Performance Comparison of Boosting Algorithms in Spices Classification Using Histogram of Oriented Gradient Feature Extraction

Muhathir[1], Reydo Trisno Pangestu[2], Ira Safira[3], Melisah[4]
[1]Faculty of Engineering, Informatics Engineering, Universitas Medan Area, Indonesia

ABSTRACT

Spice classification is an important task in the food industry to ensure food safety and quality. This study focuses on the classification of spices using the HoG feature extraction method and boosting algorithms. The objective of this research is to compare the performance of four different models of boosting algorithms, namely Adaboost Classifier, Gradient Boosting Classifier, XGB Classifier, and Light GBM Classifier, in classifying spices. The evaluation metrics used in this research are Precision, Recall, F1-Score, F2-Score, Jaccard Score, and Accuracy. The results show that the XGB Classifier model achieved the best performance, with a precision of 0.811, recall of 0.809, and F1-score of 0.809, while the Adaboost Classifier model had the lowest performance, with a precision of 0.709, recall of 0.689, and F1-score of 0.682. Overall, the results indicate a fairly good success rate in classifying spices using the HoG feature extraction method and boosting algorithms. However, further evaluation is needed to improve the accuracy of the classification results, such as increasing the number of training data or considering the use of other feature extraction methods.

Keyword : Spices, HoG, Boosting Algoritmh

Corresponding Author:
Muhathir,
Department of Informatics
Universitas Medan Area, Indonesia
Email : muhathir@uma.ac.id

Article history:
Received Mar 2, 2023
Revised Mar 14, 2023
Accepted Mar 17, 2023

1. INTRODUCTION

Indonesia is known for its vast production of spices. Historical records show that some European countries colonized the archipelago to control valuable spices when prices were high in the market. Indonesia remains an important producer and exporter of spices in the world. In 2013, Indonesian spices dominated 21.06% of the global spice market. As a spice producer, Indonesia has the potential to sell its products worldwide and improve the country's economy (Yana, 2018). Indonesian spices are a great opportunity for businesses in the plantation subsector because they are marketed globally. As a spice producer, Indonesia has the opportunity to become a global spice exporter that can contribute positively to the economy. Over the past twenty-eight years, the world's demand for spice commodities has continued to increase, with an annual demand rate of 10.21% per year. In the future, the demand for natural ingredients such as spices is expected to grow along with population growth, the economy, health demand, synthetic product prices, and awareness of environmental sustainability (Anggrasari & Mulyo, 2019). Spices and herbs are a rich source of antioxidants that can protect your body from dangerous diseases. Spices and herbs have been used as flavor enhancers, colorings, and aromas for over 2,000 years (Embuscado, 2015).

Spices have been used for a long time due to their captivating aroma and sensory quality that is used to enhance the taste of food (Ali et al., 2021). Spices are special food additives that have been used for thousands of years as flavorings, colorings, and preservatives. In addition, spices are known to offer medicinal benefits and have been used in medical practices for a long time (Biology & 2021). Cumin, nutmeg, pepper, fennel, and coriander are types of plants that use seeds. Turmeric, ginger, galangal, and fingerroot are types of spices that use rhizomes (Susiarti et al., 2021)..

In a study (Wulandari et al., 2020) at SMKN 9 Bandung, most students (47%) were not familiar with spices and herbs. With the development of digital image processing technology, it is now possible to automatically sort spices and herbs. Image classification is a way to solve this problem. The goal of image classification is to replicate human ability in understanding digital image information, so that computers can classify objects in the form of images in the same way as humans do. There are several ways to solve this problem, one of which is by using the Naïve Bayes Classification Method, and HOG (Histogram of Oriented Gradients) Feature Extraction, as discussed in a study (Muhathir & Santoso, 2020) on fruit classification..

