www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 15(5), Dec 2024, 51-63. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2024.v15i5.557

IAA-CNN: Intelligent Attendance Algorithm based on Convolutional Neural Networks

Mario Anzures García, Luz A. Sánchez Gálvez, Mariano Larios Gómez, Arturo Tapia Rodríguez, Rubén Aguirre Agustín

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ciudad Universitaria, 14 sur esquina Boulevard Valsequillo, 72570, Puebla, México.
mario.anzures@correo.buap.mx, sanchez.galvez@correo.buap.mx, mariano.larios@correo.buap.mx, arturo.tapia@alumno.buap.mx, ruben.aguirre@alumno.buap.mx

Abstract. Nowadays, virtual meetings have increased their usage both in enterprises and educational organizations, since it is necessary to join them to make decisions, conferences, classes, etc. Therefore, there are many platforms for this; such as Microsoft TEAMS, Google Classroom, Zoom, Skype, and many more; simplifying and allowing members of such organizations to work together in a shared space. However, it is possible to waste a lot of time on passing the attendance, and it is very important to carry out it in an educational environment, in which this work is focused. Consequently, an intelligent attendance pass algorithm through voice and image recognition based on convolutional neural networks is proposed. The network ResNet50 is applied to make such recognition; since this network is particularly good for image classification and object detection. Furthermore, analyzed works to pass attendance only focus on generating an assistants list, voice recognition, or image recognition. So, nothing is doing both, voice, and image recognition. Finally, a case study to probe the feasibility of the proposed algorithm is carried out.

Keywords: Intelligent Algorithm, Attendance Pass, Voice Recognition, Image Recognition,·Convolutional Neural Networks

Article Info
Received May 10, 2024.
Accepted Nov 20, 2024.

1 Introduction

Collaborative systems are a computer-based application that supports groups of people who are engaged in a common task (or goal), providing an interface to a shared environment [1, 2, 3]. This definition implies that these systems allow us to manage and control a group of users in a workspace, providing communication, collaboration, and coordination among the users of this group. In this way, collaborative systems are used in much and different environments. For example, Social Networks (Facebook, X, Instagram, WhatsApp, etc.) for entertainment; videoconferences such as ZOOM, Google Meet, and Skype; Computer Supported Collaborative Visualization applied to medical ambient, and Computer Supported Collaborative Learning (CSCL), which has been applied in classrooms since the 1970s, although the vast majority of theoretical studies related to this field date from the 1980s [4]. Where it is postulated that learning is an experience.

The most representative systems of the CSCL are the learning management systems (LMS), which include a range of services for teachers in course management, the teaching process, and interaction with users [5]; used by educational institutions and in commercial contexts for training [6]. An LMS is characterized [7] by being multiplatform, multimedia, and having restricted access; and managing information, interaction, and communication through graphical interfaces. On the other hand, it is noted that the main design characteristics of an LMS are [8]: scalability, reliability, portability, concurrency, high performance, and fast response. Furthermore, an LMS is classified as proprietary, open source, cloud-based, and hybrid [9], the most representative being: Microsoft TEAMS, Google Classroom, Blackboard Learn, MOODLE, ANGEL, Canvas, D2L, Sakai, etc. Although these LMS present an assistance pass based on the IDs of the users. Also, they do it manually, naming each of the participants and waiting for them to answer; however, is very slow and tedious, especially when it comes to large groups. As well as there are no reliable automated methods for taking attendance in virtual meetings.

The simplest model of a neural network was proposed in 1958 by Frank Rosenblatt [10], which consists of a single layer of neurons and only an output. The perceptron is initialized with random values for all weights w_n. Afterward, the values of w_n are modified according to the rule expressed in (1):

$$w_j(k+1) = w_j(k) + \eta(k)[z(k) - y(k)]x_j(k) \tag{1}$$

The parameter η(k) is an always positive parameter known as the learning rate.

Due to the limitations presented by the use of a single layer, the multilayer perceptron was proposed. This type of perceptron is composed of a layer with N input neurons, another layer with M output neurons, and a certain number of hidden layers.

The multilayer perceptron with a single hidden layer, formed by L hidden neurons, is described in (2):

$$y_j = \sum_{j=1}^{L} w_{ij} s_j = f_1\left(\sum_{j=1}^{L} w_{ij} f_2\left(\sum_{r=1}^{N} t_{jr} x_r\right)\right) \tag{2}$$

with:

1. w_{ij} is the synaptic weight that connects the output neuron *I* with the neuron *j* of the hidden layer.
2. f_1 is the activation function of the output units.
3. t_{jr} is the synaptic weight connecting the hidden neuron *j* with the input neuron.
4. f_2 is the activation function of the hidden layer units.

