

Artificial Intelligence in Foreign Object Classification in Fenceless Robotic Work Cells Using 2-D Safety Cameras

Merdan OZKAHRAMAN*, Haydar LIVATYALI

Abstract: Production systems using robotic manipulators have become common in the last few decades, and the trend is towards fenceless cells that save from space. Thus, the safety and flexibility of these systems have become more critical. The safety systems are based on either sensor data or camera images. Although the flexibility of the camera-based systems is better, conventional image processing methods are sensitive to the working environment. Artificial intelligence may be a powerful tool for them to adapt to change requirements quickly and improve accuracy and stability. In this study, a low-cost 2-D camera-based safety system was designed and installed in an experimental fenceless robotic work cell. The system controller was coupled with three alternative deep learning (ResNet-152, AlexNet, SqueezeNet) and three machine learning modules (support vector machine, random forest and decision tree). These modules were trained using photo images of ten distinct foreign objects penetrating the alarm zone. To include the ever-changing conditions of the industrial environment, disruptive effects including camera vibrations, shadows, reflections, illuminance variations etc. are included by using multiple images up to 550 for each class. Using the restricted data used for training and testing the six systems, the SqueezeNet deep learning model gave the best accuracy of 95% without any over-fitting. Despite this, machine learning-based models have been found to have 100 times faster prediction time than deep learning-based ones. Thus, the safety system can be adapted quickly to any possible changes and noise that may arise from working conditions is prevented, and time losses that may occur in industrial production may be prevented.

Keywords: artificial intelligence; image classification; robotics and automation

1 INTRODUCTION

Robot technology has been used in industrial production and many other fields for decades. Variable production demand for various products requires frequent changes on the production lines and robotic cells. Changes to the production line cause loss of time and labour [1]. An essential part of these losses comes from the safety requirements of the work cells in the production line.

Fences as well as sensor-based, and camera-based safety systems, are used in the industry. Camera-based safety systems can be considered the most advanced technology. In such systems, the structure must be readjusted due to changes in work cells. Safety systems used in reconfigurable structures, systems based on human-robot interoperability, and image processing in fenceless systems should also adapt to this flexibility [2-4]. To achieve adaptability of the safety system to the working flexibility, and to avoid the noise caused by environmental conditions, it is inevitable to include artificial intelligence algorithms in traditional image processing methods.

When the objective is to recognize and distinguish foreign objects penetrating the work cell, a system using artificial intelligence-based image processing may improve the performance of safety systems. Traditional image processing-based safety systems cannot reliably recognize friendly objects. These friendly objects may be a workpiece or an operator who is allowed to be inside the cell zone. Some additional equipment is needed for traditional systems to recognize these objects and not stop the robot arm from working. Artificial intelligence-based safety systems are much more successful in this regard. The reliability of the system will increase with the recognition of objects by the system in desired and undesired ways. However, it is known that traditional systems are affected by sources of noise such as vibration, shadows, and illuminance in the working environment. It is possible to establish a safety system that reacts quickly to future changes with increased reliability. With the

proposed design, it is aimed that the environmental noise will not affect the accuracy of the system.

Safety requirements of work cells using robot arms are specified in "ISO 10218: Robots for Industrial Environments - Safety Requirements" [5-7]. Following this standard, objects that can interfere with the robot work cell can be classified using artificial intelligence algorithms.

Artificial intelligence (AI) is defined as systems capable of interpreting data obtained from a system by imitating human intelligence with high accuracy, learning, and using the "knowledge" to serve specific purposes and tasks repeatedly [8]. AI systems perform estimation, classification, and clustering on the data and ensure that the data is processed following a given objective. By imitating nature and people, problem-solving artificial intelligence techniques such as expert systems, fuzzy logic, artificial neural networks, genetic algorithms, and algorithms, machine learning, and deep learning have been developed. AI techniques can offer different solutions in many areas such as disease diagnosis, autonomous controlled systems, renewable energy systems, industrial manufacturing, and robotic systems [9].

2 RELATED STUDIES

Deep learning is a branch of AI architecture that tries to model high-level abstractions in data using multiple processing layers. Convolutional neural network (CNN), one of the deep learning models, was used by LeCun in 1998 to recognize handwriting numbers and help object recognition [10]. Later, CNN's have emerged as a powerful classification tool. It is used frequently in object classification competitions [11]. Using the CNN architecture, AlexNet performed better than other CNN models such as VGG (Visual Geometry Group), GoogLeNet, and ResNet (residual neural network) in image recognition and classification, as it won the annual ImageNet Large-Scale Visual Recognition Challenge

(ILSVRC) in 2012. Microsoft ResNet (Microsoft Research Asia, Beijing, China) won the 2015 ILSVRC with a meagre 3.6% error rate [12, 16]. Considering AI applications in various fields, there are medical, industrial manufacturing, service, and personal use robotic systems [17, 18]. ResNet was reported to perform better than previous models in various tasks such as object detection and semantic image segmentation [19, 20]. It has gradually replaced VGGNets in the computer vision community. The actual mechanism supporting ResNets' effectiveness is not yet precise. It is also unclear whether a structure is needed anymore to train very deep networks [21].

Alternatively, AlexNet is also known to be a very powerful deep neural network [22]. It prioritizes the use of GPU-based training over the CPU, justifying its depth and the resulting need for computing power. The SqueezeNet model was designed to challenge AlexNet's efficiency and capability [23]. With almost 50 times fewer parameters, the deep neural network was able to achieve the high level of accuracy that AlexNet achieved at a greatly reduced size.

