

## TRANSFORMING ELECTRONICS EDUCATION WITH AI: A CHALLENGE-BASED LEARNING APPROACH

**José Manuel Bravo-Pacheco, Sergio Esteban-Romero, Iván Martín-Fernández,  
Manuel Gil-Martín, & Fernando Fernández-Martínez**

*Grupo de Tecnología del Habla y Aprendizaje Automático (THAU Group),  
Grupo de Innovación Docente en Ingeniería y Sistemas Electrónicos (GRIDS),  
Departamento de Ingeniería Electrónica, Information Processing and Telecommunications Center,  
E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid (Spain)*

### Abstract

The integration of Artificial Intelligence (AI) into electronics and computer science education offers new opportunities to enhance student engagement and learning outcomes. This paper presents a comprehensive approach to embedding AI-driven tools in laboratory-based courses, making complex concepts more tangible through hands-on experimentation. Developed by the GRIDS educational innovation group, this initiative leverages challenge-based learning and emerging technologies to create an interactive and adaptive learning environment. Three key AI applications serve as the foundation of this approach: Facial Emotion Recognition (FER), Human Activity Recognition (HAR), and Embedded Speech Technologies. These domains provide students with an opportunity to explore AI beyond theoretical concepts, engaging directly with machine learning models and real-world datasets. The Facial Emotion Recognition module enables students to experiment with deep learning for affective computing, using Convolutional Neural Networks (CNNs) and facial landmark detection to analyze emotions from video feeds. Through platforms like Google Colab and Raspberry Pi, students can train models with existing datasets (e.g., RAVDESS) or collect their own, gaining practical experience in AI model deployment. The Human Activity Recognition component introduces AI-powered gestures and movement analysis. By utilizing both inertial sensor data and image-based hand pose recognition, students implement deep learning architectures, such as Long Short-Term Memory (LSTM) networks and CNNs. This hands-on approach enhances their understanding of real-time AI applications in fields like robotics, healthcare, and human-computer interaction. The Embedded Speech Technologies module focuses on voice-driven interfaces, employing the Picovoice platform for speech recognition and intent detection. Using tools such as Porcupine Wake Word and Rhino Speech-to-Intent, students develop low-cost, voice-controlled systems that showcase the potential of AI for embedded applications. Across these projects, students engage in design, implementation, and optimization of AI models within resource-constrained environments, reinforcing problem-solving and critical thinking skills. The combination of service learning, design thinking, and adaptive learning methodologies, supported by platforms like Moodle, YouTube, and GitHub, ensures a scalable and accessible experience. Qualitative feedback highlights increased student motivation, deeper conceptual understanding, and stronger engagement with AI-based tools. The findings suggest that incorporating AI into laboratory settings not only enhances learning outcomes but also better prepares students for the evolving demands of AI-driven industries.

**Keywords:** *Artificial Intelligence, electronics education, hands-on learning, challenge-based learning, AI applications.*

---

### 1. Introduction

The rapid evolution of Artificial Intelligence (AI) is reshaping industries and redefining the skillsets demanded of future professionals. As this happens, new students must be prepared with not only theoretical concepts but also solid practical experience in AI technologies. Traditional methods, based on theoretical lectures and textbooks, often struggle to engage the students. To address this, there is a growing interest in project-based approaches that encourage active learning and real-world problem solving. In this context, the integration of AI-based tools into laboratory sessions is emerging as a powerful strategy to increase the student engagement and stimulate their curiosity, which is important for their academic achievement and enjoyment (Steinmayr et al., 2019).

To accomplish that, the implementation of three AI-based applications is proposed: Facial Emotion Recognition (FER), Human Activity Recognition (HAR), and Embedded Speech Technologies. The first one involves analyzing facial features such as eyes, mouth, and facial muscles to determine emotions like happiness, sadness, anger, etc. The second one focuses on determining the activity performed by an individual, like running, through inertial sensors, and recognizing hand poses, like letters of Sign Language, from images. The last one shows the implementation of speech recognition libraries on low-cost platforms such as STM32.

Along this paper, the implementation of the three systems in laboratories sessions is explained, showing their possible applications in education. In addition, the response from a set of students that used the proposed resources is described.

## 2. Description of the proposed resources

This section deeps into the implementation of the proposed systems, including different modules of an AI project lifecycle, such as data acquisition, image or signal processing, training and analysis of the results, and deployment.