The w_{ij} values of the output layer are modified according to the rule expressed in (3):

$$w_{ij}(k+1) = w_{ij}(k) + \eta[z_i(k) – y_i(k)]f'_1(h_i)s_j(k), j = 1, \dots, n+1 \tag{3}$$

The t_{jr} values of the hidden layer are modified according to the rule expressed in equation (4):

$$t_{jr}(k+1) = t_{jr}(k) + \eta \sum_{i=1}^{M} [z_i(k) - y_i(k)] \\ f_1'(h_i) w_{ij}(k) f_2'(u_j) x_r(k) \tag{4}$$

Consequently, in this paper an intelligent attendance algorithm based on convolutional neural networks, particularly, ResNet50 to carry out the recognition of voice and image, is presented. As well as a case study to demonstrate the viability of this algorithm is applied.

This paper is organized as follows: Section 2 briefly presents an introduction to convolutional neural networks. Section 3 describes the state of the art related to automated assistance in virtual environments. Section 4 explains the algorithm for intelligent attendance based on image and voice recognition by using convolutional neuronal networks. Section 5 shows the results. Finally, Section 6 outlines the conclusions and future work.

2 Convolutional Neural Network

The convolutional neural networks (CNN) [11, 12, 13, 14, 15] were proposed by LeCun [16], and present a type of artificial neural network (ANN) architecture for efficient pattern recognition in images. They are composed of convolutional layers and pooling layers, for enabling the extraction of features from input data while reducing the amount of processed information and preserving task-specific information. Two CNNs are here explained.

ResNet50 [17, 18, 19] is a 50-layer residual network, characterized by adding forward feedback every certain number of layers, and is designed to facilitate the optimization process of multi-layered neural networks. Also, the learning functions for the residual layers with references to the input layer are reformulated. Some of the uses of these networks are classification, processing, and predictions based on time series data, since there may be delays of unknown duration between important events in a time series. Furthermore, these types of networks are particularly good for image classification and object detection.

Long Short-Term Memory (LSTM) [20, 21, 22] is a recurrent neural network that can process not only individual data points but also entire data streams (such as voice or video). LSTMs were developed to address the vanishing gradient problem that can arise when training traditional recurrent neural networks.

3 State of the Art

The work [23] proposes a facial recognition system based on computer vision and machine learning to take attendance from students or employees of organizations. The system performs facial recognition of each student by taking a photograph, which is then stored on a server. The teacher can record attendance by clicking on some of the images. The system will then recognize the faces and verify their presence or absence. However, the teacher is not relieved of the roll call task, since they have to click on the images in the room, It does not work on virtual classroom platforms, and it does not use voice recognition tools, furthermore, it is a paid application.

The article [24] describes an implementation of artificial intelligence and machine learning models for emotion recognition and attendance taking. The system allows companies to generate data about their employees' emotions using facial recognition via a CNN. Offering an accuracy of 90%. Though, it does not have a user interface, it does not work on virtual classroom platforms, it does not use speech recognition tools and it is only focused on the industry.

The paper [25] carries out facial recognition in conjunction with security cameras to determine the attendance of a teacher to teach a certain class. In addition, determines the number of students who attended it; offering an accuracy of 95%. However, it does not have a user interface, does not have speech recognition, and does not work on virtual classroom platforms.

The authors of [26] record the attendance of a course by bot. It was developed as an alternative to a physical roll call system, in which a badge is swiped and each student registers using the badge's ID. Conversely, it does not use facial or voice recognition. Rather, it passes the assistance task to the student through their user account, it only works with Zoom and focuses on the medical area.

The work [27] uses facial recognition in classes through a camera placed in the middle of the room, to generate an automatic class list. It uses MultiTask Convolutional Neural Networks (MTCNN) based on facial feature extraction. Yet, it depends on the position of the camera when taking the face shots and is not intended for use on virtual platforms.

The authors of [28] identify students by voice recognition. It considers defined phrases that the user repeats and are taken as a password and unique identifier. It uses support vector machines (SVM). However, it is intended to be used in university, it only recognizes and records the information in a file; as well as it does not work with virtual platforms.

