Psychoeducative Social Robots for an Healthier Lifestyle using Artificial Intelligence: a Case-Study

Valerio Ponzi[1], Samuele Russo[2], Valerio Bianco[3], Christian Napoli[3] and Agata Wajda[4]

[1]*Department of Computer, Automation and Management Engineering, Sapienza University of Rome, via Ariosto 25 Roma 00185, Italy*
[2]*Department of Psychology, Sapienza University of Rome, via dei Marsi 78 Roma 00185, Italy*
[3]*Department of Medical Surgical Sciences and Translational Medicine, Sapienza University of Rome, Via di Grottarossa 1035, Roma 00189, Italy*
[4]*Department of Mathematics Applications and Methods for Artificial Intelligence, Faculty of Applied Mathematics, Silesian University of Technology, 44-100 Gliwice, Poland*

Abstract

Smoking is the greatest preventable cause of mortality worldwide. In this paper, we present a social experiment where mobile robot equipped with a cigarette detector alert smokers, in particular those smoking close to children. In our research, we compare different methods. In the first case we trained the cigarette detection model using a homemade dataset based on the pre-trained SSD MobileNet detection model. In the second case we analyzed how smokingNet performs applied to our task. Next to distinguish between children and adults, we take advantage of the Cascade classifier and a neural network. Both networks are built to leverage TensorFlow Lite, a mobile-friendly format that enables inference on-device. When a smoking scene is identified, the mobile robot draws near the smoker and issues a warning based on the circumstances.

Keywords

Smoking, Object Detection, Cascade Classifier, Raspberry Pi, TensorFlow Lite

1. Introduction

Smoking is the leading preventable cause of death and disability. According to WHO (World Health Organization) [1, 2] data, tobacco directly causes over 7 million deaths, while around 1.2 million are the result of non-smokers being exposed to secondhand smoke. Secondhand smoke is dangerous, especially for children, and can increase their risk of multiple health issues.

For many years it has been forbidden to smoke in the presence of children in public places except those dedicated to smokers. In recent times, a further restriction has been imposed by the Council of Ministers which has issued several legislative decrees in which the anti-smoking regulations are made even more restrictive. In addition to the medical aspects, linked to the now evident consequences associated to passive smoking, there are other damages connected to the development of fascination towards cigarettes and consequent developement of possible addiction [3]. In fact, starting from Albert Bandura's studies on social learning theory [4], it is highlighted as learning of pro-social or anti-social behaviors can also occur without direct contact with objects, or learning can also occur through indirect experiences, through the observation of other people[5, 6]. Bandura used the term modeling (imitation) to identify a learning process that is activated when the behavior of an individual who observes changes according to the behavior of another individual who acts as a model. So the behavior is the result of a process of acquiring information from other individuals. Furthermore, Bandura synthesizes a series of properties acting in a modeling situation, which influence the impact of information learned about performance: the identification that is established between model and modeled is identified as a fundamental characteristic of observational learning (or vicarious learning). . The higher it is, the more learning will have an effect on the behavior of the model. So for example, according to this theory, a child who daily observes a reference adult who smokes, will learn this behavior more easily, since he is exposed to behavior patterns that "normalize" the use of cigarettes on a daily basis. This learning theory is also called social learning, because it focuses on the identification mechanism that links observer to observe. This identification process is also linked to affective aspects, and it is often found in identifying behaviors that people adopt in certain roles or social characters. It therefore becomes essential for the child to reduce exposure to potentially dangerous behaviors as much as possible, such as that related to cigarette smoking.

The aim of this research is to develop a mobile robot capable of detecting smokers and raise a warning if those are in the proximity of youngsters.

To achieve this, we splat the objective in two main

ICYRIME 2021 @ International Conference of Yearly Reports on Informatics Mathematics and Engineering, online, July 9, 2021
ponzi@diag.uniroma1.it (V. Ponzi); samuele.russo@uniroma1.it (S. Russo); christian.napoli@uniroma1.it (C. Napoli); agata.wajda@polsl.pl (A. Wajda)
0000-0002-1846-9996 (S. Russo); 0000-0002-5775-2276 (C. Napoli); 0000-0002-1667-3328 (A. Wajda)

