

Contamination risk estimation model for respiratory diseases in monitored environments using YOLOv5

Flávio Rafael Trindade Moura¹, Diego de Lemos Brito da Silva², Diego Lisboa Cardoso¹, Karla Tereza Figueiredo Leite², Harold Dias de Mello Junior², Fernando Augusto Ribeiro Costa¹, Marcos César da Rocha Seruffo¹

¹Universidade Federal do Pará, ²Universidade do Estado do Rio de Janeiro

flavio.moura@itec.ufpa.br, lemos.diego.07@gmail.com,
dlisboacardoso@gmail.com, karla.figueiredo@gmail.com,
harold.dias@gmail.com, fernando.costa@naea.ufpa.br,
marcos.seruffo@gmail.com

Resumo. *As soluções de monitoramento de ambientes, a partir de imagens de câmeras, cresceram principalmente com a integração com modelos de aprendizado de máquina. Com a pandemia do COVID-19, várias medidas de prevenção e combate ao coronavírus foram adotadas, sendo o distanciamento social e o uso de máscaras fatores importantes no controle da propagação da doença. Este artigo apresenta uma nova abordagem baseada em redes neurais convolucionais focadas em doenças transmitidas pelo ar, que avalia a taxa de infecção do espaço monitorado, em tempo real. Para isso, através do ambiente monitorado por vídeo, são (i) identificadas pessoas utilizando máscara de forma correta ou incorreta; (ii) identificadas pessoas sem máscara; (iii) medidas as distâncias entre as pessoas. A partir dessas informações, é possível inferir (iv) um índice para medir as chances de infecção em um espaço monitorado ao longo do tempo, utilizando YOLOv5 neste ambiente de desenvolvimento.*

Abstract. *Monitoring solutions for environments, from camera images, have grown mainly with the integration with machine learning models. With the COVID-19 pandemic, several measures to prevent and combat the coronavirus have been adopted, with social distancing and the use of masking being major factors in controlling the spread of the disease. This paper presents a new approach based on convolutional neural networks focused on airborne diseases, which assesses the infection rate of the monitored space, in real time. For this, through the video monitored environment, are (i) identified people with masks used correctly or incorrectly; (ii) identified people without masks; (iii) measured the distances between people. From this information, it is possible to infer (iv) an index to measure the chances of infection in a monitored space over time, using YOLOv5 in this development environment.*

1. Introduction

Epidemiological outbreaks have gained spotlight in recent years, because there have been several cases of highly contagious diseases, such as H1N1 in 2009, Ebola in 2013 and, more recently, deserving greater prominence the pandemic caused by the SARS-CoV-2 virus originated in China [1], which broke out in early 2020 in Brazil, leading to several studies such as [2] about how the virus spreads and ways to control its contagion. From this, prevention measures for the general public were determined by the World Health Organization (WHO) [3], such as the use of masks, frequent hand hygiene, and social distancing, which are more effective ways to prevent the spread of the virus.

According to the WHO [4], by November 4, 2021, some 22 months after the first cases of COVID-19 and 11 months after the initiation of the first vaccination program

by the United Kingdom, more than 5 million confirmed deaths had been counted, in addition to an alarming number of suspected and unconfirmed deaths. And on the same date, 56 countries in various regions still reported significant increases in deaths due to COVID-19, ICU bed shortages, lack of supplies, overworked health care workers, and hospitals postponing other important procedures.

Given the situation of COVID-19 exposed, the volume of countries that, even after the development of vaccines and treatments, present crisis situations in the public administration of health is worrying. Therefore, the chances of the development of any other infectious diseases can become timely. Thus, it is crucial to maintain efforts in developing tools to ensure the measures to prevent and combat COVID-19 and possible strains that may develop, in addition to providing defensive artifices for possible endemic or pandemic crises of any other outbreaks or diseases.

