

Advancing Workplace Safety: A Proactive Approach with Convolutional Neural Network for Hand Pose Estimation in Press Machine Operations

Şuayip Aykut ATMACA, Hüseyin HAMAD, Burcu Çağlar GENÇOSMAN

¹ *Şuayip Aykut ATMACA, Orcid ID: <https://orcid.org/0000-0002-7146-641X>, aykutatmaca@trex.com.tr,

² Hüseyin HAMAD, Orcid ID: <https://orcid.org/0000-0003-3189-3246>, huseyinhamad@trex.com.tr,

³ Burcu Çağlar GENÇOSMAN, Orcid ID: <https://orcid.org/0000-0003-0159-8529>, burcucaglar@uludag.edu.tr,

* Correspondence: aykutatmaca@trex.com.tr; mobile phone +905374296371

(First received September 27, 2023 and in final form December 28, 2023)

Reference: Atmaca, Ş., A., Hamad, H., Gençosman, B., Ç. Advancing Workplace Safety: A Proactive Approach with Convolutional Neural Network for Hand Pose Estimation in Press Machine Operations. The European Journal of Research and Development, 3(4), 66-75.

Abstract

Press machine operations are integral to goods production across industries, yet worker safety faces significant risks. Machine misuse and non-compliance with safety standards contribute substantially to these incidents. This study addresses the mounting concerns regarding workplace incidents through a proactive solution—a Convolutional Neural Network (CNN) model crafted to prevent press machine misuse by monitoring workers' hand placement during operation. The model that we suggest ensures adherence to safety standards. The CNN model does not replace the role of human operators but acts as a supportive layer, providing instant feedback and intervention when deviations from safety standards are detected. In conclusion, this research endeavors to pave the way for a safer and more secure industrial environment by leveraging the capabilities of advanced technology. The proposed CNN model addresses current concerns and sets a precedent for future advancements in ensuring workplace safety across diverse industries.

Keywords: Hand-pose estimation, work safety, deep learning, CNN, hyperparameter tuning

1. Introduction

In various industries, the functioning of press machines is crucial to producing various goods [1]. However, this necessary machinery is not without risks, and the frequency of workplace incidents involving press machines has prompted concerns about worker safety over the last decade. Statistical analyses of occupational safety reports and incident databases reveal a concerning increase in the frequency and severity of events involving press machine operations [2]. Over the past ten years, these incidents have resulted in a

sizable number of injuries, ranging from mild to severe, and tragically, in some cases, fatalities.

According to data provided by the Occupational Safety and Health Administration (OSHA), there are 3933 recorded incidents only in the United States of America (USA), ranging from 1984 to 2023. Some of these incidents resulted in fractures or amputations, but sadly, 282 of these 3933 incidents resulted in fatalities [3]. Another data that the Bureau of Census provided in 1980 shows that only in the United States there are approximately 151,000 press machine operators, and yearly, 20,000 incidents end with amputation [4].

DHHS, also known as the National Institute for Occupational Safety & Health (NIOSH), submitted a publication about the injuries and amputations resulting from working with mechanical power press in March 1987. According to the paper, some of amputations occurred when the operator placed their hand in the working zone, hence misusing the equipment [5]. The publication mentions some standards for using press machines, such as the OSHA 1910.217 standard. According to the standard, every time the dual palm buttons are pressed, mechanical power presses must have a single-stroke (or anti-repeat) feature that enables the clutch to engage and the press to cycle only once, and the press must be activated using both hands; thus both palm buttons must have guards to avoid unintentional operation and space between them to stop operators from "bridging" the buttons.

When the reports of these data and the publication from DHHS are investigated, it is realized that the misuse of machines and the non-compliance with NIOSH standards by workers lead to numerous accidents. To comply with this issue, we have developed a Convolutional Neural Network (CNN) model that will stop the workers from misusing the equipment by checking if their hands are placed safely for the entire operation by following OSHA 1910.217 standard.

Research and development on the localization of hand movements has become essential in computer science and artificial intelligence (AI) [6-9]. There has been a significant advancement in hand position tracking and detection with a wide range of applications, including sensor-based [10-11] and vision-based [12-13] approaches. Estimating hand poses with computer vision applications, including virtual reality, augmented reality, sign language detection [14], and gesture recognition studies, have been increased with the development of deep learning methods [15-17].

