

# Face Mask Recognition by the Viola-Jones Method Using Fuzzy Logic

## **Serhiy Balovsyak\***

Yuriy Fedkovych Chernivtsi National University, Chernivtsi, 58012, Ukraine

E-mail: [s.balovsyak@chnu.edu.ua](mailto:s.balovsyak@chnu.edu.ua)

ORCID iD: <https://orcid.org/0000-0002-3253-9006>

\*Corresponding Author

## **Oleksandr Derevyanchuk**

Yuriy Fedkovych Chernivtsi National University, Chernivtsi, 58012, Ukraine

E-mail: [o.v.derevyanchuk@chnu.edu.ua](mailto:o.v.derevyanchuk@chnu.edu.ua)

ORCID iD: <https://orcid.org/0000-0002-3749-9998>

## **Vasyl Kovalchuk**

State institution «Scientific and methodological center for higher and pre-higher education», Kyiv, 03151, Ukraine

E-mail: [v.i\\_kovalchuk@ukr.net](mailto:v.i_kovalchuk@ukr.net)

ORCID iD: <https://orcid.org/0000-0002-5006-573X>

## **Hanna Kravchenko**

High State Educational Establishment «Chernivtsi transport college», Chernivtsi, 58000, Ukraine

E-mail: [hannakravchenko81@gmail.com](mailto:hannakravchenko81@gmail.com)

ORCID iD: <https://orcid.org/0009-0004-7609-0345>

## **Maryna Kozhokar**

Yuriy Fedkovych Chernivtsi National University, Chernivtsi, 58012, Ukraine

E-mail: [m.kozhokar@chnu.edu.ua](mailto:m.kozhokar@chnu.edu.ua)

ORCID iD: <https://orcid.org/0000-0003-1658-3549>

Received: 03 February, 2024; Revised: 13 March, 2024; Accepted: 16 April, 2024; Published: 08 June, 2024

**Abstract:** In the work, the software implementation of the face mask recognition system using the Viola-Jones method and fuzzy logic is performed. The initial images are read from digital video cameras or from graphic files.

Detection of face, eye and mouth positions in images is performed using appropriate Haar cascades. The confidence of detecting a face and its features is determined based on the set parameters of Haar cascades.

Face recognition in the image is performed based on the results of face and eye detection by means of fuzzy logic using the Mamdani knowledge base. Fuzzy sets are described by triangular membership functions. Face mask recognition is performed based on the results of face recognition and mouth detection by means of fuzzy logic using the Mamdani knowledge base. Comprehensive consideration of the results of different Haar cascades in the detection of face, eyes and mouth allowed to increase the accuracy of recognition face and face mask.

The software implementation of the system was made in Python using the OpenCV, Scikit-Fuzzy libraries and Google Colab cloud platform. The developed recognition system will allow monitoring the presence of people without masks in vehicles, in the premises of educational institutions, shopping centers, etc. In educational institutions, a face mask recognition system can be useful for determining the number of people in the premises and for analyzing their behavior.

**Index Terms:** Artificial Intelligent, Educational institutions, Fuzzy Logic, Haar cascade, Programming, Python, Viola-Jones method.

## 1. Introduction

Currently, there is a need for the development of reliable, fast and easy-to-use face mask recognition systems. Such recognition systems will allow monitoring the presence of people without masks in vehicles, in the premises of educational institutions, shopping centers, etc. Detecting people with and without masks is important for security purposes, in particular, in quarantine conditions. In addition, the use of a face mask recognition system in educational institutions will allow determining the number of people in the room and their spatial location, taking into account the results of recognition in the educational process.

Obtaining images of faces using digital video cameras is relatively simple. However, the process of recognizing faces in images with or without masks is much more complicated. This is due to different lighting conditions, different distances of people from the video cameras, people's turns relative to the video cameras, as well as due to the presence of additional objects in the frame: glasses, hats, etc. Therefore, various methods using Artificial Intelligence (AI) technologies are currently used to recognize face masks. The process of recognizing a mask on a face image is usually divided into two stages: 1) localization of the face area; 2) detection or non-detection of the mask on the face.

One of the common methods of recognizing images of faces and masks on faces is the use of Artificial Neural Networks (ANN), in particular, Convolutional Neural Networks (CNN), which are effective specifically for image processing [1]. The effectiveness of CNN for image recognition is demonstrated in [2] and training of CNN with the ImageNet architecture was carried out. It is shown that even for a large training sample (more than 1 million images), CNN training can be performed in an acceptable time. The possibility of effective recognition of a mask on a face using CNN with VGG-16, AlexNet and ResNet-50 architectures has been shown by researchers [3]. The detection of a mask on a face by CNN with the FMDNet architecture was performed by the authors [4]; for the used dataset of images, a recognition accuracy of 99% was obtained. In the work [5] considered the possibilities of face detection in images using CNN, where a person's face was highlighted in the image as a Region of Interest (ROI). A Different Scales Face Detector (DSFD) based on Faster R-CNN was used, which allow to detect faces with fairly high accuracy even at small scales. In the process of training CNN, the dataset can be divided into minibatches, which ensures high accuracy of face recognition (96.6% for Yale Datasets) [6]. When training a CNN, the input data (images) are divided into a training dataset and a validation dataset, which is designed to prevent overtraining.

Images of faces can be recognized by CNN not only in the visible range, but also in the infrared. The researchers in [7] showed that analysis of Thermal Images allows accurate Facial Expression Recognition (FER). However, infrared imaging requires special video cameras. Analysis of a person's emotional state using FER is also performed using Scale Invariant Feature Transform (SIFT) [8]. Such an analysis of a person's emotional state is especially useful when organizing the educational process, for example, for assessing students' concentration and attention. By CNN it is possible to recognize not only faces with high accuracy [9, 10], but also masks on faces. However, this requires extensive training using appropriate datasets.

Face images in most cases differ in texture from other areas of the images, which was used for face recognition in [12] for building a biometric system. Face recognition based on their texture was performed using Local Binary Pattern (LBP) and Gray Level Co-occurrence Matrix (GLCM), while three classifiers were used: Euclidean distance, Manhattan distance and Cosine distance. In the paper [13] the Discrete Complex Fuzzy Transform (DCFT) method was used to recognize face images, which consists of the stages of histogram extraction, peak points of histogram calculation and image construction. To create images, polynomial and smith fuzzy sets are used. In the classification phase, support vector machine (SVM) and k Nearest Neighborhood (KNN) classifiers are used. Due to the descriptor SURF (Speed Up Robust Features) [14] and the descriptors Histogram of Directional Gradient (HDG) and Histogram of Directional Gradient Generalized (HDGG) [15] high accuracy of image recognition is ensured. However, the proposed methods using textures, special descriptors (SURF, HDG, HDGG and DCFT) [12, 13, 14] are designed for recognizing faces, not masks on faces.

If the experimental images contain a significant level of noise (for example, Gaussian or impulse), then before face recognition, pre-processing of the images is performed by filtering noise in the spatial or frequency domains [16, 17]. Filter parameters are determined taking into account the noise level in the image [19, 20]. Image pre-processing also consists of segmentation [21], contour selection [22, 23] and contrast enhancement [24, 25, 26], which makes it possible to increase the accuracy of face recognition using CNN [27].

Detection of face regions in images is often performed by the Viola-Jones method [28], which uses Haar cascades and is one of the best in terms of accuracy and processing speed. Also, this detector has a low probability of false identification of a person. There are Haar cascades designed for detecting frontal face images and profile face images, as well as for detecting facial features (eyes, mouth). Complex use of the results of different Haar cascades potentially allows to increase the accuracy of face recognition. For example, if a face, eyes and mouth are detected simultaneously in a certain area of the image, then a face without a mask is also recognized in such an area with high confidence. An important advantage of the Viola-Jones method is the use of trained Haar cascades, which avoids a complex and lengthy training procedure (as in the case of ANN). Therefore, in the paper proposes to use the Viola-Jones method to detect (localize) faces and their features, due to which it is possible to solve the urgent task of building a system for recognizing images of masks in faces. However, due to the difficulty of detecting faces and their features in real images,

the detection confidence value can take not only the minimum or maximum, but also an arbitrary intermediate value. For this reason, the results of detecting faces and their features in images are described by means of fuzzy logic [29, 30].