In this study, CNN-based AI algorithms ResNet, AlexNet, and SqueezeNet were selected for usage in an experimental camera-based safety system. Using CNN architecture for an image-based safety system in fenceless robotic work cells was expected to improve the response speed. To compare the accuracy of the CNN-based modules, machine learning-based classification algorithms were also tested in the system.

Support vector machines (SVM) are controlled classification algorithms based on statistical data. SVM can be used in many applications, from data mining to image classification, from facial recognition to voice analysis. SVM is divided into two alternatives as to the linear and nonlinear support vector machines. The classification process is realized in linear support vector machines by choosing a suitable plane to separate the two classes. Different kernel methods classify the image since a linear plane cannot be drawn in a nonlinear support vector machine [24].

In the classification process based on artificial intelligence, logistics regression (LR) and artificial neural network (ANN) models can also be used. In a study on a data set created from X-ray images; SVM, LR and ANN models yielded accuracy levels of 90.2%, 96.2% and 96.7%, respectively [25]. A stacking method in which three models are combined was proposed, and then, the accuracy turned out to be 96.9%. Thus, it was concluded that a hybrid model was more successful than a single use of the three models.

Decision trees (DT) are one of the most popular machine learning algorithms used in both classification and regression problems. Its purpose is to create a model that predicts the value of a variable by extracting simple rules from data properties and learning these rules. The algorithm does not support missing values; hence, the missing value problem must be solved before training the machine. The random forest (RF) algorithm, on the other hand, can be used in both classification and regression problems like a decision tree. Its working principle creates multiple decision trees. When it is ready to produce a result, the average value in these decision trees is taken and the result is produced [26]. Artificial intelligence-based classification applications on robotic systems in the

literature have been reviewed. Deep learning, machine learning models, and their accuracy performances are shown in Tab. 1.

Table 1 Recent AI applications performed on robotic systems and image processing

| Model | Dataset | Accuracy and Performance | References |
|--------------------------|---------------------------|--------------------------|------------|
| GLVQ, GMLVQ, and SVM | 4000 images | up to 96% | [27] |
| SVM | 180 Frames | up to 82% | [28] |
| SVM | 284 examples | up to 93% | [29] |
| LDA, NB, NN, RF, and SVM | 11 classes | up to 98% | [30] |
| ResNet-50 | 22 classes with 30 images | up to 80% | [31] |
| SqueezeNet | 885 images | up to 90% | [32] |
| AlexNet | 3000 images | up to 84% | [33] |

Accordingly, ResNet, AlexNet and SqueezeNet deep learning algorithms gave the most promising results. SVM, RF and DT machine learning models also performed well and are generally faster. Therefore, these six models were selected in this study. The accuracies and prediction times of deep learning and machine learning models were compared based on object classification algorithms.

3 DESIGN OF THE IMAGE CLASSIFICATION SYSTEM

The data set was created with the images taken from the safety camera of an experimental table-top robot work cell. First, the data set was subjected to the pre-processing step. Next, they were used to train the artificial intelligence algorithms and the results were compared. The process steps of the designed image classification system are shown in Fig. 1 in a flow chart.

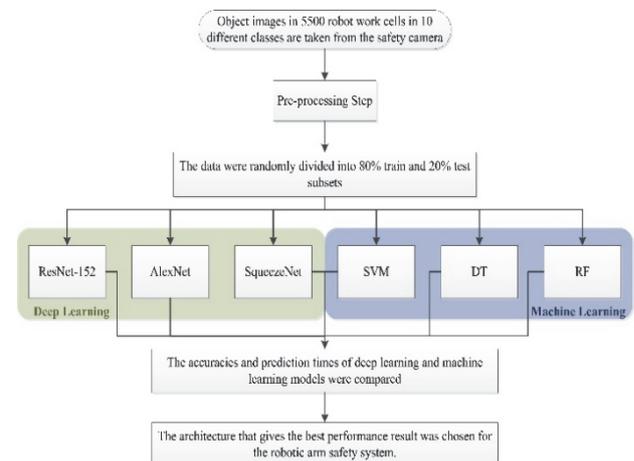


Figure 1 The process steps of the designed image classification system

The ResNet-152 model consists of 152 trainable layers, including one pooling and 151 convolutional layers [34, 35]. It is one of the most profound learning architectures offered in ImageNet with a 3×3 top pooling layer with a 7×7 core size [36, 37]. As the depth increases in classical CNN architectures, the prevention of excessive learning is realized with the residual block approach in the ResNet architecture.

For the residual Unit i , y_{i-1} is input, and f_i is trainable nonlinear mappings. The output of Unit i is recursively defined as Eq. (1) [21].

$$y_i = f_i(y_{i-1}, w_i) + y_{i-1} \quad (1)$$

where w_i denotes the trainable parameters, and f_i is often two or three stacked convolution steps. In the full pre-activation version, a convolution step's components are, in turn, a batch normalization, a rectified linear unit (ReLU) non-linearity, and a convolution layer [38, 39].

The ResNet architecture provides a deeper network with link jumping added to the feed-forward network layers. In other words, shortcut connections and properties are learned by adding x (the input value) to the $F(x)$ instead of learning from the $F(x)$ function. The ResNet learning block is shown in Fig. 2 [40]. With the residual block approach, the output is not given directly to the block, but the $x + F(x)$ results from the ReLU transaction are given. With this feature that distinguishes ResNet from the other architectures, optimization becomes more comfortable as the number of layers increases, prediction accuracy increases, and the disappearing gradient problem is solved [41].

