# IMPLEMENTATION OF HAAR CASCADE AND ADABOOST ALGORITHMS IN PHOTO CLASSIFICATION ON SOCIAL NETWORKS

Rezty Amalia Aras[1*], Andi Hutami Endang[2]
Institute Technology and Business Kalla, Makassar, Indonesia
Department of Bisnis Digital, ITB Kalla, Makassar, Indonesia[1]
Institute Technology and Business Kalla, Makassar, Indonesia[2]
E-mail address: Rezty@kallabs.ac.id[1], hutamiendang@kallabs.ac.id[2]

***Received: 06, June, 2023*** ***Revised: 08, June, 2023*** ***Accepted: 28, June, 2023***

***ABSTRACT***

Instagram is one of the fastest social networks in recent years. Instagram is a popular social media for sharing images. For example, image searches on Instagram may use certain keywords, sometimes called hashtags. There are no rules for specifying hashtags when users upload images. As such, the specified hashtag may not be relevant to the uploaded image. There are photos whose content is dominated by selfies. The study was conducted using data from Instagram, using hashtags to refine searches. Next, classify from the search results. The survey has three categories: selfies, food, and travel. Results: Two of her classification algorithms, Haar Cascade and Adaboost, were used in this study. From the study results, we can conclude that the Haar cascade has a precision rate of 0.7081/s and a detection error of 0.8816/s, while Adaboost has a precision rate of 0.7072/s and a detection error of 0.8424/s. According to the recognition results, the two algorithms can recognize and classify photos with almost the same accuracy (only 0.0392 seconds).

**Keywords: Classification, Instagram, Haar Casecade, Adaboost**

## 1. INTRODUCTION

Artificial intelligence (AI) technology is now and will continue to influence various human lives throughout the world. In companies or organizations, the adoption of this technology is predicted to continue to increase because of the various benefits it provides to humans. AI technology itself has become the main driver of new technologies such as big data, chatbots, self-driving cars, robotics and the Internet of Things (IoT). Talking about AI, social media is a pioneer of AI that is very up to date. Instagram is one of them that applies AI.

Instagram is one of the fastest-growing social media in recent years. Instagram is a popular social media for sharing images. Research results show that understanding a variety of factors improved predictions of photo popularity on social media sites (Abousaleh, Cheng, Yu, & Tsao, 2021). For example, when searching for images on Instagram, you can use certain keywords or so-called hashtags. There are no rules for hashtags when users upload images. As such, the specified hashtag may not be relevant to the uploaded image. I have an image where the majority of the content is a selfie face. As a result, the background image and the position

of the image are not fully conveyed (Priadana & Habibi, 2019). Social media image filters have been researched to change the contrast and lighting of images, remove parts of faces, and more. Cover. prosthetic glasses or animal noses (Hedman, Skepetzis, Hernandez-Diaz, Bigun, & Alonso-Fernandez, 2021). Instagram is currently believed by users who indulge the wishes of the user himself. Various photos that we encounter in one scroll on Instagram media.

The use of hashtags narrows down in searches on Instagram. Hashtags make it easy for users to find what they want on Instagram. Photos uploaded on Instagram are generally selfies, food photos and photos of tourist attractions. This causes users to get less clear information from one photo object. In research (Priadana & Habibi, 2019) stated We found that the method used resulted in 71.48% accuracy in detecting human faces. Based on human face detection results, Haar Cascade method can filter selfie face images with an accuracy value of 64.6%. Haar Cascade is one of the commonly used machine learning models as a basis for recognition applications. In 1995, Freund and Schapire jointly proposed an algorithm he called the Adaboost algorithm. The basic idea of this algorithm was to combine many weak classifiers with a particular method into one strong classifier. Then Viola et al. Applies the Adaboost algorithm to face recognition. This is a convenient level of face recognition, allowing fast face recognition to meet practical needs (Li, Zeng, Li, & Yu, 2019).

## 2. THEORY

a. Instagram

According to some theories, Instagram is a photo and video sharing application that allows users to take photos and videos, apply filters and share them on various social media.

b. Haar Cascade

Haar's algorithm uses statistical methods to detect faces (Sangeetha, Miruthula, Kavimalar, & Aakash, 2023). This method uses sample haarlike fetures (L. Zhang, Wang, & An, 2023). This classification uses fixed-size (24x24) images. Haar's method of detecting faces is to use a sliding window technique measuring 24x24 on all images and determine whether the image has a part with a shape like a face or not. Haar also has the ability to scale for detection of faces larger or smaller than the image in the classifier.