With all this in mind, one area that has a lot to offer in supporting the prevention of the spread of COVID-19 is the field of video analysis, which is a rapidly developing area and presents great results in terms of security and monitoring due to the development of computational techniques aimed at using artificial intelligence, such as models based on neural networks. With the use of these techniques it is possible to make inferences in videos that may be useful for the objective of this work.

Therefore, this paper aims to demonstrate the use of these techniques for the development of a system that determines the security level of environments monitored by video and analyzed through a model based on convolutional networks that will provide statistical information for the monitoring of the chosen location through a dashboard available via web.

From the evaluation of the prevention measures applied, such as social distancing and the use of masks, the proposed system provides a measure of risk to individuals who are present in the environment. This helps an individual to monitor his own behavior and that of others, and allows him to choose the environments that can be frequented with less risk of contamination.

The remainder of this article is organized into five sections. Section 2 presents a literature review with the most recent and important works on this application and discusses in detail the methods used as a basis for this article. In Section 3, the methodology adopted for the development of the application is presented, divided into the subsections of the database, neural network training, and application of the trained model. In Section 4 the results and discussions about the developed application, such as the accuracy of the trained neural networks and the contamination risk estimation model is shown. Section 5 concludes and indicates points to be implemented in future versions of the application.

2. Related Work

The advent of new technologies in the field of Artificial Intelligence (AI), especially in the area of image detection and recognition, has allowed for the evolution of the monitoring and security area. The latter has been used for various purposes and this includes intelligent ways to mitigate the spread of respiratory diseases such as that caused by SARS-CoV-2.

Several publications have been released with the purpose of fighting the coronavirus pandemic, such as [5] that proposed the replacement of visual inspection of mask use in public places by a mask detection system based on deep learning through the use of the YOLOv5 tool [7], such that the developed system obtained an accuracy of 97.9%. Similarly, [6] and [8] address methodologies for detecting mask use and social distancing using deep learning methods in the process, but in both works, the detections are not done simultaneously.

Unlike the previous ones, [9] does these checks together using the MobilenetV2 pre-trained model, which proposes use of an alarm if any of the security measures are not being satisfied. However, for larger scale uses would require an increase in computational resources, presenting difficulties for real-time monitoring.

The works of [10] and [11] propose the detection of people from a top perspective to use this information to determine social distance by Euclidean distance; in the first paper, a pre-trained YOLOv3 model was used and the tests were performed with the IRIS PX4 model drone on a Gazebo simulator, the system showed an accuracy of 90% for detecting various levels of crowding, while the second used region-based convolutional neural networks (Faster R-CNN) to make the detections of people, in addition to using Euclidean distance per pair of people detected in test images, with an accuracy of 96%.

On the other hand, [12] developed a hybrid model of Computer Vision and the pre-trained YOLOv4 model that is used to detect people in open and closed places using CCTV surveillance cameras, and the results obtained were superior to 3 state-of-the-art methods; also a risk analysis based on the space and time analysis of the data collected through the trajectory of movements and the rate of violation of the minimum distance by people captured in video was proposed.

Taking into consideration the objective of quantifying the risks involved in airborne virus transmission, we can see some specific studies such as in [14] that performed a field analysis in 12 salons in New York City, where estimated values were used for the airflow of the locations, quanta generation rate specific to the COVID-19 virus and the Wells-Riley equation to assess the risk of contamination in stationary and non-stationary cases; the factor of mask wearing is also taken into consideration during the risk assessment.

This paper aims to demonstrate the creation of a system capable of assessing the risk of contamination of environments by video, through the data collected by the image detection and recognition model. For this, a tactical dashboard (Figure 1) has been developed that can be made available online for practical decision making, providing statistics and pertinent information to increase the sense of security for the individuals present.

