Application of transformer nets for the discrimination of liquid cleaning products

Jiménez-Moreno Robinson[a], Espitia-Cubillo Anny Astrid[b], Rodríguez-Carmona Esperanza[b]

[a] Associate Professor, Mechatronics Engineering Program, Universidad Militar Nueva Granada, Bogotá, Colombia.
[b] Associate Professor, Industrial Engineering Program, Universidad Militar Nueva Granada, Bogotá, Colombia.
Corresponding author: anny.espitia@unimilitar.edu.co

Abstract. This paper presents an artificial intelligence algorithm based on transformer neural networks that allows the discrimination of liquid products identified with different labels and with presentations in various colors, from a camera, to facilitate the management of their subsequent handling by a computer, which allows us to realize, in manufacturing environments, the connection between the physical and the digital world. The process begins with the digitalization of the products to establish a database. Next, the parameters of the network training are defined, which are then evaluated by measuring the learning time, accuracy, and classification time, all of which are developed in a virtual environment. Thanks to the results, it is possible to conclude that even with a small amount of data, including label images that are not complete or of the best quality, the processing times do not exceed 0.5 seconds. A recognition rate of 100% accuracy is achieved, corresponding to the absence of confusion between the considered categories, given the robustness of the selected transformer network.

Keywords: Artificial intelligence, discrimination of products, labeling, transformer neural networks.

1. INTRODUCTION

Industry 4.0 is a concept implemented at different levels in manufacturing companies (Habib, Bnouachir, Chergui, & Ammoumou, 2022), some applications are based on the use of artificial intelligence, using the Faster RCNN model (Saleem, Potgieter, & Arif, 2022), deep learning systems (Vaddadi, y otros, 2022) (Wu, 2023), (Yoon, Han, & Nguyen, 2023), applying Internet of Thigs (IoT) (Bose, Mondal, Sarkar, & Roy, 2022) or with architectures such as YOLO (Qureshi, et al., 2024) (Qi & Sun, 2024) (Zheng, Chen, Cheng, Du, & Jiang, 2024).

(Mark, Rauch, & Matt, 2022) propose to understand artificial intelligence as an instrument that facilitates the execution of activities and supports decision makers, a vision shared in the present development.

The identification of standardized products can be automated, allowing for real-time metrics such as productivity, inventory levels, and product location, among others, which facilitates the timely making of informed decisions in production and logistics processes. Transformer networks (Touvron, Cord, El-Nouby, Verbeek, & Jégou, 2022) serve this purpose and have had previous successful industrial applications, which are discussed below.

For example, Ma et al. (2024) used a hierarchical transformer network to recognize industrial apparel production operations, utilizing clustering of networks to improve accuracy and reduce overhead.

For chemical processes, Wang et al. (2023) propose a prediction model based on an improved transformer network. For the optimization of petroleum processes, (Huang, et al., 2024) propose a model called Time Patch Dynamic Attention Transformer (TPDAT) that segments the data to improve the recognition of events and transient fluctuations; (Ma, Li, & Yuan, 2024) use a Transformer model to process the operation data of a well to improve production efficiency, reduce maintenance costs and increase the useful life of the equipment.

A Transformer network is also used to predict the useful life of aircraft engines, which is composed of a mechanism that dynamically extracts and weights features, yielding good performance in object detection, traffic forecasting, image segmentation, failure risk, and loss reduction (Liu, Song, & Zhou, 2022). In the energy field, (Al-Ali, et al., 2023) with a CNN-LSTM-Transformer model (Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) network, and a Transformer), they forecast solar energy production using clusters to analyze the correlation of input data along with their dynamic features. The results had higher accuracy than those obtained with models such as LSTM-CNN.

2. MATERIALS AND METHODS

As part of the methodology, five phases were defined for the development of this study. The starting point was the capture of images of each product. Secondly, the images are organized in a database. Thirdly, the network architecture to be used is selected. Then, the training parameters are defined, and finally, the results obtained are evaluated.

The process begins with the selection and digitization phase of the products of interest, which will allow the establishment of an image database.