Each feature of an object such as a capture is determined by the shape of the object, including the coordinates of the object as well as the size of the object (Munawir et al., 2020). A cascading classifier is a series of level classifiers, each used to determine whether an image subwindow object has an object of interest. Level classifiers are created using the adaptive boosting algorithm (AdaBoost). This algorithm combines the power of many weak algorithms to produce a strong classifier. The weak classifier in this case is the Haar-like feature value. The AdaBoost type used is Gentle AdaBoost (Munawir et al., 2020).

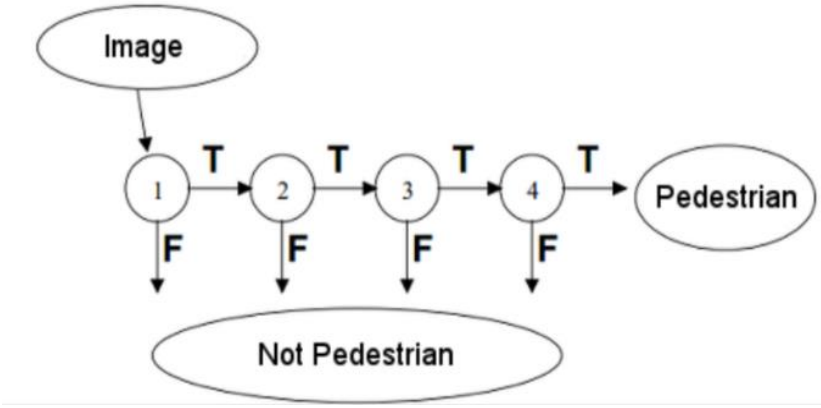



Figure 1. Classifeier model casecade

c. Haar Feature

Haar features are features based on Haar wavelets. A Haar wavelet is a single square wave (one high and one low). In two dimensions, one is bright and one is dark. In addition, it is said that the recognition of visual objects is improved by combining squares. Each Haar feature consists of a combination of black and white squares (Munawir et al., 2020).

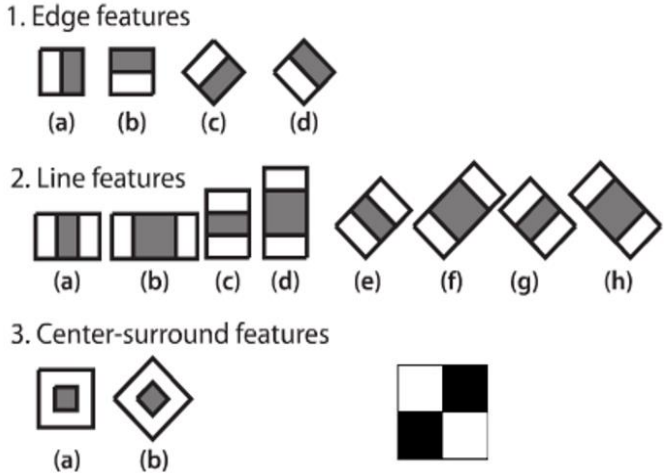



Figure 2. Types of Feature Variations on Haar

3 types of box (rectangular) feature (Zhao, U, & Luo, 2023):

1. Type two-rectangle feature (*horisontal/vertikal*)
2. Type three-rectangle feature
3. Type four-rectangle feature

The presence of hair features is determined by subtracting the dark pixel area from the average light area pixel. A feature is considered present if its delta value exceeds one or more bounding parks. The hairy feature is the difference between the sum of the grayscale values of the black box area and the white box area:

$$F(X) = SumBlack\ Rectangle - SumWhite\ Rectangle \quad (1)$$

Description: where for boxes in the Haar-like feature can be calculated quickly using the "image integral".

d. Adaboost

AdaBoost is a well-known ensemble learning method that provides an effective strategy for developing strong learners through iterative training of individual learners. AdaBoost was originally used to integrate classification and regression tree algorithms. AdaBoost implements ensemble prediction as follows:

1. The training set is distributed using the initial weights.
2. A basic learner is trained on the weighted dataset.
3. The base learner and dataset weights are updated according to the performance of the base learner in the previous iteration round.
4. Stop iterating if the stopping condition is met.
5. Combine BaseLearners using weighted voting. The process of iterative weight updates allows subsequent learners to focus more on misclassified cases. Therefore, a combination of basic learners can correct the classification bias and may give good performance. (Hu, Zhang, & Zhang, 2020). The following is a table of adaboost algorithms (Y. Zhang et al., 2019).