## 2. Theoretical Foundations of Face Mask Recognition Using the Viola-Jones Method and Fuzzy Logic

The face mask recognition system [31, 32, 33] uses the Viola-Jones method to detect (localize) the face and its features (eyes, mouth), and artificial intelligence tools [34, 35] are used for analysis detection results. Fuzzy logic is used as a means of artificial intelligence [29, 30], which allows to quantitatively assess the confidence of face detection and its features.

The Viola-Jones method is based on the use of Haar cascades (detectors) [36]. Haar detectors consist of contiguous rectangular areas of the image. They are positioned on the image, then the pixel intensities in different areas are summed, after which the difference between the sums is calculated. Such a difference is the value of a certain Haar feature (signature). At the stage of object detection using the Viola-Jones method, a window of a fixed size moves across the image, and for each area of the image over which the window passes, the Haar feature is calculated. The presence or absence of an object in the window is determined by the difference between the value of the feature and the set threshold. A significant number of Haar features are required for accurate object detection, so in the Viola-Jones method, the features are organized into a cascade classifier. The simplest rectangular Haar feature can be defined as the difference of the pixel sums of two adjacent areas inside a rectangle, which can occupy different positions and scales in the image. This type of features is called 2-rectangular. Similarly, 3-rectangular and 4-rectangular features are defined. In the original version of the Viola-Jones algorithm, only detectors (primitives) without turns were used. Later, 45-degree tilt detectors and asymmetrical configurations were proposed.

The result of the Viola-Jones method is the position of the objects in the image, which are described by the following parameters:  $x$  and  $y$  are the coordinates of the upper right corner of the rectangle (which limits the desired object, for example, a face),  $w$  is the width,  $h$  is the height of the rectangle. To localize objects in images, the Viola-Jones method uses a set of discrete scale values for the scanning window.

When detecting objects (for example, faces) using the Viola-Jones method, a certain value of the minNeighbors parameter is pre-set (in many cases, the value of the parameter is set by default). The minNeighbors parameter, which specifies how many neighbors each candidate rectangle (detected object area) must have in order to save it (the value of minNeighbors varies from 1 to 6). For each face region, the face detection confidence value  $D_f$  (which ranges from 1 to 6) is calculated as the maximum value of minNeighbors for which the face is still detected (larger  $D_f$  means more confident face recognition). Similarly, for each eye region, the eye detection confidence value  $D_e$  (which ranges from 0 to 5) is calculated as the maximum value of minNeighbors at which an eye is still detected (a larger  $D_e$  means an eye is detected with greater confidence). For each mouth region, the mouth detection confidence value  $D_s$  (which ranges from 0 to 5) is calculated as the maximum value of minNeighbors at which a mouth is still detected (a higher  $D_s$  means a more confident mouth is detected).

In the source images, the Viola-Jones method detects the face, eye and mouth regions, which allows to recognize faces without a mask and with a mask on the face with high reliability. A face without a mask is recognized with high confidence if the face and mouth are detected with high confidence. A mask is recognized with high confidence if the face and eyes are detected with high confidence and the mouth with low confidence.

Face recognition based on detected face and eye regions is performed by a fuzzy system whose input parameters are face detection confidence  $D_f$  and eye detection confidence  $D_e$ , and the output parameter is face recognition confidence  $D_{fr}$  (which lies in the range of 1 to 6).

Mamdani fuzzy knowledge base # 1 [30] has been developed for face recognition, which contains 36 rules, for example:

Rule # 1; If DfL = Low and DeL = Low, then DfrL = Low.

Rule # 6; If DfL = Below Average and DeL = Below Average, then DfrL = Below Average.

Rule # 14; If DfL = Above Average and DeL = Below Average, then DfrL = Average.

Rule # 20; If DfL = Average and DeL = Average, then DfrL = Above Average.

Rule # 30; If DfL = Below High and DeL = Above Average, then DfrL = Below High.

Rule # 33; If DfL = High and DeL = High, then DfrL = High.

In the rules of the knowledge base, the notation DfL describes the linguistic variable "Confidence of face detection", DeL – "Confidence of eye detection", DfrL – "Confidence of face recognition taking into account the eyes".

The terms of the linguistic variable DfL «Confidence of face detection» are Low, Below Average, Average, Above Average, Below High, High (which are numbered from 1 to 6), and the numerical linguistic variable is described by the face detection confidence  $D_f$  (carrier of fuzzy set). Analogous terms describe the linguistic variables DeL "Confidence of eye detection" (with numeric input variable  $D_e$ ) and DfrL "Confidence of face recognition taking into account the eyes" (with numeric output variable  $D_{fr}$ ).

The terms of the linguistic variable DfL "Confidence of face detection" are described by fuzzy sets with triangular membership functions  $\mu_{f1}(D_f)$ ,  $\mu_{f2}(D_f)$ ...  $\mu_{f6}(D_f)$  and the carrier  $D_f$  (Fig. 1). Similarly, the terms of the linguistic variable

DeL "Confidence of eye detection" are created, which are described by fuzzy sets with triangular membership functions:  $\mu_{e1}(D_e), \dots, \mu_{e6}(D_e)$ . The terms of the linguistic variable DfrL "Confidence of face recognition taking into account the eyes" are described by fuzzy sets with triangular membership functions:  $\mu_{fr1}(D_{fr}), \dots, \mu_{fr6}(D_{fr})$ .

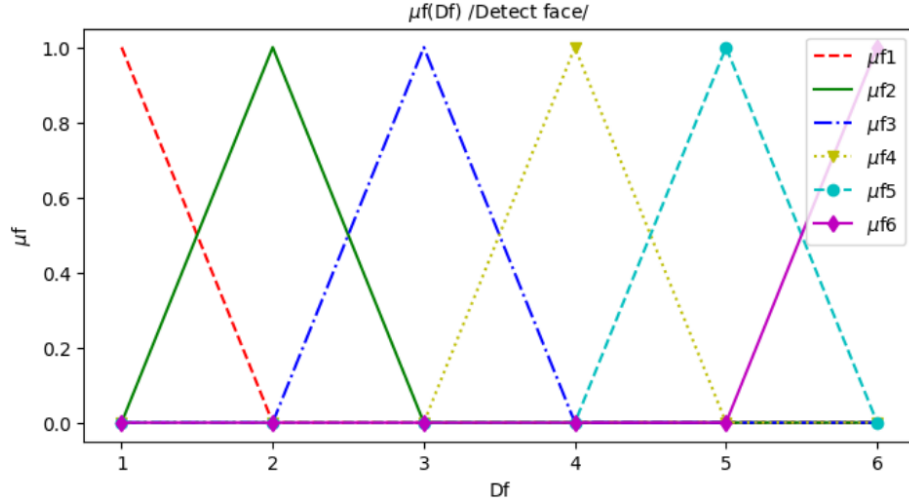


Fig. 1. Fuzzy membership functions of the linguistic variable DfL "Confidence of face detection":  $\mu_{f1}$  (red graph) – the confidence of face detection is Low;  $\mu_{f2}$  – Below Average,  $\mu_{f3}$  – Average;  $\mu_{f4}$  – Above Average;  $\mu_{f5}$  – Below High;  $\mu_{f6}$  – High

An initial image is fed to the inputs of the face recognition system, on which faces are detected with  $D_{fC}$  confidence and eyes are detected with  $D_{eC}$  confidence using the Viola-Jones method.