In the second phase, the database includes seven of the most in-demand products from a cleaning products factory, which are packaged in the same type of container and, in some cases, have the same color. The specific differentiation between each one corresponds to the product label attached to the container. The database includes digital images of the label from both the front and the side, which display at least half of the label. Also, the focal length of the camera affects the resolution of the images; sometimes, the label text is not entirely legible, which makes the recognition task difficult, so individually traditional architectures have not yielded promising results, requiring the integration of techniques such as CNN with SURF (Guacheta-Alba, Espitia-Cubillos, & Jiménez-Moreno, 2024). The database is small, as the product identification features are focused on the packaging. To achieve this, 20 images of each product are taken and divided into 75% for training (15 images), 10% for validation (2 images), and 15% for testing (3 images) per category. A sample of the database is presented in Table 1.

Table 1. Categories and samples of the database used.

Product	Training examples			
Detergent				
Dishwasher				
Cleaner				
Softener				

In third place, a learning transfer network architecture based on transformer models is selected for product identification (Dosovitskiy et al., 2021).

In the fourth phase, the parameters for the training of the network are defined. Table 2 illustrates the parameters for training, which are derived from the database and require a small mini-batch size, corresponding to 2 for the application. The input volume of the network required a previous resizing of the images to a size of 384 pixels per side, as color images (RGB), implying a depth volume of 3 for each color component. Given the predefined architecture of the network, which has 143 depth layers and is quite robust, only a few training epochs are used, as seen in case 20.

Table 2. Categories and samples of the database used.

Parameter	Value
Input volume	384 X 384 X 3
Minibatch size	2
Optimizer	SGDM
Epochs	20
Learning rate	0.0001
Number of classes	7

Finally, in the fifth phase, the results obtained are evaluated using a confusion matrix, the percentage of accuracy, the training time, and the time taken by the network to classify the products.

The transformer neural network, which allows for the discrimination of liquid products during training and testing, was implemented using an Intel Core i9 laptop computer with 24GB of RAM and an NVIDIA GeForce RTX 4080 graphics card.

3. RESULTS AND DISCUSSION

The network training time was 8 minutes 55 seconds, with an additionality by epoch 5, where an accuracy of 100% was already achieved, as shown in Figure 1.
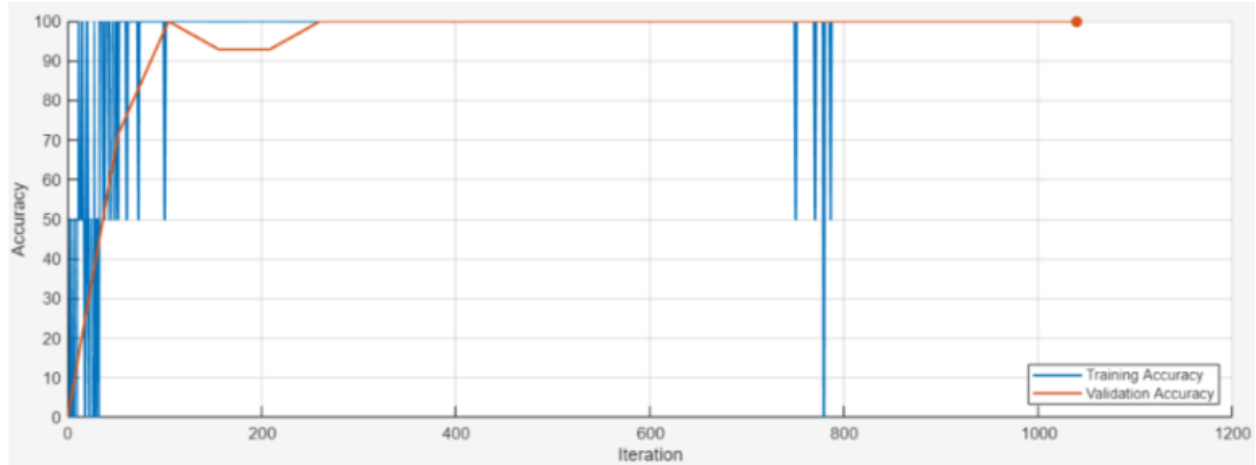
Figure 1. Progress in network training.