The basic concept of the Histogram Of Gradient (HOG) algorithm is to extract distinctive features for object detection using gradient information from each pixel. Typically, HOG features are extracted from several window sizes of an image (Ghaffari et al., 2020)(Girsang, 2021)(Tanjung & Muhathir, 2020)(Muhathir, et al., 2020). The Histogram of Oriented Gradient (HOG) feature, one of the simple computations, has attracted attention and achieved remarkable success in many computer vision tasks (Bakheet & Al-Hamadi, 2021)(Rizal, et al., 2019). In the study by Leidiyana & Warta (2022), HOG feature extraction was used for fruit classification with the SVM method..

Boosting is a method used to improve the accuracy of any algorithm that experiences overfitting due to coefficients that do not match the data points (Basha, S. M., Rajput, D. S., & Vandhan, V. 2018). Boosting has several enhancement methods that are widely used, including AdaBoost, a popular classification algorithm. In the training phase of AdaBoost, the weight distribution of the samples will increase along with the errors, and conversely, the weight distribution will decrease when the errors decrease (Sevinç, E. 2022). CatBoost is a powerful machine learning algorithm that can achieve good results in various practical tasks (Dorogush, A. V., Ershov, V., & Gulin, A. 2018). XGBoost is one of the boosting algorithm enhancement algorithms used to classify regression tree models derived from gradient boosting decision trees (Jiang, Y., Tong, G., Yin, H., & Xiong, N. 2019). LightGBM is used for efficient classification, regression, and parallel training. It is also considered a fast and efficient type of gradient boosting decision tree (GBDT) (Shaker, B., et al. 2019)..

The status of spices in the Pacific region, with a focus on ginger and vanilla, helps to understand the status of spices in Oceania. This research can assist in designing strategies to increase spice production and trade, which can positively impact the regional economy. With the development of modern times, people tend to consume fast food, which can lead to degenerative diseases based on their lifestyle. In research (Helmalia et al., 2019), spices can be used as a natural source of antioxidants that have benefits for the body.

In research (Achyunda Putra et al., 2020a), HOG feature extraction is used to detect objects such as cars. With HOG feature extraction, testing data is created from 304 x 240 pixel images. In research (Mutiara & Azizah, 2022), HOG feature extraction with the SVM method is used to classify brain tumors. This method involves three steps in the classification process: pre-processing to resize the image, feature extraction to extract information using HOG feature extraction, and training data using classification testing with SVM, resulting in an accuracy rate of 91%. HOG feature extraction has also been used to identify plant species from leaf patterns using HOG feature extraction and the KNN method (Bao et al., 2020)(Krairina, N,. et al. 2022) (Muhathir et al., 2022).

Based on previous research achievements, both Boosting method and HoG feature extraction have obtained satisfactory results. Therefore, in this study, we modeled the classification of spices using Boosting Algorithm by utilizing HoG feature extraction in classifying types of spices. This research is expected to help the community in distinguishing types of spices. The existence of this spice classification system helps young people to know the types of spices.

2. MATERIAL AND METHOD

A. Data gathering method

Pictures taken with the Vivo Y35 mobile phone camera at a distance of 30 cm were utilized to collect the data for this investigation. Prior to classification, materials must first be purchased for the spices at the nearby market, and then each spice must be photographed. This model was chosen so that the pattern of the spices looks more natural. The technique for collecting spice samples involves laying spices on HVS paper in a variety of positions, quantities, and random arrangements.

B. Data analysis

The 15 different sorts of spices (Aniseed, Cloves, Cumin, Cardamom, Candlenuts). 750 data were obtained when the data collection process was completed. 150 samples of each spice variety, or samples. Both training and testing data are included in this study's data set, which is split into two sections. Data used for testing and training are split 80:20. The accuracy of the data split information is displayed in Table 1.

Table 1. Split data

Types of Spices	the quantity of training samples	the quantity of testing samples
Aniseed	120	30
Cloves	120	30
Cumin	120	30
Cardamon	120	30
Candlenut	120	30

C. Architecture Research

Figure 1 illustrates the research methodology adopted in this study.

