Journal of Hunan University

(Natural Sciences)

Vol. 52 No. 11

November 2025

Available online at

https://jonuns.com

Open Access Article https://doi.org/10.55463/issn.1674-2974.52.11.7

Automatic box sorting system using artificial intelligence and voice interaction

Jose Miguel Bejarano[1], Robinson Jiménez-Moreno[2*], Javier Eduardo Martinez-Baquero[3]

[1] Student of Department of Mechatronics, Engineering Faculty, Universidad Militar Nueva Granada, Bogotá, Colombia,

[2] Professor of Department of Mechatronics, Engineering Faculty, Universidad Militar Nueva Granada, Bogotá, Colombia,

[3] Professor of Engineering School, Faculty of Basic Sciences and Engineering, Universidad de los Llanos, Villavicencio, Colombia,

* Corresponding author: robinson.jimenez@unimilitar.edu.co

Article history:
Received: October 12, 2025
Revised: November 20, 2025
Accepted: December 12, 2025
Published: December 30, 2025

Abstract: This article compares three artificial intelligence algorithms in the development of an automatic box sorting system using a voice-based user interaction interface. The system classifies between four different box sizes in a virtual environment where it extracts the desired size from the production line, which is identified by voice from a user interface. With the advances in artificial intelligence techniques, it is necessary to contextualize and establish the applicability of each one to generate developments that do not focus on the most recent algorithm but on the optimal one that satisfies a design requirement. By comparing techniques such as nearest neighbors, artificial neural networks, and convolutional neural networks, the effectiveness of each technique in a simple classification task is evident. Given the low computational resource usage of the nearest neighbor algorithm and its reduced response times and computational cost, the advantage of traditional algorithms over advanced pattern recognition algorithms is evident, as they can be integrated with recent voice recognition algorithms in industrial-type virtual scenarios.

Keywords: Nearest neighbors, Artificial neural networks, Convolutional neural networks, Automatic speech recognition, Box classification, Virtual environment.

采用人工智能与语音交互技术的自动箱体分拣系统

摘要： 本文比較三種人工智慧演算法在採用語音式使用者互動介面的自動箱體分揀系統開發中的應用。該系統於虛擬環境中對四種不同尺寸的箱體進行分類，並根據使用者介面所接收的語音指令，從生產線中提取所需尺寸的箱體。隨著人工智慧技術的進展，有必要對各種演算法加以情境化並確立其適用性，以推動相關系統的開發——此類開發不應僅著眼於最新的演算法，而應選擇能夠滿足設計需求的最佳方案。透過比較最近鄰法、人工神經網路及卷積神經網路等技術，可以清楚看出各方法在簡單分類任務中的效能表現。鑑於最近鄰演算法具有較低的運算資源需求、較短的回應時間以及較低的計算成本，其相較於先進的模式識別演算法所展現出的優勢顯而易見，並且能夠與當代語音識別演算法整合，應用於工業類型的虛擬場景中。

关键词：最近邻算法，人工神经网络，卷积神经网络，自动语音识别，盒式分类，虚拟环境。

1. Introduction

Pattern recognition models for classification are based on artificial intelligence algorithms that currently focus on deep learning [1][2], among which convolutional neural networks [3] stand out. The applications of pattern classification systems range from waste in vending machine containers [4], to the assessment of the toxicity of chemicals associated with plastic packaging [5], and to multi-output production chains for packaging planning [6].

Artificial intelligence is therefore the basis for current industry-oriented developments in process automation, for example, in the use of images for precision agriculture [7] and waste detection and treatment in the construction industry [8]. Many of these applications are based on robotic systems that are being simulated in virtual environments [9] for various types of developments in healthcare [10], education [11], medicine [12][13], among others.

The trend in robotic systems is toward natural voice interaction, with developments currently being made in voice-controlled medical systems [14], wheelchairs [15], and other robotic vehicles [16][17], but with a growing boom in the industry for robotic agents [18], gripping systems [19], and collaborative robotics based on interaction through natural language processing [20].

The state of the art shows the clear use of traditional and cutting-edge artificial intelligence algorithms, but it is necessary to contextualize and establish the applicability of each one to generate developments that do not focus on the most recent algorithm but on the optimal one that satisfies a design requirement, regardless of its simplicity or advanced learning structure, which is the main analysis of this applied research work.

Based on the above, this article presents a comparison of artificial intelligence algorithms, from the basics nearest neighbor technique and backpropagation neural networks to deep learning by convolutional networks, aimed at a box classification agent through voice interaction.