Because of their ability to handle complex tasks, Deep Neural Networks (DNNs), especially Convolutional Neural Networks (CNNs), have become a prominent tool in computer vision applications [18]. The filter kernels must be updated during this procedure for every input and output. Depending on the assignment, a DNN may need a significant amount of training data points—up to millions in certain situations [19].

Neethu et al. [20] provided a convolutional neural networks (CNN) classification approach to detect and identify human hand gestures. They used a CNN classifier to segment the hand's region of interest through a mask image, the segmentation of fingers, normalization of segmented finger images, and finger recognition.

Mohanty et al. [21] tried to identify the motionless hand movements while confronted with intricate backgrounds with fluctuating lighting circumstances with CNN. Three publicly available benchmark datasets—the NUS hand posture dataset with a cluttered environment, the Triesh hand posture dataset with a uniform dark background, and the Marcel hand posture dataset—were used to test their proposed model, which consists of two convolutions and pooling layers with ReLU activation function.

Studies on hand pose recognition in industry have focused on smart manufacturing and the use of VR glasses [22-24]. This study considers a hand pose estimation problem that estimates the position of workers' hands using video images to increase work safety when using a press machine in a production environment. To the best of our knowledge, there has not yet been a study that estimates the hand position before machine use to increase work safety. With this system, which is developed to prevent workplace accidents that can lead to severe financial losses if they occur, a contribution to the literature has been made. Suggestions for its use in similar machines have been developed, and how it can be disseminated in the industry has been revealed. During research, we stumbled upon various methods that we can use to train our CNN model to achieve the best accuracy to detect hand positions on the canvas as correctly as possible. The following section details the data collection process and the proposed CNN model.

2. Materials and Methods

To decrease the number of incidents that happen while using press machines, we have created an environment similar to what a press machine operator faces every day and collected images labeled 0 (Not Ok to Operate) and 1 (Ok to Operate) according to the position of the hands on the canvas.

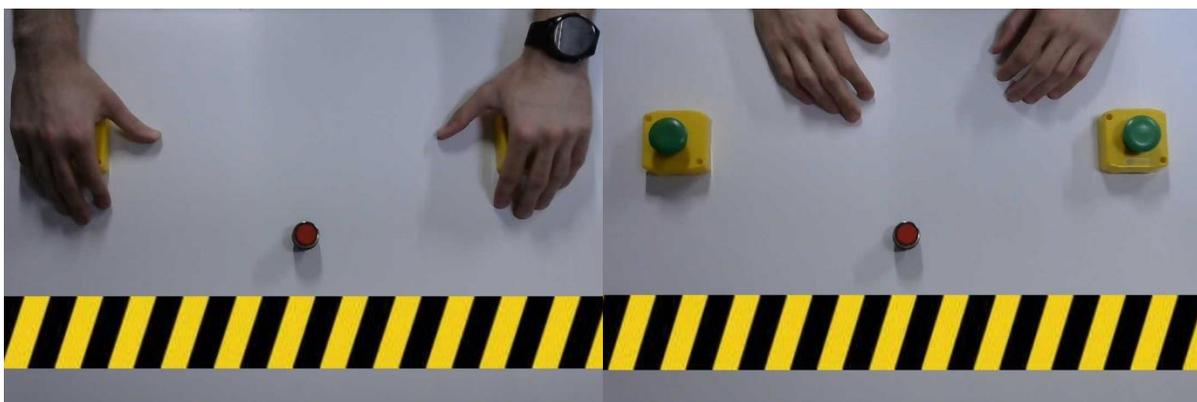


Figure 1: 'OK' and 'NOK' labelled hand positions

After collecting the images, we have developed a Convolutional Neural Network (CNN) model and trained it with the images that we collected. This model will be used by a Manufacturing Execution System (MES) to decide if the machine should operate or not by checking live from a camera feed and deciding if the operator has both hands replaced correctly on the buttons.

2.1. Data Collection and Pre-processing

We have collected a diverse dataset of 2893 images, each with a resolution of 1280x720 pixels and three-color channels. This dataset is split into training and test sets to facilitate model training and evaluation. Before inputting the data into our CNN model, we resized the images to (180, 320, 3) to match the specified input dimensions.

