received July 28, 2021
Revised August 28, 2021
Published September 28, 2021

Keywords:

Skin Disease;
KNN;
Classification;
Confusion Matrix;
Image Processing

ABSTRACT

The skin is the outermost part of the human body that is often exposed to the environment, so it is easy to experience disease disorders. Some of the skin diseases that are often contracted in humans are ulcers, herpes, and warts. Untreated skin diseases will be very annoying because of the sensation of itching so it can cause irritation and inflammation. The ability to classify skin diseases using technology is one solution. This study uses the K-Nearest Neighbour (KNN) method to detect images of skin diseases. KNN is one of the machine learning methods with a calculation method based on the proximity of k. KNN was chosen because it is fast and has high-accuracy results. The results of the research that has been carried out have obtained results of accuracy of 63%, precision of 63%, recall of 63%, and F1 Score of 63%. From the results of the study, it can be concluded that disease detection using KNN has been successfully applied and can be used in classification.

This work is licensed under a Creative Commons Attribution-Share Alike 4.0

Corresponding Author:

Ari Peryanto, Department of Informatics, Universitas Madani, Yogyakarta, Indonesia
Email: ari@umad.ac.id

1. INTRODUCTION

The skin is the largest organ in the human body that functions as the main protector against various external factors [1], [2], [3] such as infections, temperature, and chemicals. In addition, the skin also has an important role in regulating body temperature and as a sensory tool [4] A healthy skin condition is essential to support these functions. However, the skin is also susceptible to a variety of diseases, including ulcers, herpes, and warts,[5] which can interfere with skin function and aesthetics.

Ulcers are skin infections caused by bacteria, usually, Staphylococcus aureus [6] which are characterized by red, painful, and pus-filled [7] Herpes is a skin disease caused by the herpes simplex virus (HSV) and is often characterized by painful blisters around the mouth or genital area [8] Meanwhile, warts are caused by human papillomavirus (HPV) infection [9] which results in rough and irregular skin growth. These three diseases not only have an impact on physical health but can also cause psychological disorders in sufferers.

In Islam, skin diseases such as ulcers, herpes, and warts are a test from Allah as well as a form of Allah's compassion. Islam teaches humans to be patient in the face of illness and keep striving to find healing. Disease is a sin that we face with patience and trust, but in addition to patience and trust, Muslims are also obliged to seek treatment because every disease must have a cure. Always remember that Allah is *Ash-Shafi* (the Healer), and every disease has a cure. Muslims are also required to maintain cleanliness by performing ablution and

Journal of Advanced Health Informatics Research (JAHIR)
Vol. 1, No. 3, November 2023, pp. xx-xx
ISSN: 2985-6124

bathing correctly to prevent skin diseases. Muslims infected by herpes are also urged to keep their distance so as not to transmit the disease.

Identifying skin diseases is often a challenge, especially in areas with limited access to medical personnel or health facilities. Errors in diagnosis can result in improper handling, worsen the patient's condition, and increase the risk of transmission. On the other hand, manual diagnosis by dermatologists requires a lot of time and effort, especially with the high number of skin disease cases that must be treated [10]

To solve this problem, artificial intelligence technology offers an effective solution. One of the methods that can be used is the K-nearest neighbor (KNN) algorithm. KNN is a simple but effective machine learning algorithm for classifying based on feature similarities between new data and existing data [11] Using KNN, image data or features of infected skin can be analyzed to quickly and accurately determine the type of disease.

KNN is an algorithm that is easy to implement and does not require special assumptions about data distribution. This algorithm can be used to distinguish several types of skin diseases such as ulcers, herpes, and warts. KNN is suitable for small to medium datasets, which is often a challenge in medical image-based research [12] KNN can work well on data that has high feature dimensions, such as skin images with various visual characteristics.

By developing a KNN-based identification system, it is hoped that the skin disease diagnosis process can be carried out faster, more accurately, and more affordable. This will help improve the quality of health services, especially in areas with minimal medical resources.