The rest of the article is structured as follows: the next section describes the methodology employed, detailing the integration process using Python and Coppelia simulation environments. This is followed by a discussion of experimental results obtained, and finally, conclusions are drawn highlighting insights into the effectiveness and challenges of the proposed approach.

2. Methodology

With the aim of implementing an automatic box selection system from a conveyor belt based on voice command selection of box size, the methodology was structured in three phases. The first phase corresponds to the general elements of the design based on the hardware architecture used and the simulation environment. In the second phase, the database to be used and the box classification criteria are established, and phase three outlines the training of the artificial intelligence models used. Figure 1 illustrates the methodology flowchart.

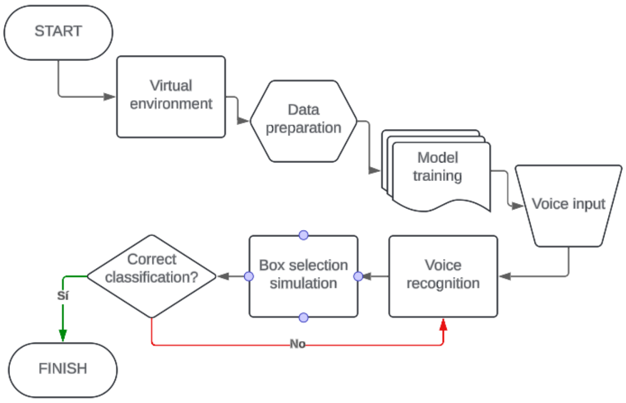

Figure 1. Methodology flowchart (Source: developed by the authors)

2.1. Hardware, Software, and Simulation Environment

The system was implemented on a laptop with an AMD Ryzen 7 5825U with Radeon G processor, 40GB of RAM, and Linux Mint 22.1 x86_64 operating system, version Cinnamon 6.4.8, noting that this operating system is based on Ubuntu 24.04 LTS.
PyCharm 2025.2.0.1 (Community Edition) was used as the development environment, and the simulation environment was carried out in CoppeliaSim Edu, Version 4.10.0 (rev 0) 64 bit.

The objective was to replicate an automated box sorting system. The construction was carried out in a modular manner, incorporating the following main elements:

1) Conveyor belt: A conveyor belt was inserted as the central element of the sorting process. The movement of the belt was controlled by an associated dummy object, in which a script was implemented to manage the speed of the conveyor based on signals received from Python. This script allowed both the continuous operation of the belt and its temporary stoppage in the event of a match between the box and the selected class.

2) Object generation (Spawner): A second dummy was included with a script for generating boxes. This spawner creates cubic objects with random dimensions within a predefined range (0.08 to 0.32 meters). The boxes were placed on the conveyor belt to simulate a continuous flow of pieces to be sorted.

3) Proximity Sensors: Proximity sensors were installed at strategic points in the environment. The first was located on the conveyor belt, allowing the arrival of a box at the sorting point to be detected. This detection activates the Dummy spawner and sends the corresponding signals to the Python script, which is responsible for obtaining the corresponding measurements and performing the sorting. Additionally, two more sensors were installed: one in the exit area and another parallel to the sorting area. Both have the function of removing detected objects, preventing the accumulation of boxes in the environment and ensuring the continuity of the flow in the simulation.

4) Ejection Actuator: A prismatic joint linear actuator was incorporated into the side section of the conveyor belt. This actuator was programmed to move in a specific direction and push the boxes off the belt in the event of a match between the class predicted by the box's RNA and the one selected by the user. Its movement was controlled by signals emitted from Python, ensuring synchronization between the sorting and the mechanical ejection action. Figure 2 shows simulation environment in Coppelia.

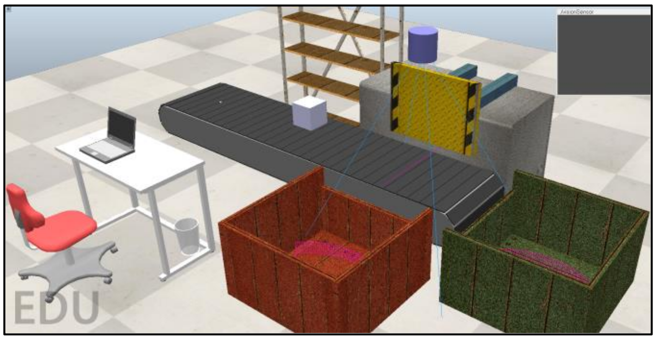

Figure 2. Simulation environment in Coppelia (Source: developed by the authors)

2.2. Training Dataset and Classification Criteria

To establish an objective reference for object classification, geometric criteria were defined based on the dimensions of the boxes generated in the simulation environment. Each box has three main measurements: length (sx), width (sy), and height (sz), which were extracted using the geometric parameter acquisition functions in CoppeliaSim. Based on these dimensions, the average of the sides was calculated using the following equation (1):

$$avg = \frac{sx + sy + sz}{3} \qquad (1)$$

This average value was considered the main parameter for determining the actual class of each box. Four classification ranges were established based on this value: small, medium, large, and extra-large boxes.