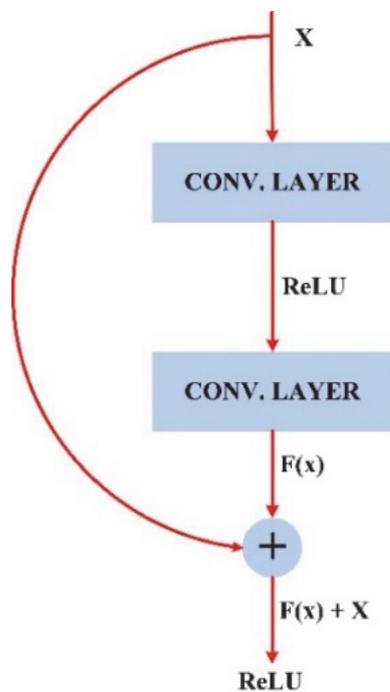


Figure 2 ResNet learning block

Fig. 3 shows a work cell using an example robot arm. Two 2-D cameras are used for the safety system of this work cell. During the robot arm's operation shown in the figure, if an unwanted object enters the working area, the robot arms will slow down and stop as specified in the standards. For this system to be defined as undesirable and acceptable objects, objects must be classified using artificial intelligence algorithms.

In the study's artificial intelligence phase, images belonging to five different classes were taken from the system, and an application was developed using the Resnet-152 deep learning architecture. The data preprocessing stage, which consists of three stages, was first performed in the application. In the first stage of data preprocessing, the images were resized to 254×254 size. Thus, all images used in the study were brought to the same size.

In the second stage, a total of 5500 photo images were divided into ten different classes. These ten classes were compared among each other. During the comparisons, questions were sought on the following effects:

- Object size
- Object brightness
- Object geometry
- Defective object corners
- Camera resolution
- The closeness of object and background colours
- Shadows
- Light reflections from the object.



Figure 3 Experimental setup

The image resolution represents the detail level of the image. This term applies to digital images, motion images, and other image types. Image resolution can be measured in several ways. The resolution of digital cameras can be defined in various ways, such as pixel, spatial, spectral, and temporal resolutions [42, 43]. The camera used in the system is placed at a height of 90 cm from the floor. Camera resolution is 1280×960 , then the actual size of a pixel is 0.142 mm^2 . Increasing the camera resolution will reduce the actual size of each pixel. The object dimensions and the corresponding number of pixels in the analyzed image are given in Tab. 2.

Figs. 4 and 5 show examples of the ten foreign objects penetrating the work zone. Image-a shows the entrance of an operator's hand. Image-b shows a large dark blue circular entrance. Image-c shows metal and rectangular objects. Image-d shows a black and large rectangle. Image-e is the entrance of a small, round, and black object. Image-f shows the entrance of a shaded and reflective rectangle. Image-g shows a shaded and reflective dark blue circular entrance. Image-h shows a round shape with irregular corners. Image-i shows a matte and shaded rectangle. Image-j shows the object in close colour with the background. The figure has been selected to simulate objects of different sizes and structures that can enter the robotic working space. It aims to learn the effect of the size, colour, geometry, brightness, and shade while classifying.

Each foreign object was represented by approximately 550 photo images. Among these photographs, 10% belonged to low-quality images. These photos were intentionally taken under vibration, dusty air, and flashing so that the images were all blurred in various ways. The effect types to be classified during learning and the visual sets to be compared are shown in Tab. 3.

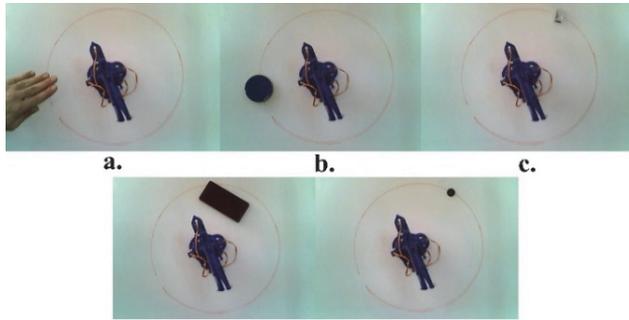


Figure 4 First classified object entries

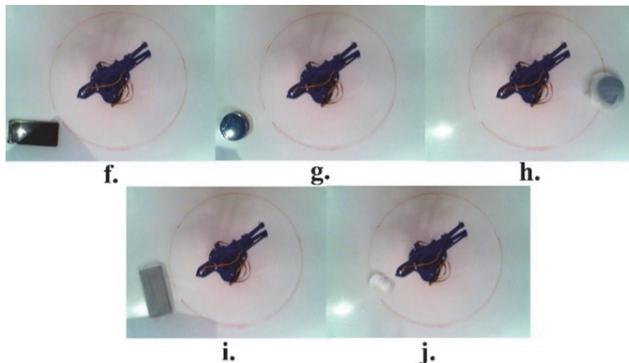


Figure 5 Second classified object entries

Table 2 Actual and pixel size of objects

| Classification | Type of Object | Actual Size / mm | Pixels (Approx.) |
|--------------------|---------------------------------|------------------|------------------|
| Circular Object | Small dark object | Ø10 | 552 |
| | Large dark object | Ø65 | 23301 |
| | Large glossy surface object | Ø65 | 23301 |
| | Large non-glossy surface object | Ø65 | 23301 |
| | Large uneven shape object | Ø65 | 23301 |
| Rectangular Object | Small glossy surface object | 25 × 10 | 1756 |
| | Large dark object | 120 × 55 | 46344 |
| | Large glossy surface object | 120 × 55 | 46344 |
| | Large non-glossy surface object | 120 × 55 | 46344 |
| | Small uneven shape object | 50 × 10 | 3511 |

In the last stage of data preprocessing, normalization was conducted by bringing all images between 0 and 1 so that the computation cost is minimized. The data were randomly divided into 80% train and 20% test subsets in the second stage. By determining the most suitable learning rate value for training, the optimization method was determined as adam, batch size value was determined as 64, and 50 epochs were trained. In the last stage, the model obtained from the training phase was evaluated on the test data.

