MASK DETECTION ANALYSIS USING HAAR CASCADE AND NAÏVE BAYES

Imam Riadi[1], Abdul Fadlil[2], Izzan Julda D.E Purwadi Putra[3]

[1] Program Studi Sistem Informasi, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 55166
[2] Program Studi Teknik Elektro, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 55166
[3] Program Studi Magister Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 55166
Email:, imam.riadi@is.uad.ac.id[1], fadlil@mti.uad.ac.id[2], izzan2008048026@webmail.uad.ac.id[3]

Corresponding Author: Izzan Julda D.E Purwadi Putra

ABSTRACT

Coronavirus Disease (COVID-19) is a new virus variant that emerged in 2019. The World Health Organization (WHO) states that 394,381,395 people have been infected with COVID-19, and 5,735,178 have died. This epidemic has been found in Indonesia since March 2020. New cases in Indonesia are still increasing every day as a whole. The Government as a policy has imposed a policy on anyone who will be required to wear a mask and also carry out physical distancing so that they can work without the maker being exposed to the virus. In the midst of a pandemic, the use of masks has increased to prevent transmission. Various types of masks are easy to find, but not all masks are recommended to avoid transmission. Among them are the N-95 masks, which are recommended to prevent transmission. This application uses the haar cascade and naive bayes methods. The pycharm edition 2021.2 tools and python 3.8 are the detection systems used in this mask. The haar cascade method is also used in detecting objects with masks or not and naive Bayes, which is used as an accuracy calculation. This study uses a dataset of 1092, which is divided into 192 positive images and 900 negative images. Accuracy results using the haar cascade method are 100% more accurate, while the nave Bayes method is 76.6% less accurate.

Keywords: *covid-19; detection; mask; haarcascade; naive bayes;*

I. INTRODUCTION

In December 2019, the acutely attention-grabbing appearance of COVID-19 took the world by storm. The outbreak was caused by a coronavirus with the name severe acute respiratory syndrome coronavirus two, abbreviated as SARS CoV-21 [1]. The first discovery in Wuhan, China, COVID-19 spread very quickly throughout the world [2]. As of June 17, 2020, the World Health Organization (WHO) recorded that this virus had infected around 8,061,550 people, and 440,290 people had lost their lives [3]. This virus was discovered in Indonesia in March 2020. Almost all areas created, such as office areas, markets, and schools, are always carried out by the task force to check everyone whether the individual is wearing a mask or not [4]. The total number of new cases in Indonesia continues to increase every day. As a policy maker, the Government has implemented regulations for anyone who wants to do it is mandatory to wear a mask and maintain physical distance so that everyone can travel from home or work without worrying about exposure to this virus. It is hoped that this application can be obeyed by every community by always wearing a mask. There are several aspects that affect compliance with this in society, namely aspects of knowledge, motivation, perception, and belief in efforts to control and prevent disease, environmental variables, quality of health instructions, and the ability to achieve access to existing sources [5]. in breathing and the most fatal, namely death is the result of exposure to this virus [6]. This virus can be transmitted through droplets when sneezing, coughing, or saliva that comes out of the human mouth. Therefore, using masks and physical distancing must be done in all aspects of life. Avoiding direct contact or keeping a distance, and wearing a mask are ways to avoid being exposed to this virus. Several aspects affect the level of community compliance, such as knowledge, motivation, perception, and belief in disease control and prevention efforts, environmental variables, quality of health instructions, and the ability to access existing sources [7]. When an individual or group wants to comply with a health analysis, but there are factors that inhibit the desire to comply with advice about the health of medical personnel, it is an act of non-compliance [8]. Non-compliance is also the same as the behaviour of an individual or group to comply with health advice or a therapeutic plan agreed upon between the individual or group and medical professionals [9]. Research on object detection is often used in robotics, medicine, surveillance, autonomous vehicles, and autonomous driving [10][11][12]. Mask detection can be done by many methods, such as Single Shot Detector, Eigenface, Local Binary Pattern (LBP), Haar-Cascade, Haar-Like Feature, and so on [13]. The Haar Cascade Viola-Jones method is the oldest method and is claimed to still have relevance for detecting faces accurately [14]. Another face detection method uses the Dlib Convolutional Neural Network and is considered quite sharp [15]. This method has been tested on face detection with barriers in the form of sunglasses, scarves, breathing apparatus, hats, and masks [16]. The haar cascade classifier is derived from the combination of black pixels and white pixels that form a box [17]. Efforts to prevent the spread of the above-

uncontrolled virus, so a study was created to detect the use of masks using the haar cascade and nave Bayes methods. The mask detection is carried out using the pycharm edition 2021 and python 3.8 tools. The system uses several libraries by upgrading pip such as NumPy, cv2, and OpenCV contrib python. Python is also used as a scripting language, as is generally dynamic programming [18].