Table 1. Adaboost Algorithm

| input: Training example set N = {(X1, Y1), (X2, Y2)…(Xn, Yn)}, number of iterations T, and weak classifier algorithm | |
|---|---|
| 1 | Initialize the weight distribution for the training examples. D1(x) = 1/n |
| 2 | For t = 1, 2 … T; |
| 3 | Train the training data set using the weight distribution Dt to get a weak classifier Gt (X). |
| 4 | Estimate the classification error rate of Gt(X) in the training set εt, and if εt is less than 0.5, proceed to the next step. Otherwise, the loop terminates (Gt (X) ≠ f(X), where f(X) is a real function |
| 5 | Calculate the weight of $G_t(x)$: $a_t = \frac{1}{2}\ In\ \frac{1-\varepsilon_t}{\varepsilon_t}$ |
| 6 | Next weight distribution : $D_{t+1}(x) = \frac{D_t(x)\exp(-a_t f(x) G_t(x))}{Z_t}$ $(Z_t\ is\ normalization\ factor)$ |
| 7 | Get the final classifier : sign $(\sum_{t=1}^{T} a_t\ G_t\ (x))$ |

## 3. METHOD

The research carried out starts from determining the object of research, analyzing the problems that have occurred so far, namely the amount of Instagram social media data to be processed into one research object. Furthermore, the formulation of problems and research objectives is carried out. At the initial research stage, the results of photo analysis on Instagram can be classified using a classification algorithm. In this study, haar casecade and adaboost were used. Furthermore, data collection is carried out. At this stage the data on Instagram with selfie objects, food objects and tourist attractions. Next, system design is carried out. The system design used in this study is php programming language. Website-based application for photo classification. Database design is done by collecting photo set data from Instagram. Furthermore, researchers tested the system from the photo set data that had been collected. The

data set used was 600 photos. 200 selfies, 200 food photos and 200 nature or scenery photos. The final step is the drawing of conclusions made by the researcher.

## 4. RESULTS AND DISCUSSION

The results of research conducted by researchers using the PHP programming language. Researchers used a computer with 16 GB of RAM, Intel Core i7 Gen 11. The test was conducted using data from Instagram with a total amount of data of 500 photos. In research using two methods, namely haar casecade and adaboost have different algorithms. Here are the results obtained from both methods:*Haar Casecade.*

```
313/313 - 2s - loss: 0.8816 - accuracy: 0.7081 - 2s/epoch - 6ms/step
```

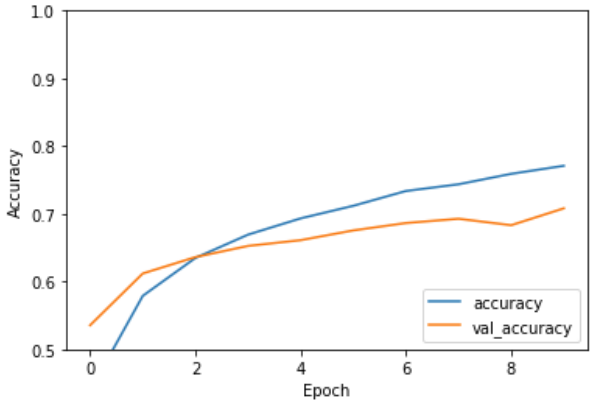



Figure 3. Haar Casecade Algorithm Graph

From the graph in figure 3 above shows detection with the Haar Casecade Algorithm with photo 313/313 accuracy rate 0.7081/sec and detection failure 0.8816/sec.