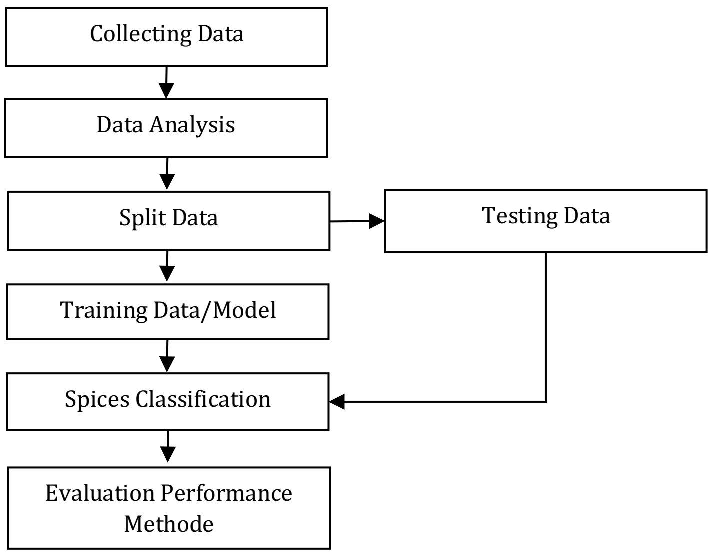

Figure 1. Architecture research (Fadlisyah & Muhathir, 2023),

Figure 1. The workflow for this study is shown, with the first step being data collection using the methods described in the data gathering method, followed by an analysis of the feasibility of the data sampled for the study. The feasibility of data acquisition is divided into two parts, training and testing, where 80% of the data is calculated and used as a model for the testing phase, and where the remaining 20% of the data is tested for similarity to or compatibility and provides classification results according to the intelligence of the training model, and in the last step evaluates the performance of the method.

D. Evaluation methode

The confusion matrix is a classifier for the amount of data that passes the test, and the confusion matrix table displays how much data fails the test. A method for assessing an object estimate model's accuracy is the confusion matrix. The prediction matrix, which will be compared with the first input class, provides details about the actual and predicted classification results.

Tabel 2 Confusion Matrix

		True Class	
		Positive	Fasle Positive
Predict Class	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

The evaluation method used in this study is Accuracy, Precision, Recall, F1-score and F2-score, Jaccard score. Accuracy is the most intuitive measure of performance and is simply the ratio of correctly predicted observations to total observations. Precision is the ratio of correctly predicted positive observations to the total number of predicted positive observations. Recall is the ratio of correctly predicted positive observations of all observations in the actual class. F1-score is a weighted average of precision and recall so this score takes into account false positives and false negatives. The weighted harmonic mean of recall and precision makes up the F2- score, which reaches its best value at 1 and its worst value at 0. A set of predicted labels for a sample is compared to the equivalent set of labels in y true using the Jaccard index or Jaccard similarity coefficient, which is defined as the size of the intersection divided by the size of the union of two label sets.

3. RESULT AND DISCUSSION

A. Sample of Spices

The images in Figure 2 showcase different spice samples, Aniseed is a plant with a sweet, licorice-like flavor used in cooking and herbal remedies for digestive problems, coughs, and respiratory ailments. Cloves are dried flower buds of an evergreen tree used in cooking, traditional medicine for their analgesic and anti-inflammatory properties, and essential oils in aromatherapy and cosmetics. Cumin is a flowering plant used as a spice in many cuisines for its warm, earthy, and slightly bitter flavor, and traditional medicine for its digestive and anti-inflammatory properties. Cardamon is a member of the ginger family and is used as a spice in many cuisines for its strong, pungent, and slightly sweet flavor and traditional medicine for its digestive and anti-inflammatory properties. Candlenut is a tree with seeds used as a food ingredient in many Southeast Asian cuisines and traditional medicine for its anti-inflammatory and analgesic properties, but it is toxic when eaten raw and must be roasted or boiled before consumption. Overall, the images in Figure 2 provide a visual representation of the diverse shapes and textures of various spice samples.