2. METHODS

Research on the application of the KNN classification method has been carried out before, for example, as carried out by [13] namely discussing X-Ray Image Pneumonia Disease in the Lungs. Research on skin classification with KNN has also been conducted by [14][15] using the texture feature, in addition [16] conducting performance research from various machine learning algorithms such as KNN with SVM and also deep learning methods using CNN. KNN involves several stages, including understanding the problems and objectives of the research, collecting data from relevant sources, and then cleaning and processing the data, including the selection of important attributes. After that, the data is divided into two parts, namely the training and testing sets, to test the model. Then, model evaluation is carried out using metrics such as accuracy, precision, and recall in the classification process. The results of the evaluation are used to identify new patterns and information that can be obtained [17] The following research methods and stages can be seen in Figure 1.

Fig. 1. Research Methods and Stages

2.1. Data Collection

This section refers to the procedures implemented to collect data relevant to the research being conducted. This data collection method is a stand-alone way of conducting data analysis, or it can be the main tool in the data analysis process. To obtain data relevant to skin diseases, the authors collected information related to this study through the Literature Study method. Images of skin diseases, ulcers, herpes, and skin were obtained from random searches through the internet among Google images, Bing images, and other image finders. The datasets collected in this study totaled 690, with details of ulcers as many as 230, herpes as many as 230, and warts as much as 230 data, then the data was divided into training data and test data with a division of 70% training data and 30% test data so that it is easy to understand can be written in table 1.

Table 1. Skin Disease Dataset

Skin Disease	Sum		
	Dataset	Train	Test
Boil	230	161	69
Herpes	230	161	69
Warts	230	161	69

2.2. Preprocessing Data

After the data is collected, the next step is to process the initial data first, to ensure that the data obtained is of good quality, so as to produce optimal results. The clarity of the dataset that will be used as training data greatly affects the accuracy of the results. Therefore, this initial data processing process is a very important and crucial step [18], [19], [20]. The stages in the initial data processing can be explained through the flow of the data processing process, as seen in Figure 2.

Fig. 2. Dataset Processing Flow

First, the image will be cut in order to obtain uniformity in size and obtain the details of the image to be used in testing and training. The image is resized using a script created using Python to obtain a size uniformity of 200x200. Then to clarify the image, it is edited manually using an image editing program such as Adobe Photoshop to remove noise that results in the image becoming out of focus.

2.3. K-Nearest Neighbors (KNN)

The algorithm used is the K-Nearest Neighbors (K-NN) method, which is included in the supervised learning method and is also classified as an instance-based learning group. This method is included in the lazy learning technique and is often used to predict or classify data. The working principle of the K-NN method is to determine the K object that is closest to the training data to the test object or data. The advantages of the K-NN method are its high accuracy, does not rely on certain assumptions about the data, as well as its sensitivity to outliers. The first step in the application of this method is to determine the value of K first [17].

2.4. Classification

Classification is a technique for grouping certain data into predefined categories, based on patterns or attributes that exist in that data [21]. The classification method is increasingly popular because of its ability to handle a wider range of data when compared to regression. The main purpose of classification is to predict the class or label of unknown data, based on the information present in previously known data [22].

Journal of Advanced Health Informatics Research (JAHIR)
Vol. 1, No. 3, November 2023, pp. xx-xx
ISSN: 2985-6124

2.5. Validation Test Results

A confusion matrix is a table used to assess the performance of classification models in machine learning and statistics. This table compares the model's prediction results with the actual values to determine the level of accuracy of the classification. From the confusion matrix, some evaluation metrics such as accuracy can be calculated using equations (1). Accuracy is used to see the percentage of data records that are classified correctly by the system [23][24].

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

To measure the extent to which the model is right in classifying positive classes can also be called Precision [24]and can be calculated using equations (2)

$$\frac{TP}{TP + FP} \quad (2)$$

The model's ability to detect all truly positive cases also known as Recall [24] is demonstrated through equations (3).

$$\frac{TP}{TP + FN} \quad (3)$$

The balance between precision and recall, especially when there is a class imbalance also known as the F1 Score [24] is shown in the equation (4).

$$2\,x\,\frac{Precission * Recall}{Precission + Recall} \quad (4)$$

2.6. Skin Diseases

Skin diseases are disorders of the outer layers of the body that are characterized by symptoms such as itching and redness. This condition can be caused by a variety of factors, including exposure to chemicals, sunlight, viral infections, weakened immune systems, microorganisms, mold, as well as a lack of attention to personal hygiene [25]. *Pityriasis versicolor* is a skin disease that is often encountered in Indonesian society. Globally, this condition is almost always present, especially in regions with tropical climates, with a prevalence of up to 50%, especially in areas with high levels of humidity and rainfall.

