A Platform for Monitoring Student Commuting in the Use of School Transport in Smart Cities - A Facial Recognition Based Approach

2023 International Conference on Smart Applications, Communications and Networking (SmartNets) | 979-8-3503-0252-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/SmartNets58706.2023.10216190

Jessé da Costa Rocha
Graduate Program in Electrical Engineering, Institute of Technology, Federal University of Pará (UFPA)
Belém, Pará, Brazil
jesse.rocha@icen.ufpa.br

Marcela Alves de Souza
Information and Communication Technology Center
Federal University of the South and Southeast of Para (Unifesspa),
Marabá, Pará, Brazil
marcela.alves@unifesspa.edu.br

Evelin Helena Silva Cardoso
Department of Computer Science
Federal Rural University of the Amazon (UFRA)
Capitão Poço, Pará, Brazil
evelin.cardoso@ufra.edu.br

Nandamudi Vijaykumar
Coordination of Applied Research and Technological Development
National Institute for Space Research (INPE)
São José dos Campos, São Paulo, Brazil
vijay.nl@inpe.br

Jasmine Priscyla Leite de Araújo
Graduate Program in Electrical Engineering, Institute of Technology, Federal University of Pará (UFPA)
Belém, Pará, Brazil
jasmine@ufpa.br

Renato Lisboa Francês
Graduate Program in Electrical Engineering, Institute of Technology, Federal University of Pará (UFPA)
Belém, Pará, Brazil
rfrances@ufpa.br

***Abstract*—This article proposes an intelligent platform for monitoring students' steps on their way to school until they leave the school to their homes. This platform can identify students and notify those responsible and competent authorities in various situations of school life, such as: entering and leaving the school bus, entering and leaving school, entering the school cafeteria, etc. The first application aims to control access to the school bus through facial recognition. In the tests carried out, the recognition system achieved excellent results in all metrics.**

Keywords—intelligent platform, facial recognition, student monitoring

I. Introduction

Since its emergence in the mid-twentieth century, the modern computer has far surpassed the human ability to perform calculations, being used efficiently as early as World War II to calculate missile trajectories and to decipher secret codes. However, a capacity that develops spontaneously in humans remained inaccessible to computers for decades: the ability to see, recognize and differentiate between objects, animals, and people. However, in recent years, the so-called computer vision has developed considerably, thanks in large part to the emergence of an artificial intelligence technique called CNN (Convolutional Neural Network).

The great milestone of this development process was the 2012 ImageNet challenge [1]. This challenge consisted of classifying a set of colored images in 1000 different categories based on training conducted with more than 1 million images. The team of Alex Krizhevsky and Geoffrey Hinton won an ImageNet Challenge 2012 achieving an accuracy of 83.6%. In the following years, the ImageNet challenge was dominated by CNNs that reached an accuracy of 96.4% in 2015.

Also in 2015, a group of Google researchers introduced FaceNet to the world, a facial recognition system that had a CNN as its core. According to the project developers, the system achieved a maximum accuracy of 99.63% using the LFW (Labeled Faces in the Wild) database with face alignment [2]. This result is considered superior to the human ability to recognize people. Since then, many companies have started offering this service, using similar technology, which has made facial recognition something present in people's daily lives.

However, this article goes a step further and presents a complete architecture of a facial recognition system in a smart city context. This architecture includes a series of services applied to the area of education and public safety using low-cost equipment and a modern web development paradigm.

Furthermore, a smart city is essentially a connected city, in which any device can connect to the network anytime, anywhere. In addition, the mobile ecosystem is currently heterogeneous and comprises a multitude of networks with different technologies, such as 5G, LTE, Wi-Fi, Bluetooth, and WiMax. The heterogeneity of a wireless environment offers the opportunity to evaluate and selects the best network among several possible options, to best meet the requirements of the different smart applications incorporated in the day-to-day life of cities, such as the smart application presented in this article. Therefore, the issue addressed in [3] will also be considered in the implementation of the architecture described in this paper to guarantee the quality of the connection and at the same time optimize energy consumption.

The next topics are distributed as follows: Sections II briefly addresses the related works, III explains the proposed architecture, IV details the first application that will be implemented, V shows the test results of the first application, and VI presents the conclusions.