tasks: cigarettes detection through a camera stream and classification between adults and teenagers. Cigarettes detection is analyzed in two different methods. In the first we started from a pre-trained SSD MobileNet model[7] with a feature pyramid network as a subnetwork. The latter is especially well-suited for mobile oriented applications since it gets rid of main memory access constraints in a large number of embedded hardware. Instead, in the second case we used SmokingNet[8] that detects smoking photos by utilizing the feature extraction capabilities of convolutional neural networks. Furthermore, the program is hosted on a Raspberry Pi [9, 10]. Once a cigarette is detected, the detection process determines the presence of people in the frame using Haar-cascade frontal face classifier[11, 12, 13], which requires significantly less hardware computation. Moreover, the image's detected face will be extracted and fed as an input to a deep neural network. The network acts as a binary classifier trained on the UTKFace dataset to distinguish children from adults.
To make models portable, both detection and classification models are written using TFlite[14] library and inference is performed on the Raspberry Pi 4. The mobile robot will approach the adult and issue a warning if there are children nearby.

2. Related Works

There has been some research on smoking detection through different methods. In [15] the authors suggest a smoking gesture based detection method. It captures changes in the orientation of a person's arm, and uses a machine learning pipeline that processes this data to accurately detect smoking gestures and sessions in real-time.
In [16] the authors proposed a machine learning method for puffing and smoking detection using data from a wrist accelerometer. More recent approaches suggest the use of latest generation techniques based on the object detection of the cigarette itself.
In [17] the author presents object detection of cigarette litter on side-walks. The system is designed to work in real-time by exploiting a lighter version of YOLOV4 [18] (Tiny-YOLOV4) in order to let the model be deployed on a mobile robot. The dataset used to train this network was specific to the littering problem, that is, detecting cigarettes butts near sidewalks. Therefore being our objective the identification of situations involving people smoking, we cannot expect to achieve high performance by applying transfer learning and fine tuning on their pretrained network.
In [19] the authors use a YOLOv2 deep-learning image based methodology for driver's cigarette object detection. The driver's images are captured by a dual-mode visible light and near-infrared camera, and the developed system judges whether or not there is a driver smoking behavior in the day and night conditions. Since their dataset is specific for foreground cigarette detection, their results are not directly transferable to our study-case.

3. Datasets

3.1. Cigarette Dataset

The dataset was realized with the objective of identifying cigarettes in a variety of situations. The background and the quality of the pictures are varied. In the images, the cigarette can be extremely large or extremely small in comparison to the entire image. The images are uploaded to Roboflow [20], a tool for annotation, allowing users to upload files, including images, annotations, and videos. It supports a wide variety of annotation formats and makes it simple to add new training data as it is collected. The format of the data set is set to TFRecords. During the annotation process, some images were discarded due to their low quality, such as cigarettes that are obscured by other objects or are barely visible to human eyes. The remaining dataset consists of 2017 images that have been divided into a training set and a test set according to radio 9:1. The training set contains 1816 images, while the test set contains 201 images. A representative sample of the dataset can be found in Figure 1.

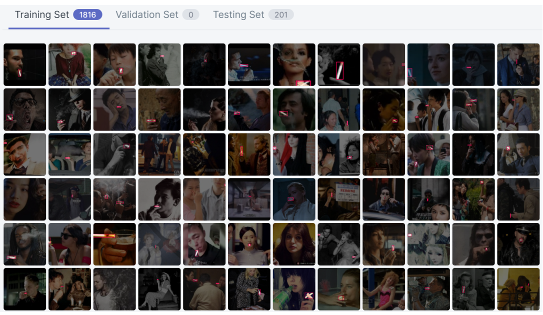


Figure 1: Samples of self-made cigarette dataset annotated on Roboflow

3.2. UTKFace Dateset

The UTKFace dataset[21] is a large-scale face collection with a wide age range (range from 0 to 116 years old). The dataset contains over 20,000 face images with age, gender, and ethnicity annotations. The images demonstrate a wide range of poses, facial expressions, illumination, occlusion, and resolution. This dataset has the potential to be used for a variety of tasks, including face detection, age estimation, age progression/regression, and landmark localization. We only use age information to classify those

Figure 2: Samples of UTKFace dataset

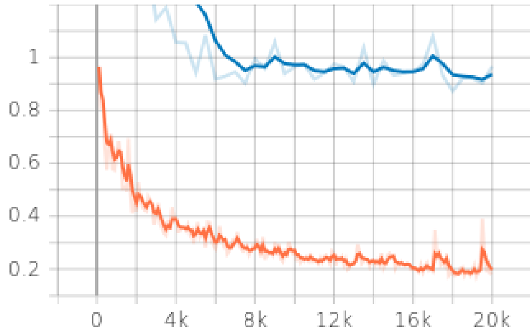


Figure 3: Total loss on training set (orange) and test set (blue).