Based on the obtained values of  $D_{fC}$  and  $D_{eC}$ , the values of the membership functions  $\mu_{f(m)}(D_{fC})$ , where  $m = 1, \dots, Q_m$  ( $Q_m = 6$ ) and  $\mu_{e(n)}(D_{eC})$ , where  $n = 1, \dots, Q_n$  ( $Q_n = 6$ ) were calculated (for the corresponding terms). Based on the fuzzy knowledge base, the values of membership functions  $\mu_{fr(k)}(D_{frC})$ , where  $k = 1, \dots, Q_k$  ( $Q_k = 6$ ), were calculated according to the formula

$$\mu_{fr(k)}(D_{frC}) = \min(\mu_{f(m)}(D_{fC}), \mu_{e(n)}(D_{eC})), \quad (1)$$

where  $m$  is number of the membership function  $\mu_f$ ,  $n$  is number of the membership function  $\mu_e$ ,  $k$  is number of the membership function  $\mu_{fr}$ .

Membership function numbers are determined according to the rules of the fuzzy knowledge base. For example, for rule # 14: DfL = Above average ( $m = 4$ ), DeL = Below average ( $n = 2$ ), DfrL = Average ( $k = 3$ ), the values of membership functions  $\mu_{fr3}(D_{frC})$  were calculated

$$\mu_{fr3}(D_{frC}) = \min(\mu_{f4}(D_{fC}), \mu_{e2}(D_{eC})). \quad (2)$$

After that, the composition of the fuzzy membership functions  $\mu_{fr(k)}(D_{fr})$ , where  $k = 1, \dots, Q_k$ , is performed, as a result of which the fuzzy set  $\mu_{fUa}(D_{fr})$  is calculated according by formula

$$\mu_{fUa}(D_{fr}) = \max(\mu_{fr1}(D_{fr}), \dots, \mu_{fr(k)}(D_{fr}), \dots, \mu_{fr6}(D_{fr})). \quad (3)$$

On the basis of the fuzzy set  $\mu_{fUa}(D_{fr})$ , its defuzzified value  $D_{frC}$  is calculated by the method of the gravity center according to the formula

$$D_{frC} = \frac{\sum_{k=1}^{Q_k} D_{fr}(k) \cdot \mu_{fUa}(D_{fr}(k))}{\sum_{k=1}^{Q_k} \mu_{fUa}(D_{fr}(k))}. \quad (4)$$

where  $D_{fr}(k)$  are discrete values of the  $D_{fr}$  parameter (from 1 to 6).

Thus, based on the parameters of the detected face, eyes and mathematical model (1-4), the face recognition (with confidence  $D_{frC}$ ) is performed by fuzzy logic. During such fuzzy inference, it is taken into account that if eyes are detected within a face region with high confidence, then such a face is recognized with high confidence. Conversely, if eyes are detected within a face region with low confidence, then such a face is recognized with low confidence.

Face mask recognition based on recognized face regions (using fuzzy inference) and detected mouth regions (obtained by the Viola-Jones method) is performed by a fuzzy system whose input parameters are face recognition confidence  $D_{fr}$  and mouth detection confidence  $D_s$ , and the output parameter is mask recognition confidence  $D_m$  (which lies in the range from 0 to 5).

Mamdani fuzzy knowledge base # 2 has been developed for mask recognition, which contains 36 rules, for example:

- Rule # 1; If DfrL = Low and DsL = Medium, then DmL = Low.
- Rule # 6; If DfrL = High and DsL = Below Average, then DmL = Below Average.
- Rule # 14; If DfrL = Low and DsL = Low, then DmL = Medium.
- Rule # 20; If DfrL = Medium and DsL = Low, then DmL = Above Medium.
- Rule # 30; If DfrL = Above Medium and DsL = Low, then DmL = Below High.
- Rule # 36; If DfrL = Lower High and DsL = Low, then DmL = High.

In the rules of the knowledge base, the notation DfrL describes the linguistic variable "Confidence of face recognition taking into account the eyes", DsL – "Confidence of mouth detection", DmL – "Confidence of mask recognition".

The linguistic variable terms are the values Low, Below Average, Medium, Above Average, Below High, High (which are numbered from 1 to 6). For the linguistic variable DsL "Confidence of mouth detection" the input variable is  $D_s$ . For the linguistic variable DmL "Confidence of mask recognition" the output variable is  $D_m$ .

The terms of the linguistic variable DsL "Confidence of mouth detection" are described by fuzzy sets with triangular membership functions  $\mu_{s1}(D_s), \mu_{s2}(D_s) \dots \mu_{s6}(D_s)$  and the carrier  $D_s$ . Similarly, the terms of the linguistic variable DmL "Confidence of mask recognition" are created, which are described by fuzzy sets with triangular membership functions  $\mu_{m1}(D_m), \dots, \mu_{m6}(D_m)$ .

Mouth regions are detected on the input image using the Viola-Jones method with  $D_{sC}$  confidence. Based on the predefined face recognition confidence  $D_{fr}$  and mouth detection confidence  $D_{sC}$ , the values of membership functions  $\mu_{fr(m)}(D_{frC})$ , where  $m = 1, \dots, Q_m$  ( $Q_m = 6$ ) and  $\mu_{s(n)}(D_{sC})$ , where  $n = 1, \dots, Q_n$  ( $Q_n = 6$ ) were calculated (for the corresponding terms). Based on the rules of the fuzzy knowledge base, the values of the membership functions  $\mu_{m(k)}(D_{mC})$ , where  $k = 1, \dots, Q_k$  ( $Q_k = 6$ ), are calculated according to the formula

$$\mu_{m(k)}(D_{mC}) = \min(\mu_{fr(m)}(D_{frC}), \mu_{s(n)}(D_{sC})) , \quad (5)$$

where  $m$  is number of the membership function  $\mu_{fr}$ ,  $n$  is number of the membership function  $\mu_s$ ,  $k$  is number of the membership function  $\mu_m$ .

Membership function numbers are determined according to the rules of the fuzzy knowledge base. For example, for rule #20: DfrL= Medium ( $m = 3$ ), DsL = Low ( $n = 1$ ), DmL = Above Medium ( $k = 4$ ), the values of membership functions  $\mu_{m4}(D_{mC})$  were calculated

$$\mu_{m4}(D_{mC}) = \min(\mu_{fr3}(D_{frC}), \mu_{s1}(D_{sC})) . \quad (6)$$

After that, the composition of the fuzzy membership functions  $\mu_{m(k)}(D_m)$ , where  $k = 1, \dots, Q_k$ , is performed, as a result of which the fuzzy set  $\mu_{mUa}(D_m)$  is calculated according by formula

$$\mu_{mUa}(D_{fr}) = \max(\mu_{m1}(D_m), \dots, \mu_{m(k)}(D_m), \dots, \mu_{m6}(D_m)) . \quad (7)$$

On the basis of the fuzzy set  $\mu_{mUa}(D_m)$ , its defuzzified value  $D_{mC}$  is calculated by the method of the gravity center according to the formula

$$D_{mC} = \frac{\sum_{k=1}^{Q_k} D_m(k) \cdot \mu_{mUa}(D_m(k))}{\sum_{k=1}^{Q_k} \mu_{mUa}(D_m(k))} . \quad (8)$$

where  $D_m(k)$  are discrete values of the  $D_m$  parameter.

Thus, on the basis of the parameters of the recognized face, the detected mouth and the mathematical model (5-8), the mask recognition (with confidence  $D_{mC}$ ) is performed by fuzzy logic. During such fuzzy inference, it is taken into account that if a mouth is detected within a face area with high confidence, then a mask is recognized with low confidence on such a face. Conversely, if a mouth is detected within a face area with low confidence, then a mask on such a face is recognized with high confidence.