- Class 0 = avg < 0.14
- Class 1 = 0.14 ≤ avg < 0.20
- Class 2 = 0.20 ≤ avg < 0.26
- Class 3 = avg ≥ 0.26

Once the classification criterion by average side length had been defined, the data was recorded to form the training and validation set for the neural network. Each box generated in the simulation environment was analyzed using the proximity sensor responsible for capturing its dimensions (sx, sy, sz).

The measurements extracted were processed in a Python script that calculated the average value of the sides and automatically assigned the corresponding actual class according to the previously established ranges. This data, together with the class label, was stored in a CSV file. In total, a dataset consisting of 300 boxes was taken.

2.3. Artificial intelligence models

Three methods were used to compare artificial intelligence techniques in line with the evolution of algorithms, where the most basic algorithm used is the nearest neighbor algorithm, followed by artificial neural networks, and finally a deep learning algorithm based on convolutional neural networks. These algorithms are described below:

1) Training and Use of the Artificial Neural Network

The artificial neural network was trained in Python 3.12, using the TensorFlow/Keras libraries to build the model, NumPy to handle matrix structures and vectorized operations, and scikit-learn to normalize data using a standard scaler (StandardScaler).

The neural network implemented corresponded to a feedforward model composed of fully connected dense layers. The architecture consisted of an input layer with three neurons (one for each dimension of the box), three hidden layers (128, 64, and 32 neurons) with ReLU (Rectified Linear Unit) activation functions, and an output layer with four neurons activated by Softmax, corresponding to the four established classes.

The model was compiled using the Adam optimizer with a default learning rate of 0.001. The maximum number of epochs was set to 500, with a batch size of 16 samples per iteration.

Training was performed on the training set, reserving 20% of it for internal validation. A stopping technique was implemented, configured to monitor validation accuracy and stop training if no improvements were observed for 20 consecutive epochs. In addition, automatic restoration of the weights corresponding to the best validated performance was activated. In the network training, an accuracy of 95.24% was achieved in the test set.

2) KNN Algorithm Training

Similar to ANN training, the K-Nearest Neighbors (KNN) classification model was implemented using the same dataset as the ANN, and also used the scikit-learn library for data normalization.

The classifier was configured with k = 5 neighbors, seeking a balance between robustness against noise and pattern detection capability. The metric used to determine proximity was Euclidean distance, which calculates the square root of the sum of the squared differences between the dimensions of the boxes, which is intuitive in three-dimensional geometric spaces.

In terms of voting, a uniform weighting scheme was adopted, meaning that all neighbors have the same influence in determining the class of the analyzed sample. The classification process is performed by comparing the input sample with all the vectors stored in the database, selecting the five closest ones, and assigning the majority class. Although this approach does not generate a parametric model, its structural simplicity allows for efficient implementation and competitive performance.

3) CNN Algorithm Training

For the implementation of the convolutional neural network (CNN), a sample of 400 images was taken for each class using a visionSensor provided by CoppeliaSim, which by default already has an image size of 224 x 224. The images were captured using Python code, which took the images together with the measurements of the boxes and classified them into folders to avoid errors in the creation of the training dataset.

The ImageDataGenerator module was used for image manipulation and data preprocessing, which allowed the pixel values to be scaled to the range [0,1] and the dataset to be divided into training and validation with a ratio of 80% and 20%, respectively.

As this was a controlled and ideal environment, the CNN architecture was designed in a simple way with the aim of obtaining a functional model without requiring long training times. The first layer corresponded to a two-dimensional convolution with 16 3 x 3 kernels, using the ReLU (Rectified Linear Unit) activation function. Subsequently, a MaxPooling layer with a size of 2 x 2 was included for dimensionality reduction. The second convolutional layer consisted of 32 3 x 3 kernels, also with ReLU activation, followed again by a 2 x 2 MaxPooling layer.