Table 3 Feature of compared objects

| Classification | Type of Effect | Compared Objects |
|--------------------|---|------------------|
| Operator's Hand | Object geometry | a |
| | Shadow or light falling on the object | |
| Circular Object | Object size | b, e, g, h |
| | Object brightness | |
| | Object geometry | |
| | Defective object corners | |
| | Shadow or light falling on the object | |
| Rectangular Object | Object size | c, d, f, i, j |
| | Object brightness | |
| | Object geometry | |
| | Defective object corners | |
| | Object colour and background colour are close to each other | |
| | Shadow or light falling on the object | |

In the image classification algorithm, three deep learning and three machine learning modules were trained using the image data set. Deep learning and machine learning architectures were compared in terms of accuracy and response time. Firstly, Resnet-152, SqueezeNet, and AlexNet deep learning architectures were applied. In the next stage, Support Vector Machine (SVM), Random Forest (RF) and Decision Tree (DT) machine learning algorithms were used. The data set used in the study consists of images. Therefore, training operations were conducted by transforming the images used in machine learning algorithms into a one-dimensional vector (first-order tensor). The training phase for Resnet-152 is shown in Fig. 6.

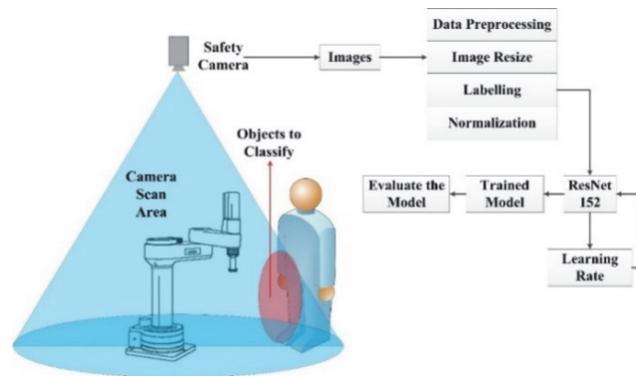


Figure 6 Workflow diagram

In the study, artificial intelligence algorithms are written in the Python programming language. The training process is compiled on cloud services. The training process was completed by using the GPU processor in the GoogleColab system and the Tensorflow library. Once the best performing AI architecture is discovered, it may be programmed into the GPU of the image-based safety system of a given robotic work cell.

4 RESULTS AND DISCUSSIONS

The learning rate value for Resnet-152 was determined to select the model with the highest accuracy. Fig. 7 shows that the optimal values of the learning rate curve are between 0.00001 and 1. Training on the images was performed at 0.001 which produces the minimum loss within this range.

A confusion matrix was generated to determine the accuracy of the model after training (Fig. 8). It shows a

comparison of the ten class visuals among each other according to Tab. 3, and high accuracy is obtained. These confusion matrices are used to determine whether there is any over-fitting in the models. It is seen that the installed system and architectural harmony are sufficient in terms of work safety.

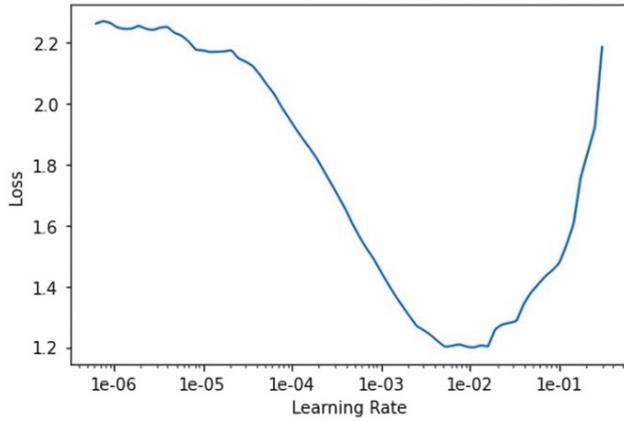


Figure 7 Learning rate chart for Resnet-152 architecture

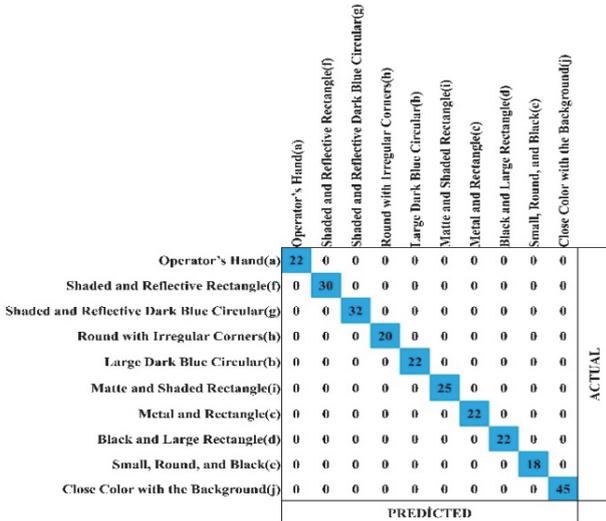


Figure 8 The ten-object confusion matrix of the Resnet-152 model

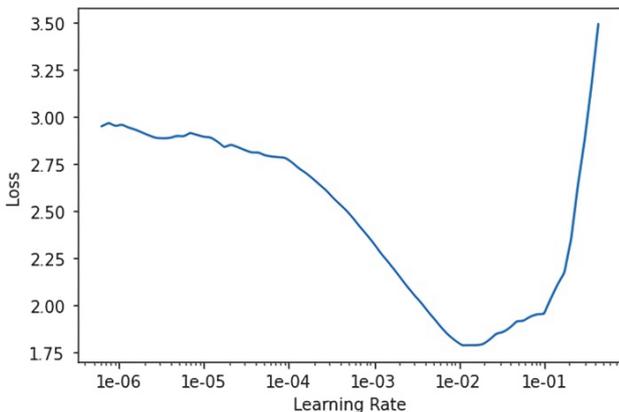


Figure 9 Learning rate chart for SqueezeNet architecture

Next, the training process was conducted using the SqueezeNet deep learning architecture. By using the learning rate optimization method, starting from a small rate, the training was continued until a high learning rate with small batch sizes and loss values recorded for each batch was obtained. The recorded learning rate recordings