### 3. Software Implementation of Face Mask Recognition

The fuzzy face and face mask image recognition system was software implemented in Python using the Google Colab cloud service (in the Jupyter Notebook) [31, 32, 33] based on mathematical models (1-4) and (5-8). Fuzzy triangular membership functions for linguistic variables are generated by the "trimf" method of the scikit-fuzzy library. The operation of the recognition system is described by the following algorithm (Fig. 2).

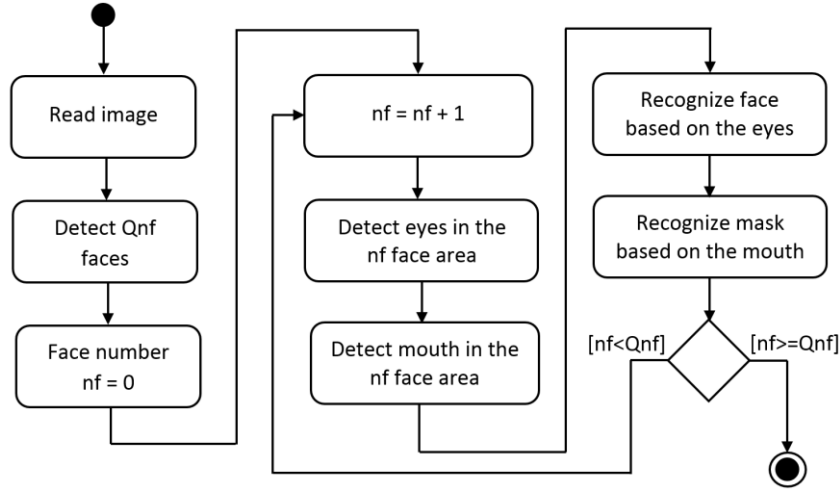


Fig. 2. Activity diagram of face and face mask image recognition system

The initial (input) data for the program are digital images which read from a video camera or from graphic files. Such images may contain masked or unmasked faces. The initial color image is read as a three-dimensional array  $f_{RGB}(i, k, c)$ , where  $i = 0, \dots, M-1$ ;  $k = 0, \dots, N-1$ ;  $M$  – image height in pixels,  $N$  – image width in pixels,  $c = 0, \dots, 2$  – color channel number (Red, Green, Blue). Color image  $f_{RGB}$  are converted to grey image  $f$ , which are processed by software as rectangular matrices  $f = (f(i, k))$ , where  $i = 0, \dots, M-1$ ,  $k = 0, \dots, N-1$ .

On the obtained images  $f$ , the Viola-Jones method detects  $Q_{nf}$  of rectangular areas of faces. The following Haar cascades [36] were used for detection in the image of the face, eyes, and mouth:

1. haarcascade\_frontalface\_default.xml (Detection of frontally turned faces).
2. haarcascade\_profileface.xml (Detection of faces in profile).
3. haarcascade\_eye.xml (Eye detection).
4. haarcascade\_smile.xml (Mouth or smile detection).

In the loop, all faces with numbers  $n_f$  from 1 to  $Q_{nf}$  are processed. To detect frontally rotated faces, the face\_cascade classifier is created based on the Haar cascade (using the opencv library imported into the program as cv2) by the command:

```
face_cascade = cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
```

Detection of areas of faces is performed by the command:

```
face_coord = face_cascade.detectMultiScale(f, scaleFactor, minNeighbors),
```

where  $f$  is the initial grey image;

scaleFactor is the parameter that determines how much the size of the scanning window is reduced for each image scale (for example, scaleFactor = 1.1);

minNeighbors is the parameter that specifies how many neighbors each candidate rectangle must have in order to save it (varies from 1 to 6).

For a detected rectangular area, 4 parameters are recorded in face\_coord:  $x$  – coordinates of the upper left corner of the area by width,  $y$  – coordinates of the upper left corner of the area by height,  $w$  – width of the area,  $h$  – height of the area. The face detection confidence  $D_f$  is calculated as the maximum value of minNeighbors at which a face is still detected.

Within the rectangular area of each face ( $h_f \times w_f$  pixels), eyes and mouth are detected by the Viola-Jones method using the corresponding Haar cascades. For each eye region, the value of eye detection confidence  $D_e$  is calculated, and for each mouth region, the value of mouth detection confidence  $D_s$  is calculated (similarly to face detection).

On the basis of the developed mathematical model (1-4), face recognition is performed taking into account the eyes using Mamdani fuzzy knowledge base #1. Similarly, on the basis of the mathematical model (5-8), a fuzzy inference is performed for mask recognition using Mamdani fuzzy knowledge base # 2.

The output data (results) of the software are the coordinates of the rectangles that correspond to the faces in the image, as well as the face recognition confidence  $D_{fC}$  (4) and mask recognition confidence  $D_{mC}$  (8) for each such rectangle.

#### 4. Results and Discussion

The developed software reads initial images from a video camera or from graphic files. Let's consider example # 1 of facial mask recognition using the developed program.

First, a face is detected in the image with high confidence  $D_{fC} = 6$  (Fig. 3), and then eyes are detected with also high confidence  $D_{eC} = 5$  (Fig. 4).

Therefore, the initial data of the face recognition system are as follows:  $D_{fC} = 6$ , the linguistic variable DfL "Confidence of face detection" takes the value "High" (Fig. 1);  $D_{eC} = 5$ , the linguistic variable DeL "Confidence of eye detection" takes the value "High". On the basis of the initial data, fuzzy inference is performed using the rules of the developed Mamdani knowledge base #1. At the same time, rule # 33 is fulfilled:

If DfL = High and DeL = High, then DfrL = High.

By defuzzifying the "High" value of the linguistic variable DfrL "Confidence of face recognition taking into account the eyes", a defuzzified value of face recognition confidence  $D_{frc} = 5.7$  (maximum value of 6) was obtained (Fig. 5, Fig. 6a). Therefore, the initial data of the mask recognition system is as follows:  $D_{frc} = 5.7$ , the linguistic variable DfrL "Confidence of face recognition taking into account the eyes" takes the values "Below High" and "High". The mouth is not detected in the image, so  $D_{sC} = 0$ , the linguistic variable DsL "Confidence of mouth detection" takes the value "Low". On the basis of the initial data, a fuzzy inference is performed using the rules of the developed knowledge base Mamdani # 2. At the same time, rule # 36 is fulfilled:

If DfrL = Lower High and DsL = Low, then DmL = High.

By defuzzifying the "High" value of the linguistic variable DmL "Confidence of mask recognition", a defuzzified value of mask recognition confidence  $D_{mC} = 4.7$  (with a maximum value of 5) (Fig. 6b) was obtained, which means the presence of a mask.