3. RESULTS AND DISCUSSION

After the dataset is preprocessed data, the next step is to start the classification process using KNN. The results of image classification using the KNN algorithm showed that the model was able to identify objects with an accuracy of 63%. Based on the classification report obtained, the model managed to correctly classify 63% of the positive class (recall), while the precision reached 63%, indicating that most of the positive predictions were correct. With an F1 score of 63%, the model shows a good balance between precision and recall in image classification tasks. The classification report of the research that has been carried out is shown in Figure 3.

	precision	recall	f1-score	support
bisul	0.62	0.58	0.60	69
herpes	0.67	0.65	0.66	69
kutil	0.60	0.65	0.62	69
accuracy			0.63	207
macro avg	0.63	0.63	0.63	207
weighted avg	0.63	0.63	0.63	207

Fig. 3. Classification Report

Figure 4 is the result of the confusion matrix obtained from the results of the study. In the matrix, each row represents a predicted category, while each column represents the actual category (or vice versa) [26]. If the highest value in the confusion matrix is diagonally from the top left to the bottom right, then the model used can be said to have good performance [27]. The explanation of the confusion matrix results in this study is that for the ulcer category, the correct results were obtained of 40, 12 were selected for herpes and 17 were selected for warts, then for the herpes category the correct results were obtained of 45, 11 were selected for ulcers and 13 were selected for warts, while for the wart category, the correct results were 45, 14 for ulcers and 10 for herpes.

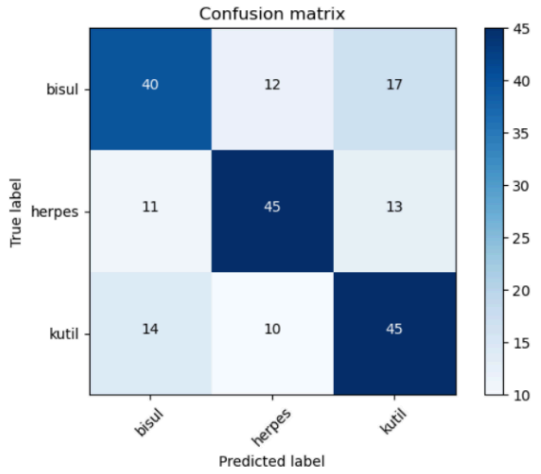

Fig. 4. Confusion Matrix

4. CONCLUSION

In this study, the K-Nearest Neighbors (KNN) algorithm has been applied to classify the image of ulcers, herpes, and warts. Based on the results of the evaluation using the confusion matrix, the KNN model managed to achieve an accuracy level of 63%, showing its ability to recognize patterns in skin disease images quite well, although there are still some classification errors between categories.

One of the main advantages of KNN is its ease of implementation and its ability to adapt to datasets without the need for a complex training process. However, this model has some limitations, such as being less

than optimal in handling large datasets and susceptible to noise, which can affect the accuracy of classification results.

To improve the performance of the model in future studies, it is recommended to explore various optimization techniques, such as determining a more optimal k-value, applying better feature extraction methods, or combining KNN with other algorithms to improve accuracy and efficiency in the classification process.

Overall, the KNN model developed in this study has the potential to be the basis for the development of an image-based automatic detection system for skin diseases. This system can provide benefits for medical personnel in helping the initial diagnosis process faster and more efficiently.

REFERENCES

[1] M. Ahn *et al*., "Protective Barriers Provided by the Epidermis," *Biomater. Res.*, vol. 27, no. 1, pp. 1–13, 2023, doi: 10.3390/ijms24043145.

[2] M. Ahn *et al*., "3D biofabrication of diseased human skin models in vitro," *Biomater. Res.*, vol. 27, no. 1, pp. 1–13, 2023, doi: 10.1186/s40824-023-00415-5.

[3] A. T. Slominski, R. M. Slominski, C. Raman, J. Y. Chen, M. Athar, and C. Elmets, "Neuroendocrine signaling in the skin with a special focus on the epidermal neuropeptides," *Am. J. Physiol. - Cell Physiol.*, vol. 323, no. 6, pp. C1757–C1776, 2022, doi: 10.1152/AJPCELL.00147.2022.

[4] A. Srivastav, Y. Srivastav, A. Hameed, and M. I. Ahmad, "Prevention and cure of dermatology disorders using herbal medications: summary," *Int. J. Indig. Herbs Drugs*, vol. 9, no. 1, pp. 1–14, 2024, doi: 10.46956/ijihd.v9i1.568.

