

Artificial Intelligence Based Emergency Identification Computer System

Diana Velychko¹, Halyna Osukhivska¹, Yuri Palaniza²,
Nadiia Lutsyk¹, Łukasz Sobaszek^{3*}

¹ Computer Systems and Networks Department, Faculty of Computer Information Systems and Software Engineering, Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine

² Department of Radio Engineering Systems, Faculty of Applied Information Technologies and Electrical Engineering, Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine

³ Department of Information Technology, Faculty of Mathematics and Information Technology, Lublin University of Technology, Lublin, Poland

* Corresponding author's e-mail: l.sobaszek@pollub.pl

ABSTRACT

The use of Artificial Intelligence is currently being observed in many areas of life. In addition to assisting in intellectual work, solving complex computational problems, or analyzing various types of data, the aforementioned techniques can also be applied in the process of providing security to people. The paper proposes an emergency identification system based on Artificial Intelligence that aims to provide timely detection and notification of dangerous situations. The proposed solution consider the position of a person “hands up” as an emergency situation that will indicate a potential danger for a person. Because people in the face of potential danger are mostly forced to raise their hands up and this pose attracts attention, emphasizes the emotional reaction to certain events and is usually used as a sign of risk or as a means of subjugation. The system should recognize the pose of a person, detect it, and consequently inform about the threat. In this paper, an AI based emergency identification system was proposed to detect the human pose “hands up” for emergency identification using the PoseNet Machine Learning Model. The assumption consists that the utilization only of 6 key points made allows reducing the computing resources of the system since the conclusion is made taking into account a smaller amount of data. For the study, a dataset of 1510 images was created for training an Artificial Intelligence model, and the decisions were verified. Supervised Machine Learning methods are used to classify the definition of an emergency. Alternative methods: Support Vector Machine, Logistic Regression, Naïve Bayes Classifier, Discriminant Analysis Classifier, and K-nearest Neighbours Classifier based on the accuracy were evaluated. Overall, the paper presents a comprehensive and innovative approach to emergency identification for quick response to them using the proposed system.

Keywords: artificial intelligence, deep learning, detection, emergency system, computer system.

INTRODUCTION

In today's world, the increasing number of terrorist attacks, assaults, thefts, and other emergencies that pose a threat to human security has become a serious challenge for law enforcement agencies, security services, businesses and citizens. Existing systems are not always able to provide timely detection and notification of a dangerous situation. A computerised system that

identifies emergencies can be an effective tool for preventing and minimising the consequences of emergencies. The use of artificial intelligence will allow real-time recognition of emergencies and automatic notification.

Papers [1–4] describe the systems that allow the identification of various dangerous situations. The authors of study [1] propose a hazard detection technology based on the analysis of multimodal data from several sensors on a smartphone and

smartwatch, as well as contextual data from several people and places over time. The proposed Multi-log Data Analysis (MLDA) helps to detect dangerous situations that are abnormal by analysing the contextual state of people, objects, and places.

The system described in [2] utilizes non-invasive and non-intrusive sensors, RFID tags, and GPS to collect and analyze real-time physiological signals in order to determine when a child is in danger. The system makes assumptions about the state of danger based on the verification of specific biometric responses to certain situations, using a self-learning algorithm developed for this architecture. In the paper [3], the authors propose a model that uses artificial intelligence to detect objects (objects) and human actions that could provoke a dangerous situation, as well as the relationship between them, in CCTV images. There are also systems that are used to determine the emergence of human conditions at home.

The authors of [4] proposed a system that allows the detection of dangerous situations, such as a person falling, based on posture detection. The various security automation systems and technologies are described in the paper [5]. The authors analysed different systems, such as Intelligent acoustic and vibration recognition/alert systems for security breaching detection, proximity danger identification, and perimeter protection, home security system using Arduino UNO, Low Power Accelerometer Based Intrusion and Tamper Detector, IoT based Intruder Detection System Using GSM, Identification of Damaging Activities for Perimeter Security, Smart IoT Security System for Smart-Home Using Motion Detection and Facial Recognition, Ambient Intelligence and IoT Based Decision Support System for Intruder Detection, Theft Detection System using PIR Sensor, Ambient Intelligence and IoT Based Decision Support System for Intruder Detection, Theft Detection System using PIR Sensor, etc.