After the convolutional stages, the feature maps were flattened to be processed in dense layers. A fully connected layer of 64 neurons with ReLU activation was included, along with a 30% Dropout layer, used to mitigate overfitting during training. Finally, an output layer with 4 neurons and Softmax activation was added, which allowed for multi-class classification corresponding to the four defined box categories.

The model was compiled using the Adam optimizer with a learning rate of 0.001. The selected loss function was categorical cross-entropy, appropriate for multi-class classification problems. Model training was carried out for 20 epochs, using a batch size of 32 samples.

Finally, the model's performance was evaluated by calculating the confusion matrix and a classification report generated with the scikit-learn and seaborn libraries. These results allowed us to analyze the precision, recall, and F1-score metrics for each class, as well as the overall accuracy of the model

4) Automatic speech recognition

The wav2vec algorithm conducts audio recognition and converts it to text. This self-supervised approach creates speech representations from raw audio inputs without traditional preprocessing steps like cepstral coefficients or spectrogram generation.

Wav2Vec employs a multilayer CNN to transform input waveforms into low-dimensional latent representations. Subsequently, a Transformer network captures contextual dependencies between audio segments.

Table 1 lists audio capture parameters for recognizing object labels, enabling voice-driven interaction within the Coppelia simulation environment.

Table 1. Voice recording (Source: developed by the authors)

Parameter	Setting Value
Recording time	3 seg
Sample frequency	22050
Number of bits	16
Number of channels	1 (monophonic)

3. Results

To validate the artificial intelligence algorithms used with the simulation environment, integration was carried out using CoppeliaSim's remote API (ZMQ Remote API), which enabled bidirectional communication between the simulation processes and a user interface under Python code. Through this link, signals from the proximity sensors, including the identifiers of the detected boxes, were received in real time. Subsequently, the geometric parameters of each object (dimensions sx, sy, sz) were consulted to be processed by the previously trained model.

The Python code was structured into independent modules: one dedicated to classification using the selected model, another responsible for the graphical interface developed in Tkinter, and a third block for voice recognition, implemented with the SpeechRecognition library and regular expressions to interpret user commands. In this way, the operator can select the desired box class using only voice commands. Once a box has been passed, we will obtain the data from it as well as the classification assigned by the model and whether or not it matches the user's selection, as shown in Figure 3.

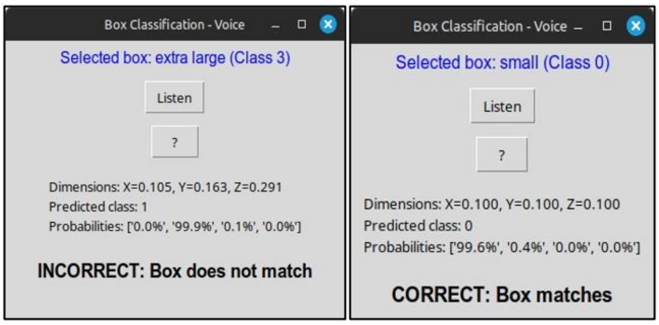

Figure 3. Matlab application commanding robotic manipulator in Coppelia simulator (Source: developed by the authors)

During the integration of voice recognition, certain limitations were identified in the accuracy of the Speech Recognition library. In several cases, the system did not correctly interpret the instruction on the first utterance, forcing the command to be repeated. This behavior is mainly related to the model's sensitivity to background noise, pronunciation, and the quality of the microphone used. Failure to properly detect a class result in a validation error, as shown in figure 4.

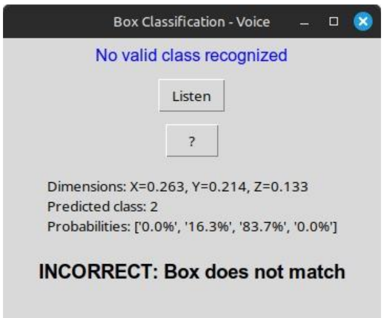

Figure 4. Help button (Source: developed by the authors)

Figure 5 illustrates a case of correct selection, where the right side shows the selection actuator in operation, removing the desired box from the conveyor belt. Figure 6 shows an enlargement to validate the online selection and the dimensions of the box.