This study uses the haar cascade algorithm, a machine learning model used as the basis for object detection applications in an image or video using OpenCV [19]. The data used is taken from the Kaggle dataset. The data in this study is a comparison between people who wear masks and those who do not. Objects that use the dataset go through a training process which is then detected by the algorithm to find out whether the person is wearing a mask or not.

II. METHODOLOGY

The main step carried out in the research of the mask detection system using the Haar cascade, and Naive Bayes methods is to carry out a literature study carried out to analyze and find the right system to be applied as well as to carry out the objectives of the research and solve existing problems. Figure 1. shows a simple process in the detection system, from image acquisition using the camera, then the image being processed and producing output.

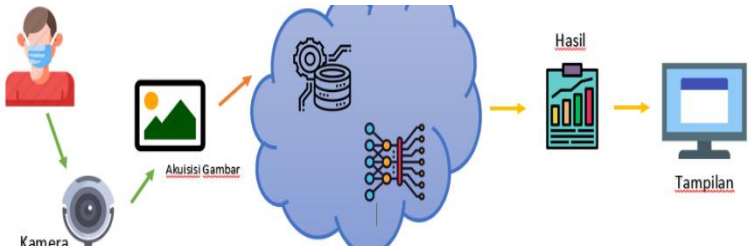

Figure 1. Mask Image Processing Flow

Classification is a data analysis task that is formed on a model (classifier) in order to formulate an object to a category or class that has been described previously [20]. There are two steps in the classification process. The first stage is training, in which the classification algorithm creates a classifier model by analyzing or studying a number of training data. The second stage is testing; here, the classifier model that has been formed will be used to classify the test data. The following is Figure 2. Explanation of the classification modelling framework.

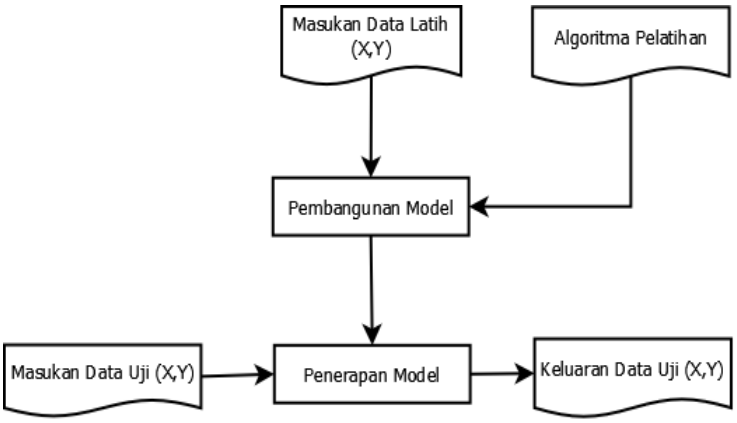

Figure 2. Classification modelling framework

Figure 2 shows the training data (X, Y), which is the model design data. In this case, the model is used as a class prediction from the test data (X, Y), so class Y is obtained. The model that has been built during training is used to predict new and unknown data labels.

The design of the mask detection analysis system using the haar cascade and Naive Bayes methods. The Haar algorithm utilizes statistical methods in the face detection process. Haar-like features are used when detecting objects in a digital image. Initially, image processing was carried out by observing the RGB value for each pixel, unfortunately, this method was found to be ineffective. Then, Viola and Jones did a development that led to the creation of the Haar-Like feature. The Haar-Like feature processes images in the form of boxes, which in a box consists of a number of pixels, each square which is then processed and looks for differentiating values that indicate

dark areas and bright areas [21]. These values will then be the basis for image processing [22]. The flow in detecting faces using Haar Cascade is shown in Figure 3.