In this article, we propose to use a different idea to identify emergency situations, namely, by the specific behaviour of people in the room, the “hands up” posture, since people in potential danger are mostly forced to raise their hands up as a sign of risk or as a means of submission. This pose attracts attention, emphasises an emotional reaction to certain events, and is usually used when there is a threat. In the context of a crime, the “hands up” pose is a serious warning signal and can cause fear and anxiety among people at the

scene, which may indicate robbery, intimidation or detention, as well as that a person is being coerced. Therefore, it is advisable for the system to detect an emergency situation based on the “hands up” pose. In this paper, we propose a system that detects the “hands up” pose and, based on this, informs about an emergency. The system recognises the “hands up” pose using artificial intelligence.

Artificial intelligence is already commonly used for human pose recognition. Several researchers have studied the problem of human pose recognition and have determined the most effective algorithms for this purpose. Their research is described in the literature [6–10]. These studies mainly consider different algorithms and verify the accuracy of the determination of key joints in the human body. However, these studies do not take into account the particular posture that a person adopts for a given set of data, i.e., for a given set of joint coordinates. For some purposes, it is not necessary to have information about a person’s posture at any given time, but it is necessary for the model to identify and report a particular posture that may signal danger, i.e., the “hands up” posture.

Therefore, the task is to develop a system that recognizes a specific pose – “hands up”. Such a system can be used in different premises for different industries, such as banks, shops, schools, and other institutions. In addition, the use of such a system can reduce security costs, as it replaces the human factor in the process of monitoring the video stream.

System architecture and algorithm

The generalized architecture of the AI-based emergency identification system is shown in Figure 1. The main task performed by the system is to identify a dangerous situation for a person, namely, to determine the person’s “hands up” pose.

According to the proposed architecture, the first step is to acquire an image from the video stream of the video surveillance system, for example. Next, the video is recorded and transmitted, i.e., “captured”, converted into digital signals, processed, adjusted, and transmitted to the server. The video information is displayed and access to the module is provided on a smartphone, tablet, or PC via a browser, and the emergency situation is identified using artificial intelligence. It is proposed to store data (i.e., received video recordings and emergency history) in the cloud or on physical

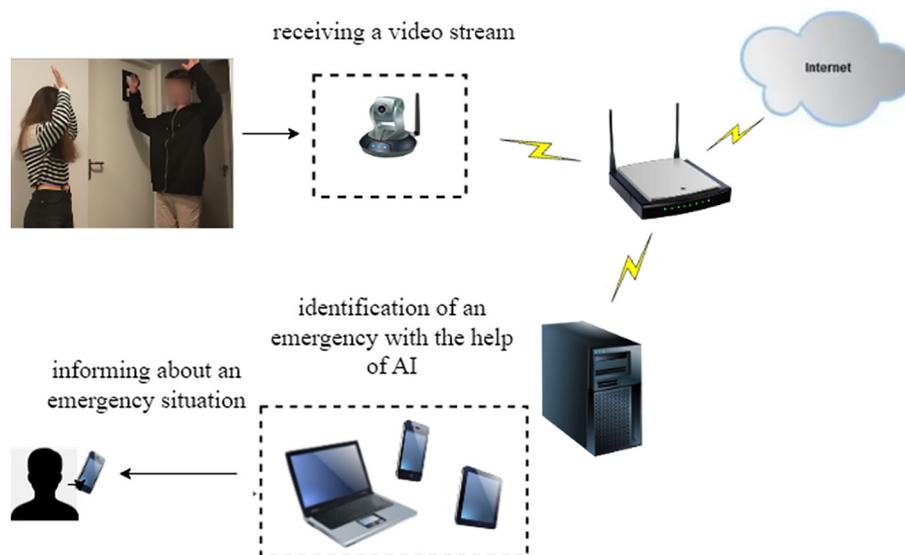


Fig. 1. General architecture of the system

server storage. It is proposed to use a Telegram bot to send a message about an emergency.

This architecture allows you to access video content at any time, implement the function of identifying an emergency (hands up), and provide the ability to inform about an emergency. A pose detection algorithm is used to detect an emergency. This component of the system analyses the received video data and determines the pose of a person's "hands up". Pose recognition uses pose and orientation to predict and track the location of a person or object. Accordingly, pose detection allows applications to estimate the spatial position of a body ("pose") in an image or video. Pose detection is performed by finding key points on a person or object.

Assessing body posture is a challenging task because the appearance of the body is dynamically changing, and the angle of view and external context also have an impact. In the field of human posture estimation, there are two main approaches: bottom-up and top-down methods [7]. Bottom-up methods involve first estimating each individual body joint and then grouping them to determine the overall pose. Top-down methods first detect a person in the image using a person detector and then estimate the body joints within the bounding boxes of the detected person [8]. On the other hand, top-down methods start with running a person detector to identify the person in the image and then estimate the body joints within the detected bounding boxes [6]. The ultimate goal of human posture estimation is to predict the positions of the different body parts and joints

in photographs or videos. Accurate estimation of body position is important for identifying human actions since movement patterns are often influenced by body posture [6].