II. Related Work

The FaceNet system was a milestone in the history of facial recognition and even today remains a major reference in this area. [2] exposes the main characteristics of this system. According to that article, FaceNet is a unified system for facial verification, recognition, and clustering. In this context, verifying means examining whether two images are

of the same person. To recognize, in turn, is to say who is the person that appears in a given image. Clustering, on the other hand, is to find similar people in a set of images. FaceNet uses a CNN to build a vector of the most significant points on the face and then uses the Euclidean distance to perform the three tasks already mentioned. In this way, verification becomes the task of examining whether a given Euclidean distance is within a threshold. The recognition turns into a classification process using the KNN (K Nearest Neighbor) algorithm. Clustering, on the other hand, becomes a simple task that can be done with the aid of the k-means algorithm. The databases used in the tests were LFW and YouTube Faces. The first is an image database and the second is a video database. The results were as follows: with LFW, without face alignment, 98.87 accuracy. With LFW, with face alignment, 99.63 accuracy. With YouTube Faces, using 100 frames, 95.12% accuracy. With YouTube Face, using 1000 frames, 95.18% accuracy.

The HyperFace project, presented in [4], is a system based on a CNN capable of detecting human faces, locating the main landmarks of the face, estimating the position of the person's head and differentiating between male and female genders. HyperFace assumes that the learning of a CNN is hierarchical, that is, the shallower layers detect simpler features and can be better used to locate reference points and to estimate the position of a head. The deeper layers, in turn, apprehend more general characteristics and can be better used in face detection and gender distinction tasks. Therefore, the researchers built a standard CNN and then extracted features from different layers of this network. But to realize multitasking learning, the extracted features from different layers were merged into a new CNN. This new network, finally, was trained simultaneously with different error functions, one for each type of task. The authors reached two important conclusions: 1) Merging intermediate layers improved the performance of all tasks. 2) All tasks benefited from using the multitasking learning framework.

[5] proposes an integrated IoT (Internet of Things) environment based on RFID (Radio-Frequency Identification) and sensor networks. They also propose an architecture to articulate the varied uses of RFID and sensor networks in different areas, such as: health, education, security, climate monitoring, etc. As an addition, they suggest a monitoring and processing center planned to use cloud computing. The proposed architecture consists of four layers, namely: 1st) Perception layer (devices, sensors, and RFID tags). 2nd) Data conversion layer (Middle ware). 3rd) Network layer (routers, gateways, and switches). 4th) Application layer (intelligent services and applications from different domains). The main challenges pointed out by the developers of this system are related to compatibility and security. They conclude that the said system could serve the general public and can be implemented in different regions.

The study presented in [6] analyzes the influence of people's age and the passage of time in the facial recognition process. It used the COTS-A and FaceNet systems to carry out the recognitions, along with a database consisting of images of children and adolescents from 2 to 18 years old. The best results were obtained with the combination of the two systems. For a one-year time lapse between the training image set and the test image set, a recognition accuracy of 90.18% was achieved. The authors also indicated an 80% recognition accuracy with a time lapse of 2.5 years between the training image set and the test data set.

There are studies [7] showing means to improve the training of unbalanced classes. It indicates two traditional strategies: resampling and cost-sensitive learning. The first consists of undersampling the majority class or oversampling the minority class. The second assigns a higher cost to the minority class misclassification. The authors propose a new method that divides the database into clusters within classes and between classes, which results in balanced class boundaries. Experiments were done with facial recognition and prediction of an attribute on a face. Results similar to the traditional ones were found.

The system proposed here presents an accuracy very close to the system created by [2] with the advantage of being a complete system (with local, web and mobile application) in the context of a smart city and not just an isolated facial recognition system. The proposed system also uses a simpler and more efficient network than the one presented in [4]. In relation to the work published in [5], the present work innovates by replacing RFID monitoring with facial recognition which, in this context, proves to be safer and more reliable. It was also decided to use the resampling technique in the system tests, which is simpler and presents results like the technique proposed by [7]. Finally, the work presented by [6] warns of the need to annually update the images used for system training, which will be done when students enroll.

III. Overview Of The Student Path Monitoring Platform

Fig. 1 shows the overview of the student trajectory monitoring platform. It is possible to see the actors involved, the main applications and the technologies used. From left to right, the “User Registration” appears first. It is a web application developed with the React.js framework that will be used by system administrators to register students from public schools. In this register, the personal data of each student and some photos captured on the spot by a webcam will be inserted. Following just below, the “Model Training” appears, which is a microservice developed with the Python language. This microservice takes photos of students as input and makes use of a series of artificial intelligence algorithms to create a model capable of recognizing the faces of registered students. Next, the model training result is stored in MongoDB, which is a non-relational database.