Figure 2 Spice Samples (a) Aniseed, (b) Cloves, (c) Cumin, (d) Cardamom, (e) Candlenuts.

B. Histogram of Oriented Gradient Feature Detection

Here are the detection results obtained from the histogram of oriented gradient. This method is used to identify objects in an image by analyzing the gradient distribution of pixels in the image. In the detection results, we can see how objects in the image can be clearly identified based on the gradient characteristics of pixels found by the histogram of oriented gradient method. With this technique, we can more accurately recognize objects in the image and facilitate further analysis or data processing.

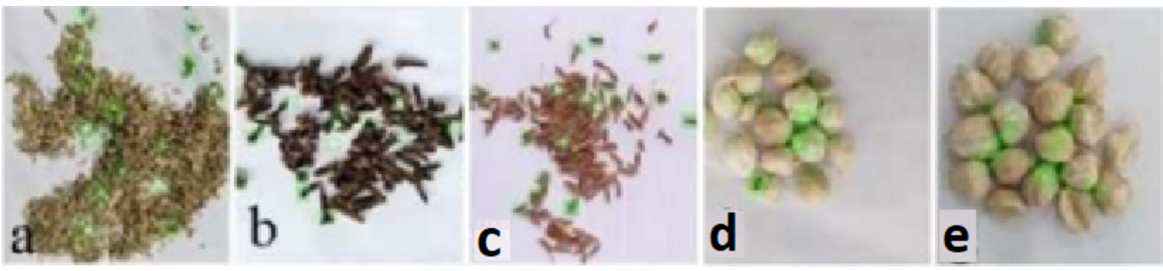

Figure 3 Spice Samples (a) Aniseed, (b) Cloves, (c) Cumin, (d) Cardamom, (e) Candlenuts

C. Experiment

In this study, a modified version of the boosting algorithm was tested by implementing the experimental design presented in Figure 1. Four trial models of the boosting algorithm, namely Adaboost Classifier, Gradient Boosting Classifier, XGB Classifier, and Light GBM Classifier, were used to carry out the experiment. The classification results were then analyzed and presented in the form of a confusion matrix, as depicted in Figure 4.