In addition, the posture estimation can be performed in both 2D and 3D. 2D human pose estimation involves using visual data, such as photographs and videos, to approximate the spatial positioning or location of key points on the human body [9–10]. Such models exist for human pose estimation: kinematic, planar, and volumetric. After analyzing different models for human pose estimation, it was found that the kinematic model, also known as the skeleton-based model, is the most suitable for detecting unusual human behaviour such as the "hands up" posture. Additionally, most techniques use an N-joints rigid kinematic model, which represents the human body as an object with joints and limbs that provide information on body kinematic structure and shape [6, 9]. When assessing a person's pose, the key points will be joints such as elbows, knees, wrists, etc.

In this paper propose to detect the human pose "hands up" for emergency identification using the PoseNet machine learning model. PoseNet is a Deep Learning TensorFlow model that determines a person's pose by estimating body parts defined as key points (17 in this model), ranging from eyes, and ears, to ankles and knees, including wrists, elbows, shoulders, and nose [11]. These points are connected to form the skeletal structure of the body. All the points and the pose itself are also assessed for authenticity [9, 10].

The AI-based emergency recognition function (“hands up” pose) is implemented using the following algorithm: storyboarding the video received from the camera, recognising key points (joints), determining the coordinates of six informative points, normalising the data, and recognising the “hands up” pose.

The flowchart of the artificial intelligence based emergency identification computer system is shown in Figure 2. The video stream received is first pre-processed, and then divided into separate frames (storyboarded), then in each frame, using Deep Neural Networks, the presence of a person is determined and key points (joints) are recognised. If there is no person in the frame, the next frame is analysed. The next step is to identify informative points and normalise them, and then, based on the location of these points, the trained model decides whether the person is in a “hands up” pose. If the system detects this pose, an emergency notification is sent, if not, the next frame is analysed.

The main informative points used to recognise the “arms up” pose are: “left_shoulder”, “right_shoulder”, “left_elbow”, “right_elbow”, “left_wrist”, “right_wrist”. In other words, the system needs 6 key points (classes) to make a decision instead of 17 and recognises the “arms up” pose according to their location. This solution reduces the computing resources of the system since the conclusion is made taking into account a smaller amount of data (i.e., instead of 17 points in one image frame, only 6 are analysed), and thus the computational complexity of the task is significantly lower.

DATASET AND METHODS

Since no data set of people with a given pose was found on the Internet, for the testing of the system, therefore, 1510 images were formed, of which 930 are in the “hands up” pose, 580 are poses with other hand positions. They were obtained from videos in which people of different ages and genders, in different rooms and at different distances from the camera, make hand movements, including raising their hands up, imitating the pose that the system should recognise as an emergency.

Figure 3 shows an image from the generated dataset used for training with key points identified using the PoseNet model. The proposed system, using PoseNet, recognises people in images

and then recognises key points (joints). The next step is to obtain the coordinates of 6 informative points and normalise them for further analysis. These points are: “left_shoulder”, “right_shoulder”, “left_elbow”, “right_elbow”, “left_wrist”, and “right_wrist”. The data is normalised in the following way: the point is chosen in the middle