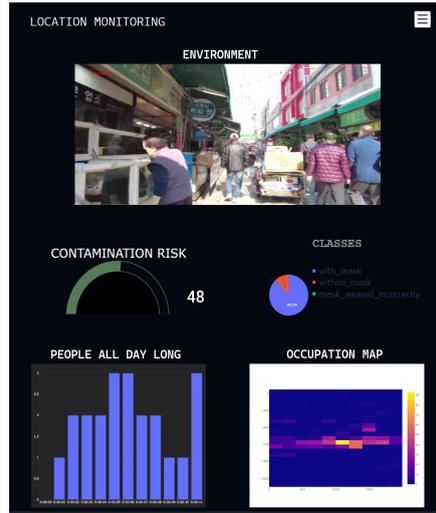


Fig. 1. Tactical dashboard containing the estimated contamination risk, count of detected classes, people detected during detection and heat map to monitor the occupation of monitored space.

3. Methodology

For this paper, neural network models were developed based on the requirement of online evaluation, which requires high processing speed, to enable reading environments in real time. Therefore, for each available version of YOLOv5, neural networks were trained using the database created.

Thus, inferences are made, as described in Figure 2-A, about the distance between people in the environment and the mask use, according to the database classifications described in Section 3.1. Then, the data obtained are used to estimate the risk of local contamination, generating several graphs to simplify the visualization of information, as exemplified in Figures 2-B and 2-C.

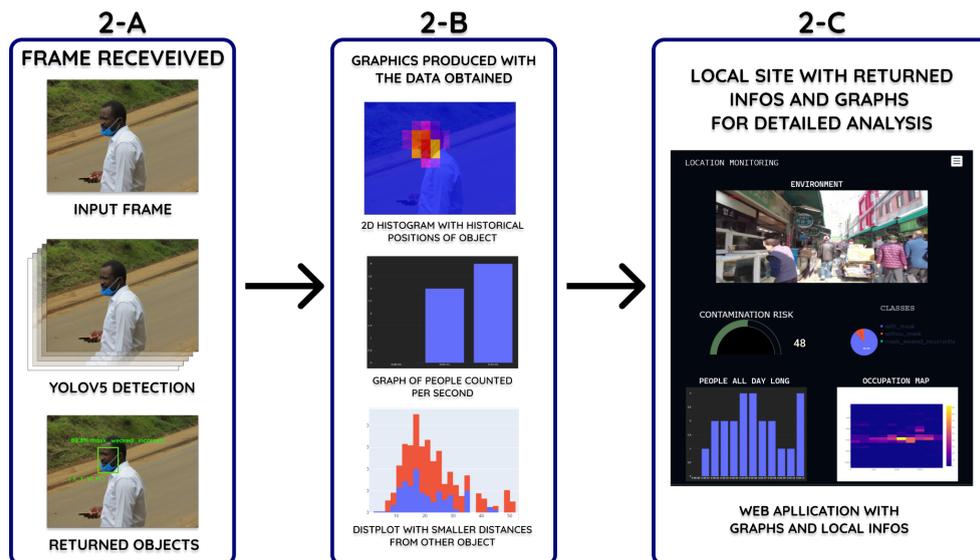


Fig. 2. Schematic of the analysis process and display of results for each frame. Such that in the first block the detection process with Yolo is described, in the second block the graphics generated to visualize the information obtained and in the third block the tactical dashboard accessible online.

3.1. Database

The database used in this article was produced independently, combining several datasets available in open repositories of face mask detection¹²³⁴ with people of different ethnicities and age groups, including the data created by [12]. The images that make up the dataset were separated into 3 classes, as shown in Table 1:

Table 1. Containing descriptions of each class of the dataset.

<i>with mask</i>	People wearing masks correctly.
<i>without mask</i>	People without masks.
<i>mask used incorrectly</i>	People using a mask incorrectly (not covering the nose and/or mouth).

There are about 14,864 annotations of 9,947 images, made in the Roboflow visual computing tool, and the images are separated into 70% for training, 20% for validation and 10% for testing. This same tool was used to pre-process the images, using the data augmentation methods, where rotated and mirrored copies of the images are created to increase the size of the dataset. These strategies were adopted so that the database would provide the neural networks with enough data to improve the generalization ability of the models.