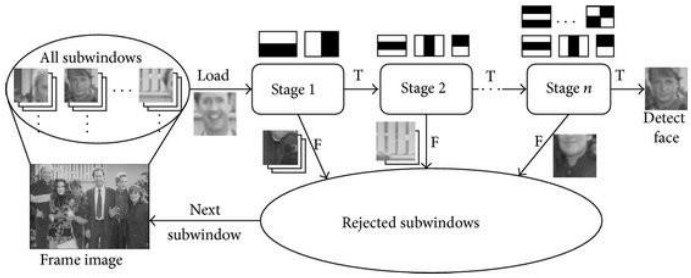

Figure 3. Face detection flow using Haar Cascade

By using datasets in the system, the data are trained on the tools pycharm edition 2021.3 and python 3.8. Pycharm is an Integrated Development Environment (IDE) that focuses on developing projects using the Python programing language. Pycharm was developed by JetBrains; Pycharm is a cross-platform IDE, meaning that it can run on various information systems such as Windows, Linux, and Mac [23].

2.1 Data Test

Testing training data using images is divided into 2 types of images, namely:

1. Positive data depicted people wearing masks as many as 192 images of masks.
2. Negative data depicted people not wearing masks as many as 900 images without masks.

2.2 Pre-Processing

A grayscale image is a collection of pixels containing two colours, black and white. The colour black has a value (range) from 0 (black) to 255 (white). Converting an RGB image to grayscale can divide the total value in pixels divided by 3. The grayscale image and equation can be seen in Figure 4.

$$Gray = (R + G + B) \div 1 \quad (1)$$

Is known :

R = Red (Red)

G = Green (Green)

B = Blue (Blue)

2.3 ROI Mask

A region of Interest is created to mark the area on the face that uses a mask. In the design, the author describes a blue rectangle that has a thickness of 2 and a blue line, as shown in Figure 4.

Figure 4. Region of Interest marking the mask area on the face

2.4 Naive Bayes

This system is also added with the Naive Bayes classification method (2). This method is a statistical classification that can be used as a predictor of the possibility of something in the future. This method is based on Bayes' theorem, which has an advantage in classifying classes similar to creating a decision tree. This method is considered capable of creating a higher accuracy and speed value when testing on more total datasets [24][25][26].

$$P\ (X|C) = \frac{P\ (X|C)\ P(C)}{P\ (X)} \quad (2)$$

Is known:
X = unkown class data
C = Hipotesis data x which is the more specific class
P (C|X) = *posterior probability*
P (C) = *prior probability*
P (X|C) = probability based on hypothesis condition
P (X) = probability c

2.5 Confusion Matrix

Comparison of original data and predicted data, consisting of True Positive, True Negative, False Positive, and False Negative. Data is declared TP when the original data and predictive data are positive. TN if the predicted data and the original data are negative. FP if the predicted data is negative and the original data is positive. FN if the predicted data is positive and the original data is negative. Accuracy parameters in equation (3) are obtained in the Confusion Matrix.

TABLE 1
CONFUSION MATRIX

		Prediction	
Ac-		P	N
Darkl	P	TP	FN
	N	FP	TN

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2.6 Application Flowchart

The system analysis is shown in Figure 5. The system starts by entering an image in the form of an image or webcam video then the image is converted from an RGB image to a grayscale. Feature extraction processes and produces image output, and then the process again detects the presence of a mask; when the mask is found, the ROI will mark the part of the face that uses the mask with ROI. The ROI process is completed if the mask is detected, or the process is completed if the face does not detect the mask image.

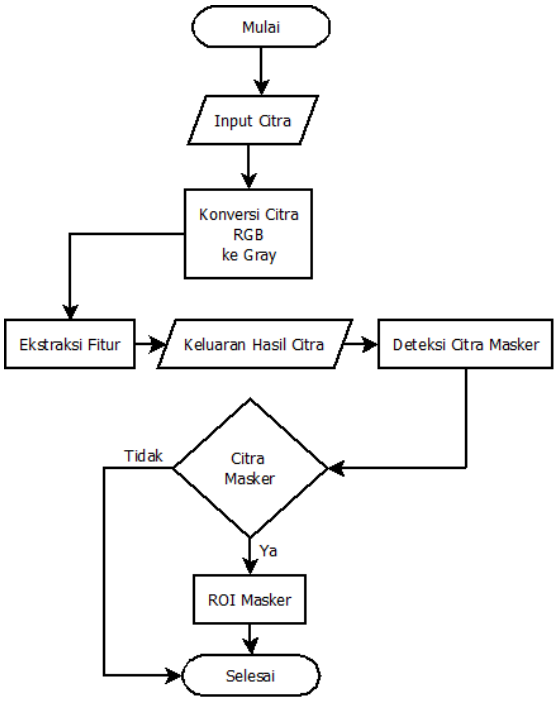

Figure 5. Flowchart of the process on the system

2.8 Tools and Materials

The equipment used is an Acer Inside E14 laptop, a Xiaomi Pocophone F1 smartphone camera and a USB connector. Supporting tools for this system are pycharm edition 2021 as well as python 3.8. The mask detection analysis tool using the Haar Cascade and Naive Bayes methods can be seen in Table 1.