were drawn, and the final learning rate value was determined by taking the midpoint of the region where the downward slope was the sharpest. Fig. 9 shows that the learning rate varies between 0.000001 and 1. The model was trained by a learning rate of 0.004, where the loss is closest to zero. After the training phase, a confusion matrix was created to determine the accuracy of the model (Fig. 10). Accordingly, the SqueezeNet architecture detects objects on the data set with 99% accuracy.

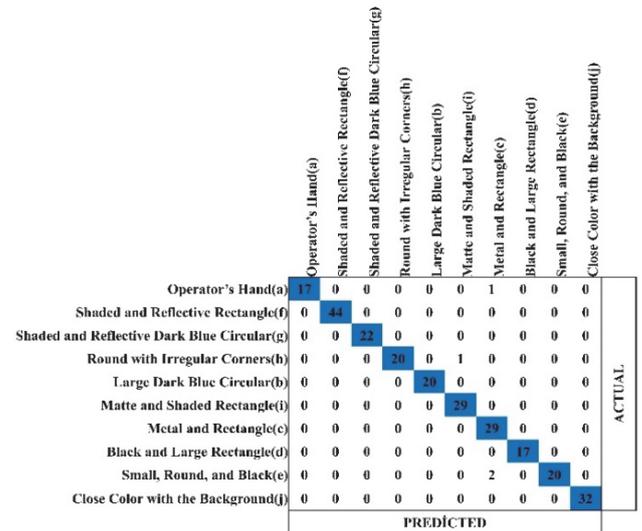


Figure 10 The ten-object confusion matrix of the SqueezeNet model

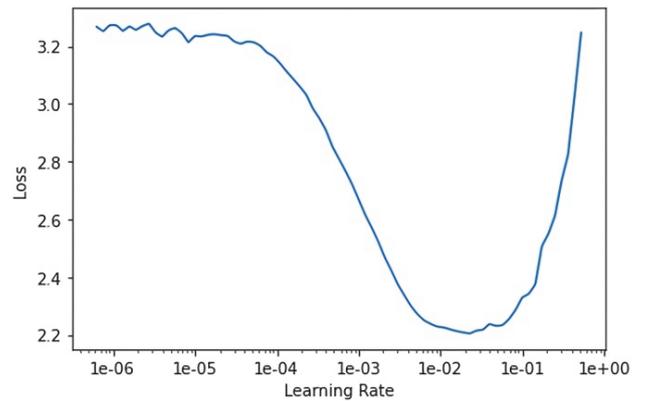


Figure 11 Learning rate chart for AlexNet architecture

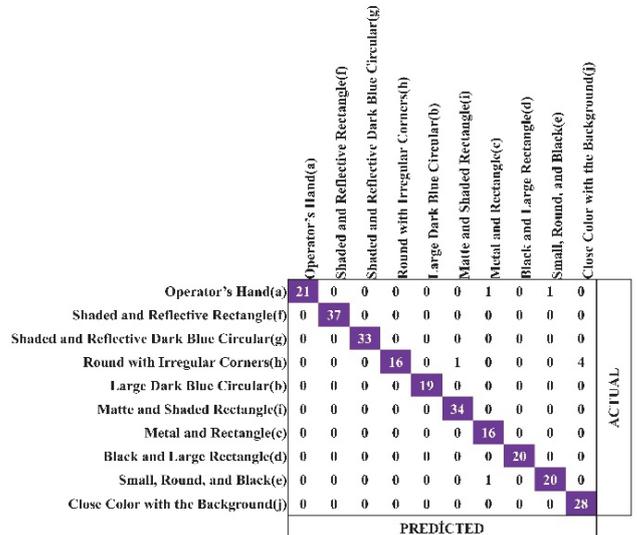


Figure 12 The ten-object confusion matrix of the AlexNet model

The training was performed on the dataset with AlexNet as the third deep learning model. Fig. 11 shows that the learning rate varies between 0.000001 and 1. The model was trained at a learning rate at a minimum loss of 0.005. After the training, a confusion matrix was again created to determine the accuracy of the model (Fig. 12). Consequently, the AlexNet module detected objects with an accuracy of 97% on the data set.