The paper [29] explicates a software project on MatLab for speech recognition using Euclidean distances to feature extraction. Nevertheless, it requires many samples to obtain accurate results, it limits its expansion to mobile or web platforms because it uses MatLab, it does not offer any user interface and it is not oriented to virtual platforms.

The authors of [30] developed a facial recognition-based attendance management system for education based on CNN. Here, the face recognition dataset is trained to the proposed CNN model. Using the Open CV face recognition approach, an input image will be processed, a face will be detected and then a spreadsheet will also be utilized to record attendance. Nonetheless, it does not have speech recognition, neither it is not oriented to virtual platforms.

The paper [31] makes use of facial recognition technologies for a class attendance system. This is designed to detect and recognize faces in real-time from classroom’s live streaming video. At the end of the session, the attendance information is automatically sent via mail to the respective faculty member. Though, it is not oriented to virtual platforms, and it employs voice recognition.

The authors of [32] propose a system that make used of various algorithms such as histogram of oriented Gradient (HOG), CNN and SVM for recognizing the face, after this, the reports of attendance are going to be created, maintained, and stored in excel format. Yet, it only makes this sort of recognition, and it is not oriented to virtual platforms.

The paper [33] has implemented Deep Learning model Convolutional Neural Network architecture for face detection to build a smart attendance system that will detect the faces of all the staff members and the attendance is marked automatically. This system has an accuracy of 90%. Nonetheless, it is not oriented to virtual platforms or to detect voice.

The authors of [34] develop a Smart Attendance Management System (SAMS) that is a web-based application, to provide attendance of students using face recognition, in realtime. This recognition is constructed on novel CNN architecture. However, it does not have speech recognition, and it focuses on meeting students physically.

The paper [35] monitors the students' attendance in the classroom using CCTV with a biometric facial recognition system. For this reason, the Principle Component Analysis (PCA), Eigen face value detection, CNN are the methods being used in this work. Which is not oriented to virtual meetings or speech recognition.

The authors of [36] propose an algorithm for face detection and recognition based on CNN, which outperforms the traditional techniques. In order to validate the efficiency of the proposed algorithm, a smart classroom for the student’s attendance using face recognition has been proposed. Although the proposed system achieved 97.9% accuracy on the testing data, it focuses not on virtual platforms.

In the paper [37] an unobtrusive face recognition based on a smart classroom attendance management system using the high definition rotating camera for capturing the faces of students is proposed. This system uses the MaxMargin Face Detection (MMFD) technique for the face detection and the model is trained using the Inception-V3 CNN technique for the students’ identification. But, it is not oriented to virtual platforms or to detect speech.

The authors of [38] introduce a smart and efficient system for attendance using face detection and face recognition. This system takes attendance in colleges or offices using real-time face recognition with the help of CNN. However, it is not oriented to virtual meetings or speech recognition.

In conclusion, different works were investigated in the areas of facial recognition, user identification by voice, and automated attendance taking for virtual environments; finding that most applications only use facial recognition and generate an output file; as well as those that work with cameras focused on face-to-face meetings.

Therefore, the proposal of this work innovates in the area of assistance systems by focusing on platforms for remote videoconferences, as well as with the integration of image and voice recognition, when most of the analyzed systems only use one of these.

4 Intelligent Attendance Algorithm

In this section, an intelligent algorithm to pass students' attendance in an LMS based on CNN for recognition of voice and images, is presented. Since this algorithm automates and reduces time in the aforementioned process.

Figure 1 shows the algorithm that establishes and regulates the operation of the automatic attendance of students in a virtual course, carrying out the acquisition of images and voice recordings that will be sent to CNN, which predict the attendance or non-attendance of a specific student. The algorithm consists of:

1. **Start**.

2. **Selection of LMS.** In this case, Microsoft TEAMS has been designated to aplicate this algorithm; since it is an LMS more used in virtual courses and our institution.

3. **Specification of Models.** It is necessary to create the models for storing the data —images and audio of students— both for the training set and processing set. A model is the single, definitive source of information about the data; containing the essential fields and behaviors of the same being stored. Typically, each model maps to a single database table. For this work, Each model is a Python class, and each attribute in the model represents a database field. The following models were created:

   a. *ImageModel*. It represents the profile images of the students of the course (in .jpg format).
   b. *AudioModel*. Which characterizes the raw audio (without preprocessing) of the speakers (in .wav format).
   c. *Student*. This denotes the students of the dataset in use.
   d. *NumpyModel*. It symbolizes the speakers' Mel coefficient arrays.
   e. *Endpoint*. It is an API Endpoint.
   f. *MLAlgorithm*. This represents the machine learning algorithm that you want to use for a specific dataset.
   g. *MLAlgorithmStatus*. It characterizes the state of the algorithm over time, it is used to know if it is active or inactive.
   h. *MLRequest*. This is used to save the interactions between the client and the server with the algorithms.