### 2.1. Face emotion recognition (FER)

As mentioned previously, this system is focused on identifying and interpreting human emotions based on facial expressions. To achieve this, it analyses a set of landmarks extracted from relevant face points. These systems have wide application fields such as education (tracking the attention of the students), public safety (lie detectors), or employment (helping recruiters).

The final system can detect the facial emotion that a user is expressing in real time, showing the prediction on screen. It has different stages: data acquisition, image processing to acquire 478 landmarks, artificial intelligence-based model training, and system deployment for real-time testing.

Concerning data acquisition, the project allows the students to work both with existing datasets found on the internet or custom ones recorded by them. In this sense, they can play with the changes between using large datasets, which are usually less accurate to a specific desired functionality, and small datasets, which are usually more focused on the task, or even combining them, leading to more robust systems.

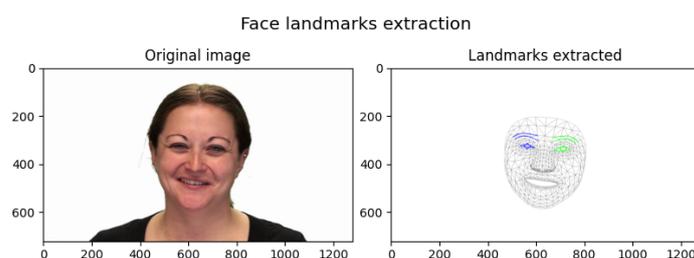
The RAVDESS dataset (Livingstone & Russo, 2018) is proposed as a baseline. The dataset includes recordings of 24 actors expressing emotions in both audio and video formats, with video frames labeled by emotion. Specifically, seven emotions are recorded: calm, happy, sad, angry, fearful, surprise, and disgust.

On the other hand, the students are allowed to record their own dataset for training the systems. In this sense, they can decide the number of classes to record, their class names and the number of images to save at each class. This personalization of the process allows them to test different combinations and to see their effect in the model. This recording is done using a Raspberry Pi equipped with a camera module.

Regarding the image processing, the system extracts a set of face landmarks for doing the classification. This reduces the amount of irrelevant information provided to the model, resulting in a smaller and easier approach to train models. The landmarks are extracted from the images using the MediaPipe library (Google, n.d.). Specifically, it takes an image in RGB format and returns an array with 478 points, normalized between 0 and 1 with respect to the size of the input image. The landmark extraction process can be seen in Figure 1.

While the landmarks are normalized with respect to the image size, they still include biases that can interfere in the classification such as the position of the face in the image or the distance from the camera. By normalizing this data, students can observe the impact of these techniques in the result of the system, learning interactively its importance. As a baseline, a normalization for these two cases is also proposed.

Figure 1. Landmark extraction process. The image is extracted from the RAVDESS dataset.



Once the data is loaded and preprocessed, the training starts. In this phase, the students iterate the data through a model and, by reducing the loss between the predictions dropped by the model and the real classes, it finds characteristic patterns in the landmarks. Since the process is very demanding on computer resources, a Google Colab (Google, n.d.) notebook is used. Two different models are proposed as initial options: one Convolutional Neural Network (CNN) (LeCun et al., 1998) and one Visual Transformer (ViT) (Dosovitskiy et al., 2020). Once the training is finished, the models obtained can be compared and downloaded. The comparison is made with the accuracy metric and 95% confidence interval. Students can experience comparing their different approaches and selecting the best solution found for their application.

Finally, the students can easily implement their recently trained models to make a real-time demo of the task. A Raspberry Pi 5 with 4GB of RAM is used for the deployment of the demo. Once the model is loaded, the deployment script can be run. This script enables the camera of the device and shows the recording on the screen. With every frame, the landmarks of the face are extracted, and a square is drawn around the detected user, informing about the detected emotion. A demonstration of the system is available online at: <https://youtu.be/HEggPs9-L8M>

## 2.2. Human activity recognition (HAR)

Commonly recognized daily activities in the literature include walking, running, sitting, and standing. However, it is also possible to recognize activities directly related to sport such as performing squats or steps. In parallel, HAR also covers the detection of specific hand poses, which can enhance the development of human-computer interaction systems, thereby improving user experiences across diverse domains. Hence, two systems are proposed for HAR: Human Activity Recognition from acceleration signals and Hand Pose Recognition from images. As in the previous system, these applications include four stages: data acquisition, signal processing, model training and demo deployment. Moreover, it also relies on combining both a Raspberry PI and Google Colab for a balance in low-resource devices and high-power computation.