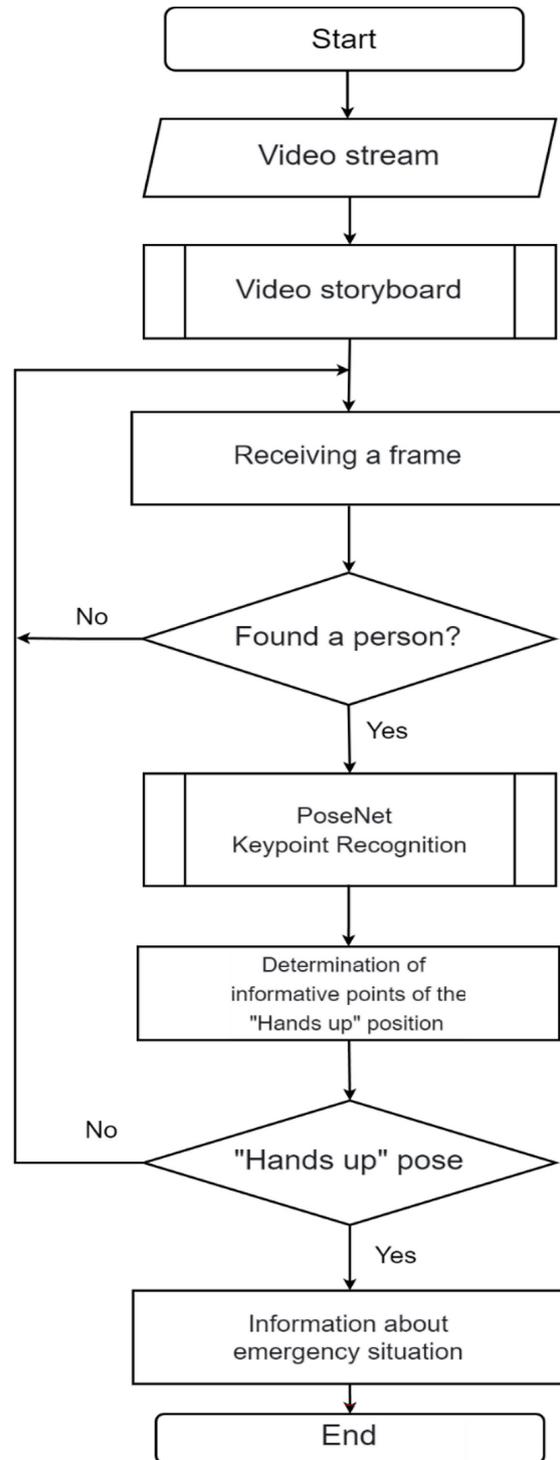


Fig. 2. The flowchart of the artificial intelligence based emergency identification computer system

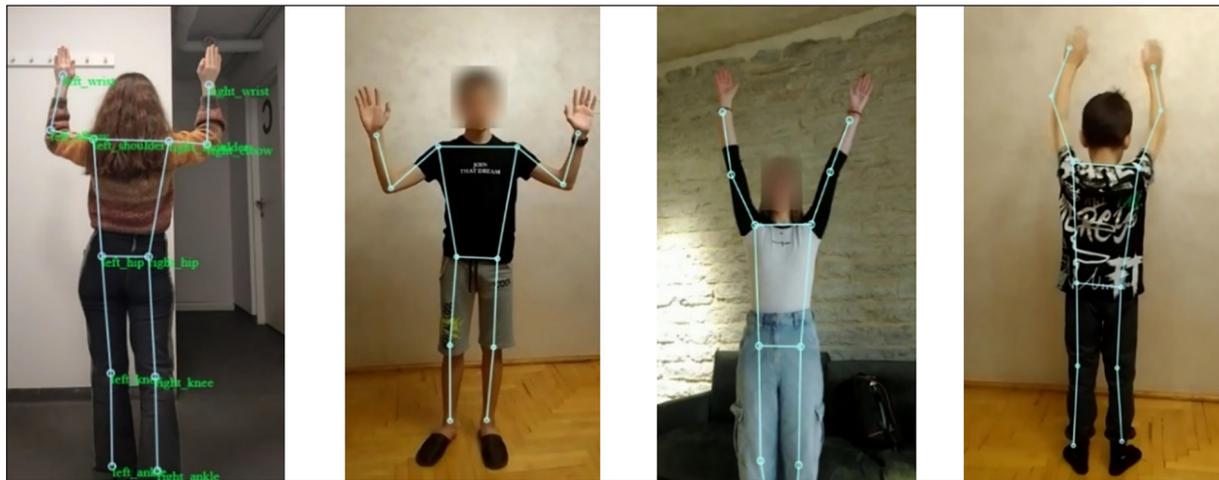


Fig. 3. Example of images with key points identified

between the shoulders, and the distance from the middle point to all other points is calculated to find the furthest points in two dimensions (x and y). Then the furthest points are fit into the square in such a way that they lie on the sides of the square. Thus, the distances between the furthest points in both dimensions are used as a scaling factor for all coordinates. As a result, all points fit into a square with side lengths equal to 1, so that the value of any coordinate is in the range between 0 and 1. Figure 4 shows the visual representation of the selected points and a point for normalisation – a midpoint.

As a result of normalisation, the coordinates are found in the range from 0 to 1, so the height of the person, the distance from the camera, and the placement of the person in a video or a picture do not affect the prediction result. The graphs of normalised coordinates, when hands are in the “hands up” position and in other positions, are presented in Figure 5. Basically, the input consists of 12 variables (x and y coordinates of those

6 points), and the output is class 0 if the posture is predicted to be “hands up”, and 1 if the posture is considered as any other.