At the core of the architecture is the Back-end Server that performs the main functions that make possible the coordinated functioning of the various parts of the system. Among these functions, the following stand out: controlling access to the database, sending notifications to users, and communicating with web services. The Back-end Server uses the JavaScript language and the Node.js framework. Next to the Server is the “System Management” which is a web application also implemented with the React.js framework. “System management” oversees the entire system and, among other things, creates administrator profiles.

At the top center of Fig. 1, the first application can be observed: “App for controlling access to the school bus through facial recognition”. This application will be the first to be deployed and aims to monitor the use of the school bus. How this application works will be detailed in a later topic.

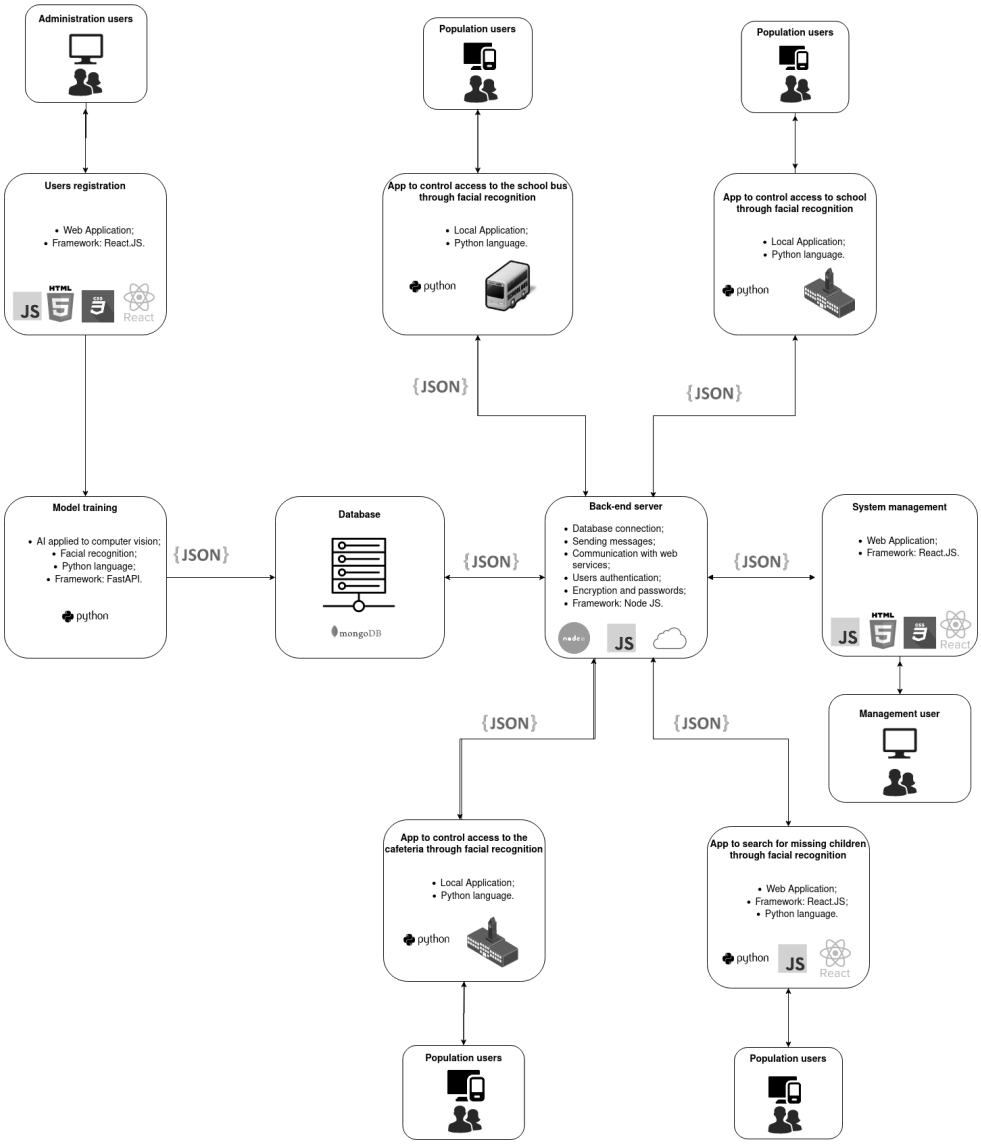

Fig. 1. Facial recognition system architecture.