*a. Adaboost*

```
313/313 - 2s - loss: 0.8424 - accuracy: 0.7072 - 2s/epoch - 6ms/step
```

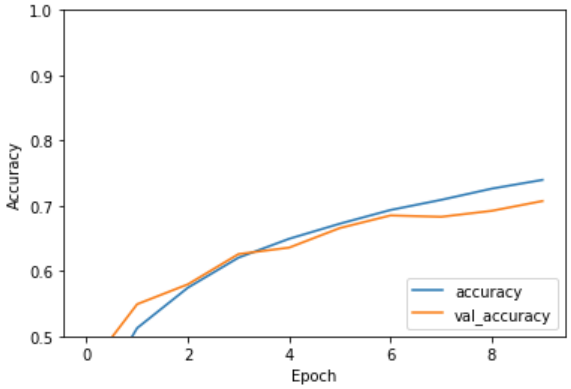



Figure 4. Adaboost Algorithm Graph

From the graph in figure 4 shows the detection with the Adaboost algorithm with the same photo used as when testing the Adaboost algorithm with an accuracy level of 0.7072/sec and a detection failure of 0.8424/sec. The application of both algorithms in the application is seen as below:

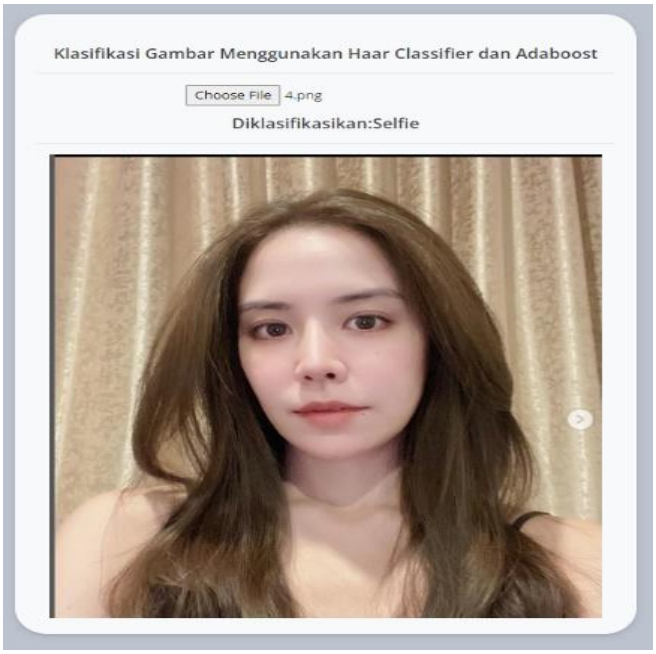



Figure 5. Classification with the use of Selfie harstags

Figure 5 is one of the data with the use of selfie harstags performed in the test with the use of two lagorithms that use can be detected. The selfies used in the test were 200 photos with pre-processing and processing stages. Photos are used with various backgrounds. The selfie category that is successfully detected if the selfie photo is a photo that does not use accessories such as glasses, masks, and hats. Another thing that the system cannot recognize is the position at the time of the selfie. In the table below are some examples that work and cannot be recognized by the system:

Table 2. Selfie Detection Results

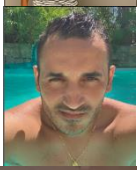
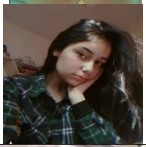
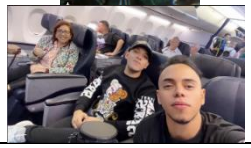
| Image | Categorically | Description |
|---|---|---|
|  | Selfie | Identified |
|  | Selfie | Identified |
|  | Selfie | Identified |
|  | Not Selfie | Not |
|  | Not Selfie | Not |


Figure 6. Classification with the use of Food harstags

Figure 6 is one of the data with the use of Food harstags performed in a test with the use of two lagorithms that use can be detected. In the food photo category, the same data was used in selife, which was 200 total photos with preprocessing and processing stages. But at the stage with the food classification category, it is slightly different from selfies. Because the pattern is recognized by more than one and the type of food photos used have different shapes, the preprocessing stage the author uses several steps to get maximum detection results. One of them is the use of RGB at the preprocessing and processing stages. Some data was successfully recognized. Here's some recognized and unrecognized data:

Table 3. Food Detection Results

| Image | Categorically | Description |
|---|---|---|
|  | Food | Identified |
|  | Food | Identified |
|  | Not Food | Not |