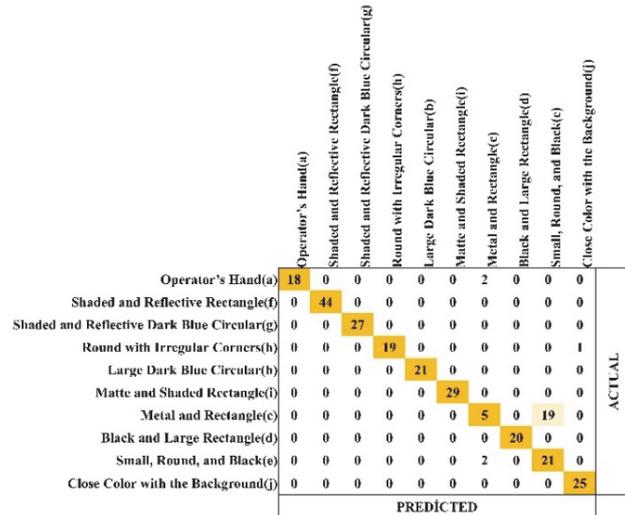


Figure 13 The ten-object confusion matrix of the SVM model

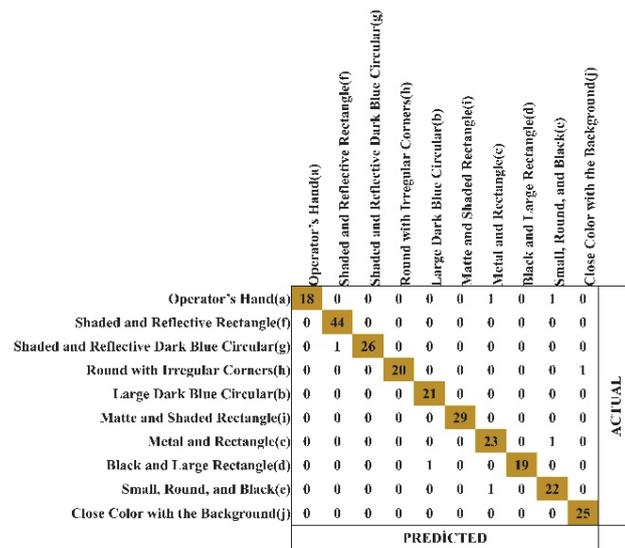


Figure 14 The ten-object confusion matrix of the RF model

In the machine learning test phase, the same data set was used to train Support Vector Machine (SVM), Random Forest (RF) and Decision Tree (DT) modules. The confusion matrix obtained from the training results of the SVM module is given in Fig. 13. Accordingly, the SVM algorithm detected objects on the data set with an accuracy of 91%. The confusion matrix of the trained RF machine learning module is given in Fig. 14. As a result, the RF algorithm detected objects on the data set with an accuracy of 98%.

Finally, the DT machine learning algorithm was tested. The confusion matrix of the trained module is given in Fig. 15, and the DT algorithm detected objects on the data set with an accuracy of 98%. The results obtained from all artificial intelligence models used in the study are compared in Tab. 4. It shows that SqueezeNet is the most

successful algorithm with an accuracy of 99%. AlexNet followed it with 97% and ResNet-152 gave 95% accuracy. The fact that all three deep learning architectures used in the study achieved over 95% performance shows that the data set and the architectures are compatible.

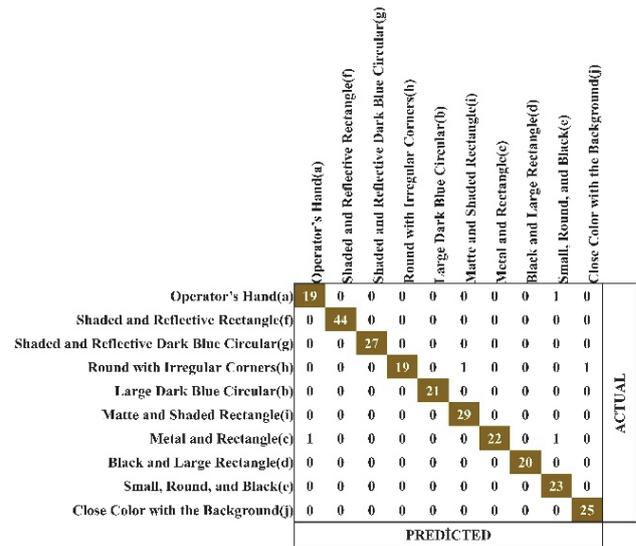


Figure 15 The ten-object confusion matrix of the DT model

Table 4 Feature of compared objects

| Model | Accuracy | Precision | Sharpness | F1-Score |
|------------|----------|-----------|-----------|----------|
| Resnet-152 | 95% | 96% | 95% | 95% |
| AlexNet | 97% | 97% | 96% | 96% |
| SqueezeNet | 99% | 99% | 99% | 99% |
| SVM | 91% | 91% | 91% | 90% |
| RF | 98% | 98% | 98% | 98% |
| DT | 98% | 97% | 97% | 98% |

Among the machine learning algorithms, RF and DT were found to be equally the most successful architectures with an accuracy of 98% where SVM could produce an accuracy of 91%. The main reason why RF is more successful than DT and SVM is that it has a more integrated structure.

In the next stage of the study, the prediction times of deep learning and machine learning algorithms on the test data are calculated and the results are compared in Tab. 5. Accordingly, the prediction times of machine learning algorithms on data set are much faster than deep learning algorithms. The prediction time of SqueezeNet, which is the most successful deep learning model, is 3.94 seconds, while the most successful DT algorithm among machine learning algorithms is the model that gives a predictive response to the test data in the least time with 0.015 seconds. For this reason, it has been determined that the data set used in the study gives the most successful result with the DT, which is one of the machine learning algorithms.

The main reason why the response times of the deep learning architectures used in the study are much higher than the machine learning algorithms is due to the high computational cost of the response times of the architectures used in the deep learning algorithms. ResNet-152 has the highest prediction time among deep learning architectures. This is because of the higher number of layers than the AlexNet and SqueezeNet architectures. Among the machine learning algorithms, the main reason why the SVM machine learning algorithm has a higher

response time as compared to RF and DT architectures is the computational cost arising from vectorial calculations.

Table 5 Prediction times of deep learning and machine learning algorithms on the data set

| Model | Time / sec. |
|------------|-------------|
| ResNet-152 | 4.39 |
| AlexNet | 3.91 |
| SqueezeNet | 3.94 |
| SVM | 2.39 |
| RF | 0.035 |
| DT | 0.015 |

The deep learning-based SqueezeNet model reached higher accuracy due to the more successful feature extraction. Besides, the machine learning-based RF and DT algorithms approached similar accuracy with a negligible difference of 1%. Nevertheless, it has been proven that machine learning-based algorithms are 100 times faster than deep learning-based algorithms. Therefore, DT was found to perform with the given image data the best.