At the top right is the application “School access control app through facial recognition”. This application uses the Python language and runs on a local device. It has several hardware and software similarities with the first application. The main difference is in the interaction with other actors such as the municipal education department, the Guardianship Council, and the Prosecution Office. The education department will receive a report on student attendance. The Guardianship Council and the Prosecution Office, upon prior registration, will obtain a report with the names of students who have been absent three times or more in a month. In this way, the referred application will not only be a control of the entrance of students, but also be an instrument to mitigate school dropout.

At the bottom left one can see the “App for controlling access to the cafeteria through facial recognition”. This is also an application developed in Python that runs on a local device. In addition to controlling access to the cafeteria, this application, through the Back-end Server, notifies those responsible when a student enters the cafeteria to eat school lunch and can also be used by the education department to control the lunch stock.

Finally, at the bottom right appears the “App to search for missing children through facial recognition”. This is a web application programmed in Python that makes use of the model trained with the images of students from public schools and through security cameras installed in public places of the Smart City to seek to recognize missing children.

IV. APPLICATION OF SCHOOL BUS ACCESS CONTROL THROUGH FACIAL RECOGNITION

Fig. 2 shows all the steps of the school bus access control system. Initially a person tries to get on the bus. At that moment, a camera strategically placed inside the bus captures the image of that person's face. Then, the captured image is processed by an intelligent algorithm written in the Python language and embedded in a Raspberry Pi 4 installed on the bus. The algorithm capable of recognizing faces checks if it belongs to a previously registered student. If the recognition is positive, the turnstile is released for the student to enter. At the same time, the Raspberry Pi sends the information about the student boarding the bus to a server on the Internet, containing a timestamp with date and time. Subsequently, the Internet service sends the student's entry information to the person in charge who was also previously registered. The person in charge receives the notification directly in the application developed for smartphones, which today is the main way to access the Internet in Brazil, widely used by practically all social classes. Finally, Fig. 2 illustrates the application for the Internet that manages the facial recognition system and whose main function is to register students and their respective guardians.