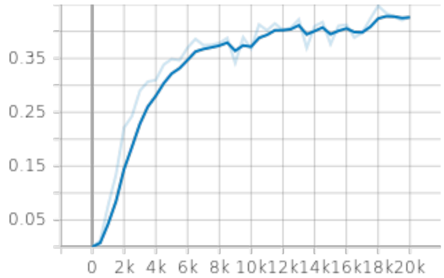


Figure 4: mAP of all objects on the test set.

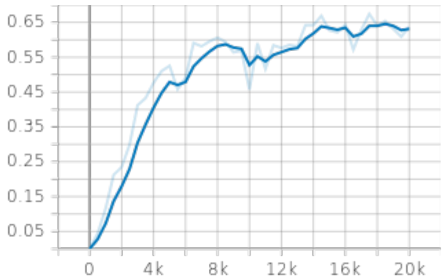


Figure 5: mAP of large objects on the test set.

under the age of 18 as children and those over the age of 18 as adults.

4. Cigarette detection

4.1. SSD MobileNet

To achieve a trade-off between speed and accuracy, in the first case we used the SSD MobileNet V2 object detection model with the FPN-lite feature extractor, shared box predictor, focal loss and a 640x640 training image size. SSD[22] is a multi-category single-shot detector that is substantially faster and more accurate than the initial version of YOLO. In the literature, there is a variety of more accurate but slower techniques such as Faster R-CNN. However, since our main focus is to deploy a real time application, we give more importance to the speed of the model. The core of SSD is predicting category scores and box offsets for a fixed set of default bounding boxes using small convolutional filters applied to feature maps. MobileNet V2[7] significantly improves the performance of mobile models on a variety of tasks and benchmarks. It is based on an inverted residual structure in which shortcut connections are made between thin bottleneck layers. FPN[23] exploits the inherent multi-scale, pyramidal hierarchy of deep convolutional networks to construct feature pyramids with marginal extra cost. It is a top-down architecture with lateral connections is developed for building high-level semantic feature maps at all scales.

4.2. Training

The entirety of the training process was conducted on Google Colab[24]. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. With Colab Pro, we are able to train our model with a K80 GPU for up to 24 hours. The training procedure is carried out by means of 20,000 epoches.

The model performs poorly on small objects in comparison to large objects. This is partly due to the SSD MobileNet's features and also to the compressed input size. Large objects have an accuracy of up to 0.65, while medium and small objects have an accuracy of only 0.57 and 0.29, respectively. This results in a decrease in the overall accuracy of all images.

4.3. SmokingNet

A different method for cigarette detection is using SmokingNet. SmokingNet was announced in 2018, it detects smoking photos by utilizing the feature extraction capabilities of CNN. The convolution kernels of the CNN

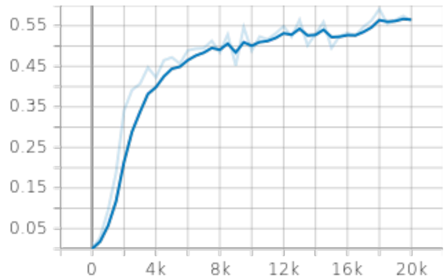


Figure 6: mAP of medium objects on the test set.

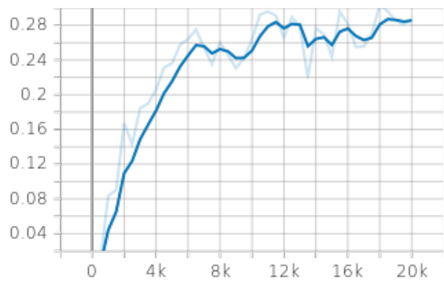


Figure 7: mAP of small objects on the test set.

convolutional layers have been used to extract local features of a given image, and the features extracted by the first convolutional layer directly affect the feature fusion of the deep network. Based on the shape characteristics of cigarettes, convolution kernels of four sizes are included in the first convolutional layer of SmokingNet. This method can detect smoking images by utilizing only the information of human smoking gestures and cigarette image characteristics without requiring the real detection of cigarette. This model achieves an accuracy and recall of 0.9.