[5] K. A. and S. A., "A Herbal Approach towards Skin Diseases: An Updated Review," *Int. J. Newgen Res. Pharm. Healthc.*, vol. 1, no. 1, pp. 93–100, Jun. 2023, doi: 10.61554/ijnrph.v1i1.2023.21.

[6] G. Y. C. Cheung, J. S. Bae, and M. Otto, "Pathogenicity and virulence of Staphylococcus aureus," *Virulence*, vol. 12, no. 1, pp. 547–569, 2021, doi: 10.1080/21505594.2021.1878688.

[7] P. Del Giudice, "Skin infections caused by staphylococcus aureus," *Acta Derm. Venereol.*, vol. 100, no. 100-year theme Cutaneous and genital infections, pp. 208–215, 2020, doi: 10.2340/00015555-3466.

[8] K. Madavaraju, R. Koganti, I. Volety, T. Yadavalli, and D. Shukla, "Herpes Simplex Virus Cell Entry Mechanisms: An Update," *Front. Cell. Infect. Microbiol.*, vol. 10, no. January, pp. 1–18, 2021, doi: 10.3389/fcimb.2020.617578.

[9] S. García-Oreja, F. J. Álvaro-Afonso, D. Sevillano-Fernández, A. Tardáguila-García, M. López-Moral, and J. L. Lázaro-Martínez, "A non-invasive method for diagnosing plantar warts caused by human papillomavirus (HPV)," *J. Med. Virol.*, vol. 94, no. 6, pp. 2897–2901, 2022, doi: 10.1002/jmv.27514.

[10] C. Flohr and R. Hay, "Putting the burden of skin diseases on the global map," *Br. J. Dermatol.*, vol. 184, no. 2, pp. 189–190, 2021, doi: 10.1111/bjd.19704.

[11] O. K. Pal, "Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm," *Proc. Int. Conf. Electron. Commun. Inf. Technol. ICECIT 2021*, no. April, 2021, doi: 10.1109/ICECIT54077.2021.9641120.

[12] W. Chaimae, L. S. Mohammed, K. Maha, K. Mohamed, and B. Said, *KNN Classification of Kolb Learning Styles: A Comparative Study on Balanced and Unbalanced Datasets*, no. Ml. Atlantis Press International BV, 2023. doi: 10.2991/978-94-6463-360-3_6.

[13] F. Antony, H. Irsyad, and M. E. Al Rivan, "KNN Dan Gabor Filter Serta Wiener Filter Untuk Mendiagnosis Penyakit Pneumonia Citra X-RAY Pada Paru-Paru," *J. Algoritm.*, vol. 1, no. 2, pp. 147–155, 2021, doi: 10.35957/algoritme.v1i2.893.

[14] M. A. Araaf, K. Nugroho, and D. R. I. M. Setiadi, "Comprehensive Analysis and Classification of Skin Diseases based on Image Texture Features using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 31–40, 2023, doi: 10.33633/jcta.v1i1.9185.

[15] S. Jagtap, M. Karnavat, R. Khairnar, R. Kolge, and J. R. Panchal, "Skin Disease Analysis and Prediction using Machine Learning," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 2, pp. 58–62, 2024, doi: 10.48175/IJARSCT-22113.

[16] K. Vayadande, A. A. Bhosle, R. G. Pawar, D. J. Joshi, P. A. Bailke, and O. Lohade, "Innovative approaches for skin disease identification in machine learning: A comprehensive study," *Oral Oncol. Reports*, vol. 10, no. April, p. 100365, 2024, doi: 10.1016/j.oor.2024.100365.

[17] A. Yogianto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk

Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.

[18] A. Barry, L. Han, and G. Demartini, "On the Impact of Data Quality on Image Classification Fairness," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, in SIGIR '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 2225–2229. doi: 10.1145/3539618.3592031.

[19] L. Budach *et al.*, "The Effects of Data Quality on Machine Learning Performance," vol. 1, no. 1, 2022, doi: 10.48550/arXiv.2207.14529.

[20] D. Berrabah and Y. Gafour, "Improve Image Classification Using Data Optimization," *Eurasia Proc. Sci. Technol. Eng. Math.*, vol. 26, pp. 262–271, 2024, doi: 10.55549/epstem.1409569.