3.2. Training

To create an expert model for mask identification, the YOLOv5 algorithm was used and trained with the database mentioned in the previous topic. After the training process, the models with the best weight settings pointed out in the validation stage, where a portion of the image base is separated and presented to the model for accuracy evaluation, were selected and saved to be used later in the application.

3.3. Application

Object detection is done using the model with the best weights in the YOLOv5 training step. Frame-by-frame analysis of the video chosen for testing is done in a later step, following the sequence in Figure 3.

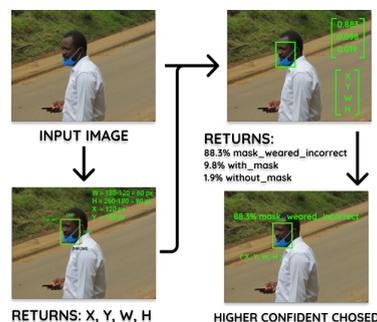


Fig. 3. Method of detection and analysis for every frame with YoloV5.

¹ <https://www.kaggle.com/mloey1/medical-face-mask-detection-dataset>

² <https://www.kaggle.com/datasets/andrewmvd/face-mask-detection>

³ <https://humansintheloop.org/resources/datasets/medical-mask-dataset/>

⁴ <https://www.kaggle.com/datasets/prithwirajmitra/covid-face-mask-detection-dataset>

The frame is analyzed by the neural network and two outputs are obtained: the position vectors of the detected faces and their respective degrees of confidence for each established class, where the class with the highest reliability is selected. With this, the risk of contamination is estimated, and then the data is visualized through the Dash⁵ and Plotly⁶ libraries, which structure the tactical dashboard of the monitored space in real time.

Furthermore, through the Streamlink⁷ library, it is possible to analyze any environment monitored by video and accessible via Uniform Resource Locator (URL), using local parameters provided by the user and allowing convenience in structuring a network that provides access to information to any user in a simplified manner.

After determining the position of each person present in the frame, it is necessary to determine the distances in which people are from each other. To do this, the distance between the centers of the bounding boxes surrounding the people in the frame is used, but the output value is measured in pixels. The following equation, which uses the width of a person's face (defined as 15cm [13]) as a reference, was used to convert the distances from pixels to meters:

$$Dist = \frac{\sqrt{(X_{c1} - X_{c2})^2 + (Y_{c1} - Y_{c2})^2} * 0.15}{W_1 + W_2}, \quad (1)$$

where X_c = horizontal position of the center of the object, Y_c = vertical position of the center of the object, W = width of the object and 0.15 is a predefined face width, retired from [13].

Regarding the quantification of the risk of contamination, the methodology was followed, mentioned in Section 2, established in [15], reformulated in [14] and schematized in Figure 4. Then, for the applied methodology, the estimation of the airflow rate is fundamental. In cases of environments with sensors, to determine the airflow rate in real time it is possible to perform online inference of the risk of contamination. However, in cases where sensing is not possible, the following equation can be used to estimate the airflow rate per person found in the scene:

$$Q = \left(\frac{N}{C_s - C_o} \right) \times 10^6, \quad (2)$$

where Q = outdoor airflow rate per person (m^3/s), N = CO2 generation rate per person, assumed as $0.0000052 m^3/s$ for a female aged 21 to 30 years performing light work [14], C_s = CO2 average concentration in the space (ppm), and C_o = CO2 concentration in outdoor air (the average measured outdoor CO2 concentration is 419 ppm).

Once this value is obtained, it is multiplied by the number of people identified in the location to determine the total airflow rate.