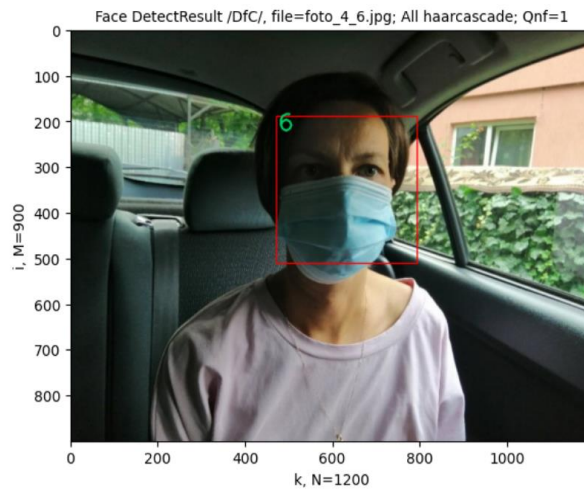


Fig. 3. Detection of faces in the image, face detection confidence  $D_{fC} = 6$  (maximum value is 6)

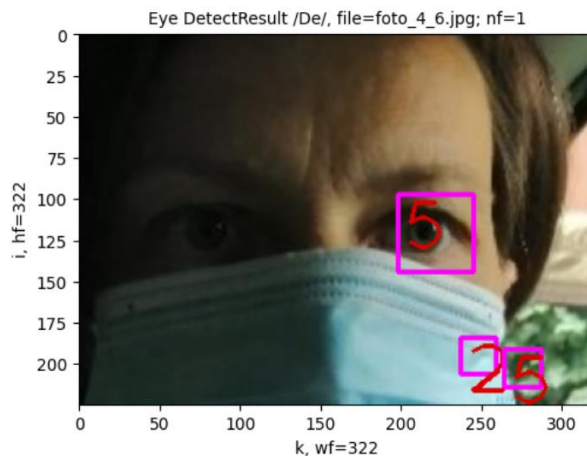


Fig. 4. Detection of eyes in the image, eye detection confidence  $D_{eC} = 5$  (maximum value is 5)

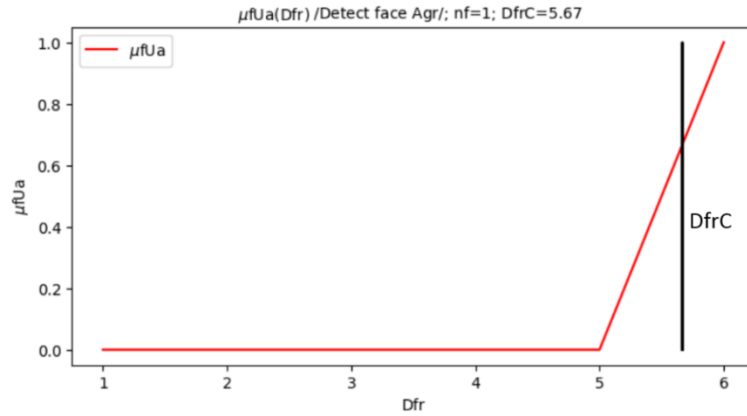


Fig. 5. Fuzzy inference for recognizing a face with number  $nf=1$ ; calculation of  $D_{frC}$  by aggregation of membership function  $\mu fUa$  using the center of gravity method ( $D_{frC}$  – face recognition confidence)

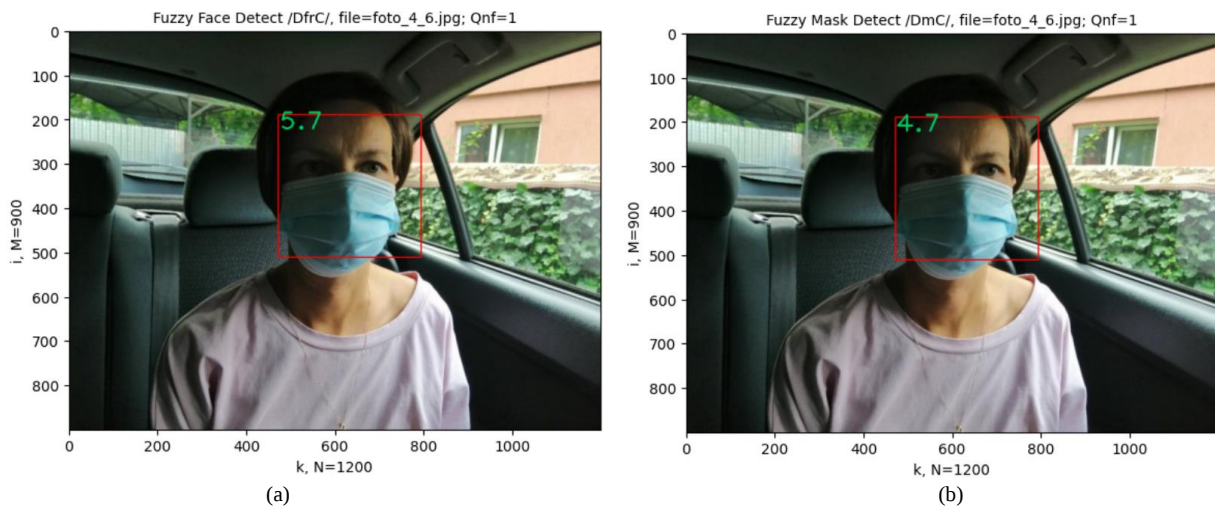


Fig. 6. Recognition of face with confidence  $D_{frC} = 5.7$  (a) and recognition of mask with confidence  $D_{mC} = 4.7$  (b)

Let's consider example # 2 of facial mask recognition using the developed system.

First, the face is detected in the image with high confidence  $D_{fC} = 6$  (Fig. 7), and then eyes are detected with confidence  $D_{eC} = 5$  (Fig. 8a) and the mouth is detected with confidence  $D_{sC} = 1$  (Fig. 8b).

Therefore, the initial data of the face recognition systems are as follows:  $D_{fC} = 6$ , the linguistic variable  $D_{fL}$  "Confidence of face detection" takes the value "High";  $D_{eC} = 5$ , the linguistic variable  $D_{eL}$  "Confidence of eye detection" takes the value "High". On the basis of the initial data, fuzzy inference is performed using the rules of the developed Mamdani knowledge base #1. At the same time, rule # 33 is fulfilled:

If  $D_{fL} = \text{High}$  and  $D_{eL} = \text{High}$ , then  $D_{frL} = \text{High}$ .

By defuzzifying the "High" value of the linguistic variable  $D_{frL}$  "Confidence of face recognition taking into account the eyes", a defuzzified value of face recognition confidence  $D_{frC} = 5.7$  (maximum value of 6) was obtained (Fig. 9a).

Therefore, the initial data of the mask recognition system are as follows:  $D_{frC} = 5.7$ , the linguistic variable  $D_{frL}$  "Confidence of face recognition taking into account the eyes" takes the values "Below High" and "High";  $D_{sC} = 1$ , the linguistic variable  $D_{sL}$  "Confidence of mouth detection" takes the value "Below Average". On the basis of the initial data, a fuzzy inference is performed using the rules of the developed knowledge base of Mamdani # 2. At the same time, rule # 6 is fulfilled:

If  $D_{frL} = \text{High}$  and  $D_{sL} = \text{Below Average}$ , then  $D_{mL} = \text{Below Average}$ .

By defuzzifying the "Below Average" value of the linguistic variable  $D_{mL}$  "Confidence of mask recognition", a defuzzified value of mask recognition confidence  $D_{mC} = 1.0$  (with a maximum value of 5) (Fig. 9b) was obtained, which means the absence of a mask.