4.4. Comparison between SSD MobileNet model and SmokingNet

Table 1
Performance of SSD MobileNet model and SmokingNet

Model	Precision	Recall
`SSD MobileNet`	0.43	0.46
`SmokingNet`	0.90	0.90

As we can see in table 1 the results obtained by SmokingNet are much better than the other. However, this is also due to the type of research that is carried out to understand whether or not the cigarette is present in the image. This method is effective in not very crowded situations. Instead, in the presence of a large number of people the amount of gestures to be analyzed becomes too heavy from a computational point of view and too many elbow movements are misleading.

5. Age classification

5.1. Cascade Classifier

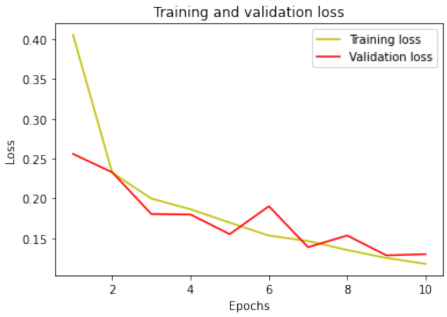
Object detection with Haar feature-based cascade classifiers is a powerful object detection technique proposed in 2001 by Paul Viola and Michael Jones[11]. It is a method for combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image. It is a machine learning-based technique that involves training a cascade function on a large number of positive and negative images. It is then applied to other images in order to detect objects. OpenCV includes pretrained cascade model for the frontal face, eye, body, and even the smile. For our research, we used the default Haar cascade frontal face model.

5.2. Age classifier

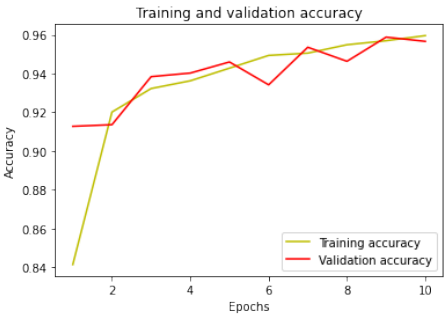
Image classification is a classical problem in computer vision which is the task of assigning a label to an input image, from a fixed set of categories. This is a fundamental problem in Computer Vision, and despite its simplicity, it has a wide range of practical applications. With the UTKFace dateset, we trained a binary classifier to distinguish children and adults, which allows the mobile robot to give different warnings. After training a CNN model with only 10 epochs, the model can achieve over 95 percent accuracy. The results can be observed in Figure 8. In subfigure 8.a we can see the loss tendency through the epoches. Subfigure 8.b shows the accuracy over the training and validation set. We can appreciate the absence of overfitting at the end of the training process.

6. Mobile Robot

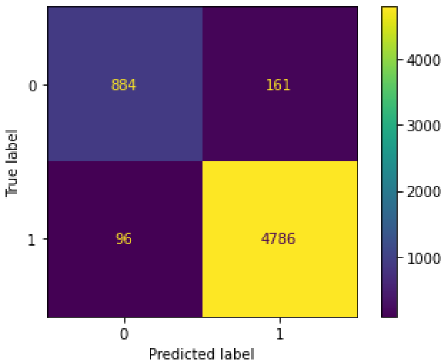
The mobile robot chosen for this research is the Sapienza's robot MARRtino. MARRtino is a ROS-based low-cost differential drive robot platform that comes in many shapes. MARRtino has been designed to be easy-to-build and easy-to-program, but at the same time it uses professional software based on ROS. It is thus suitable to implement and experiment many typical Robotics and Artificial Intelligence tasks, such as smart navigation, spoken human-robot interaction, image analysis, etc. It uses a differential wheeled robot, hence its movement is based on two separately driven wheels placed on either side of the robot body. It can thus change its direction



(a) Training and validation loss of age classifier.



(b) Training and validation accuracy of age classifier.



(c) Confusion matrix of predication and ground truth on test set.