⁵ <https://dash.plotly.com>

⁶ <https://plotly.com>

⁷ <https://pypi.org/project/streamlink/>

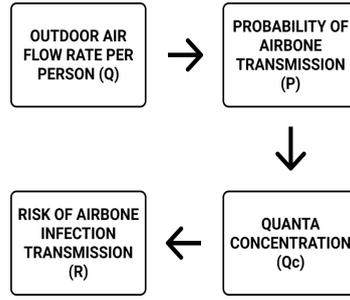


Fig. 4. Schematic of the contamination risk quantification method.

The next step in quantifying the contamination risk is to determine the probability of contamination at the location under stationary conditions. For this, the Wells-Riley equation is used [16]

$$P \times 100 = 1 - \exp\left(-\frac{I \cdot q \cdot iR \cdot t}{Q}\right), \quad (3)$$

where P = probability of airborne infection transmission, I = number of infected individuals (assumed as one (1) in this study), q = quanta generation rate (2.3666/min for SARS-CoV-2, as reported by [11]), iR = inhalation rate (0.016 m³/min), t = time (min), and Q = outdoor airflow rate (m³/min).

However, in real situations, people carrying the virus may be in transit in the location used as reference. Thus it is necessary to consider non-stationary conditions of contagion, through Equation 4, which updates the concentration of *quanta* in relation to the contact time between the analyzed people:

$$q_c = \frac{q}{Q} \left[1 - \exp\left(\frac{-Q \cdot t}{V}\right)\right], \quad (4)$$

where q_c = quanta concentration (quanta/m³), q = quanta generation rate (2,3666/min), Q = outdoor airflow rate (m³/min), t = time (min), and V = volume of the local (m³).

It is also worth mentioning that the use or not of masks on people captured in the video is considered, which significantly impacts the possibility of air transmission. In [17], a reduction in quanta generated ranging from 56% to about 76% was observed in more controlled cases, therefore, an average reduction of 60% was considered, close to the minimum observed, in the rate of generation of individual quanta for a person wearing a mask, and then an average was generated according to the number of individuals in the place.

Finally, the final risk is calculated, with Equation 5, which takes into account all the terms calculated previously, directly or indirectly, to determine a percentage of contamination risk:

$$R(\%) = 100 \times [1 - \exp(-P \cdot t \cdot q_c)], \quad (5)$$

where R = risk of airborne infection transmission, P = probability of airborne infection transmission, q_c = quanta concentration (quanta/m³), t = time (hour).

4. Results and discussion

Two factors were crucial to the application's performance: i) the detection of people and the classification regarding the use of masks, described in Section 4.1 about the

hyperparameters used, the accuracy of the models trained, and aspects taken into account to produce the dataset; and ii) the contamination risk estimation model, described in Section 4.2 through the results obtained from a one-week case study.

4.1. Detection and classification

For detection and classification of people regarding mask use, the YOLOv5 model was used, as described in Sections 3.1 and 3.2, trained with a customized image bank with three classes: with mask, without mask, and incorrectly used masks. Additionally, the neural network versions available from Ultralytics⁸ were trained and implemented: YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5xl, YOLOv5s6, YOLOv5m6 and YOLOv5l6 in order to determine the best models in terms of their accuracy.

For these versions used, we standardized the input hyperparameters for training the neural networks, as described in Table 3:

Table 3. Hyperparameters used in trained versions of the YOLOv5 neural network.

Hyperparameter	Description	Value
Image size	Image size in pixels	400x400
Batch size	Number of samples to be worked on before updating the internal model parameters	64
Epochs	Number of times the learning algorithm will run on the entire training data set	100

The best accuracies were obtained with versions YOLOv5s6 and YOLOv5m6, with rates significantly higher than with versions YOLOv5l6 and YOLOv5xl. As for the accuracy values for each class presented in the output, Figures 5 and 6 detail the results for versions YOLOv5s6 and YOLOv5m6. Therefore, even for the lowest accuracy values for a class, the models still manage to have a relatively high average accuracy value.