2.2. Convolutional Neural Network Architecture

The CNN architecture is designed with three convolutional layers, each employing ReLU activation functions. The number of neurons progressively increases (32, 64, and 64), and each convolutional layer is followed by a MaxPooling layer to downsample the spatial dimensions of the feature maps. A flattening layer succeeds the final convolutional layer to convert the 3D output to a 1D vector. Subsequently, two dense layers, one of which has 64 and the other has 2 units, lead to the output layer with two units and a softmax activation function, suitable for binary classification. The General CNN structure is shown in Figure 2.

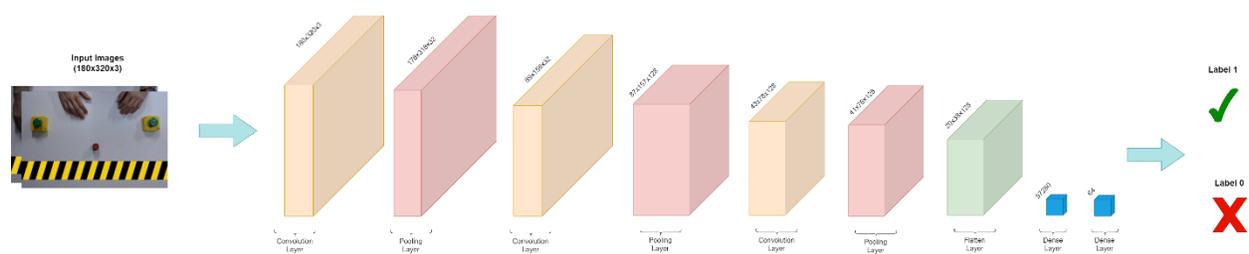


Figure 2: CNN Architecture Demonstration

2.3. Hyperparameter Tuning for Performance

We use hyperparameter tuning with different optimizer values to get more efficient model parameters. Tuning operation is based on accuracy value. Table 1 shows the alternative values of the hyperparameter tuning operation.

Table 1: Alternative values of parameter tuning

Variables	Values
Batch Size	[25, 30, 60, 90]
Learning Rate	[0.01, 0.001, 0.0001]
Neuron Count (Step: 32)	(32, 256)

The best-tuned values are identified as follows: batch size is set to 40, learning rate is configured at 0.001. The model architecture consists of convolutional layers with the ReLU activation, followed by MaxPooling layers. The first convolutional layer employs 32 filters, the second has 64 filters, and the final layer has 64 filters. Each convolutional layer is followed by a MaxPooling layer. The feature maps are flattened, leading to a dense layer with 64 neurons and softmax activation. The output layer is configured with 2 neurons and softmax activation, tailored for a binary classification task.

2.4. Model Compilation and Training

We utilize the Adam Optimizer and sparse categorical cross-entropy loss function for model compilation, with accuracy as the evaluation metric. The training process includes running the model on the training images and labels for 100 epochs while the batch size is 40. We employ the validation set during the training to monitor and prevent overfitting. Early stopping, implemented as a callback, ensures that training will cease if the validation accuracy does not show improvement. Figure 3 shows the architecture of the CNN model, including the best parameters.

2.5. Class Labels

Our binary classification task is defined by the class labels "0" and "1," representing the distinct categories in the study. While "0" represents the negative situation, the value "1" represents the positive situation.

The system that we recommend, additionally the current press machine systems, provides to run after detecting hand position and existence using image processing. The main reason for the application is that in manufacturing, one of the buttons can be bypassed by different techniques (putting stuff on one of the buttons, tying it up with a string, etc.). This situation could cause the operator to push the button with one hand and intervene in the workpiece that is processed in the press machine with the other hand. Thanks to our deep learning model, besides the current button-controlled system, the system is not allowed to run until the model validates the hand position and existence.