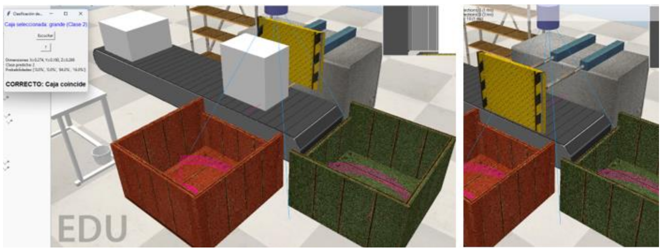

Figure 5. Help button (Source: developed by the authors)

Figure 6. Sorting progress in Coppelia virtual environment (Source: developed by the authors)

In some cases, synchronization problems were detected between CoppeliaSim events and the instructions sent from Python. For example, unexpected actuator activations occurred without a box reaching the sorting area, indicating that at certain times residual signals were generated or not reset correctly in the client-server exchange.

4. Discussion

The three-dimensional box classification experiment was conducted with a total of 500 test samples, evaluating the performance of the nearest neighbor classifier (KNN), the artificial neural network (ANN) and the CNN. These three artificial intelligence algorithms correspond to different eras of technological advancement and involve differentiable learning architectures, where it is evident that each one manages to solve the classification objective. Table 2 summarizes the main indicators obtained for three models, including accuracy, average classification time, and the dispersion of the measured times. Although the level of accuracy of each model can be considered sufficient, there is no evidence of a correlation between the complexity of the architecture and accuracy, with CNN being the most complex, followed by RNA and finally KNN. Another important aspect derives from the computational cost, while KNN has a training time of only 0.5 milliseconds, for RNA the training process lasted 13.80 seconds and CNN is in the order of tens of minutes. However, the most relevant time is that of inference by class, where KNN is much faster and has a lower standard deviation, which makes it stand out as the most favorable model to use.

Table 2. Algorithm comparison (Source: developed by the authors)

Modelo	Exactitud (%)	Tiempo promedio (ms)	Desviación estándar (ms)
RNA	94.20	43.98	9.94
KNN	89.40	1.29	0.20
CNN	85.00	44.59	7.83

This allows us to identify that traditional algorithms, thanks to their low computational cost, are preferable for classification cases that involve the need to reduce inference time in a real-world expansion. It should be clarified that the level of accuracy of the convolutional network can be improved by increasing the network depth, which entails a greater digital weight and therefore inference time. This does not improve the results obtained nor does it modify the decision argument in choosing the best algorithm to implement in this application. Robust classification architectures such as convolutional networks and transformer networks currently offer strong benefits in the development of automation and robotics systems, but the criteria defined here allow their selection as explained in [21], consistent with what has been obtained here, where the most recent architecture in the state of the art is not optimal for the application in question.

5. Conclusion

Three classification algorithms were compared that, in controlled environments, allow high levels of accuracy to be obtained, since the best overall corresponding to the nearest neighbor algorithm. It was concluded that this algorithm is the most functional for the proposed validation case, given its low computational cost and the fact that it does not need to be very accurate due to the low level of confusion between classes.

State-of-the-art pattern recognition algorithms such as convolutional networks achieve very high levels of accuracy, and their selection as the application algorithm must be accompanied by other criteria such as computational cost and classification time, which can balance the functional requirements against less accurate algorithms that are also used in classification systems.

In practical applications outside the simulation environment, the variability of lighting conditions can affect the levels of accuracy achieved, but in this case, no changes in the selection of the classification architecture are foreseen.

Future work includes using depth cameras to extract the dimensions of the boxes in order to establish variations in the classification architectures.

Declarations

Author Contributions

Conceptualization, formal analysis and writing—review and editing Jiménez-Moreno R. Bejarano Jose and Martinez Javier; methodology, validation, investigation supervision, project administration, funding acquisition, and data curation Jiménez-Moreno R.; Software, writing—original draft preparation and visualization Bejarano Jose and Jiménez-Moreno R. All authors have read and agreed to the published version of the manuscript.

Funding

Product derived from the research project titled " Design of a human-robot interaction model using deep learning algorithms” INV-ING-3971 financed by the vice-rector for research of the Universidad Militar Nueva Granada, year 2024.

Acknowledgements

The authors thank the Universidad Militar Nueva Granada and Universidad de los llanos for the time and resources available for the development of this article where they are associate professors.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committe of the Universidad Militar Nueva Granada (project INV-ING-3971, date of approval: 15/05/2024).

Conflicts of Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

References

[1] CAO Y., XIANG W., WEI J., CAO S., TIAN X., ZHONG J., FANG X., LUO B., LYU H. and LI XIANGKUI. CrossFusionSleepNet: A multimodal deep learning model for automatic sleep stage classification. *Biomedical Signal Processing and Control*, 2026,112(PartA): 108538. https://doi.org/10.1016/j.bspc.2025.108538.