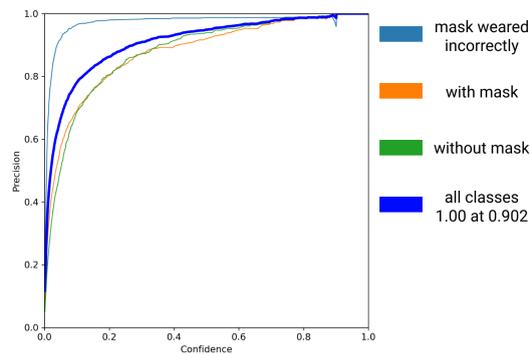


Fig. 5. Graph Precision Vs Confidence for each class of neural network based in YOLOv5s6, where precision refers to the ability to get the same results on different reads and confidentiality refers to the chance the results are correct.

⁸ <https://ultralytics.com>

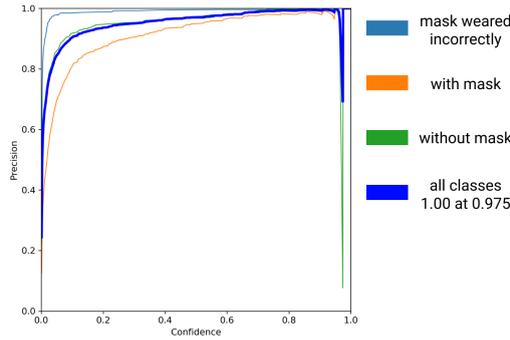


Fig. 6. Graph Precision Vs Confidence for each class of neural network based in YOLOv5m6, where precision refers to the ability to get the same results on different reads and confidentiality refers to the chance the results are correct.

However, the blue curve observed in Figure 6, which represents Accuracy x Confidence for all classes, demonstrates low accuracy of the YOLOv5m6 neural network for high confidences. By analyzing the confusion matrices, presented in Figure 7, an accuracy above 0.9 (90%) is verified for all classes, as observed in the ratio of predicted to true for each class.

Figure 7 represents a confusion matrix of the set of classes trained on the YOLOv5s6 model. There one can observe the good accuracy obtained with the model by looking at the main diagonal of the confusion matrix where all the accuracies are greater than 90%. More information about the results is available in the repository [18]. The application used in this paper is open and can be accessed through [19].

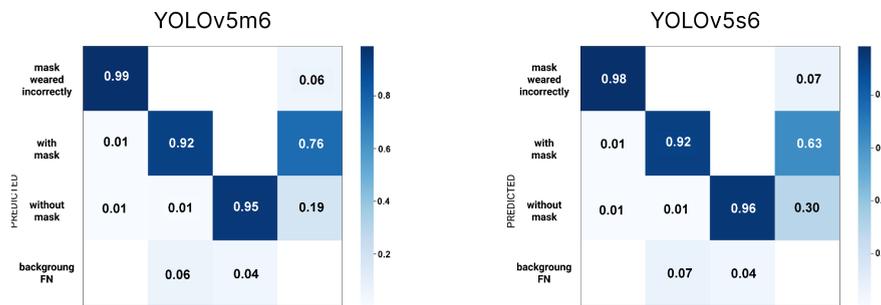


Fig. 7. Confusion matrices for the outputs of the YOLOv5s6 and YOLOv5m6 models.

4.2. Risk Quantification

For this stage of the project a case study was performed using the methodology proposed in [4] and presented in Section 3.3 of this article. So, for this study, it was used as a parameter an office room with a volume of 180 m³ that is used in a typical way and daily in a journey of 8 hours/day. The monitoring was done for 5 days, taking into consideration the occupation time, the number of people present in the room, and whether or not a mask was worn. The final result is presented as a calculated contamination risk rate for the environment, in percentage values.

Observing the results obtained in Figures 9, 10 and 11, we can initially notice that, as time passes, the risk of contamination increases regardless of the number of people and the use of masks. Because the *quanta* concentration at the location increases to very high values and, consequently, the chances of contamination also increase.