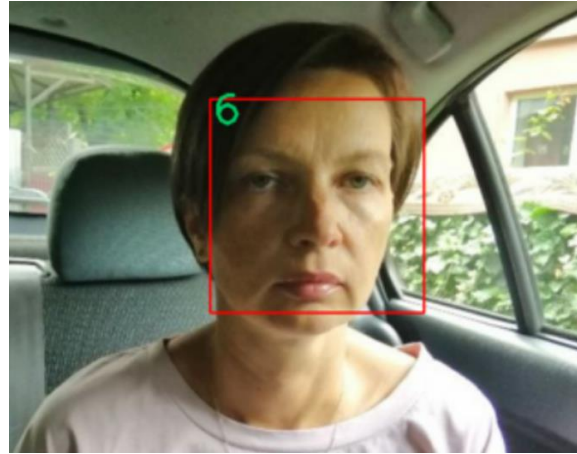


Fig. 7. Detection of faces in the image, face detection confidence  $D_{fC} = 6$  (maximum value is 6)

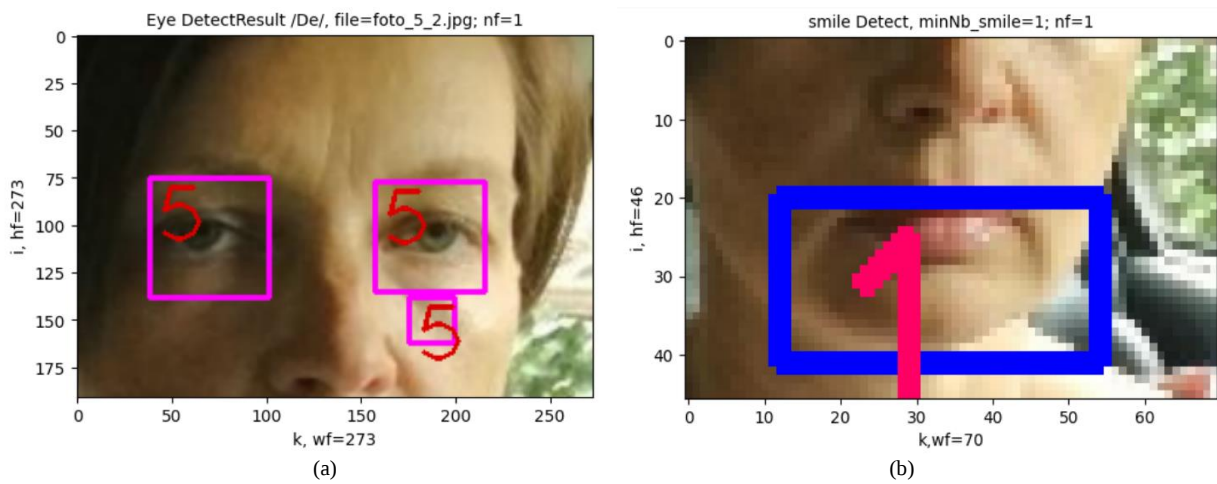


Fig. 8. Eye detection with confidence  $D_{eC} = 5$  (a) and mouth detection with confidence  $D_{sC} = 1$  (b)

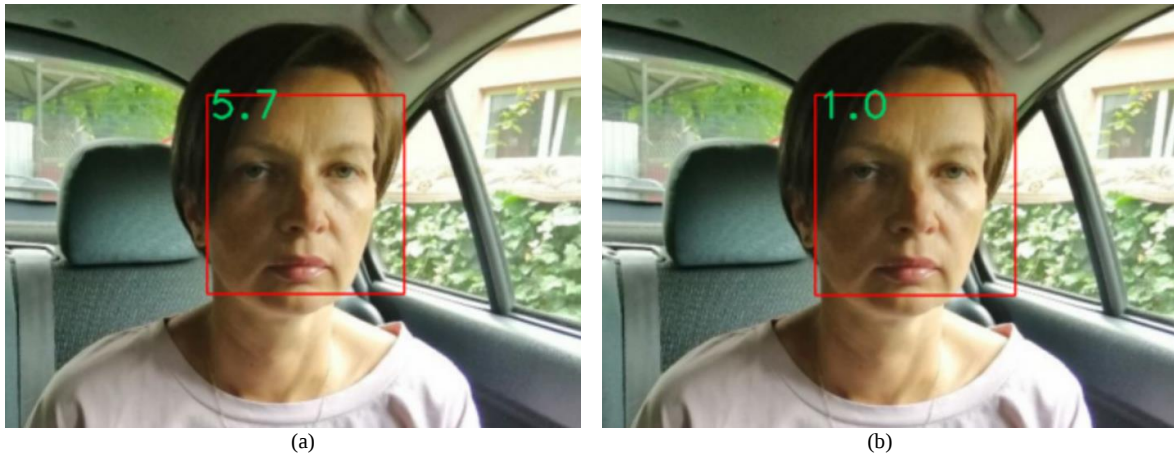


Fig. 9. Recognition of face with confidence  $D_{fC} = 5.7$  (a) and recognition of mask with confidence  $D_{mC} = 1.0$  (b)

Consider example # 3 of detecting faces (Fig. 10) and eyes (Fig. 11) of participants in the educational process [37] using the developed system. Detection is performed with a fixed value of the minNeighbors parameter of the Viola-Jones method, so some areas are mistakenly detected as faces (area # 4 in Fig. 10). However, eyes are not detected in such area # 4 (Fig. 11), so only face areas #1-#3 (Fig. 10) can be recognized with high confidence. On the areas of faces #1-#3 (Fig. 10), the mouth is detected only for face #1. Therefore, the mask is not recognized on face #1, and the mask is recognized on faces #2, #3.

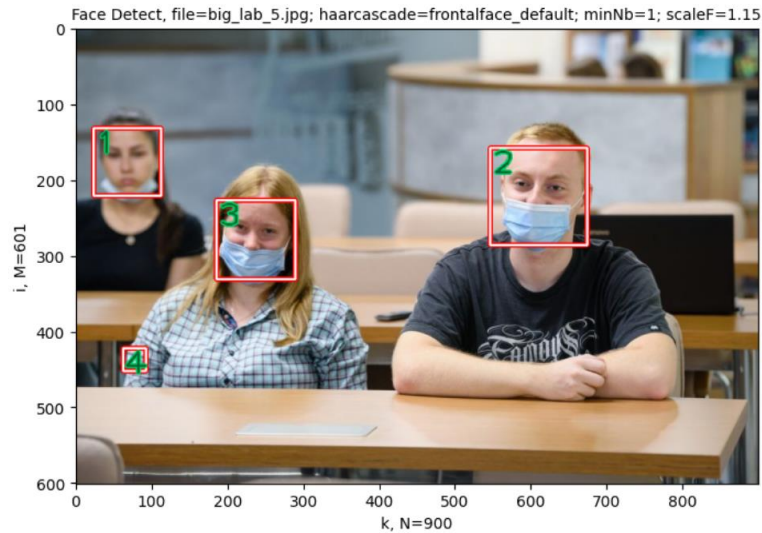


Fig. 10. Detection of faces in the image with fixed parameters of the Viola-Jones method

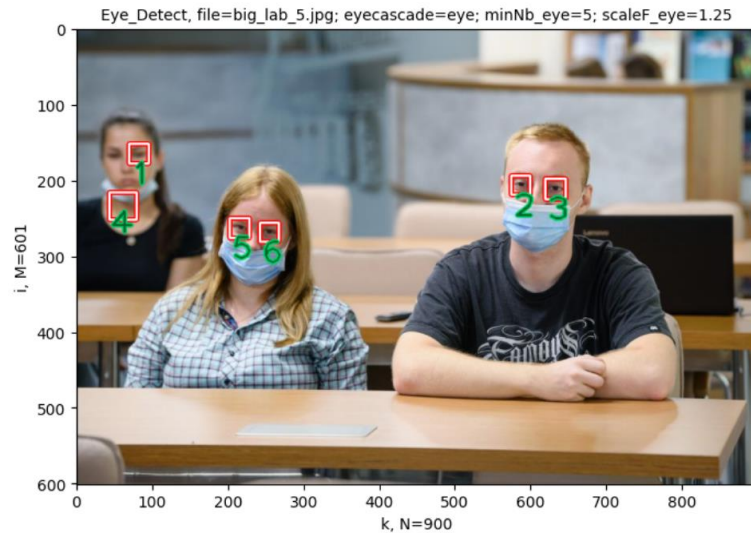


Fig. 11. Detection of eyes in the image with fixed parameters of the Viola-Jones method

The time of face mask recognition in the developed software is an order of magnitude higher than when using the Viola-Jones method with fixed parameters (since the detection is performed for 6 values of the minNeighbors parameter). Therefore, such recognition can be applied to individual images or frames of a video stream. This is acceptable, in particular, when recognizing images of participants in the educational process [38, 39], when recognition should be carried out with a period of several seconds or more.