The confusion matrix analysis shows that the level of accuracy reached, corresponding to 100%, is reflected in a set of tests (3 images per category). Figure 2 illustrates that there is no confusion between classes, which, given the small number of images used, demonstrates the network's remarkable capacity to process the identification of each product according to the learning requirements specified by the database.
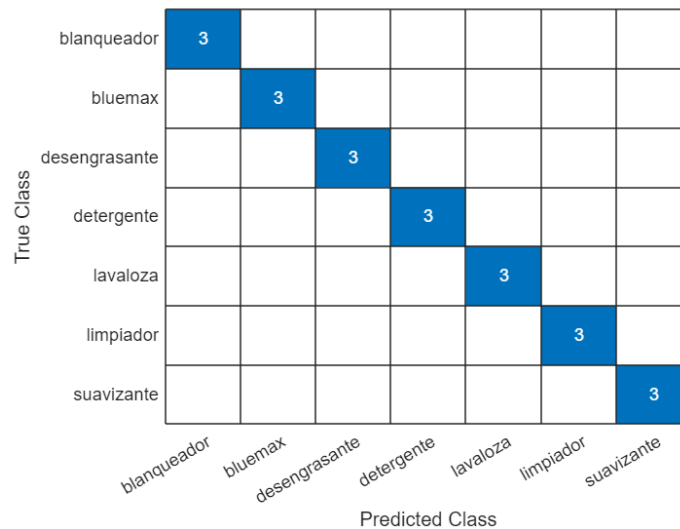

Figure 2. Network test confusion matrix.

Additionally, other images outside the initial set are used for final testing and measurement of the response time in classifying the network. This result is presented in Table 3, which highlights changes in background color, product location, and size, allowing for a clear identification of each class with an average time of 0.467 seconds. This demonstrates that, even with an architecture as deep as the one used, the response time is fast for a product manufactured on automated production lines, for shipping, inventory, or readiness.

Table 3. Categories and samples of the database used.

Validation	Imagen	Classification time
1	suavizante 	0.442301 seconds
2	bluemax 	0.492344 seconds

The results achieved are contrasted with other works that share the objective of classifying products using artificial intelligence that require the integration of models (Guacheta-Alba, Espitia-Cubillos, & Jiménez-Moreno, 2024) for the recognition of similar databases, which evidences the impact of the use of transformer networks in industry-oriented applications, having adequate time for product manipulation once it has been identified.

In this regard, Espitia Cubillos, Jiménez Moreno, & Rodríguez Carmona (2025) point out that to select the most convenient model according to the end user's objectives in an industrial production environment, training times and, in this case, the depth of the network are not relevant, but criteria such as accuracy and classification times are.

CONCLUSIONS

The evidence obtained in the present study suggests that the use of robust transformer architectures does not necessitate an extensive database, even when they are applied to incomplete and blurred images. With the training criteria established, despite the depth of the selected transformer network, training the network took less than 9 minutes and allows processing in under 0.5 seconds, which is useful for automation applications that require both agile implementation and execution. The 100% accuracy level achieved is remarkable, all the more so when it was reached after only five epochs.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest within this research, authorship, and/or publication of this article.

REFERENCES

Al-Ali, M., Hajji, Y., Said, Y., Hleili, M., Alanzi, A., Laatar, A., & Atri, M. (2023). Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model. *Mathematics*, 11(3), 676. https://doi.org/10.3390/math11030676

Bose, R., Mondal, H., Sarkar, I., & Roy, S. (2022). Design of smart inventory management system for construction sector based on IoT and cloud computing. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 2, 100051. https://doi.org/10.1016/j.prime.2022.100051

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & al., M. D. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929

Espitia Cubillos, A. A., Jiménez Moreno, R., & Rodríguez Carmona, E. (2025). Deep learning architectures for location and identification in storage systems. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 592-601. https://doi.org/10.11591/ijai.v14.i1.pp592-601

Guacheta-Alba, J. C., Espitia-Cubillos, A. A., & Jiménez-Moreno, R. (2024). Automated Box Classification in a Virtual Industrial Environment Using Machine Vision Algorithms. *12th International Conference on Control, Mechatronics and Automation (ICCMA)* (págs. 305-310). London, UK: IEEE. https://doi.org/10.1109/ICCMA63715.2024.10843920