[2] GUO B.E., SHEN Y., ZHOU Z.F., LIU X., WEI Y.X. and YANG L. Advanced deep learning for automatic classification of fired bullets from standard-issue firearms. *Science & Justice*, 2025,65(6): 101335. https://doi.org/10.1016/j.scijus.2025.101335.

[3] NASIR S., SHEIKH S.A. and MALIK F.M. Automatic modulation classification using convolutional neural network and support vector machine. *Digital Signal Processing*, 2025,(64):105249. https://doi.org/10.1016/j.dsp.2025.105249

[4] ARNELA M., VIDAÑA-VILA E., FANTINELLI A., MOÑUX-BERNAL A., VAQUERIZO-SERRANO J. and SOCORÓ J.C. Generation of ultrasonic and audible sound waves for the automatic classification of packaging waste in reverse vending machines. *Waste Management*, 2025,204: 114934. https://doi.org/10.1016/j.wasman.2025.114934.

[5] HOSSAIN MD.M. and ROY K. The development of classification-based machine-learning models for the toxicity assessment of chemicals associated with plastic packaging. *Journal of Hazardous Materials,* 2025,(484): 136702, https://doi.org/10.1016/j.jhazmat.2024.136702.

[6] YILDIZ S.N., OKAY F.Y., ISLAMOV A. and ÖZDEMIR S. Improved Chain-based Multi-Output Classification for Packaging Planning. *Procedia Computer Science, 2024,*231: 697-702. https://doi.org/10.1016/j.procs.2023.12.159.

[7] TEYE E., AMUAH C.L.Y., LAMPTEY F.P., OBENG F. and NYORKEH R. Artificial intelligence for honey integrity in Ghana: A feasibility study on the use of smartphone images coupled with multivariate algorithms. *Smart Agricultural Technology,*2024,(8):100453. https://doi.org/10.1016/j.atech.2024.100453.

[8] DODAMPEGAMA S., HOU L., ASADI E., ZHANG G. and SETUNGE S. Revolutionizing construction and demolition waste sorting: Insights from artificial intelligence and robotic applications. *Resources, Conservation and Recycling,* 2024,202:107375. https://doi.org/10.1016/j.resconrec.2023.107375.

[9] MCENTEGGART F., RAMASUBRAMANIAN A.K, ZEINALI M. and PAPAKOSTAS N. Optimising robot motion planning through the integration of diverse process simulation tools. *Procedia CIRP,* 2025,136: 918-923. https://doi.org/10.1016/j.procir.2025.08.156.

[10] AUGENSTEIN T.E., NAGALLA D., MOHACEY A., CUBILLOS L.H., LEE M.H., RANGANATHAN R. and KRISHNAN C. A novel virtual robotic platform for controlling six degrees of freedom assistive devices with body-machine interfaces. *Computers in Biology and Medicine,*2024,1781:108778. https://doi.org/10.1016/j.compbiomed.2024.108778.

[11] FARIAS G., FABREGAS E., PERALTA E., TORRES E. and DORMIDO S. A Khepera IV library for robotic control education using V-REP. *IFAC-PapersOnLine, 2017,*50(1): 9150-9155. https://doi.org/10.1016/j.ifacol.2017.08.1721.

[12] MARINI N., MARCHESIN S., FERRIS L.B., PÜTTMANN S., WODZINSKI M., FRATTI R., PODAREANU D., CAPUTO A., BOYTCHEVA S., VATRANO S., FRAGGETTA F., NAGTEGAAL I., SILVELLO G., ATZORI M. and MÜLLER H. Automatic labels are as effective as manual labels in digital pathology images classification with deep learning. *Journal of Pathology Informatics*,2025,18:100462. https://doi.org/10.1016/j.jpi.2025.100462.

[13] KA H. Voice-Controlled Vision-based Semi-Autonomous Assistive Robotic Manipulation Assistance. *Archives of Physical Medicine and Rehabilitation*, 2015,96(10):e10-e11. https://doi.org/10.1016/j.apmr.2015.08.028.

[14] CHAKRADEO V.K., MALHOTRA K., LEE M.R. and NATHAN C.A.O. Navigational Surgery with Voice: Controlled Robotic Assist for Endoscopic Approach to the Pituitary. *Otolaryngology - Head and Neck Surgery*, 2005,133(2),Supplement:P153. https://doi.org/10.1016/j.otohns.2005.05.344.

[15] BAKOURI M. Development of Voice Control Algorithm for Robotic Wheelchair Using MIN and LSTM Models. *Computers, Materials and Continua*, 2022,73(2): 2441-2456. https://doi.org/10.32604/cmc.2022.025106.