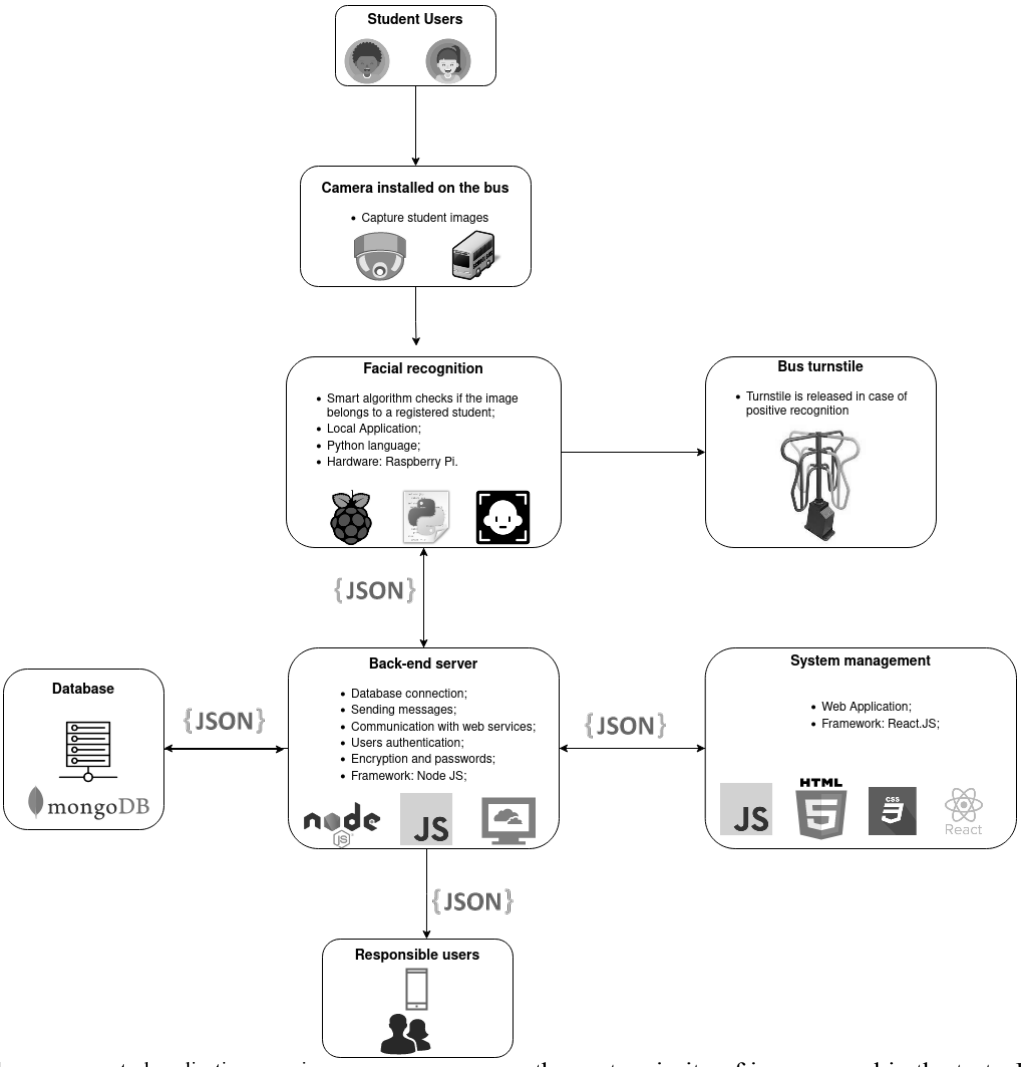

Fig. 2. School bus access control application overview.

Fig. 3 shows the four main steps of the facial recognition process [2]. In the first step, the HOG algorithm is used to search the input image for a pattern corresponding to a human face. If this pattern is found, a rectangle is drawn to delimit the space where the face is located. Fig. 4 (left) shows an example output from the first step, with the face delimited. The image belongs to the LFW public dataset.

In the second step, the HOG and SVM algorithms are used together. The objective of this step is to detect, in the previously delimited space of the face, 68 reference points. Fig. 4 (right) shows an example of the second step output, with the reference points marked in gray.

In the third step, the 68 facial landmarks points are used as input to a previously trained artificial neural network. It is a CNN of the ResNet type [1]. This neural network generates in its output a facial descriptor that is a numerical vector of 128 positions that geometrically describes the main features of the face. This step also saves a file with the descriptors of each image and another file with the respective image labels.

In the fourth step, the KNN algorithm is used, which works as follows: Initially, a facial descriptor is generated for the test image. Then, the Euclidean distance between the test image descriptor and each of the training image descriptors is calculated [8]. Finally, the smallest distance is selected and if this distance is within a previously established threshold, the test image receives the same label as the corresponding training image. In successive experiments, a threshold of 0.5 proved to be sufficient to accurately determine the identity of the vast majority of images used in the tests. Fig. 4 shows an example of the fourth step output. The number in the image represents the Euclidean distance.

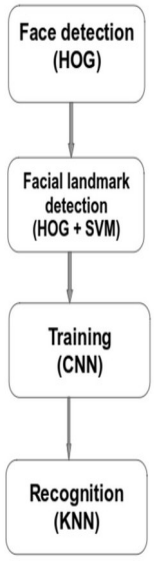

Fig. 3. Recognition process pipeline.

Fig. 4. Example of detected face, facial landmarks and recognized face

V. RESULTS

The LFW database was used in the first test of the facial recognition system. This base was chosen because it is widely used in benchmarks and because it is frequently cited in publications in facial recognition. It is a public database created by the University of Massachusetts from images of famous people taken from the Internet. The base consists of images of politicians, sportsmen, artists, and businessmen from different parts of the world, men, and women of different ethnic groups. In most of the images people are not posing for the photos, but just being themselves. Altogether there are 13233 images of 5749 people, however, most people only have one or two images.

Initially, a subset was separated with those with 7 or more images, which totals 158 people. Then this subset was divided into two categories, one for training and the other for testing. The training set has between 5 to 10 images of each person. The test set, on the other hand, has 2 other images of each of the 158 people. All images were labeled with the person's first name, last name, and a sequential number. The result obtained in this test was the following: 310 correct answers from a total of 316 images tested, which means an accuracy of 98.10%. This result is very close to the one presented in [2].

Table I presents the main metrics related to the first test. To facilitate the visualization, only the first 16 classes and the final averages are shown. The weighted average accuracy (precision) for all classes is 99%. The weighted average recall for all classes is 98%. Finally, the weighted average f1-score for all classes is 98%.

Precision represents the proportion of true positives and false positives, so a high value of precision presupposes a low value of false positives.

The recall, in turn, indicates the proportion between true positives and false negatives, so a high recall implies a low value of false negatives.

On the other hand, the f1-score is a harmonic average of the two previous metrics that will be high only if both accuracy and recall are high.

TABLE I. SUMMARY OF THE MAIN CLASSIFICATION METRICS.