## 6. Conclusion and Future Research Directions

The software implementation of the face mask recognition system using the Viola-Jones method was performed. To improve the accuracy of face and mask recognition, Haar cascades are used for face, eye and mouth detection. Object detection was performed at different values of the minNeighbors parameter for the Viola-Jones method, due to which the confidence of object detection was determined.

A mathematical model of fuzzy inference (1-4) and Mamdani fuzzy knowledge base #1 are developed, which is designed for face recognition based on face detection confidence and eye detection confidence. In the process of fuzzy inference, fuzzy sets are used, which are described by fuzzy triangular membership functions. Mathematical model of fuzzy inference (5-8) and Mamdani fuzzy knowledge base #2, which is designed for mask recognition based on face recognition confidence and mouth detection confidence, are also developed.

The fuzzy face and face mask image recognition system was software implemented in Python using the Google Colab cloud service. The initial data for the program are digital images read from a video camera or from graphic files. The results of the program are the coordinates of the rectangles that correspond to the faces in the image, as well as the face recognition confidence  $D_{frc}$  (4) and mask recognition confidence  $D_{mc}$  (8) for each such rectangle. Processing of test images showed generally correct recognition of faces and masks on faces.

Due to the detection of not only faces, but also eyes and mouths, as well as due to the implementation of fuzzy inference, the time of face mask recognition in the developed software is an order of magnitude greater than when using the Viola-Jones method with fixed parameters. The decrease in the speed of image recognition is compensated by higher accuracy, since the complex use of the results of various Haar cascades and fuzzy inference allowed to increase the accuracy of recognition of faces and face masks. The speed of the developed software is sufficient, in particular, for recognizing images of participants in the educational process.

The developed system for recognizing images of faces and masks on faces can be used not only for applied purposes, but also in STEM education when studying the principles of image recognition [40].

## Acknowledgment

We would like to express our gratitude to the reviewers for their precise and succinct recommendations that improved the presentation of the results obtained.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] S. Sen, and K. Sawant, "Face Mask Detection for Covid\_19 Pandemic Using Pytorch in Deep Learning," *IOP Conf. Ser. Mater. Sci. Eng.* 2021, doi: 10.1088/1757-899X/1070/1/012061.
- [2] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.
- [3] Walid Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," *Signal, Image and Video Processing*, vol. 16, pp. 605–612, 2021.
- [4] J.V.B. Benifa, C. Chola, A.Y. Muaad, M.A.B. Hayat, M.B. Bin Heyat, R. Mehrotra, F. Akhtar, H.S. Hussein, D.L.R. Vargas, Á.K. Castilla, and et al. "FMDNet: An Efficient System for Face Mask Detection Based on Lightweight Model during COVID-19 Pandemic in Public Areas," *Sensors*, vol. 23, pp. 1-15, 2023, doi: 10.3390/s23136090.
- [5] Wenqi Wu, Yingjie Yin, Xingang Wang and De Xu, "Face Detection With Different Scales Based on Faster R-CNN," *IEEE Transactions on Cybernetics*, vol. 49, no. 11, pp. 4017-4028, 2019.
- [6] Deepa Indrawal, and Archana Sharma, "Multi-Module Convolutional Neural Network Based Optimal Face Recognition with Minibatch Optimization", *International Journal of Image, Graphics and Signal Processing*, vol. 14, no. 3, pp. 32-46, 2022, doi:10.5815/ijigsp.2022.03.04.
- [7] Yomna M. Elbarawy, Neveen I. Ghali, and Rania Salah El-Sayed, "Facial Expressions Recognition in Thermal Images based on Deep Learning Techniques", *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 10, pp. 1-7, 2019, doi: 10.5815/ijigsp.2019.10.01.
- [8] Htwe Pa Pa Win, Phyo Thu Thu Khine, and Zon Nyein Nway, "Emotion Recognition from Faces Using Effective Features Extraction Method", *International Journal of Image, Graphics and Signal Processing*, vol. 13, no. 1, pp. 50-57, 2021, doi: 10.5815/ijigsp.2021.01.05.
- [9] Leslie Ching Ow Tiong, Song Tae Kim, and Yong Man Ro, "Multimodal Facial Biometrics Recognition: Dual-stream Convolutional Neural Networks with Multi-feature Fusion Layers," *Image and Vision Computing*, vol. 12, pp. 2020.
- [10] G. Guo, H. Wang, Y. Yan, J. Zheng, and B. Li, "A Fast Face Detection Method via Convolutional Neural Network," *Neurocomputing*, vol. 395, pp. 128–137, 2020, doi: 10.1016/j.neucom.2018.02.110.
- [11] J. Zhang, X. Wu, S.C.H. Hoi, and J. Zhu, "Feature Agglomeration Networks for Single Stage Face Detection," *Neurocomputing*, vol. 380, pp. 180–189, 2020, doi: 10.1016/j.neucom.2019.10.087.
- [12] Marwa Y. Mohammed, "Face Recognition based Texture Analysis Methods," *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 7, pp. 1-8, 2019, doi: 10.5815/ijigsp.2019.07.01.
- [13] Turker Tuncer, Sengul Dogan, and Erhan Akbal, "Discrete Complex Fuzzy Transform based Face Image Recognition Method", *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 4, pp. 1-7, 2019, doi: 10.5815/ijigsp.2019.04.01.
- [14] Surbhi Gupta, Kutub Thakur, and Munish Kumar, "2D-human face recognition using SIFT and SURF descriptors of face's feature regions," *The Visual Computer*, vol. 37, pp. 447–456, 2021.
- [15] Farid Ayeche, and Adel Alti, "HDG and HDGG: an extensible feature extraction descriptor for effective face and facial expressions," *Pattern Analysis and Applications*, vol. 24, pp. 1095–1110, 2021.
- [16] DaeEun Kim, Dosik Hwang, *Intelligent Imaging and Analysis*. Switzerland, Basel: MDPI, 2020.
- [17] R. Gonzalez, R. Woods, *Digital image processing*. 4th edition, Pearson/ Prentice Hall, NY, 2018.
- [18] Serhiy V. Balovsyak, and Khrystyna S. Odaiska, "Automatic Highly Accurate Estimation of Gaussian Noise Level in Digital Images Using Filtration and Edges Detection Methods," *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, vol. 9, no. 12, pp. 1-11, 2017, doi: 10.5815/ijigsp.2017.12.01.
- [19] S.V. Balovsyak, and Kh. S. Odaiska, "Automatic Determination of the Gaussian Noise Level on Digital Images by High-Pass Filtering for Regions of Interest," *Cybernetics and Systems Analysis*, vol. 54, no. 4, pp. 662-670, 2018, doi: 10.1007/s10559-018-0067-3.
- [20] S.V. Balovsyak, O.V. Derevyanchuk, I.M. Fodchuk, O.P. Kroitor, Kh.S. Odaiska, O.O. Pshenychnyi, A. Kotyra, and A. Abisheva, "Adaptive oriented filtration of digital images in the spatial domain," *Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*, vol. 11176, pp. 111761A-1–111761A-6, 2019, doi: 10.1117/12.2537165.