RESULTS AND DISCUSSION

Thus, the generated dataset used for the study consists of 1510 images: 930 are “hands up” (“0”) and 580 are other poses, i.e., not “hands up” (“1”). For further classification of the achieved coordinates, the methods of machine learning are utilised. This paper considered an approach that takes into account that there are two classes of data (the location of the hands to determine the pose): the “hands up” pose (an emergency identifier), and any other pose. Therefore, the task of binary classification is essentially solved. For this purpose, we used the support vector method, i.e., SVM (Support Vector Machine), and for comparison, logistic regression, Naïve Bayes classifier, Discriminant Analysis classifier, and K-nearest

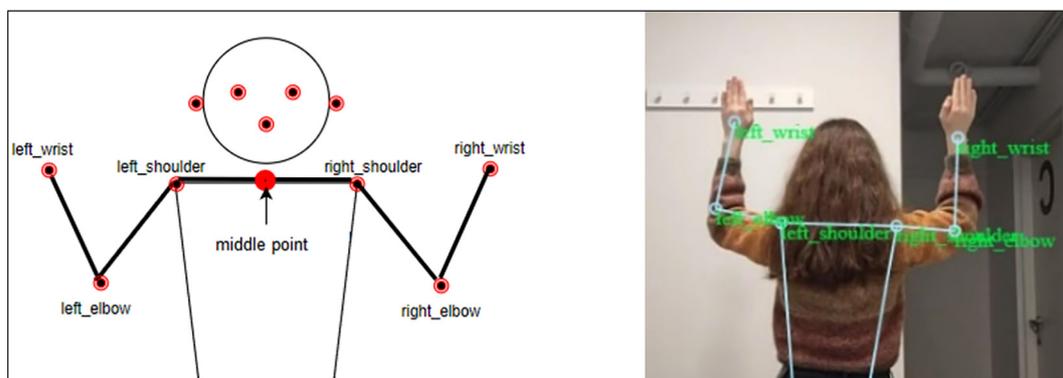


Fig. 4. Visualisation of 6 chosen points and a normalisation point

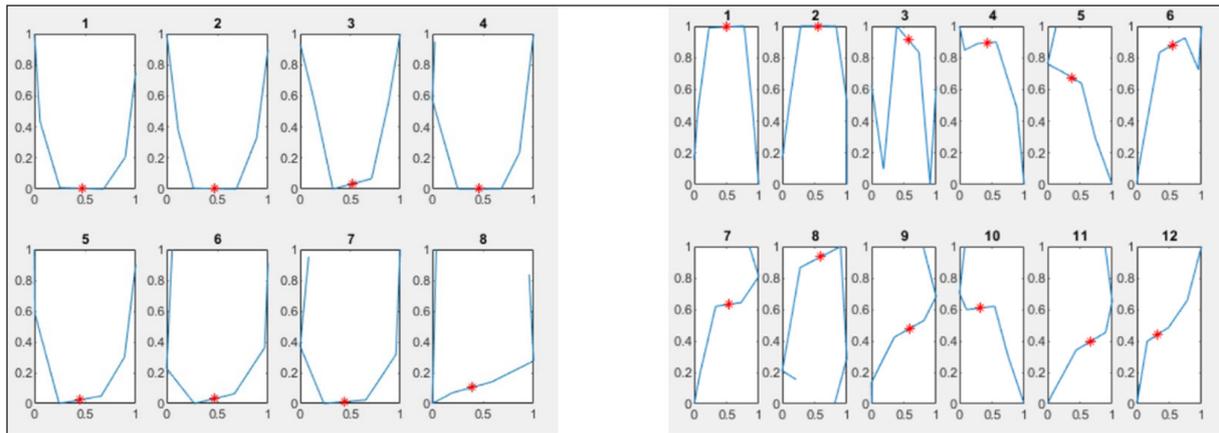


Fig. 5. Graphs of normalised coordinates when hands are in (a) position “hands up”, (b) other positions

Neighbours classifier. The methods are used for the classification of postures based on 6 normalised coordinates. The implementation of these methods was done in MATLAB.

The classifier is first “trained” on objects from the training set (from the generated dataset), which are predefined with class labels. Then, the trained algorithm predicts the class label from the test sample for each object [12]. The SVM algorithm uses a set of functions that are defined as a kernel (a function that is provided to the machine learning algorithm). In this case, we used Gaussian kernels. By maximising the distance between points and finding the best hyperplane, SVM divides the data into different categories. Given a dataset where each element belongs to one of two categories (“hands” up and other), the SVM is trained and then the classifier is tested.