The RF model, which gives the highest accuracy, precision and sharpness in image recognition consistently, is the second fastest model at 0.035 seconds. This response time may also be acceptable in an industrial environment where a more robust behaviour with a better F1 score is expected. On the other hand, SVM performed at a pace in the same order of magnitude with the deep learning modules, however, with a much lower F1-score of 90%.

5 CONCLUDING REMARKS

In the study, the foreign object detection capability of an experimental robotic safety system was improved by using six artificial intelligence models constructed on an image data set consisting of 5500 still images that belong to 10 different classes. An image data set of foreign objects created for ten objects that may interfere with robotic cell work zones was used to train the ResNet-152, AlexNet, SqueezeNet, SVM, DT and RF models. As a result, the accuracy of the models ranges from 91% to 99%. Although the deep learning-based SqueezeNet model gave the highest accuracy at 99%, machine learning-based RF and DT models gave a very close accuracy of 98%. However, in the prediction time ranking, the lowest time among deep learning-based models is 3.91 sec. with AlexNet. Among the machine learning-based models, the DT is the model with the lowest prediction time at 0.015 sec. Therefore, it can be said that a machine learning-based model, especially DT and RF models, are the most suitable alternatives for a fenceless robot cell safety system. Evaluations of the confusion matrices proved that there was no over-fitting in none of the models. With the high accuracy classification achieved by the RF and DT, the machine learning model improved the robot cell safety criteria in terms of work area interferences.

In this application, which is considered as an alternative to traditional image processing algorithms, the problems due to variability and disruptive effects of industrial production environments are covered. Disruptive factors such as vibration, illuminance changes, dust, and pollution that occur in the industrial production environment are equalized by the learning algorithms. Image attributes such as colour, proximity to ground tone, penetration speed, and machine shadows affect the model. Consequently, disruptive effects and image attributes that

reduce the accuracy and stability of the safety system adversely are absorbed by the statistical classification approach of artificial intelligence. Besides, the adaptation time to changes has also decreased significantly. This way, the desired and foreign objects can be identified, and access to a fenceless robot workspace can be controlled. Therefore, a relatively more reliable and flexible safety system is presented compared to sensor-based and traditional image processing-based systems.

Although AI was implemented using a wide range of data for various environments and conditions, distant hue proximity, penetration rate, or shadows can still affect the model's accuracy. It is not possible to eliminate these problems; however, expanding the data set to include more objects with varying characteristics will gradually improve the system's performance.

Acknowledgements

The work presented in this article is conducted as a part of M. Ozkahraman's doctoral thesis. Gratitude goes to Dr. Cuneay Yilmaz of Yildiz Technical University and Dr. Z. Yagiz Bayraktaroglu of Istanbul Technical University for their support and guidance.

6 REFERENCES

- [1] Takata, S. & Hirano, T. (2011). Human and robot allocation method for hybrid assembly systems. *CIRP Annals*, 60(1), 9-12. <https://doi.org/10.1016/j.cirp.2011.03.128>
- [2] Koren, Y. et al. (1999). Reconfigurable manufacturing systems. *CIRP Annals*, 48(2), 527-540. [https://doi.org/10.1016/S0007-8506\(07\)63232-6](https://doi.org/10.1016/S0007-8506(07)63232-6)
- [3] Krüger, J. et al. (2005). Image-based 3D surveillance for flexible man-robot-cooperation. *CIRP Annals-Manufacturing Technology*, 54(1), 19-22. [https://doi.org/10.1016/S0007-8506\(07\)60040-7](https://doi.org/10.1016/S0007-8506(07)60040-7)
- [4] Krüger, J. et al. (2009). Cooperation of human and machines in assembly lines. *CIRP Annals*, 58(2), 628-646. <https://doi.org/10.1016/j.cirp.2009.09.009>
- [5] Weitschat, R. et al. (2018). Safe and efficient human-robot collaboration part I: Estimation of human arm motions. *IEEE International Conference on Robotics and Automation*, 2018. <https://doi.org/10.1109/ICRA.2018.8461190>
- [6] Oberer, S. & Schraft, R. D. (2007). Robot-dummy crash tests for robot safety assessment. *IEEE International Conference on Robotics and Automation*, 2007. <https://doi.org/10.1109/ROBOT.2007.363917>
- [7] ISO 10218-1 (2011). Robots and robotic devices - Safety requirements for industrial robots - Part 1 Robots, International Organization for Standardization.
- [8] Kaplan, A. & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25. <https://doi.org/10.1016/j.bushor.2018.08.004>
- [9] Arıkan, A. et al. (2018). Control method simulation and application for autonomous vehicles. *IEEE International Conference on Artificial Intelligence and Data Processing*, 2018. <https://doi.org/10.1109/IDAP.2018.8620918>
- [10] LeCun, Y. et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>
- [11] Russakovsky, O. et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211-252. <https://doi.org/10.1007/s11263-015-0816-y>