TABLE 2
EQUIIPMENT & TOOLS

No.	Equipment & Tool	Description
1.	Laptop Acer Inspire	*Windows 10*, 64 Bit, 6 GB RAM
2.	Xiaomi Pocophone F1	108 MP camera
3.	*PyCharm Community edition* 2022	*Tools Coding*
4.	*Python* 3.8 *module docs* 8	*Interpreter*
5.	3 jenis masker	Blue, White, Black

III. RESULT AND DISCUSSION

The author of this study uses a set of datasets which are divided into two positive and negative variables. Positive is a collection of photos with masks and negatives without masks. The dataset is obtained from the Kaggle dataset. Two methods were also used in this study Haar Cascade and Naive Bayes. The datasets are shown in Figures 6 and 7. The number of datasets for testing and training and respondents.

Figure 6. Positive dataset set using masks

Figure 7. A collection of negative datasets that do not use masks

This study was conducted to classify face recognition using masks as well as faces that do not use masks with the code "CascadeClassifier". In path code = 'myhaar.xml', this stage is the process of reading the image after the image is done with training data to recognize and enter the image from the camera or an image extraction can be equated to the image in the dataset using the haar cascade algorithm that has gone through the training process. The source code for calling the dataset is shown in Figure 8.

```
face_cascade = cv2.CascadeClassifier('myhaar.xml')
cap = cv2.VideoCapture(0)
```

Figure 8. Data Path dataset

3.1 System Test

First, a normal test of the source code in python with a laptop webcam is carried out in real-time. Figure 9 shows the image results after processing in the system, showing that the system works by marking the face with the mask.

Figure 9. Image result system

3.2 Comparison of System Testing

The system will be compared with two different classification methods, namely Cascade and Naïve Bayes. The test data were prepared with 23 training data for Naïve Bayes, 15 for Haar Cascade and three lighting conditions bright (1300 Lux), normal (400 Lux), and dark (50 Lux). Tables 3 and 5 are the results of system testing using 2 methods.

No.	Mask Color	Lighting Conditions	Lux	Result
1.	White	Bright	1300	Detected
2.	White	Normal	400	Detected
3.	White	Dark	50	Not Detected
4.	Blue	Bright	1300	Detected
5.	Blue	Normal	400	Detected
6.	Blue	Dark	50	Not Detected
7.	Black	Bright	1300	Detected

8.	Black	Normal	400	Detected

TABLE 3
TESTING WITH THE CASCADE METHOD

No.	Mask Color	Lighting Conditions	Lux	Result
9.	Black	Dark	50	Not Detected
10.	Light Green	Bright	1300	Detected
11.	Light Green	Normal	400	Detected
12.	Light Green	Dark	50	Not Detected
13.	Red	Bright	1300	Detected
14.	Red	Normal	400	Detected
15.	Red	Dark	50	Not Detected

TABLE 4
CASCADE METHOD CONFUSION MATRIX

Actual	Prediction	
	P	N
P	10	5
N	0	0

The confusion matrix in Table 4 shows the TP of the Cascade classification model of 10 and the TN of 5 with an accuracy of 100% with equation (5).

$$accuracy = \frac{10+5}{10+5+0+0} \quad (4)$$

$$\frac{15}{15} \times 100 = 100\ \% \quad (5)$$

TABLE 5
TESTING WITH NAÏVE BAYES METHOD

No.	Mask Color	Lighting Conditions	Lux	Result
1.	White	Bright	1300	Detected
2.	White	Normal	400	Detected
3.	White	Dark	50	Not Detected
4.	Light Green	Bright	1300	Detected
5.	Blue	Normal	400	Detected
6.	Blue	Dark	50	Not Detected
7.	Black	Bright	1300	Detected
8.	Black	Normal	400	Detected
9.	Green	Dark	50	Not Detected
10.	Black	Normal	1300	Not Detected
11.	Light Green	Normal	400	Detected
12.	Light Blue	Normal	50	Detected
13.	Red	Normal	400	Not Detected
14.	Blue	Bright	1300	Not Detected
15.	Dark Blue	Dark	50	Detected
16.	Red	Bright	1300	Not Detected
17.	Black	Dark	50	Not Detected
18.	Red	Dark	50	Not Detected
19.	Light Green	Dark	50	Detected
20.	Black	Dark	50	Not Detected