Figure 8: Performance of our age classifier

by varying the relative rate of rotation of its wheels and hence does not require an additional steering motion. On the front of the mobile robot, a 480p webcam is installed, which is essential for our objective of recognizing a smoking scene by detecting cigarettes, adults, and children in its view. The webcam can rotate horizontally and vertically and has two degrees of freedom. After recognizing a smoking scene, the speaker in the center of the mobile robot will issue a warning. All of the sensors and components are connected to a drive board that extends

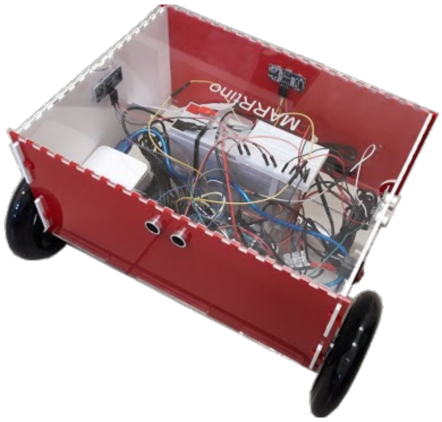


Figure 9: MARRtino robot

all of our Raspberry Pi's connections. As a result, the Raspberry Pi can get all the measurements from sensors and control motors and speakers accordingly.

6.1. Raspberry Pi

The brain of our mobile robot is a Raspberry Pi 4B. The Raspberry Pi is a small, powerful, and low-cost embedded device. The Raspberry Pi 4B uses a Broadcom BCM2711 SoC with a 1.5 GHz 64-bit quad-core ARM Cortex-A72 processor, with a 1 MB shared L2 cache. The Raspberry Pi Foundation, in collaboration with Broadcom, developed a series of miniature single-board computers (SBCs) in the United Kingdom. Initially, the Raspberry Pi initiative was geared toward promoting the teaching of fundamental computer science in schools and impoverished countries. The first model achieved greater popularity than planned, selling outside of its intended market for applications such as robots. It is widely utilized in a variety of fields, including weather monitoring, due to its inexpensive cost, modular construction, and open architecture. Due to its support for HDMI and USB devices, it is commonly utilized by computer and electronic hobbyists.

6.2. Depth Camera

There are numerous types of depth cameras, which vary in terms of how they receive world data or how that data is processed in order to present it in a useful format. The sensors can differ in a variety of ways, including acquisition method, resolution, and range. Stereo sensors attempt to replicate human vision by utilizing two cameras addressing the scene with a certain amount of separation between them. The images from these cameras are gathered and then utilized to extract and match visual features (important visual information) in order to create what is known as a disparity map between the

Figure 10: Astra Pro (Our RGB-D camera)

cameras' viewpoints. Time of Flight (ToF) sensors illuminate the entire image and determine depth based on the time required for each photon to return to the sensor. This means that each pixel corresponds to a single beam of light projected by the device, resulting in increased data density, less shadows cast by objects, and simplified calibration (no stereo matching). By contrast, structured light (SL) sensors make use of a predetermined pattern projected into the scene by the IF sensor. The deformation of the pattern is then used to generate the depth map. In this research, we choose a depth camera with a structured light sensor, a specific variant of the Orbbec Astra Pro. Actually, the Astra Pro has a higher RGB resolution camera as well as a depth camera. Astra Pro was created to be largely compatible with the existing OpenNI library. Through a Python binding for OpenNI2, we are able to obtain both RGB and depth information from the camera.

The initial mobile robot was equipped with a standard monocular camera that is unable of determining the depth of a scene. We did, however, investigate the possibility using an RGB-D camera. Due to mechanical constraints, we were unable to directly mount the camera on the mobile robot with screws, but we devised a method for attaching the camera to the mobile robot's base. It is obvious that when testing alone with the RGB-D camera, the depth of the detected item can be easily determined. The camera features a depth sensor in addition to the standard color sensor. The depth sensor can be used to determine the proximity of an object to the camera. As a result, once our object has been detected, it is straightforward to locate and calculate the distance to the identical object on the depth map.

7. Implementation of the Whole Structure

Prior to combining everything, we need to convert object detection model to TensorFlow Lite in order to deploy it on the Raspberry Pi. TensorFlow Lite is a free, open-source deep learning framework that enables the deployment of TensorFlow models on mobile devices. It is optimized for machine learning on-device. After conversion, we can use TensorFlow Lite on our mobile robot to create predictions based on the input data. As illustrated in the plots, the TensorFlow Lite model is still capable of doing high performance cigarette detection on the test photos when the cigarettes are sufficiently visible.