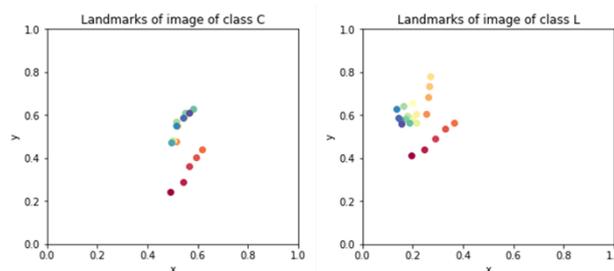
A Raspberry PI platform is used for recording data for both systems. This platform is equipped with a camera and a Sense HAT add-on board (with an LSM9DS1 IMU that contains an accelerometer).

For the recording of the physical activity signals, the subjects used a hip belt to attach the Raspberry Pi at the lumbar during the data collection. Though the system is flexible to capture acceleration data from the activities the students are interested in, the default configuration invited the students to capture data from four activities related to sport: running, step, squat and circles. The recording of the hand poses is done using the camera. The subjects located the Raspberry Pi in front of them while performing different hand poses. For this system, the students were invited to perform hand poses related to some letters from the Sign Language Alphabet.

The acceleration signals captured from the sensor were segmented into windows. The default configuration asked the students to capture enough time to obtain 50 examples of 4-second windows with a 1-second overlap per class, which were directly used as input to the model.

Regarding the hand pose system, a set of 21 hand landmarks are extracted from the images using the MediaPipe library. Figure 2 shows the points that were obtained after processing the images for two hand poses (“C” and “L”).

Figure 2. Hand poses for the Sign Language Alphabet.



For the physical activity system, the processed acceleration signals were used as an input of a simple Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) network to model the motion pattern in the time domain. This model is designed to process sequential data consisting of three acceleration axes. In case of the hand pose system, the x and y coordinates of the 21 landmarks were used as an input of a simple CNN. The models' performance was evaluated using the accuracy metric with 95% confidence intervals. As the training process of these systems is resource-demanding, the stage is also performed through Google Colab platform.

For both applications, the trained model is loaded in the Raspberry Pi to offer the possibility for real-time testing. This way, the students can observe how the system performs in real life examples. The students can observe the predicted class in a terminal connected to the Raspberry Pi. However, the 8x8 RGB LED matrix can also be used to display a different letter and color depending on the predicted class. A demonstration of the system is available online at: <https://youtu.be/yy-fC5ERWAs>

For the second system, students could witness how the system accurately classifies the hand poses they had chosen beforehand. As the students executed the hand poses, they could view the predicted class on a screen simultaneously. A demonstration of the system is available online at: <https://youtu.be/URrn2O6BPt4>

### 2.3. Embedded speech technologies

Enabling speech recognition on low-cost platforms like STM32 microprocessors opens up a wide range of practical applications in various domains and can revolutionize user interactions, making them more intuitive, hands-free, and efficient. However, it also presents challenges due to limited computational resources and memory constraints. To address this, we explore the technologies provided by Picovoice (Picovoice Inc., n.d.).

Picovoice is a platform designed with developers in mind, empowering you to create voice-related applications according to your preferences. Diverging from Amazon or Google, Picovoice's AI operates solely on the device, offering enhanced accuracy. Specifically, the way Picovoice works is by listening until the configured wake-word is said. When this happens, the library processes the following phrase to decode the intention of the message based on a set of predefined actions. In this case, the full interface consists of three different steps: audio recording, sample processing and intention processing. The deployment of the system is done in a STM32-F411E-DISCO platform.

The interface uses the Integrated Interchip Sound (I2S) peripheral to connect the board with a MEMS microphone. This microphone records samples in PDM format. Since Picovoice requires PCM samples, the audio is converted from PDM to PCM.

We also want to ensure that the software is processing the samples at the same time as we keep recording. To make this possible the system implements a double buffer, one for reading and one for writing, and, by alternating them, both tasks can be performed at the same time.

For processing the obtained samples, Picovoice library makes two principal actions. First, it keeps checking if the wake-word has been said. If it is the case, the library automatically processes the following phrase and returns the intention of the voice command. The intention's parameters are returned as a set of previously configured options which is highly customizable from its web.