21.	White	Dark	59	Detected
22.	Black	Bright	1300	Not Detected
23.	Blue	Bright	1300	Not Detected

TABLE 6
NAÏVE BAYES METHOD CONFUSION MATRIX

		Prediction	
Actual		P	N
	P	11	12
	N	7	0

The confusion matrix in Table 2 shows that the TP of the nave Bayes classification model is 11, the TN is 12, and the FP is 7. The accuracy obtained is 76.6% with equation 7.

$$accuracy = \frac{11+12}{11+12+7+0} \quad (6)$$

$$\frac{23}{30} \times 100 = 76{,}6\ \% \quad (7)$$

IV. CONCLUSION

Analysis of mask detection using the haar cascade and nave Bayes methods results that the haar cascade classifier method can work superior in terms of accuracy compared to the nave Bayes method. The system can recognize whether a person is using a mask or not by marking facial objects with Region of Interest. The system cannot work properly when the light intensity is in dark conditions. The position of the object, the distance of the object, and the background also affect the level of effectiveness of the system. This study obtained the results of the accuracy of the haar cascade classifier method of 100% and the nave Bayes method of 76.6%. For further research, it is expected to add sample data and use different methods such as the Histogram of Oriented Gradient or use a deep learning system.

REFERENCE

[1] M. Anwari, "Training On Making Face Shields That Are Easy And Economical As A Tool To Protect Themselves From Covid-19," pp. 144–147, 2021, [Online]. Available: http://proceeding.mbunivpress.or.id/index.php/bamara

[2] M. Gerakan, P. Juta, and V. I. Abdullah, "Peran Perempuan Dalam Pemutusan Mata Rantai Covid - 19," *J. Abdi Mas TPB*, vol. 3, no. January 2020, pp. 94–99, 2021, [Online]. Available: www.abdimastpb.unram.ac.id

[3] A. Thariq and R. Y. Bakti, "Sistem Deteksi Masker dengan Metode Haar Cascade pada Era New Normal COVID-19," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 241, 2021, doi: 10.26418/justin.v9i2.44309.

[4] V. K. Pandey, V. K. Gupta, and S. Kumar, "Face Mask Detection Using Convolutional Neural Network," *Proc. - 2021 3rd Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2021*, pp. 951–954, 2021, doi: 10.1109/ICAC3N53548.2021.9725689.

[5] Ahmadi, A. C. Saputra, and A. Lestari, "Rancang Bangun Aplikasi New Normal Covid- 19 Deteksi Penggunaan Masker," *J. Teknol.*

Inf., vol. 15, no. 2, pp. 199–209, 2021, [Online]. Available: https://e-journal.upr.ac.id/index.php/JTI/article/view/3291/2706

[6] I. Putra and I. Dana, "DAMPAK COVID – 19 TERHADAP PEREKONOMIAN INDONESIA," *E-Jurnal Manaj. Univ. Udayana*, vol. 5, no. 11, p. 249101, 2016.

[7] Y. Ibrahim, E. Okafor, and B. Yahaya, "Optimization of RBF-SVM hyperparameters using genetic algorithm for face recognit," *Niger. J. Technol.*, vol. 39, no. 4, pp. 1190–1197, 2021, doi: 10.4314/njt.v39i4.27.

[8] A. S. Prihantana and S. S. Wahyuningsih, "Hubungan Pengetahuan dengan Tingkat Kepatuhan Pengobatan pada Pada Pasien Tuberkulosis di RSUD dr. Soehadi Prijonegoro Sragen," *Farm. Sains dan Prakt.*, vol. II, no. 1, p. 47, 2016, [Online]. Available: http://journal.unimma.ac.id/index.php/pharmacy/article/view/188%0Ahttps://journal.unimma.ac.id/index.php/pharmacy/article/download/188/135/

[9] D. Wulandari, "Analisis Faktor-Faktor yang Berhubungan dengan Kepatuhan Pasien Tuberkulosis Paru Tahap Lanjutan Untuk Minum Obat di RS Rumah Sehat Terpadu Tahun 2015," *J. Adm. Rumah Sakit*, vol. 2, no. 1, pp. 17–28, 2015.