Furthermore, the data from the first hours of measurement reveal that in the first hours of analysis the wearing of a mask has a strong influence on the risk of infection. For example, when looking at Figure 11, during the first two hours of day 5, it can be seen that the risk of infection is only 8%, even if there are 8 people in the office.

On the other hand, it is also possible to see that the lack of masking greatly increases the risk of infection. This can be seen using Figure 9 as an example, where although the crowding during the first two hours of the day is the same as in Figure 11, the infection risks differ due to the number of people wearing masks. Therefore, at the time when only 5 people were wearing masks, the risk of contamination rises to 18.3%. Yet another notable point in the data is the fact that even with a relatively small number of people, if a large proportion of the individuals are without masks, the chance of contamination increases considerably. Similar to this, in the case presented in Figure 10, the first risk calculation showed a rate of 33.7% for a number of 7 people with only 2 wearing masks.

Therefore, Figures 9, 10 and 11 demonstrate consistent results regarding the general knowledge about the modes of transmission and ways to mitigate the spread of respiratory diseases.

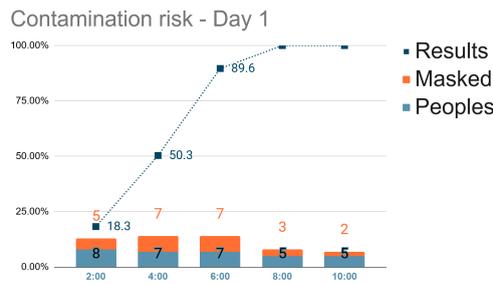


Fig. 9. Graph of contamination risk for the analyzed Scenario for the first day.

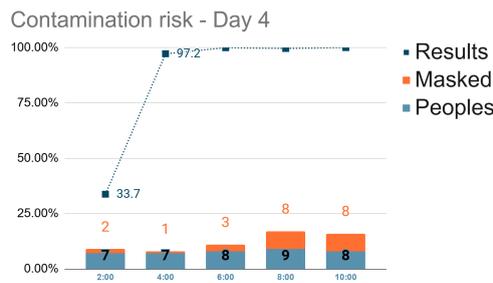


Fig. 10. Graph of contamination risk for the analyzed scenario for the fourth day.

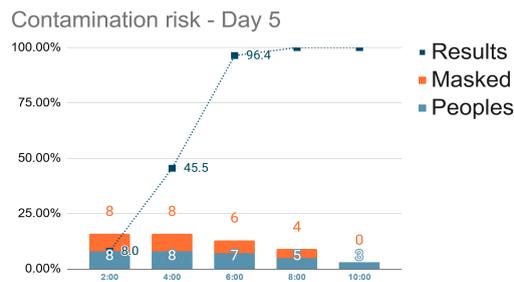


Fig. 11. Graph of contamination risk for the analyzed Scenario for the fifth day.

5. Conclusion

In this work two main points were analyzed: detection and classification of people regarding the use of masks and the definition of a contamination risk factor in a monitored environment and, for both objectives, the methods applied proved satisfactory.

In relation to the classification regarding the use of the mask, the neural network models, built with YOLOv5 and trained with the custom database, presented accuracy above 90% for all classes, besides presenting a relatively fast processing, which allows it to be applied for real-time classification.

In quantifying the environment's risk, the results obtained were in line with what is known about the main risk factors that can lead to the contamination of individuals, such as not wearing masks and a large concentration of people in small spaces, indicating that it can be a useful tool for assessing the hazard of the monitored site.

Thus, considering the results and the characteristics presented in the case studies, the environment control dashboard proved to be a fully viable tool to assist in decision making regarding the safety of the monitored people.

Regarding future works, the dashboard could be updated with models capable of detecting people with profile and back. As for quantification, the application could be used in combination with other tools, such as thermal cameras, to help detect sick people, instead of using empirically defined values.