	Precision	Recall	F1-score	Support
Abdullah Gul	1.00	1.00	1.00	2
Adrien Brody	1.00	1.00	1.00	2
Alejandro Toledo	1.00	1.00	1.00	2
Alvaro Uribe	1.00	1.00	1.00	2
Amelie Mauresmo	1.00	1.00	1.00	2
Andre Agassi	0.67	1.00	0.80	2
Andy Roddick	1.00	0.50	0.67	2
Angelina Jolie	1.00	1.00	1.00	2
Ann Veneman	1.00	1.00	1.00	2
Anna Kournikova	1.00	1.00	1.00	2
Ari Fleischer	1.00	1.00	1.00	2
Ariel Sharon	1.00	1.00	1.00	2
Arnold Schwarzenegger	0.67	1.00	0.80	2
Atal Bihari	1.00	1.00	1.00	2
Bill Clinton	1.00	1.00	1.00	2
Bill Gates	1.00	1.00	1.00	2
Accuracy			0.98	316
Macro avg	0.98	0.97	0.97	316
Weighted avg	0.99	0.98	0.98	316

In Fig. 5 there are some examples of people who were recognized by the proposed system. It is possible to see that the system recognizes faces at different angles and with different expressions.

In a second test, the system was trained with photos from personal files and then tested using a webcam on an Acer Nitro 5 laptop with an Intel core i7 processor and 8 GB of RAM, the same used in the first test. In both tests, each recognition occurred in a fraction of a second, showing that the system is efficient and effective.

Fig. 5. Examples of recognized people.

Finally, the facial recognition system was tested on a Raspberry Pi 4 with an ARM Cortex-A72 processor, 4 GB of RAM and a camera module developed for the Raspberry Pi. In this test, the recognition was made with the help of the camera module and used to activate a relay module that in the final product will activate the electric lock of the school bus turnstile. Also in this test, the recognition was performed in a fraction of a second and the relay module was activated correctly.

VI. Conclusions

The recognition system proposed here achieved, in the tests carried out, excellent results in all metrics. It also showed to be able to recognize people with different facial expressions and at different angles. The recognition takes fractions of a second, which demonstrates that the referred system can be incorporated into the school routine without causing delays in boarding buses or entering the school. The hardware used has small dimensions and low power consumption and can be installed almost anywhere. It already has both Wi-Fi and Bluetooth connections and can also receive a module with GPS and GSM enabling the tracking of school buses in real time.

The monitoring platform can be accessed through an Internet browser or through mobile devices allowing both parents and school management to monitor students' journeys from home to school and from school to home. This information may also be automatically passed on to other competent authorities such as the Prosecution Office and the Guardianship Council. Thus, the system can be used as a tool to mitigate school dropout and the disappearance of children and adolescents.

Finally, the first application, that is, the application of access control to the school bus through facial recognition, is in line with all the requirements presented by the stakeholders and is at maturity level TRL 6, that is, it has already been demonstrated the validity of critical prototype functions in relevant environments.

Acknowledgment

This work was supported by the Coordination for the Improvement of Higher Education Personnel—CAPES, the National Council for Scientific and Technological Development— CNPq, the Municipal Fund for Sustainable Development of Canaã dos Carajás —FMDS, and the Support Program for Qualified Production—PROPESP/UFPA (PAPQ) (notice 02/2022). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] F. Chollet, Deep learning with python, Shelter Island, 2018.

[2] D. Schroff, D. Kalenichenko, J. Philbin, “Facenet: a unified embedding for face recognition and clustering”, IEEE explore, pp. 815-823, 2015.

[3] T. Coqueiro, J. Jailton , T. Carvalho, and R. Francês,"A fuzzy logic system for vertical handover and maximizing battery lifetime in heterogeneous wireless multimedia networks", Hindawi Wireless Communications and Mobile Computing, Volume 2019.

[4] R. Ranjan, V. Patel, R. Chellappa, “Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 121-135, 2019.

[5] V. Jerald, A. Rabara, D. Bai, “Internet of things (IOT) based smart environment integrating various business applications”, International Journal of Computer Applications, pp. 32-37, 2015.

[6] D. Deb, N. Nain, K. Jain, "Longitudinal study of child face recognition", International Conference on Biometrics, pp. 225-232, 2018.

[7] K. Huang, Y. Li, C. Loy, and X. Tang, “Deep imbalanced learning for face recognition and attribute prediction”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2781-2794, 2020.

[8] F . Provost, T. Fawcett, Data science para negócios. Alta Books, 2016.