Figure 6 shows the mismatch matrix of SVM with Gaussian kernels and ROC graph. The

matrix displays the total amount of data in each cell, with rows corresponding to the correct class and columns to the predicted class [13]. The diagonal corresponds to the correctly classified classes (930 for the “hands up” pose, 580 for the other). The line at the bottom of the figure shows the percentage of all data that belongs to the correctly and incorrectly classified class. The column in the right corner of the figure shows the percentage of all data belonging to each class that is correctly or incorrectly classified, i.e., accuracy [13, 14]. A ROC curve is a graph showing the performance of a classification model at all classification thresholds [15, 16, 17]. The “steepness” of ROC curves is particularly significant because it is desirable to maximise true positive rates while minimising false positive rates [13].

For comparison, machine learning models based on the following classification methods were built: Logistic Regression; Naive Bayes;

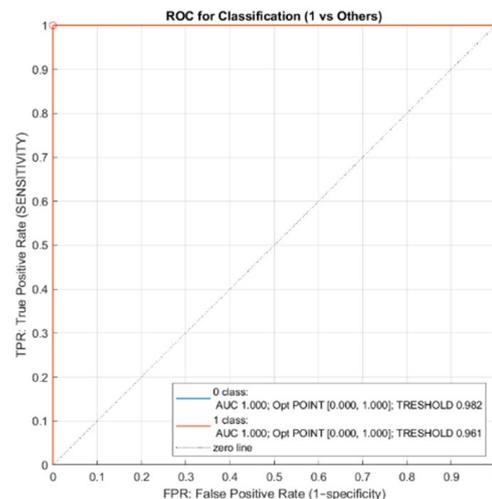
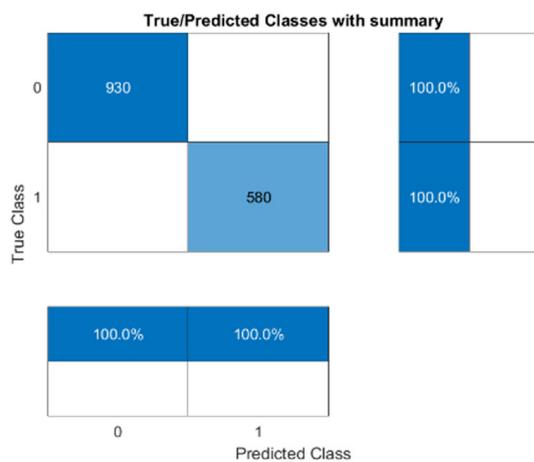