Habib, F. E., Bnouachir, H., Chergui, M., & Ammoumou, A. (2022). Industry 4.0 concepts and implementation challenges: Literature Review. *2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM),* (págs. 1-6). Rabat, Morocco. https://doi.org/10.1109/WINCOM55661.2022.9966456

Huang, T., Qian, H., Huang, Z., Xu, N., Huang, X., Yin, D., & Wang, B. (2024). A time patch dynamic attention transformer for enhanced well production forecasting in complex oilfield operations. *Energy*, 309, 133186. https://doi.org/10.1016/j.energy.2024.133186

Liu, L., Song, X., & Zhou, Z. (2022). Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture. *Reliability Engineering & System Safety*, 221, 108330. https://doi.org/10.1016/j.ress.2022.108330

Ma, Y., Li, X., & Yuan, C. (2024). Intelligent prediction of oil well working conditions based on Transformer. *Journal of Physics: Conference Series*, 2901, 012022. https://doi.org/10.1088/1742-6596/2901/1/012022

Ma, Y., Wang, X., Yuan, J., Zhang, L., Chen, J., & Fen, K. (2024). Clothing Detection Action Recognition Based on Hierarchical Transformer Networks. *16th International Conference on Communication Software and Networks (ICCSN),* (págs. 199-206). Ningbo, China. https://doi.org/10.1109/ICCSN63464.2024.10793382

Mark, B. G., Rauch, E., & Matt, D. T. (2022). Systematic selection methodology for worker assistance systems in manufacturing. *Computers & Industrial Engineering*, 166, 107982. https://doi.org/10.1016/j.cie.2022.107982

Qi, Y., & Sun, H. (2024). Defect Detection of Insulator Based on YOLO Network. *9th International Conference on Electronic Technology and Information Science (ICETIS)*, (págs. 232-235). Hangzhou, China. https://doi.org/10.1109/ICETIS61828.2024.10593675

Qureshi, A., Butt, A., Alazeb, A., Mudawi, N., Alonazi, M., Almujally, N., . . . Liu, H. (2024). Semantic Segmentation and YOLO Detector over Aerial Vehicle Images. *Computers, Materials & Continua*, 80(2). https://doi.org/10.32604/cmc.2024.052582

Saleem, M. H., Potgieter, J., & Arif, K. M. (2022). Weed Detection by Faster RCNN Model: An Enhanced Anchor Box Approach. *Agronomy*, 12(7), 1580. https://doi.org/10.3390/agronomy12071580

Touvron, H., Cord, M., El-Nouby, A., Verbeek, J., & Jégou, H. (2022). Three things everyone should know about vision transformers. En S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Edits.), *Computer Vision–ECCV,* 13684, 497-515. Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-20053-3_29

Vaddadi, S., Srinivas, V., Reddy, N., Girish, H., Rajkiran, D., & Devipriya, A. (2022). Factory Inventory Automation using Industry 4.0 Technologies. *022 IEEE IAS Global Conference on Emerging Technologies (GlobConET),* (págs. 734-738). Arad, Romania. https://doi.org/10.1109/GlobConET53749.2022.9872416

Wang, S., Sun, H., Wang, Y., & Luo, X. (2023). Prediction of key chemical parameters based on improved Transformer. *4th International Conference on Computer Engineering and Application (ICCEA),* (págs. 855-859). Hangzhou, China. https://doi.org/10.1109/ICCEA58433.2023.10135471

Wu, B. (2023). Motion Control Algorithm for Automatic Welding of Complex Intersecting Line Joints Based on Deep Learning. *2023 International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII),* (págs. 352-356). Melbourne, Australia. https://doi.org/10.1109/ICMIII58949.2023.00073

Yoon, J., Han, J., & Nguyen, T. (2023). Logistics box recognition in robotic industrial de-palletising procedure with systematic RGB-D image processing supported by multiple deep learning methods. *Engineering Applications of Artificial Intelligence*, 123(B), 106311. https://doi.org/10.1016/j.engappai.2023.106311

Zheng, H., Chen, X., Cheng, H., Du, Y., & Jiang, Z. (2024). MD-YOLO: Surface Defect Detector for Industrial Complex Environments, *Optics and Lasers in Engineering*, 178, 108170. https://doi.org/10.1016/j.optlaseng.2024.108170