*P-ISSN : 2088-6705*
*E-ISSN : 2621-5608*

| Image | Categorically | Description |
|---|---|---|
|  | Food | Identified |
|  | Not Food | Not |
|  | Not Food | Not |


Figure 7. Classification with the use of tourist harstags

Figure 7 with the use of a tourist harstag performed in a test with the use of two lagorithms that use can be detected. The tourist photos used in the test were 200 photos with pre-processing and processing stages. At the food category stage, the same thing is done as when the selfie category. Here some of the data used is recognizable and unrecognizable:

Table 4. Tourism Detection Results

| Image | Categorically | Description |
|---|---|---|
|  | Tourist | Identified |
|  | Tourist | Identified |
|  | Not Tourist | Not |
|  | Not Tourist | Not |
|  | Tourist | Identified |
|  | Not Tourist | Not |

## 5. CONCLUSIONS AND SUGGESTIONS

The results of the study with the application of two algorithms, namely Haar casecade and Adaboost with the data used in total are 600 photos originating from Instagram with narrowing the search using Harstag and dividing three categories, namely the Selfie category, Food category and travel category can be concluded the accuracy level of 0.7081 / sec and detection failure 0.8816/sec from Haar casecade algorithm, while on Adaboost algorithm with accuracy level 0.7072/sec and detection failure 0.8424/sec. According to the detection results, the two algorithms can detect and classify photos with a degree of accuracy, not much different, only 0.0392 secound.

## 6. ACKNOWLEDGEMENTS

The author congratulates the Ministry of Education and Culture, Research and Technology for being financially supported to carry out research for novice lecturers.

## REFERENCES

Abousaleh, F. S., Cheng, W. H., Yu, N. H., & Tsao, Y. (2021). Multimodal Deep Learning Framework for Image Popularity Prediction on Social Media. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(3), 679–692. https://doi.org/10.1109/TCDS.2020.3036690

Hedman, P., Skepetzis, V., Hernandez-Diaz, K., Bigun, J., & Alonso-Fernandez, F. (2021). *On the Effect of Selfie Beautification Filters on Face Detection and Recognition*. 1–7.

Hu, X., Zhang, P., & Zhang, Q. (2020). A Novel Framework of CNN Integrated with Adaboost for Remote Sensing Scene Classification. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2643–2646. https://doi.org/10.1109/IGARSS39084.2020.9324261

Li, Z. Z., Zeng, Q. H., Li, X. D., & Yu, Y. (2019). Face detection technology based on combining skin color model with improved adaboost algorithm. *2019 IEEE 4th International Conference on Signal and Image Processing, ICSIP 2019*, 381–384. https://doi.org/10.1109/SIPROCESS.2019.8868565

Munawir, M., Fitria, L., Hermansyah, M., Prathivi, R., Kurniawati, Y., Pramono, A., … Krisnadi, I. (2020). Sistem Presensi Kelas Menggunakan Pengenalan Wajah Dengan Metode Haar Cascade Classifier. *Electronics Letters*, *56*(1), 1–13. https://doi.org/10.24176/simet.v11i1.3754

Priadana, A., & Habibi, M. (2019). 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT 2019) : 13-15 March 2019, Platinum Adisucipto Yogyakarta Hotel & Conference Center. *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, 6–9.

Sangeetha, T., Miruthula, V., Kavimalar, C., & Aakash, V. (2023). Face Mask Detection and Social Distancing Using Machine Learning with Haar Cascade Algorithm. *Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022*, 987–999.

Zhang, L., Wang, J., & An, Z. (2023). Vehicle recognition algorithm based on Haar-like features and improved Adaboost classifier. *Journal of Ambient Intelligence and Humanized Computing*, *14*(2), 807–815.

Zhang, Y., Ni, M., Zhang, C., Liang, S., Fang, S., Li, R., & Tan, Z. (2019). Research and application of adaboost algorithm based on SVM. *Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2019*, (Itaic), 662–666. https://doi.org/10.1109/ITAIC.2019.8785556

Zhao, W., U, K., & Luo, H. (2023). An image super-resolution method based on polynomial exponential function and non-uniform rectangular partition. *The Journal of Supercomputing*, *79*(1), 677–701. https://doi.org/10.1007/s11227-022-04691-1