Fig. 6. The confusion matrix and ROC for SVM classifiers using a Gaussian Kernel

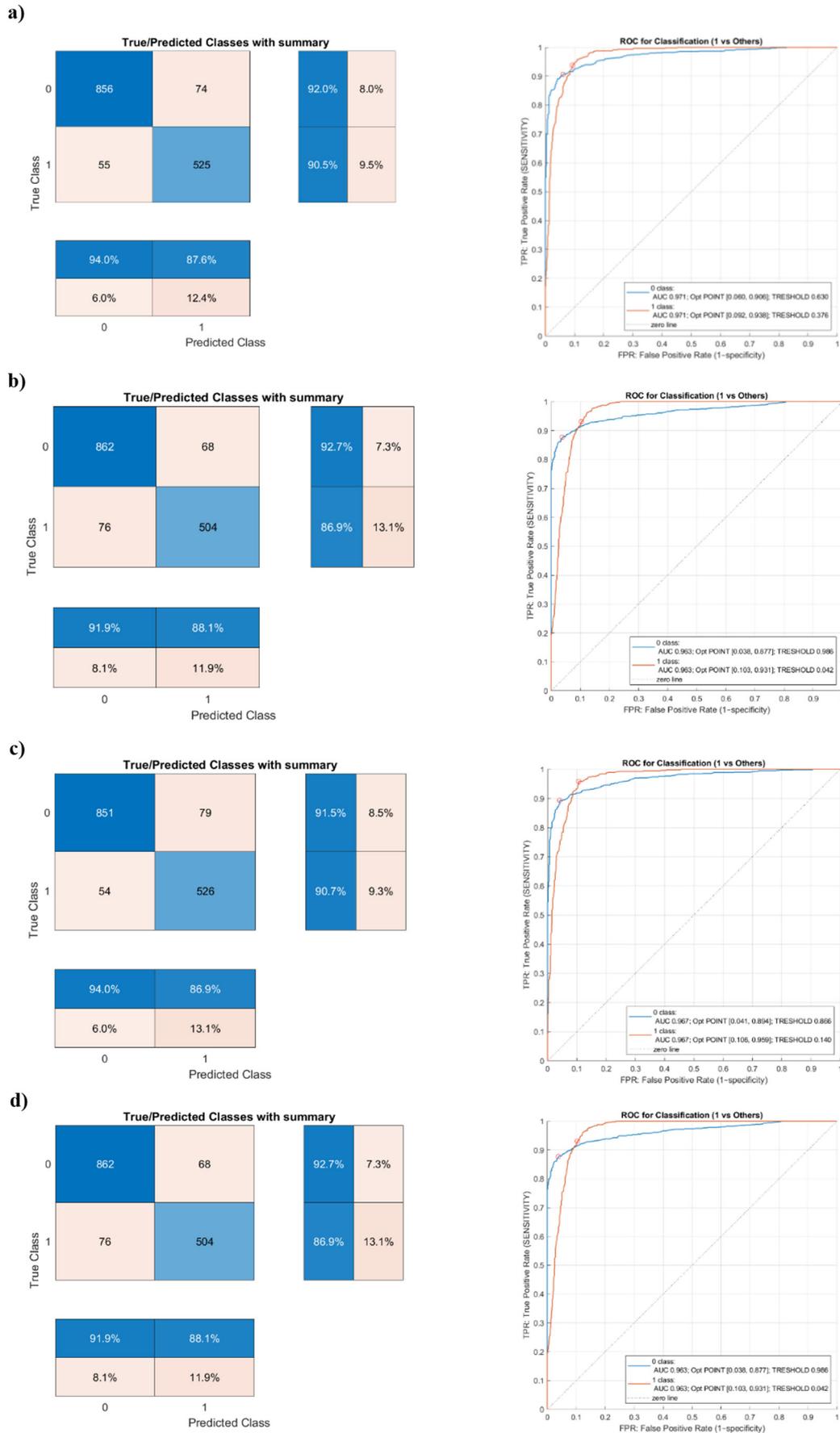


Figure 7. Confusion matrices and ROC for classification by: (a) logistic regression; (b) naive bayes; (c) discriminant analysis; (d) K-Nearest Neighbors

Table 1. The results of human pose recognition

Name of classification method	"Hands up" pose		Other poses	
	Total: 930		Total: 580	
	Correctly classified	Incorrectly classified	Correctly classified	Incorrectly classified
SVM classifiers using a Gaussian Kernel	930	0	580	0
Logistic regression	856	74	525	55
Naive bayes	862	68	504	76
Discriminant analysis	851	79	526	54
K-Nearest neighbors	862	68	504	76

Table 2. The comparison of the prediction results

Name of classifier	Recall	Precision
SVM classifiers using a Gaussian Kernel	100 %	100 %
Logistic regression	94.0 %	92.0 %
Naive bayes	91.9 %	92.7 %
Discriminant analysis	94.0 %	91.5 %
K-Nearest neighbors	91.9 %	92.7 %

Discriminant Analysis; and K-Nearest Neighbours, and their accuracy were evaluated. In general, all investing using the studied methods is done in a similar way. The resulting Confusion matrices and ROC for classification by logistic regression, Naive Bayes, Discriminant Analysis, and K-Nearest Neighbours are shown in Figure 7. The results of human pose recognition by different machine learning methods are presented in Table 1. The comparison of the prediction results by 5 models is presented in Table 2.

Analysing the results, we can conclude that SVM Classifier is ideal for this type of problem, as our dataset shows a prediction accuracy of 100%. The other methods showed slightly lower results, in particular, the accuracy of the K-Nearest Neighbours method for classifying the "hands up" pose in the emergency detection system was the same and amounted to 92.7%, the logistic regression method was 92%, and Discriminant Analysis was 91.5%. Nevertheless, the classification results obtained are high in each case. Therefore, machine learning methods can be effectively used for classification in order to recognise emergencies in the proposed system

CONCLUSIONS

The article proposes an architecture and algorithm for the operation of an artificial

intelligence-based system for identifying emergencies, which is capable of detecting atypical (emergency) situations by determining the "hands up" posture of a person in a room. It is justified to identify an emergency situation using the PoseNet model by six key points, which reduces the computing resources of the system. Supervised machine learning methods are used to classify the definition of an emergency: SVM (Support Vector Machine), Logistic Regression, Naïve Bayes classifier, Discriminant Analysis classifier, and K-nearest Neighbours classifier, and their accuracy were evaluated. The SVM method with Gaussian kernels showed the best result. Further research is needed to optimize the system's performance and evaluate its effectiveness in real-world scenarios. The proposed computer system can facilitate the more efficient work of security services. It can automatically detect emergencies and report them. This reduces the need for constant video stream monitoring and increases the accuracy of identifying dangerous situations, ensuring prompt response to potential dangers and reducing risks from them. Such a system can be used in organizations, enterprises, and institutions of various purposes, such as banks, schools, shopping and entertainment establishments, the service sector, etc., as well as a „smart home" system component. In addition, using such a system can reduce security costs, as it replaces the human factor in the monitoring video stream process.