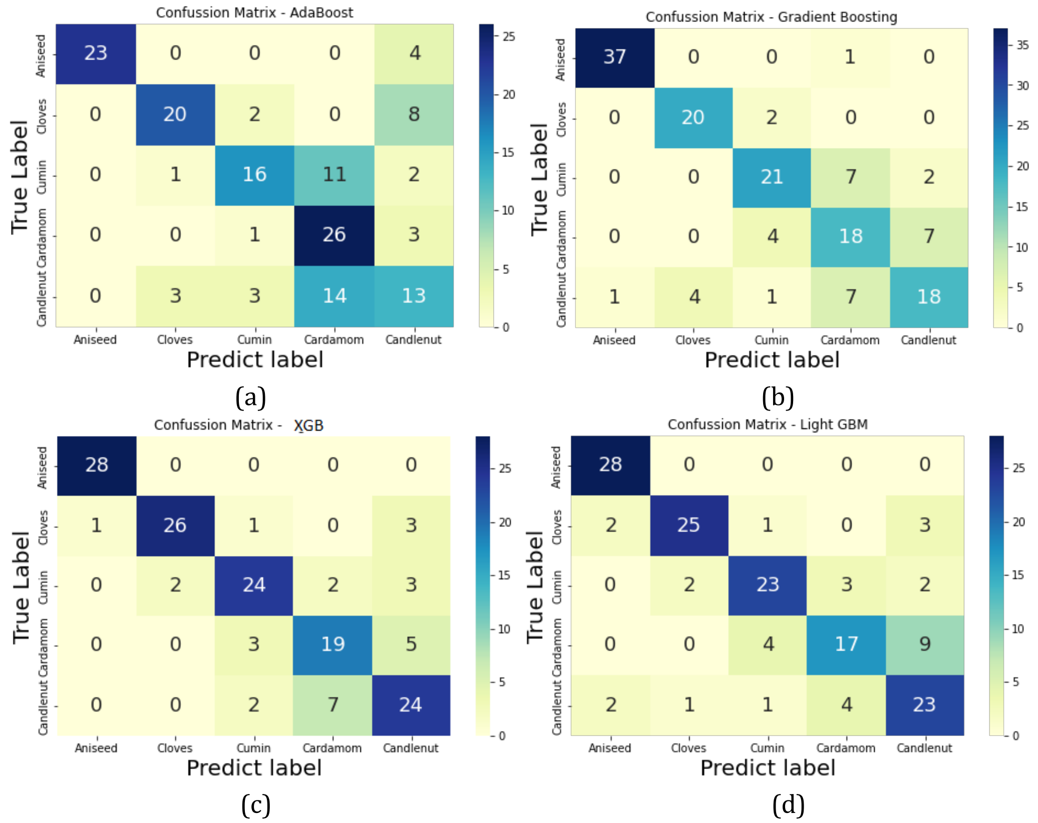

Figure 4. Confusion Matrix of Variations Boosting Algoritmh (a) Adaboost Classifier, (b) Gradient Boosting Classifier, (c) XGB Classifier (d) Ligth GBM Classifier

The classification results from the confusion matrix cannot be concluded in this study, hence there is a need for evaluation to determine the success achieved in classifying.

Table 3. The performance of Adaboost Classifier

	Precision	*Recall*	*F1-Score*	*F2-Score*	*Jaccard Score*
Aniseed	1	0.736842	0.848485	0.777778	0.73684211
Cloves	0.703704	0.863636	0.77551	0.826087	0.63333333
Cumin	0.857143	0.6	0.705882	0.638298	0.54545455
Cardamom	0.521739	0.827586	0.64	0.740741	0.47058824
Candlenut	0.464286	0.419355	0.440678	0.427632	0.2826087
Accuracy			0.68		

Table 4. The performance of Gradient Boosting Classifier

	Precision	*Recall*	*F1-Score*	*F2-Score*	*Jaccard Score*
Aniseed	0.973684	0.973684	0.973684	0.973684	0.94871795

Cloves	0.833333	0.909091	0.869565	0.892857	0.76923077
Cumin	0.75	0.7	0.724138	0.709459	0.56756757
Cardamom	0.545455	0.62069	0.580645	0.604027	0.40909091
Candlenut	0.666667	0.580645	0.62069	0.596026	0.45
Accuracy			0.76		

Table 5. The performance of XGB Classifier

	Precision	*Recall*	*F1-Score*	*F2-Score*	*Jaccard Score*
Aniseed	0.965517	1	0.982456	0.992908	0.96551724
Cloves	0.928571	0.83871	0.881356	0.855263	0.78787879
Cumin	0.8	0.774194	0.786885	0.779221	0.64864865
Cardamom	0.678571	0.703704	0.690909	0.698529	0.52777778
Candlenut	0.685714	0.727273	0.705882	0.718563	0.54545455
Accuracy			0.806		

Table 6. The performance of Ligth GBM Classifier

	Precision	*Recall*	*F1-Score*	*F2-Score*	*Jaccard Score*
Aniseed	0.875	1	0.933333	0.972222	0.875
Cloves	0.892857	0.806452	0.847458	0.822368	0.73529412
Cumin	0.793103	0.766667	0.779661	0.771812	0.63888889
Cardamom	0.708333	0.566667	0.62963	0.590278	0.45945946
Candlenut	0.621622	0.741935	0.676471	0.714286	0.51111111
Accuracy			0.773		

D. Discussion

Based on the performance measurements of the four classification models tested in this study, it can be seen that the XGB Classifier model shows the best performance with a precision of 0.811, recall of 0.809, and an F1-score of 0.809. Meanwhile, the model with the lowest performance is the Adaboost Classifier with a precision of 0.709, recall of 0.689, and an F1-score of 0.682.