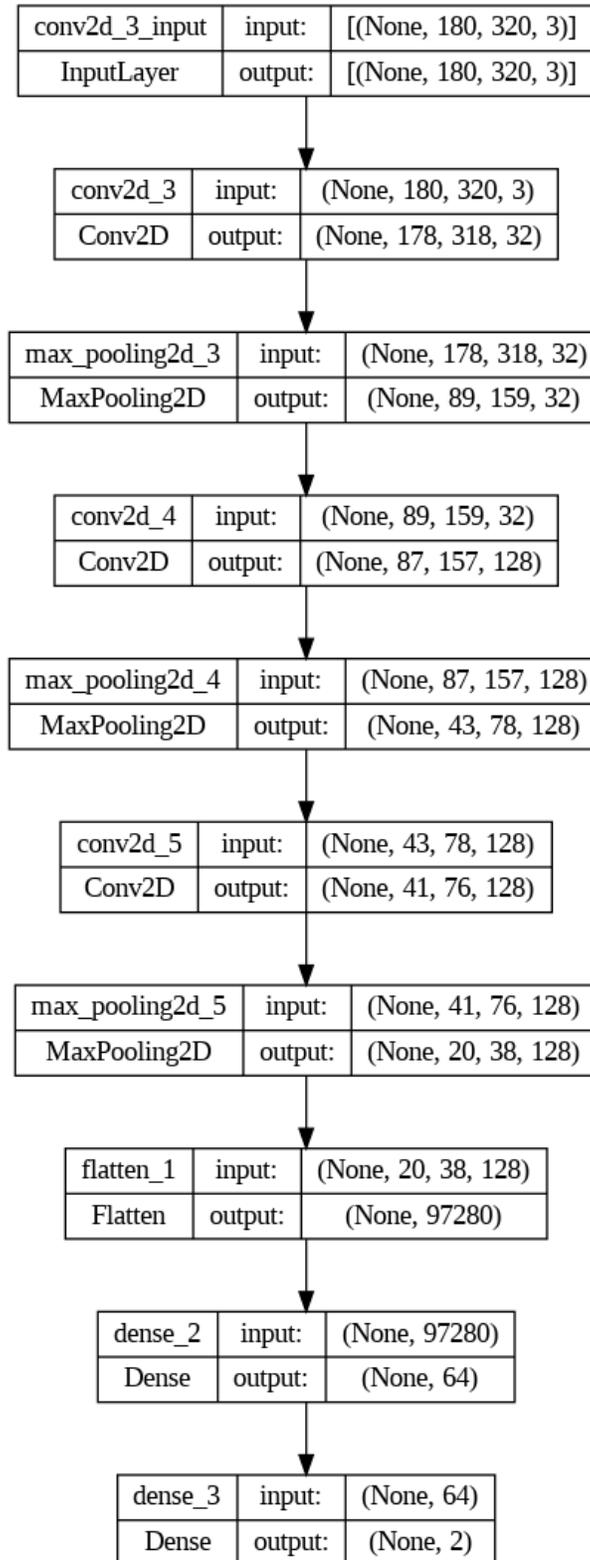


Figure 3: CNN Architecture Layers and Parameters

3. Results

We have collected 2893 images and have used 80% of them for training and 20% for testing. In addition, we have performed the validation process with a discrete dataset of 275. Our model has demonstrated exceptional performance across various metrics, presented in equations (1-4), showcasing its robustness and accuracy. The evaluation results are shown in Table 2.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (3)$$

$$F_1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (4)$$

Table 2: Evaluation Results

Metric	Value
Accuracy	0.9854
Precision	0.9785
Recall	0.9927
F1 Score	0.9856
ROC-AUC Score	0.9854

These metrics highlight the effectiveness of our approach in accurately predicting hand positions. The high accuracy and precision indicate a low rate of misclassifications, while the elevated recall suggests a minimal number of false negatives. The F1 score, balancing the precision and recall, reinforces the model's overall reliability. Moreover, the confusion matrix presented in Table 3 proves the success of the CNN model.

Table 3: Confusion Matrix

	Predicted Positive	Predicted Negative
Actually Positive	137	1
Actually Negative	3	134

The ROC-AUC score in Figure 4 further validates the model's ability to distinguish between different hand positions, with a score of 98.54% signifying a high level of discriminatory power.