Our task is to determine whether one or more persons are smoking and whether children are near the smoker. Therefore, the object detector will first determine the number of smokers and children in the area. For example, if it detects a cigarette, two adults and a child, that means there are a smoker and a child there. Then the robot will approach the smoker and attempt to convince them to stop smoking or inform them that passive smoking is dangerous for children.

8. Human-Robot Interaction

In this section we see how the interaction of the robot affects people. Social psychology has shown how explicit prohibitions can lead individuals to develop a conduct opposite to what is required, whereby an explicit prohibition to do something causes the subject to disregard that prohibition and to take the prohibited behavior. For this reason, a very important aspect was the construction of a dataset that contained "kind messages" that could be sent to people to dissuade them from smoking in the presence of children. The characteristics of the message that we considered relevant were the following: 1) it did not have to contain an explicit prohibition; 2) the sentence had to be short and understandable; 3) the sentence had to have a content that could be judged plausible by the subject; 4) the sentence had to use direct but gentle language. Eg. some of the phrases used could contain a message like this: "Please smoke away from here because there is a child" - "Kindly, do not smoke in this area because there are children too close". To evaluate the "quality" of the answers, and the effectiveness of the robot, the subjects were subjected to a questionnaire that contained 3 types of questions: how did they evaluate the relevance of the robot with respect to the task for which it was programmed; how they assessed the robot's kindness in being able to persuade them not to smoke in front of the children; and how they assessed the robot's ability to persuade them to smoke in general. Follow a table (Table 2) and a chart with the results of the social experiment (Figure 11).

Table 2
The questions proposed to the users after the social experiment

N.	Question	% YES	% NO
1	Was the robot pertinent?	77	23
2	Was the robot polite and kind?	83	17
3	Will you smoke near children?	35	65
4	Will you quit smoking?	29	71

9. Conclusion

In this research, we deployed a mobile robot to deal with the exposure of youngsters to second-hand smoke. To detect cigarettes, we compared a custom SSD MobileNet model trained on our home-made dataset with SmokingNet. Next, to discriminate between children and adults, we use a pretrained Cascade classifier as a face detector and a CNN model trained on the UTKFace dataset as an age classifier. The age classifier is only initialized when a mobile robot comes across a cigarette and the Cascade classifier detects faces. This approach improved the accuracy of distinguishing children from adults. In the literature, researchers focused on achieving high performance related to the cigarette detector. It can be shown that these models required a significant amount of computational time. The models chosen were structured in order to be mobile friendly and succeeded to be deployed on a Raspberry Pi. In the first model of cigarette detection one of the main drawbacks was the performance of the detector related to really small size images. This may also need some additional data augmentation on the dataset. The limited performance of SSD MobileNet chosen lead to choose SmokingNet which turns out to be much more precise. We also discovered that the Raspberry Pi's performance is insufficient for running a high-performance object detector. When running inference with TensorFlow Lite, the frames per second is only 0.55 which a bit slow for a real-time detection. It's difficult to strike a balance between efficiency and performance. A more powerful computer, such as the Jetson Nano, will be able to run a deeper neural network, which will undoubtedly increase efficiency and performance. In this paper we have shown how, even with few computational resources available, satisfactory results can be achieved for important and large-scale problems such as the second-hand smoking towards teenagers and children. In conclusion, from the answers to the questions, highlighted by the chart, it emerges how the use of the robot can be a good "facilitator" to communicate messages inviting them not to smoke in the presence of children. Probably this role of facilitator is made possible by the "sympathy" that the robot can arouse in the majority of the people involved in this study. In fact, as regards the questions related to how the robot was perceived, it has a very high percentage of positive answers. There is also a significant percentage of affirmative answers on the robot's ability to persuade people not to smoke in the presence of children. However, it would be necessary to consider that these latter responses could be conditioned by the sympathy aroused by the robot, rather than by a sincere intent to change one's lifestyle. However, when asked about the effectiveness of the robot in convincing people to quit smoking in general, there was a considerable number of negative responses, which leads us to think that cigarette addiction is very high and certainly requires further strategies to convincing people to change their lifestyle in a stable and radical way.
It would also be interesting to repeat the experiment with a control group and an interview at least six months apart.