[10] T. Septiana, N. Puspita, M. Al Fikih, and N. Setyawan, "Face Mask Detection Covid-19 Using Convolutional Neural Network (Cnn)," *Semin. Nas. Teknol. dan Rekayasa 2020*, vol. 3, pp. 27–32, 2020.

[11] M. A. Al-masni *et al.*, "Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 1230–1233. doi: 10.1109/EMBC.2017.8037053.

[12] S. Varma and M. Sreeraj, "Object detection and classification in surveillance system," in *2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Dec. 2013, pp. 299–303. doi: 10.1109/RAICS.2013.6745491.

[13] N. Setyawan, D. Nur Fajar, and K. Hidayat, "Perencanaan Jalur Robot Sepak Bola Ummiros Menggunakan Algoritma a*," *Semin. Nas. Teknol. dan Rekayasa (SENTRA)2019*, pp. 2527–6042, 2019.

[14] Sik-Ho Tsang, "Review: SSD — Single Shot Detector (Object Detection)," 2018. https://towardsdatascience.com/review-ssd-single-shot-detector-object-detection-851a94607d11 (accessed Mar. 14, 2021).

[15] A. Priadana and M. Habibi, "2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT 2019) : 13-15 March 2019, Platinum Adisucipto Yogyakarta Hotel & Conference Center.," *2019 Int. Conf. Artif. Intell. Inf. Technol.*, pp. 6–9, 2019.

[16] D. R. Tobergte and S. Curtis, "Practical Python and OpenCV(3rd)," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.

[17] M. Syarif and Wijanarto, "Deteksi Kedipan Mata Dengan Haar Cascade Classifier Dan Contour Untuk Password Login," *Techno.com*, vol. 14, no. 4, pp. 242–249, 2015.

[18] J. W. Shipman, "Tkinter 8.5 reference: a GUI for Python," *Computer (Long. Beach. Calif).*, pp. 1–118, 2013, [Online]. Available: tcc-doc@nmt.edu

[19] G. Aprilian Anarki, K. Auliasari, and M. Orisa, "Penerapan Metode Haar Cascade Pada Aplikasi Deteksi Masker," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 179–186, 2021, doi: 10.36040/jati.v5i1.3214.

[20] R. A. Asmara, B. S. Andjani, U. D. Rosiani, and P. Choirina, "Klasifikasi Jenis Kelamin Pada Citra Wajah Menggunakan Metode Naive Bayes," *J. Inform. Polinema*, vol. 4, no. 3, p. 212, 2018, doi: 10.33795/jip.v4i3.209.

[21] M. Kim, D. Lee, and K. Y. Kim, "System architecture for real-time face detection on analog video camera," *Int. J. Distrib. Sens. Networks*, vol. 2015, 2015, doi: 10.1155/2015/251386.

[22] P. Purwanto, B. Dirgantoro Ir, and A. S. Nugroho Jati, "Implementasi Face Identification Dan Face Recognition Pada Kamera Pengawas Sebagai Pendeteksi Bahaya Implementation of Face Identification and Face Recognition on Security Camera As Threat Detector," vol. 2, no. 1, p. 718, 2015.

[23] Jet Brains, "PyCharm: the Python IDE for Professional Developers by JetBrains." https://www.jetbrains.com/pycharm/. [Accessed: 25- Jun-2020 (accessed Mar. 15, 2022).

[24] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, E. Purnamasari, D. Palupi Rini, P. Studi Magister Ilmu Komputer, F. Ilmu Komputer, and U. Sriwijaya Palembang, "Seleksi Fitur menggunakan Algoritma Particle Swarm Optimization pada Klasifikasi Kelulusan Mahasiswa dengan Metode Naive Bayes," *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 469–475, 2017.

[25] A. Kesumawati and D. T. Utari, "Predicting patterns of student graduation rates using Naïve bayes classifier and support vector machine," *AIP Conf. Proc.*, vol. 2021, no. October 2018, 2018, doi: 10.1063/1.5062769.

[26] S. Modak And A. C. Mondal, "A Comparative Study Of Classifiers' Performance For Gender Classification," *Int. J. Innov. Res. Comput. Commun. Eng.*, Vol. 2, No. 5, Pp. 4214–4222, 2014.