- [21] Ihor Tereikovskiy, Zhengbing Hu, Denys Chernyshev, Liudmyla Tereikovska, Oleksandr Korystin, and Oleh Tereikovskiy, "The Method of Semantic Image Segmentation Using Neural Networks," *International Journal of Image, Graphics and Signal Processing*, vol. 14, no. 6, pp. 1-14, 2022, doi: 10.5815/ijigsp.2022.06.01.
- [22] O.V. Derevyanchuk, H.O. Kravchenko, Y.V. Dereviachuk, and V.V. Tomash, "Recognition images of broken window glass," *Proceedings of SPIE*, vol. 12938, pp. 210-213, 2024, doi: 10.1117/12.3012995.
- [23] V.M. Kramar, O.V. Pugantseva, A.V. Derevyanchuk, "Spatial confinement, self-polarization and exciton-phonon interaction effect on the location of exciton line in lead iodide nanofilms," *Low Temperature Physics*, vol. 40, no. 8, pp. 766–770, 2014, doi: 10.1063/1.4892649.
- [24] S.V. Balovsyak, O.V. Derevyanchuk, and I.M. Fodchuk, "Method of calculation of averaged digital image profiles by envelopes as the conic sections," *Advances in Intelligent Systems and Computing*, Hu Z., Petoukhov S., Dychka I., He M. (Eds.), Springer International Publishing, vol. 754, pp. 204-212, 2019, doi: 10.1007/978-3-319-91008-6\_21.
- [25] S.V. Balovsyak, O.V. Derevyanchuk, H.O. Kravchenko, O.P. Kroitor, and V.V. Tomash, "Computer system for increasing the local contrast of railway transport images," *Proc. SPIE, Fifteenth International Conference on Correlation Optics*, vol. 12126, pp. 121261E1-7, 2021, doi: 10.1117/12.2615761.
- [26] S. Balovsyak, Kh. Odaiska, O. Yakovenko, and I. Iakovlieva, "Adjusting the Brightness and Contrast parameters of digital video cameras using artificial neural networks," *Proceedings of SPIE*, vol. 12938, pp. 129380I-1 - 129380I-4, 2024, doi: 10.1117/12.3009429.
- [27] Raveendra K, Ravi J, and Khalid Nazim Abdul Sattar, "Face Recognition Using Modified Histogram of Oriented Gradients and Convolutional Neural Networks," *International Journal of Image, Graphics and Signal Processing*, vol. 15, no. 5, pp. 60-76, 2023, doi: 10.5815/ijigsp.2023.05.05.
- [28] Paul Viola, and Michael J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [29] A. R. Fayek, "Fuzzy Logic and Fuzzy Hybrid Techniques for Construction Engineering and Management", *Journal of Construction Engineering and Management*, vol. 146, no. 7, pp. 1-12, 2020. doi: 10.1061/(ASCE)CO.1943-7862.0001854.
- [30] D.S. Hooda, and Vivek Raich, *Fuzzy Logic Models. An Introduction*. U.K., Oxford, Alpha Science International Ltd, 2017.
- [31] Face Detection with Python using OpenCV, URL: <https://www.datacamp.com/community/tutorials/face-detection-python-opencv> (Last accessed: 20.11.2023).
- [32] OpenCV People Counter, URL: <https://www.pyimagesearch.com/2018/08/13/opencv-people-counter/> (Last accessed: 21.11.2023).
- [33] Face Recognition System Using Google Colab, URL: <https://levelup.gitconnected.com/face-recognition-system-using-google-colab-cccc1d56f5f3> (Last accessed: 21.11.2023).
- [34] Vasyl Lytvyn, Olga Lozynska, Dmytro Uhryn, Myroslava Vovk, Yuriy Ushenko, and Zhengbing Hu, "Information Technologies for Decision Support in Industry-Specific Geographic Information Systems based on Swarm Intelligence," *International Journal of Modern Education and Computer Science*, vol. 15, no. 2, pp. 62-72, 2023. doi: 10.5815/ijmecs.2023.02.06.
- [35] Oleh Prokipchuk, Victoria Vysotska, Petro Pukach, Vasyl Lytvyn, Dmytro Uhryn, Yuriy Ushenko, and Zhengbing Hu, "Intelligent Analysis of Ukrainian-language Tweets for Public Opinion Research based on NLP Methods and Machine Learning Technology," *International Journal of Modern Education and Computer Science*, vol. 15, no. 3, pp. 70-93, 2023. doi: 10.5815/ijmecs.2023.03.06.
- [36] 4-Cascade\_classification.ipynb, URL: [https://colab.research.google.com/github/computationalcore/introduction-to-opencv/blob/master/notebooks/4-Cascade\\_classification.ipynb](https://colab.research.google.com/github/computationalcore/introduction-to-opencv/blob/master/notebooks/4-Cascade_classification.ipynb). (Last accessed: 20.11.2023).
- [37] Zaporizhzhia State Medical and Pharmaceutical University, URL: [https://mphu.edu.ua/new\\_3368.html](https://mphu.edu.ua/new_3368.html) (Last accessed: 22.11.2023).
- [38] O.V. Derevyanchuk, V.I. Kovalchuk, V.M. Kramar, H.O. Kravchenko, D.V. Kondryuk, A.V. Kovalchuk, and B.V. Onufriichuk, "Implementation of STEM education in the process of training of future specialists of engineering and pedagogical specialties," *Proceedings of SPIE*, vol. 12938, pp. 214-217, 2024, doi: 10.1117/12.3012996.
- [39] V. Kovalchuk, A. Androsenko, A. Boiko, V. Tomash, O. Derevyanchuk, "Development of Pedagogical Skills of Future Teachers of Labor Education and Technology by means of Digital Technologies," *International Journal of Computer Science and Information Security*, vol. 22 (9), pp. 551-560, 2022, doi: 10.22937/IJCSNS.2022.22.9.71.
- [40] S. Balovsyak, O. Derevyanchuk, H. Kravchenko, Y. Ushenko, and Z. Hu, "Clustering Students According to their Academic Achievement Using Fuzzy Logic," *International Journal of Modern Education and Computer Science*, vol. 15, no. 6, pp. 31-43, 2023, doi: 10.5815/ijmecs.2023.06.03.

## Authors' Profiles



**Serhiy Balovsyak:** Graduated from Chernivtsi State University (1995). In 2018, he defended his doctoral dissertation in the specialty "Computer systems and components".

Currently position – associate professor at the Department of Computer Systems and Networks of Yuriy Fedkovych Chernivtsi National University, Ukraine.

Research Interests: digital processing of signals and images, programming, pattern recognition, artificial neural networks.



**Oleksandr Derevyanchuk:** Received the Master of Engineering degree (1999) and Ph.D. of Physics and Mathematics (2014) at the Yuriy Fedkovych Chernivtsi National University. He is a Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Professional and Technological Education and General Physics, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine.

Research Interests: digital processing of signals and images, programming, pattern recognition, artificial neural networks.



**Vasyl Kovalchuk:** Studied at Yuri Fedkovych Chernivtsi State University and got specialist degree in Professional education (1993-1998). In 2014 defended doctoral dissertation “Theoretical and methodological principles of developing pedagogical skills of vocational schools professional training masters in postgraduate education” and became Doctor of Pedagogical Sciences, Professor. Research Interests: include professional and higher education development, introducing innovative teaching technologies in the educational process, teachers’ pedagogical skills development, introducing digital technologies in the educational process, emotional intelligence development, leadership.



**Hanna Kravchenko:** PhD student at the of the Department of Professional and Technological Education and General Physics, Physical, Technical and Computer Sciences Institute of Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine.

Research Interests: digital processing of signals and images, programming, pattern recognition, artificial neural networks.



**Maryna Kozhokar:** Associate Professor at the Department of Physical culture. PhD in Pedagogical Sciences (2012).

Current position – Docent, Deputy head of the department for educational and methodological and research work.

Research Interests: Data Mining and Analysis, Computer Vision and Pattern Recognition, Physical culture, Biophysics.

**How to cite this paper:** Serhiy Balovsyak, Oleksandr Derevyanchuk, Vasyl Kovalchuk, Hanna Kravchenko, Maryna Kozhokar, "Face Mask Recognition by the Viola-Jones Method Using Fuzzy Logic", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.16, No.3, pp. 39-51, 2024. DOI:10.5815/ijigsp.2024.03.04