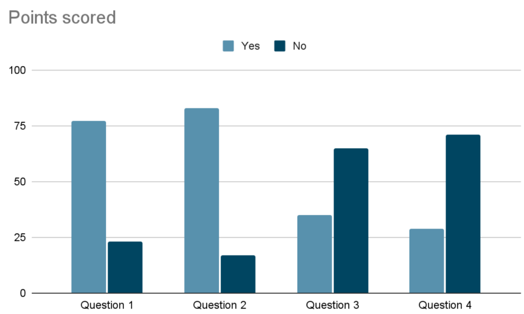


Figure 11: Column chart showing the percentage of people's responses to the robot.

References

[1] World health organization, 2022. URL: https://www.who.int/.
[2] C. Pasquarella, L. Veronesi, C. Napoli, P. Castiglia, G. Liguori, R. Rizzetto, I. Torre, E. Righi, P. Farruggia, M. Tesauro, et al., Microbial environmental contamination in italian dental clinics: A multicenter study yielding recommendations for standardized sampling methods and threshold values, Science of the total environment 420 (2012) 289–299.
[3] P. Caponnetto, et al., The effects of physical exercise on mental health: From cognitive improvements to risk of addiction, International Journal of Environmental Research and Public Health 18 (2021). doi:10.3390/ijerph182413384.
[4] A. Bandura, Model of Causality in Social Learning Theory, Springer US, Boston, MA, 1985, pp. 81–99. URL: https://doi.org/10.1007/978-1-4684-7562-3_3. doi:10.1007/978-1-4684-7562-3_3.
[5] S. Russo, C. Napoli, A comprehensive solution for psychological treatment and therapeutic path plan-

ning based on knowledge base and expertise sharing, volume 2472, 2019, p. 41 – 47.

[6] S. Illari, S. Russo, R. Avanzato, C. Napoli, A cloud-oriented architecture for the remote assessment and follow-up of hospitalized patients, volume 2694, 2020, p. 29 – 35.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[8] D. Zhang, C. Jiao, S. Wang, Smoking image detection based on convolutional neural networks, in: 2018 IEEE 4th International Conference on Computer and Communications (ICCC), IEEE, 2018, pp. 1509–1515.

[9] Raspberry, 2022. URL: https://www.raspberrypi.com/for-home/.

[10] M. Woźniak, D. Połap, C. Napoli, E. Tramontana, Application of bio-inspired methods in intelligent gaming systems, Information Technology and Control 46 (2017) 150 – 164. doi:10.5755/j01.itc.46.1.13872.

[11] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, volume 1, Ieee, 2001, pp. I–I.

[12] J. Starczewski, S. Pabiasz, N. Vladymyrska, A. Marvuglia, C. Napoli, M. Wózniak, Self organizing maps for 3d face understanding, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9693 (2016) 210 – 217. doi:10.1007/978-3-319-39384-1_19.

[13] N. Brandizzi, V. Bianco, G. Castro, S. Russo, A. Wajda, Automatic rgb inference based on facial emotion recognition, volume 3092, 2021, p. 66 – 74.

[14] M. A. ..., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.

[15] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, E. Kalogerakis, Risq: Recognizing smoking gestures with inertial sensors on a wristband, MobiSys 2014 - Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services 2014 (2014). doi:10.1145/2594368.2594379.

[16] Q. Tang, D. Vidrine, E. Crowder, S. Intille, Automated detection of puffing and smoking with wrist accelerometers, ICST, 2014. doi:10.4108/icst.pervasivehealth.2014.254978.

[17] Identification of cigarette litter with the use of outdoor mobile robots (2021). URL: https://projekter.aau.dk/projekter/files/419098381/Master_Thesis_Mathiebhan.pdf.

[18] A. Bochkovskiy, C. Wang, H. M. Liao, Yolov4: Optimal speed and accuracy of object detection, CoRR abs/2004.10934 (2020). URL: https://arxiv.org/abs/2004.10934. arXiv:2004.10934.

[19] Deep learning based driver smoking behavior detection for driving safety (2020). URL: http://www.joig.net/uploadfile/2020/0318/20200318051129839.pdf.

[20] Roboflow, https://roboflow.com, 2022. URL: https://roboflow.com.

[21] S. Y. Zhang Zhifei, Q. Hairong, Age progression/regression by conditional adversarial autoencoder, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21–37.

[23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.

[24] Colab, https://colab.research.google.com, 2022.