[16] MEGHANA M., USHA KUMARI CH., STHUTHI PRIYA J., MRINAL P., ABHINAV VENKAT SAI K., PRASHANTH REDDY S., VIKRANTH K., SANTOSH KUMAR T. and PANIGRAHY A.K. Hand gesture recognition and voice controlled robot. *Materials Today: Proceedings,2020,*33(Part7):4121-4123. https://doi.org/10.1016/j.matpr.2020.06.553.

[17] PARTHA SARADI V. and KAILASAPATHI P. Voice-based motion control of a robotic vehicle through visible light communication. *Computers & Electrical Engineering*,2019,76:154-167. https://doi.org/10.1016/j.compeleceng.2019.03.011.

[18] KADRI I., SELOUANI S.A., GHRIBI M., GHALI R. and MEKHOUKH S. LLM-driven agent for speech-enabled control of industrial robots: A case study in snow-crab quality inspection. *Results in Engineering*, 2025,27: 106660. https://doi.org/10.1016/j.rineng.2025.106660.

[19] GUO Y., XU W., PRADHAN S., BRAVO C. and BEN-TZVI P. Personalized voice activated grasping system for a robotic exoskeleton glove. *Mechatronics*, 2022,83: 102745. https://doi.org/10.1016/j.mechatronics.2022.102745.

[20] BARATTA A., CIMINO A., LONGO F. and NICOLETTI L. Digital twin for human-robot collaboration enhancement in manufacturing systems: Literature review and direction for future developments. *Computers & Industrial Engineering,2024,*187:109764. https://doi.org/10.1016/j.cie.2023.109764.

[21] JIMÉNEZ-MORENO R. and ESPITIA-CUBILLOS A. A. Comparative Performance Analysis of Transformer and Convolutional Networks for Machine Vision-Oriented Mobile Robots. *Journal of Hunan University Natural Sciences*,2025,52(2):113-121. https://doi.org/10.55463/issn.1674-2974.52.2.11

参考文:

[1] CAO Y., XIANG W., WEI J., CAO S., TIAN X., ZHONG J., FANG X., LUO B., LYU H. and LI XIANGKUI. CrossFusionSleepNet: 一种用于自动睡眠阶段分类的多模态深度学习模型，*《生物医学信号处理与控制》*. *《生物医学信号处理与控制》*, 2026,112(部分 A): 108538. https://doi.org/10.1016/j.bspc.2025.108538.

[2] GUO B.E., SHEN Y., ZHOU Z.F., LIU X., WEI Y.X. and YANG L. 基于深度学习的先进技术，实现对制式枪械发射弹头的自动分类。. *科学与司法*, 2025,65(6): 101335. https://doi.org/10.1016/j.scijus.2025.101335.

[3] NASIR S., SHEIKH S.A. and MALIK F.M. 基于卷积神经网络与支持向量机的自动调制分类. *数字信号处理*, 2025,(64):105249. https://doi.org/10.1016/j.dsp.2025.105249

[4] ARNELA M., VIDAÑA-VILA E., FANTINELLI A., MOÑUX-BERNAL A., VAQUERIZO-SERRANO J. and SOCORÓ J.C. 利用超声波与可听声波实现自动售货机包装废弃物的自动分类. *废物管理*, 2025,204: 114934. https://doi.org/10.1016/j.wasman.2025.114934.

[5] HOSSAIN MD.M. and ROY K. 基于分类的机器学习模型在塑料包装相关化学品毒性评估中的开发. *危险物质杂志*, 2025,(484):136702, https://doi.org/10.1016/j.jhazmat.2024.136702.

[6] YILDIZ S.N., OKAY F.Y., ISLAMOV A. and ÖZDEMIR S. 改进的基于链的多输出分类法在包装规划中的应用. 计算机科学进展 *,2024,*231:697-702. https://doi.org/10.1016/j.procs.2023.12.159.

[7] TEYE E., AMUAH C.L.Y., LAMPTEY F.P., OBENG F. and NYORKEH R. 加纳蜂蜜完整性的人工智能应用：基于智能手机图像与多元算法结合的可行性研究. *智能农业技术*,2024,(8):100453. https://doi.org/10.1016/j.atech.2024.100453.

[8] DODAMPEGAMA S., HOU L., ASADI E., ZHANG G. and SETUNGE S. 革新建筑与拆除废料分拣：人工智能与机器人应用的洞见. *资源、保护与回收,* 2024,202:107375. https://doi.org/10.1016/j.resconrec.2023.107375.