Finally, when the intention has been decoded, the system must process it as a normal programming flow to implement the desired functionalities. As a demo, a system was built to control a set of LED lights, however, the possibilities are endless. A demonstration of the system is available online at: [https://youtu.be/4zYQIVsk\\_SM](https://youtu.be/4zYQIVsk_SM)

## 3. Teaching methodology and results

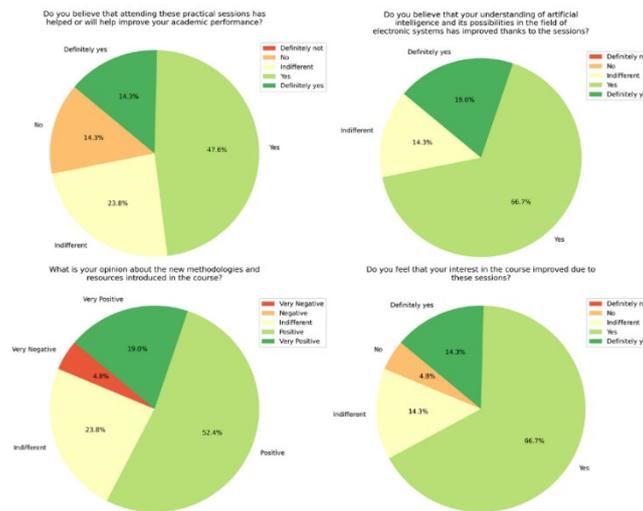
To assess the students' acceptance, motivation, and involvement, we followed a methodology based on qualitative interviews with a select group of students that experienced the integration of the systems, and observations during the laboratory sessions.

For example, the students' response of the new HAR resources showed great acceptance. Students shared that they were afraid at the beginning when the resources were presented. However, the novelty of the technology and its application over a real system sparked their curiosity to achieve their own systems based on artificial intelligence. Furthermore, the students expressed their delight about having the opportunity to put theoretical concepts to practical use.

Professors noted that students' body language and expressions evolved positively while using the resources when they observed that the systems correctly classified their own activities or hand poses. Moreover, professors observed that students actively engaged in problem-solving when using different deep learning approaches to solve the recognition challenges. In addition, students engaged in collaborative efforts with their peers throughout the practical lessons.

Figure 3 shows the feedback given by students after using the HAR project in laboratories. As can be seen, the students' interest improved using these projects, also helping them to achieve a higher understanding of AI systems. Moreover, survey results suggest that these practical sessions and resources helped to improve their academic performance.

Figure 3. Results of polls taken to students about the proposed methodology.



#### 4. Conclusions

Our initiative of incorporating practical sessions based on AI systems as tools to enhance students' academic performance in electronics education appears advantageous. These experiences suggest that it is possible to boost the students' problem-solving skills in an enjoyable environment with a interactive and engaging systems related to Facial Emotion Recognition, Human Activity Recognition or Embedded Speech Technologies.

While we acknowledge the limitations of a smaller sample size, the qualitative findings provide a rich narrative that complements and enriches our understanding of the educational impact of these technologies. Further research, with larger samples or diverse student groups, can continue to explore the qualitative dimensions of this innovative approach to learning.

#### Acknowledgements

This work was funded by Project ASTOUND (101071191 — HORIZON-EIC-2021-PATHFINDERCHALLENGES-01) of the European Commission and by the Spanish Ministry of Science and Innovation through the projects GOMINOLA (PID2020-118112RB-C22), TRUSTBOOST (PID2023-150584OB-C21) and BeWord (PID2021-126061OB-C43), funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR". Sergio Esteban-Romero research was supported by the Spanish Ministry of Education (FPI grant PRE2022-105516). Iván Martín-Fernández research was supported by the Universidad Politécnica de Madrid (Programa Propio I+D+i).

#### References

- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. Retrieved April 24, 2025, from <https://arxiv.org/abs/2010.11929>
- Google. (n.d.). *Google Colaboratory*. Retrieved April 24, 2025, from <https://colab.research.google.com/>
- Google. (n.d.). *MediaPipe*. Retrieved April 24, 2025, from <https://mediapipe.dev/>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>
- Livingstone, S. R., & Russo, F. A. (2018). *The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)*. Retrieved April 24, 2025, from <https://zenodo.org/record/1188976>
- Picovoice Inc. (n.d.). *Picovoice Platform Documentation*. Retrieved April 24, 2025, from <https://picovoice.ai/>
- Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic achievement – Replicating and extending previous findings. *Frontiers in Psychology*, 10, 1730. <https://doi.org/10.3389/fpsyg.2019.01730>