Overall, it can be said that the classification results from the confusion matrix show a fairly good level of success in classifying types of spices using HOG feature extraction and Boosting algorithm. However, further evaluation of model performance is still needed, such as increasing the amount of training data or considering the use of other methods for feature extraction, in order to improve the accuracy of classification results.

Table 7. The performance of Variations Boosting Algoritmh

	Precision	*Recall*	*F1-Score*	*F2-Score*	*Jaccard Score*	*Accuracy*
Adaboost Classifier	0.709374	0.689484	0.682111	0.682107	0.533765386	0.68
Gradient Boosting Classifier	0.753828	0.756822	0.753744	0.755211	0.62892144	0.76
XGB Classifier	**0.811675**	**0.808776**	**0.809498**	**0.808897**	**0.695055402**	**0.806**
Ligth GBM Classifier	0.778183	0.776344	0.77331	0.774193	0.643950716	0.773

4. CONCLUSION

The results show that all four models tested achieved relatively high accuracy, with the XGB Classifier performing the best with an accuracy of 80.6%. In terms of precision, recall, and F1-score, the XGB Classifier also outperformed the other models. The Adaboost Classifier had the lowest performance with a precision of 0.709, recall of 0.689, and F1-score of 0.682. The Jaccard Score measures similarity between two sets of data and is used to evaluate the accuracy of clustering algorithms. In this study, the Jaccard Score ranged from 0.534 to 0.696, indicating that there is some overlap between the predicted and actual class labels. Overall, the results suggest that the HOG feature extraction method combined with Boosting algorithms can effectively classify different types of spices. However, there is still room for improvement in terms of increasing the sample size and considering other feature extraction methods. Further evaluations could be done to determine if these models can be applied to other datasets or if they need to be retrained.

REFERENCES

Achyunda Putra, F. A. I., Utaminingrum, F., & Mahmudy, W. F. (2020a). HOG Feature Extraction and KNN Classification for Detecting Vehicle in The Highway. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 14(3), 231. https://doi.org/10.22146/ijccs.54050

Ali, A., Wu, H., Ponnampalam, E. N., Cottrell, J. J., Dunshea, F. R., & Suleria, H. A. R. (2021). Comprehensive profiling of most widely used spices for their phenolic compounds through lc-esi-qtof-ms2 and their antioxidant potential. Antioxidants, 10(5). https://doi.org/10.3390/antiox10050721

Anggrasari, H., & Mulyo, J. H. (2019). The Trade Of Indonesian Spice Comodities In International Market. Agro Ekonomi, 30(1). https://doi.org/10.22146/ae.41665

Embuscado, M. E. (2015). Spices and herbs: Natural sources of antioxidants - A mini review. Journal of Functional Foods, 18, 811–819. https://doi.org/10.1016/j.jff.2015.03.005

Bakheet, S., & Al-Hamadi, A. (2021). A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve bayesian classification. Brain Sciences, 11(2), 1–15. https://doi.org/10.3390/brainsci11020240

Basha, S. M., Rajput, D. S., & Vandhan, V. (2018). Impact of gradient ascent and boosting algorithm in classification. International Journal of Intelligent Engineering and Systems (IJIES), 11(1), 41-49.

Biology, M. L.-A. of the R. S. for C., & 2021, undefined. (2021). Role of Spices in Food for Immune Boosting. Annalsofrscb.Ro, 25(3), 5841–5853

Bao, T. Q., Kiet, N. T. T., Dinh, T. Q., & Hiep, H. X. (2020). Plant species identification from leaf patterns using histogram of oriented gradients feature space and convolution neural networks. Journal of Information and Telecommunication, 4(2), 140–150. https://doi.org/10.1080/24751839.2019.1666625.