[21] A. Vajpayee, P. P. Kaur, A. Sharma, and A. Tiwari, "An Extensive Analysis of Pattern Identification and Categorization," in *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 2024, pp. 1–4. doi: 10.1109/ACROSET62108.2024.10743229.

[22] A. Alshaibanee, "Predicting Class Label Using Clustering-Classification Technique: A Comparative Study," *J. Kufa Math. Comput.*, vol. 10, no. 1, pp. 1–12, 2023, doi: 10.31642/jokmc/2018/100101.

[23] Jude Chukwura Obi, "A comparative study of several classification metrics and their performances on data," *World J. Adv. Eng. Technol. Sci.*, vol. 8, no. 1, pp. 308–314, 2023, doi: 10.30574/wjaets.2023.8.1.0054.

[24] A. M. El-Assy, H. M. Amer, H. M. Ibrahim, and M. A. Mohamed, "A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data," *Sci. Rep.*, vol. 14, no. 1, pp. 1–19, 2024, doi: 10.1038/s41598-024-53733-6.

[25] A. Sajwan, A. Belwal, N. Sharma, and Y. Joshi, "A Concise Review on Skin Disorders," *Saudi J. Med. Pharm. Sci.*, vol. 8, no. 5, pp. 252–256, 2022, doi: 10.36348/sjmps.2022.v08i05.007.

[26] M. Paluszek and S. Thomas, "Duplex-Hierarchy Representation Learning for Remote Sensing Image Classification," *Sensors*, vol. 24, no. 4, pp. 1–14, 2024, doi: 10.1007/978-1-4842-5124-9_11.

[27] M. M. Islam, M. A. Talukder, M. A. Uddin, A. Akhter, and M. Khalid, "BrainNet: Precision Brain Tumor Classification with Optimized EfficientNet Architecture," *Int. J. Intell. Syst.*, vol. 2024, no. 1, p. 3583612, 2024, doi: https://doi.org/10.1155/2024/3583612.

panunen_fix_080225.docx

ORIGINALITY REPORT

18%	14%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ejournal.ptti.web.id Internet Source	6%
2	apnic.foundation Internet Source	1%
3	Kuldeep Vayadande, Amol A. Bhosle, Rajendra G. Pawar, Deepali J. Joshi, Preeti A. Bailke, Om Lohade. "Innovative approaches for skin disease identification in machine learning: A comprehensive study", Oral Oncology Reports, 2024 Publication	1%
4	publikasi.dinus.ac.id Internet Source	1%
5	Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, Indra Kishor. "Recent Advances in Sciences, Engineering, Information Technology & Management - Proceedings of the 6th International Conference “Convergence2024” Recent Advances in Sciences, Engineering, Information Technology & Management, April 24–25, 2024, Jaipur, India", CRC Press, 2025 Publication	1%
6	eprints.unmer.ac.id Internet Source	1%
7	Ning Li, Mingliang Wang, Gaochao Yang, Bo Li, Baohua Yuan, Shoukun Xu. "DENS-YOLOv6: a small object detection model for garbage	1%

	detection on water surface", Multimedia Tools and Applications, 2023 Publication	
8	Ima Kurniastuti, Teguh Herlambang, Tri Deviasari Wulan, Dike Bayu Magfira et al. "Improved Edge Detection using Morphological Operation to Segmentation of Fingernail Images", Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2025 Publication	1 %
9	www.researchgate.net Internet Source	1 %
10	dc.swosu.edu Internet Source	1 %
11	pure.tue.nl Internet Source	1 %
12	S. Prasad Jones Christydass, Nurhayati Nurhayati, S. Kannadhasan. "Hybrid and Advanced Technologies", CRC Press, 2025 Publication	1 %
13	journal.uad.ac.id Internet Source	1 %
14	time.news Internet Source	1 %
15	www.e-journal.citakonsultindo.or.id Internet Source	1 %
16	www2.mdpi.com Internet Source	<1 %
17	journal.unnes.ac.id Internet Source	<1 %
18	dokumen.pub Internet Source	<1 %

19	Mohammad Imran, Peter Moyle, Danielle Kamato, Yousuf Mohammed. "Advances in, and prospects of, 3D preclinical models for skin drug discovery", Drug Discovery Today, 2024 Publication	<1 %
20	www.atlantis-press.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		