4. **Serialization.** Serializers allow us to convert complex data, such as query sets and model instances, into native data types that can then be easily represented in JSON, XML, or other content types. Serializers also provide deserialization, allowing parsed data to be converted back to complex types, after first validating the incoming data. For this reason, a model was designed for each serializer.

5. **View**. It is a method that takes a request and returns a response, therefore, for each of the data that needed to be obtained from the client, a view was developed. Once the serializers have been created, it proceeds to design the views for each one. The views developed were the following:

   a. *Student_APIView*.
      - Get. It gets all students from the database.
      - Post. This adds students to the database.

   b. *Student_APIView*:Detail.
      - Get_object. Which gets the student that matches our primary key from the database (for internal data use only).
      - Get. This obtains a student from the database.
      - Put. It updates or creates a student in the database.
      - Delete. This deletes a student from the database.

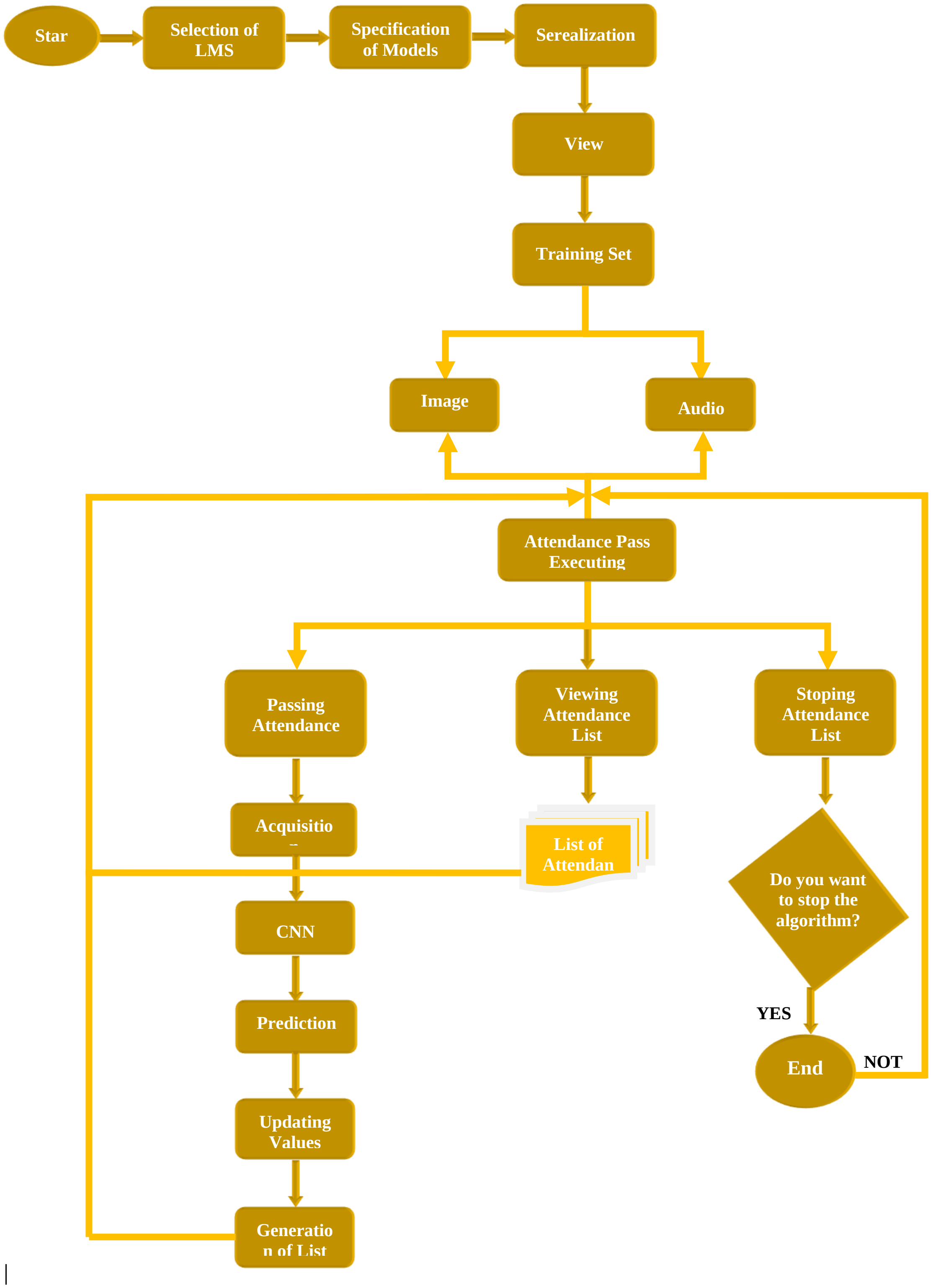


Figure 1. Intelligent Attendance Algorithm.

c. *Image*_APIView.

- Get. It gets all students' profile images from the database.
- Post. This adds students' profile images in .jpg format and size 128x128 to the database.

- *d.* *Image_*APIView:Detail.
  - Get_object. Which gets a profile image that matches our primary key from the database (for internal data use only).
  - Get. This obtains the URL of the profile image from the database, together with the class (student) it represents.
  - Put. It updates or creates a new Image model using the serializer that it defined before in the database.
  - Delete. This deletes an element of the image model from the database.
- *e.* *Numpy_*APIView:
  - Get It gets all Mel coefficients saved in the database.
  - Post. This adds Mel coefficients in .npy format to the database.
- *f.* *Numpy_APIView*:Detail:
  - Get_object. Which gets the numpy array that matches our database primary key (exclusively for internal data use).
  - Get. This obtains obtains the numpy array of the database.
  - Put. It updates or creates a numpy array of the database.
  - Delete. This deletes a numpy array from the database.
- *g.* *Audio_*APIView:
  - Get It gets all the audio of the speakers in the database.
  - Post. This adds audio recordings in .wav format to the database.
- h. *Audio_APIView_Detail.*
  - Get_object. Which gets the audio that matches our database primary key (for internal data use only).
  - Get. This obtains audio from the database.
  - Put. It updates or creates audio from the database.
  - Delete. This deletes audio from the database.