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Ghaffari, S., Soleimani, P., Li, K. F., & Capson, D. W. (2020). Analysis and Comparison of FPGA-Based Histogram of Oriented Gradients Implementations. IEEE Access, 8, 79920–79934. https://doi.org/10.1109/ACCESS.2020.2989267

Girsang, N. D. (2021, February). Classification Of Batik Images Using Multilayer Perceptron With Histogram Of Oriented Gradient Feature Extraction. In *Proceeding International Conference on Science and Engineering* (Vol. 4, pp. 197-204).

Helmalia, A. W., Putrid, P., & Dirpan, A. (2019). Potensi Rempah-Rempah Tradisional Sebagai Sumber Antioksidan Alami Untuk Bahan Baku Pangan Fungsional). Canrea Journal: Food Technology, Nutritions, and Culinary Journal, 2(1), 26–31. https://doi.org/10.20956/canrea.v2i1.113

Jiang, Y., Tong, G., Yin, H., & Xiong, N. (2019). A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. IEEE Access, 7, 118310-118321.

Khairina, N., Sibarani, T. T. S., Muliono, R., Sembiring, Z., & Muhathir, M. (2022). Identification of Pneumonia using The K-Nearest Neighbors Method using HOG Fitur Feature Extraction. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *5*(2), 562-568.

Muhathir, M., & Santoso, M. H. Muliono. R(2020a). Analysis Naïve Bayes In Classifying Fruit by Utilizing Hog Feature Extraction. Journal of Informatics and Telecommunication Engineering, 4(1), 151–160. https://doi.org/10.31289/jite.v4i1.3860

Muhathir, M., Sibarani, T. T. S., & Al-Khowarizmi, A. K. (2020). Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction. *Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal*, *1*(1).

Muhathir, M., Muliono, R., & Hafni, M. (2022). Image Classification of Autism Spectrum Disorder Children Using NaÃ¯ve Bayes Method With Hog Feature Extraction. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *5*(2), 494-501.

Mutiara, T. A., & Azizah, Q. N. (2022). Klasifikasi Tumor Otak Menggunakan Ekstraksi Fitur HOG dan Support Vector Machine. Jurnal Infortech, 4(1), 45–50.

Rizal, R. A., Sihotang, J. S., & Gultom, R. (2019, November). Comparison of SURF and HOG extraction in classifying the blood image of malaria parasites using SVM. In *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)* (pp. 1-6). IEEE.

Sevinç, E. (2022). An empowered AdaBoost algorithm implementation: A COVID-19 dataset study. Computers & Industrial Engineering, 165, 107912

Shaker, B., Yu, M. S., Song, J. S., Ahn, S., Ryu, J. Y., Oh, K. S., & Na, D. (2021). LightBBB: computational prediction model of blood–brain-barrier penetration based on LightGBM. Bioinformatics, 37(8), 1135-1139

Susiarti, S., Rahayu, M., Ningsih, D. Q. W., Arifa, N., & Setiawan, M. (2021). Tanaman Rempah dan Masakan Tradisional di Kelurahan Nanggewer Mekar, Cibinong, Kabupaten Bogor. Jurnal Masyarakat Dan Budaya, 23(3), 337–353. https://doi.org/10.14203/jmb.v23i3.1434.

Tanjung, J. P., & Muhathir, M. (2020). Classification of facial expressions using SVM and HOG. *Journal of Informatics and Telecommunication Engineering*, *3*(2), 210-215.

Yana. (2018). Study Jenis Rempah – Rempah dan Pemanfaatannya di Pasar Tradisional Angso Duo. Skripsi, 125.

Wulandari, I., Yasin, H., & Widiharih, T. (2020). Klasifikasi Citra Digital Bumbu Dan Rempah Dengan Algoritma Convolutional Neural Network (CNN). Jurnal Gaussian, 9(3), 273–282. https://doi.org/10.14710/j.gauss.v9i3.27416