References

- [1] WORLD HEALTH ORGANIZATION et al. WHO-convened global study of origins of SARS-CoV-2: China Part. 2021.
- [2] GÜNER, HATİCE RAHMET; HASANOĞLU, İmran; AKTAŞ, Firdevs. COVID-19: Prevention and control measures in community. *Turkish Journal of Medical Sciences*, v. 50, n. SI-1, p. 571-577, 2020.
- [3] Advice for the public: Coronavirus disease (COVID-19). World Health Organization, 2021. Available in: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>. Accessed on: 22 of Dec. 2021.
- [4] WHO Generals and Directors Speeches. Opening Remarks at the Media Briefing on COVID-19; WHO Generals and Directors Speeches: 4 November 2021.
- [5] YANG, Guanhao et al. Face Mask Recognition System with YOLOV5 Based on Image Recognition. In: **2020 IEEE 6th International Conference on Computer and Communications (ICCC)**. IEEE, 2020. p. 1398-1404.
- [6] D. K. Reddy N, G. Jeevan Kumar, and G. R. Krishna, "Social Distance Monitoring And Face Mask Detection System For Covid-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 12, pp. 2200–2206, 2021.
- [7] Official YOLOv5 PyTorch page: https://pytorch.org/hub/ultralytics_yolov5/. Accessed on: 24 of Dec.2021.

- [8] S. Meivel, K. Indira Devi, S. Uma Maheswari, and J. Vijaya Menaka, "Real time data analysis of face mask detection and social distance measurement using Matlab," *Mater. Today Proc.*, no. xxxx, 2021.
- [9] S. Bhutada, N. Sahithi, M. Mounika, and M. Revathi, "SOCIAL DISTANCING AND MASK DETECTOR BASED ON," vol. II, no. May, pp. 81–87, 2021.
- [10] P. Somaldo, F. A. Ferdiansyah, G. Jati, and W. Jatmiko, "Developing Smart COVID-19 Social Distancing Surveillance Drone using YOLO Implemented in Robot Operating System simulation environment," *IEEE Reg. 10 Humanit. Technol. Conf. R10-HTC*, vol. 2020-December, 2020, doi: 10.1109/R10-HTC49770.2020.9357040.
- [11] M. Rezaei and M. Azarmi, "Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic," *Appl. Sci.*, vol. 10, no. 21, pp. 1–29, 2020, doi: 10.3390/app10217514.
- [12] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19," *Smart Heal.*, vol. 19, pp. 1–5, 2021, doi: 10.1016/j.smhl.2020.100144.
- [13] M. F. Catapan, "ANÁLISE ANTROPOMÉTRICA DA CABEÇA HUMANA PARA DIMENSIONAMENTO DE CAPACETES BALÍSTICOS," 2014.
- [14] HARRICHANDRA, Amelia; IERARDI, A. Michael; PAVILONIS, Brian. An estimation of airborne SARS-CoV-2 infection transmission risk in New York City nail salons. **Toxicology and industrial health**, v. 36, n. 9, p. 634-643, 2020.
- [15] Riley EC, Murphy G and Riley RL (1978) Airborne spread of measles in a suburban elementary school. *American Journal of Epidemiology* 107: 421–432.
- [16] RILEY, E. C.; MURPHY, G.; RILEY, R. L. Airborne spread of measles in a suburban elementary school. **American journal of epidemiology**, v. 107, n. 5, p. 421-432, 1978.
- [17] DHARMADHIKARI, Ashwin S. et al. Surgical face masks worn by patients with multidrug-resistant tuberculosis: impact on infectivity of air on a hospital ward. **American journal of respiratory and critical care medicine**, v. 185, n. 10, p. 1104-1109, 2012.
- [18] Image Dataset repository used in this project: <https://drive.google.com/drive/folders/1sHDgVl0dyaJcyYty85Z32XwhCwfzuUa7?usp=sharing>.
- [19] Repository page of the code for this project: <https://github.com/Horusprg/Monitoramente-de-ambientes>.