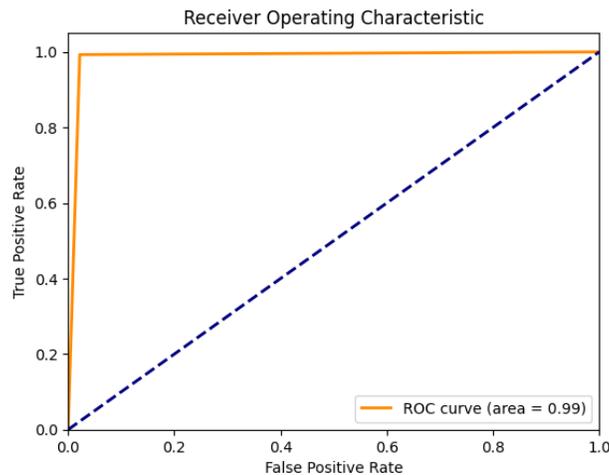


Figure 4: The ROC-AUC Graph

These results not only underscore the success of our hand position estimation model but also position it as a robust solution for applications demanding precise and reliable hand tracking.

4. Discussion and Conclusion

Occupational safety is one area of industry where artificial intelligence applications are being used extensively on a daily basis. They play a significant role in mitigating human-induced errors, particularly in industrial settings such as press operations, where the distinction lies in the potential for accidents to result in organ loss or fatality. These AI applications function as valuable support systems, contributing to the prevention of catastrophic incidents and enhancing overall workplace safety.

By analyzing real-time photos of operators' hand placements from the camera above the press machine, a system has been built to support the current system in press machines. It determines whether or not the operators are positioned in a safe working position. The safe working position is achievable by the high accuracy rate our planned and produced artificial intelligence application has achieved. This strategy aims to prevent any potential work-related accidents and operators utilizing the machine in a hazardous manner.

- S. S. Rautaray and A. Agrawal, 'Vision based hand gesture recognition for human computer interaction: a survey', *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, Jan. 2015, doi: 10.1007/s10462-012-9356-9.
- [13] F. Al Farid et al., 'A Structured and Methodological Review on Vision-Based Hand Gesture Recognition System', *J. Imaging*, vol. 8, no. 6, Art. no. 6, Jun. 2022, doi: 10.3390/jimaging8060153.
- [14] 'S. Ameen and S. Vadera, "A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images," *Expert Systems*, vol. 34, no. 3, 2017. doi:10.1111/exsy.12197'.
- [15] J. Wang, T. Liu, and X. Wang, 'Human hand gesture recognition with convolutional neural networks for K-12 double-teachers instruction mode classroom', *Infrared Phys. Technol.*, vol. 111, p. 103464, Dec. 2020, doi: 10.1016/j.infrared.2020.103464.
- [16] A. Erol, G. Bebis, M. Nicolescu, R. Boyle, and X. Twombly, 'Vision-based hand pose estimation: A review', *Comput. Vis. Image Underst.*, vol. 108, pp. 52–73, Oct. 2007, doi: 10.1016/j.cviu.2006.10.012.
- [17] D. Sarma and M. K. Bhuyan, 'Methods, Databases and Recent Advancement of Vision-Based Hand Gesture Recognition for HCI Systems: A Review', *SN Comput. Sci.*, vol. 2, no. 6, p. 436, 2021, doi: 10.1007/s42979-021-00827-x.
- [18] 'I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016'.
- [19] 'F. Chollet, *Deep learning with Python*. Manning Publications Company, 2017'.
- [20] 'Neethu, P. S., Suguna, R., & Sathish, D. (2020). An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks. *Soft Computing*, 24, 15239-15248'.
- [21] 'Mohanty, Aparna & Rambhatla, Sai & Sahay, Ram. (2017). *Deep Gesture: Static Hand Gesture Recognition Using CNN*. 10.1007/978-981-10-2107-7_41'.
- [22] 'Industry use of virtual reality in product design and manufacturing: a survey | *Virtual Reality*'. Accessed: Nov. 22, 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s10055-016-0293-9>
- [23] B. Wang, P. Zheng, Y. Yin, A. Shih, and L. Wang, 'Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective', *J. Manuf. Syst.*, vol. 63, pp. 471–490, Apr. 2022, doi: 10.1016/j.jmsy.2022.05.005.
- [24] S. Phuyal, D. Bista, and R. Bista, 'Challenges, Opportunities and Future Directions of Smart Manufacturing: A State of Art Review', *Sustain. Futur.*, vol. 2, p. 100023, Jan. 2020, doi: 10.1016/j.sftr.2020.100023.