- i. *GE2EFitView*.
  - Get. It begins the training process of the speaker recognition neural network.
- j. *GE2EPredictView*.
  - Post. This requests a prediction from the speaker recognition neural network, receiving a wav audio recording as input.
- k. *ResNetFitView*.
  - Get. It begins the training process of the image recognition neural network.
- l. *ResNetPredictView*.
  - Post. This requests a prediction from the image recognition neural network, receiving as input an image that is automatically scaled to 128x128 in jpg format.

6. **Training_Set**. In order to carry out the CNN training both audio and image, the students of the Projects I+D 1 university course were considered. 40 of 44 students sent the image of his/her Microsoft TEAMS profile and audio with the phrase “I am in the Projects I+D 1 course”. These data were stored in a database of images, and audio, respectively. Such data were used as the training set of the image and audio CNN to recognize the course members accurately. So, this process is divided into two parts: Image Preprocessing, and Audio Preprocessing.

   a. *Image.* An image in JPEG format is taken as input, which is converted into an image of 128x128 pixels. Subsequently, the image is cast into an array (128, 128, 3), which is then used as input for the neural network. To predict who owns an image, it uses the ResNet50 neural network.

   b. *Audio*. The Discrete Fourier Transform is obtained, which will be useful to obtain Mel's cepstral coefficients. Also, it is possible for time domain processing to calculate the envelope of the audio samples, allowing the audio signal to be normalized. In this part, sounds are grouped by similarity based on descriptors. As well as timbre characteristics are also obtained through the centroid and Mel's cepstral coefficients. Due to the dimensionality of the embedded vector, once it processed the voice audio, it was concluded that ResNet50 would be able to give us a good result for speaker recognition. This is because the embedded vector also generates a two-dimensional matrix, so it can be assumed that it is possible to interpret it as an image.

7. **Attendance Pass Executing.** In order to execute the intelligent attendance algorithm on virtual platforms an extension in Microsoft Teams was developed. A Once a professor starts a course on Microsoft TEAMS, he/she can perform the attendance pass —which he/she has previously installed on his/her computer, laptop,

or mobile device. For illustrative purposes, in Figure 2 the advanced software engineering course is shown; in which students 24 attendance that day. In this figure can see the attendance pass system menu with three options: Pass Attendance, Viewing Attendance List, and Stoping Attendance Algorithm.