REFERENCES

1. Park S.Y., Kwon E., Byon S., Shin W., Jung E., Lee Y. Danger detection technology based on multimodal and multilog data for public safety services. *Etri Journal* 2021; 44(2): 300–312. <https://doi.org/10.4218/etrij.2020-0372>
2. Lopes N.V., Santos H., Azevedo A.I. Detection of dangerous situations using a smart internet of things system. In: Rocha A., Correia A., Costanzo S., Reis L. (eds) *New contributions in information systems and technologies. Advances in Intelligent Systems and Computing*, vol. 354. Springer, Cham, 2015. https://doi.org/10.1007/978-3-319-16528-8_36
3. Jang S., Battulga L., Nasridinov A. Detection of dangerous situations using deep learning model with relational inference. *Journal of Multimedia Information System* 2020; 7(3): 205–214. <https://doi.org/10.33851/jmis.2020.7.3.205>
4. Prati A. An intelligent surveillance system for dangerous situation detection in home environments. *Unipr*, 2016. https://www.academia.edu/21940139/An_Intelligent_Surveillance_System_for_Dangerous_Situation_Detection_in_Home_Environments
5. Yadav, A., Thaker, N., Makwana, D., Waingankar, N., Upadhyay, P. Intruder Detection System: A Literature Review 2021; *Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021)*, Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3866777>
6. Dang Q., Yin J., Wang B., Zheng W. Deep learning based 2D human pose estimation: A survey. *Tsinghua Science & Technology* 2019; 24(6): 663–676. <https://doi.org/10.26599/tst.2018.9010100>
7. Gong W., Zhang X., González J., Sobral A., Bouwmans T., Tu C., Zahzah E. Human pose estimation from monocular images: a comprehensive survey. *Sensors* 2016; 16(12): 1966. <https://doi.org/10.3390/s16121966>
8. Odemakinde E. Human Pose Estimation with Deep Learning – Ultimate Overview in 2023. *Viso.ai*, 2023, <https://viso.ai/deep-learning/pose-estimation-ultimate-overview/>
9. Toshpulatov M., Lee W., Lee S., Roudsari A.H. Human Pose, Hand., Mesh Estimation using Deep Learning: a Survey. *The Journal of Supercomputing* 2022; 78(6): 7616–7654. <https://doi.org/10.1007/s11227-021-04184-7>
10. University of Toronto Machine Intelligence Team. Sparse R-CNN: towards more efficient object Detection models. *Medium*, 2021, <https://utorontomist.medium.com/sparse-r-cnn-towards-more-efficient-object-detection-models-fb244178998f>
11. Pandey S. Posenet model in ML. *OpenGenus IQ: Computing Expertise & Legacy*, 2022, <https://iq.opengenus.org/posenet-model/>
12. Hastie, T., Tibshirani, R., Friedman, J.H. *Data Mining, Inference, and Prediction*, Second Edition. Springer, New York; 2009, <https://doi.org/10.1007/978-0-387-84858-7>
13. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters* 2006; 27(8): 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
14. Machine Learning | Google for Developers. (n.d.). Google for Developers. <https://developers.google.com/machine-learning/crash-course>
15. Alyamani A., Yasniy O. Classification of EEG signal by methods of machine learning. *Applied Computer Science* 2020; 16(4): 56–63. doi:10.23743/acs-2020-29
16. Shabliy N., Lupenko S., Lutsyk N., Yasniy O., Malyshevska O. Keystroke Dynamics Analysis using Machine Learning Methods. *Applied Computer Science* 2021; 17(4): 75–83. <https://doi.org/10.23743/acs-2021-30>.
17. Yasniy O., Lutsyk N., Demchuk V., Osukhivska H., Malyshevska O. The prediction of structural properties of Ni-Ti shape memory alloy by the supervised machine learning methods. *ITTAP 2023*: 73–78. <https://ceur-ws.org/Vol-3628/short1.pdf>