[9] MCENTEGGART F., RAMASUBRAMANIAN A.K, ZEINALI M. and PAPAKOSTAS N. 通过整合多样化工艺仿真工具优化机器人运动规划. *CIRP 会议论文集,* 2025,136: 918-923. https://doi.org/10.1016/j.procir.2025.08.156.

[10] AUGENSTEIN T.E., NAGALLA D., MOHACEY A., CUBILLOS L.H., LEE M.H., RANGANATHAN R. and KRISHNAN C. 一种新型虚拟机器人平台，用于通过人机接口控制六自由度辅助装置. *生物学与医学中的计算机应用*,2024,1781:108778. https://doi.org/10.1016/j.compbiomed.2024.108778.

[11] FARIAS G., FABREGAS E., PERALTA E., TORRES E. and DORMIDO S. 基于 V-REP 的机器人控制教育用 Khepera IV 库. *国际会计师联合会在线论文库, 2017,*50(1): 9150-9155. https://doi.org/10.1016/j.ifacol.2017.08.1721.

[12] MARINI N., MARCHESIN S., FERRIS L.B., PÜTTMANN S., WODZINSKI M., FRATTI R., PODAREANU D., CAPUTO A., BOYTCHEVA S., VATRANO S., FRAGGETTA F., NAGTEGAAL I., SILVELLO G., ATZORI M. and MÜLLER H. 在基于深度学习的数字病理图像分类中，自动标签与人工标签同样有效. *病理学信息学杂志*,2025,18:100462. https://doi.org/10.1016/j.jpi.2025.100462.

[13] KA H. 语音控制视觉辅助半自主机器人操作辅助系统. *物理医学与康复档案*, 2015,96(10):e10-e11. https://doi.org/10.1016/j.apmr.2015.08.028.

[14] CHAKRADEO V.K., MALHOTRA K., LEE M.R. and NATHAN C.A.O. 语音导航手术：垂体内镜手术的可控机器人辅助技术. *耳鼻喉科 - 头颈外科*, 2005,133(2),Supplement:P153. https://doi.org/10.1016/j.otohns.2005.05.344.

[15] BAKOURI M. 基于 MIN 和 LSTM 模型的机器人轮椅语音控制算法开发. *计算机、材料与连续体*, 2022,73(2): 2441-2456. https://doi.org/10.32604/cmc.2022.025106.

[16] MEGHANA M., USHA KUMARI CH., STHUTHI PRIYA J., MRINAL P., ABHINAV VENKAT SAI K., PRASHANTH REDDY S., VIKRANTH K., SANTOSH KUMAR T. and PANIGRAHY A.K. 手势识别与语音控制机器人. *今日材料：会议论文集,2020,*33(Part 7):4121-4123. https://doi.org/10.1016/j.matpr.2020.06.553.

[17] PARTHA SARADI V. and KAILASAPATHI P. 基于可见光通信的机器人车辆语音运动控制. *计算机与电气工程*,2019,76:154-167. https://doi.org/10.1016/j.compeleceng.2019.03.011.

[18] KADRI I., SELOUANI S.A., GHRIBI M., GHALI R. and MEKHOUKH S. 基于 LLM 的语音控制工业机器人智能体：以雪蟹品质检测为例的研究. *工程学成果*, 2025,27: 106660. https://doi.org/10.1016/j.rineng.2025.106660.

[19] GUO Y., XU W., PRADHAN S., BRAVO C. and BEN-TZVI P. 用于机器人外骨骼手套的个性化语音激活抓取系统. *机电一体化*, 2022,83: 102745. https://doi.org/10.1016/j.mechatronics.2022.102745.

[20] BARATTA A., CIMINO A., LONGO F. and NICOLETTI L. 数字孪生在制造系统中提升人机协作的应用：文献综述与未来发展方向. *计算机与工业工程, 2024,*187:109764. https://doi.org/10.1016/j.cie.2023.109764.

[21] JIMÉNEZ-MORENO R. and ESPITIA-CUBILLOS A. A. 面向机器视觉的移动机器人中卷积神经网络与变压器网络的性能比较分析。《湖南大学学报（自然科学版）》,2025,52(2):113-121. https://doi.org/10.55463/issn.1674-2974.52.2.11

Manuscript Information
Word count: 5,537 words (excluding references).

Peer-Review Record
Fast-track status: Not fast-tracked.
First-round reviews received: 3 reports.
Revision cycles completed: 3 rounds.
Final version submitted: December 12, 2025