- [12] He, K. et al. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *IEEE International Conference On Computer Vision*. <https://doi.org/10.1109/ICCV.2015.123>
- [13] Krizhevsky, A. I. et al. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105. <https://doi.org/10.1145/3065386>
- [14] Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.
- [15] Szegedy, C. et al. (2015). Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2015.7298594>
- [16] Han, S. S. et al. (2018). Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, 138(7), 1529-1538. <https://doi.org/10.1016/j.jid.2018.01.028>
- [17] Karabegović, I. & Doleček, V. (2017). The role of service robots and robotic systems in the treatment of patients in medical institutions. *Advanced Technologies, Systems, and Applications*. https://doi.org/10.1007/978-3-319-47295-9_2
- [18] Oliveira, D. A. B. et al. (2021). A Review of Deep Learning Algorithms for Computer Vision Systems in Livestock. *Livestock Science*. <https://doi.org/10.1016/j.livsci.2021.104700>
- [19] Dai, J. et al. (2016). Instance-aware semantic segmentation via multi-task network cascades. *IEEE Conference on Computer Vision and Pattern Recognition*. <http://doi.org/10.1109/CVPR.2016.343>
- [20] Chen, L. C. et al. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [21] Wu, Z. et al. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90, 119-133. <https://doi.org/10.1016/j.patcog.2019.01.006>
- [22] Lokku, G. et al. (2021). A Robust Face Recognition model using Deep Transfer Metric Learning built on AlexNet Convolutional Neural Network. *IEEE International Conference on Communication, Control and Information Sciences*, 1, 1-6. <https://doi.org/10.1109/ICCIsc52257.2021.9484935>
- [23] Ullah, A. et al. (2021). Comparative Analysis of AlexNet, ResNet18 and SqueezeNet with Diverse Modification and Arduous Implementation. *Arabian Journal for Science and Engineering*, 1-21. <https://doi.org/10.1007/s13369-021-06182-6>
- [24] Khan, A. et al. (2021). A Spectrogram image-based Network Anomaly Detection System using Deep Convolutional Neural Network. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2021.3088149>
- [25] Taspınar, Y. S., Cinar, I., & Koklu, M. (2021). Classification by a stacking model using CNN features for COVID-19 infection diagnosis. *Journal of X-Ray Science and Technology*, 1-16. <https://doi.org/10.3233/XST-211031>
- [26] Anmala, J. & Turuganti, V. (2021). Comparison of the Performance of Decision Tree (DT) Algorithms and ELM model in the Prediction of Water Quality of the Upper Green River watershed. *Water Environment Research*. <https://doi.org/10.1002/wer.1642>
- [27] Losing, V. et al. (2015). Interactive online learning for obstacle classification on a mobile robot. *IEEE International Joint Conference on Neural Networks*, 1-8. <https://doi.org/10.1109/IJCNN.2015.7280610>
- [28] Hernández, A. C. et al. (2016). Object classification in natural environments for mobile robot navigation. *IEEE International Conference on Autonomous Robot Systems and Competitions*, 217-222. <https://doi.org/10.1109/ICARSC.2016.55>
- [29] Himmelsbach, M. et al. Real-time object classification in 3D point clouds using point feature histograms. *IEEE International Conference on Intelligent Robots and Systems*, 994-1000. <https://doi.org/10.1109/IROS.2009.5354493>
- [30] Spiers, A. J. et al. (2016). Single-grasp object classification and feature extraction with simple robot hands and tactile sensors. *IEEE Transactions on Haptics*, 9(2), 207-220. <https://doi.org/10.1109/TOH.2016.2521378>
- [31] Ayub, A. & Wagner, A. R. (2020). Tell me what this is: few-shot incremental object learning by a robot. *IEEE International Conference on Intelligent Robots and Systems*, 8344-8350. <https://doi.org/10.1109/IROS45743.2020.9341140>
- [32] Jiang, Y. et al. (2020). Robotic Grasp Detection Using Lightweight CNN Model. *IEEE Chinese Control And Decision Conference*, 1034-1038. <https://doi.org/10.1109/CCDC49329.2020.9164710>
- [33] Abbas, M. et al. (2020). AlexNet based Real-Time Detection and Segregation of Household Objects using Scorbot. *IEEE 4th International Conference on Computational Intelligence and Networks*, 1-6. <https://doi.org/10.1109/CINE48825.2020.234392>
- [34] He, K. et al. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2016.90>
- [35] Bang, S. et al. (2019). Encoder-decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering*, 34(8), 713-727. <https://doi.org/10.1111/mice.12440>
- [36] Wang, H. & Xia, Y. (2018). Chestnet: A deep neural network for classification of thoracic diseases on chest radiography.
- [37] Son, H. et al. (2019). Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Automation in Construction*, 99, 27-38. <https://doi.org/10.1016/j.autcon.2018.11.033>
- [38] Ioffe, S. & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*. <https://doi.org/10.5555/3045118.3045167>
- [39] Nair, V. & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. *International Conference on Machine Learning*. <https://doi.org/10.5555/3104322.3104425>
- [40] Farooq, A. et al. (2017). A deep CNN based multi-class classification of Alzheimer's disease using MRI. *IEEE International Conference on Imaging systems and techniques*. <https://doi.org/10.1109/IST.2017.8261460>
- [41] Korfiatis, P. et al. (2017). Residual deep convolutional neural network predicts MGMT methylation status. *Journal of digital imaging*, 30(5), 622-628. <https://doi.org/10.1007/s10278-017-0009-z>
- [42] CIPA DCG-001-2005 (2005). Guideline for Noting Digital Camera Specifications in Catalogs.
- [43] CIPA DC-003-2003 (2003). Resolution measurement methods for digital cameras. Standard of the Camera & Imaging Products Association, Japan.

Contact information:**Merdan OZKAHRAMAN**

(Corresponding author)
Isparta University of Applied Sciences, Department of Mechatronics Engineering,
Isparta, 32260 Turkey
E-mail: merdanozkahraman@isparta.edu.tr

Haydar LIVATYALI

Yildiz Technical University, Department of Mechatronics Engineering,
Istanbul, 34349 Turkey
E-mail: hlivatya@yildiz.edu.tr