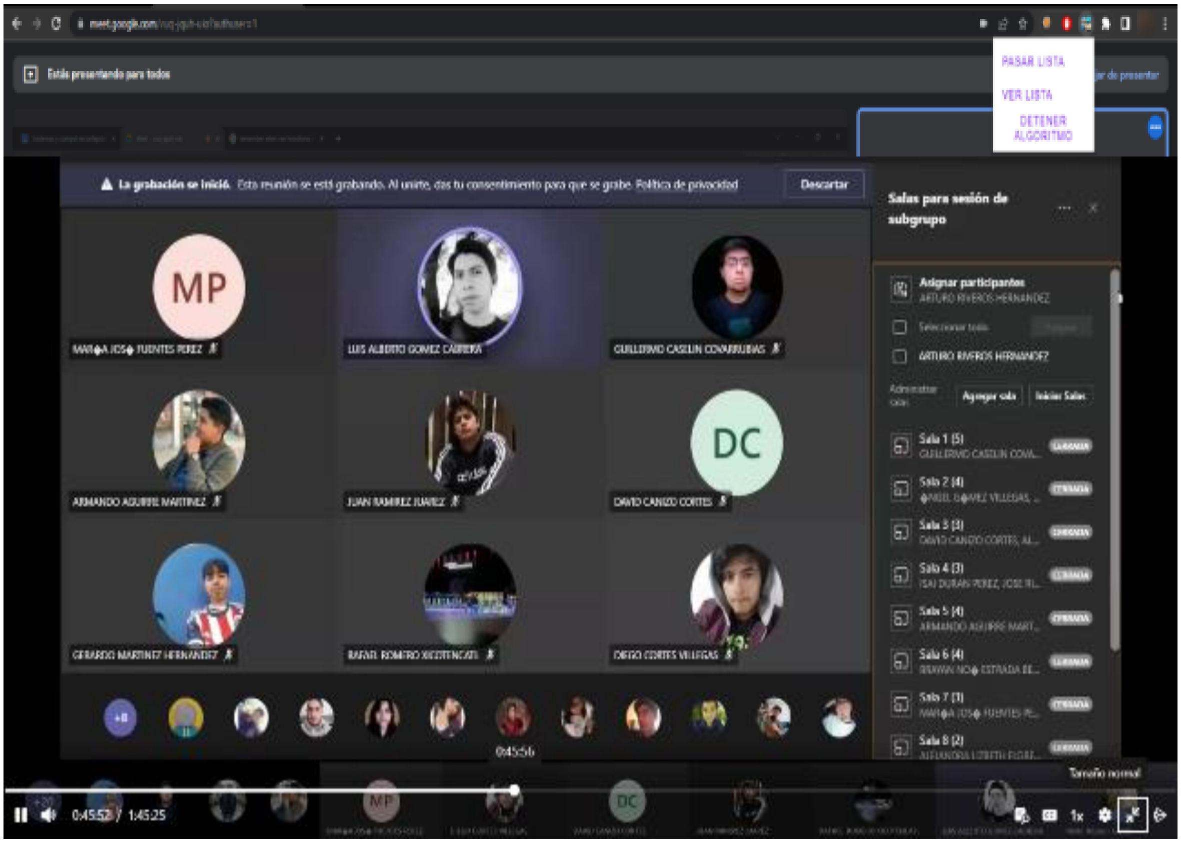


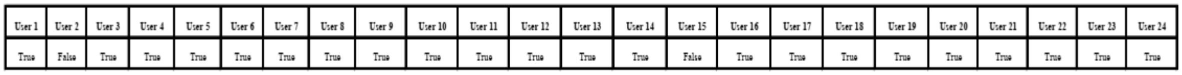
Figure 2. Intelligent Attendance System running in Microsoft Teams..

a. *Passing attendance*. In this option, several processes are carried out.
   - Acquisition. A scraping process is carried out both for obtaining images and audio. Thus, a Scrapping process is started to detect the profile images of each user and as a result, an array is generated. Furthermore, an audio recording process is started for 3 seconds, to detect the people who are speaking.
   - CNN. Once the acquisition of images and voice is applied, the array and recording are sent to the CNNs, correspondently.
   - Prediction. A prediction is generated based on the array and recording obtained in the step above, by each student.
   - Updating values. The CNN predictions both audio recognition and voice recognition are received; in case the certainty is greater than 0.9 the student's attendance is considered.
   - Generation of list. The predictions of both CNNs are combined, and an attendance final list is generated.

b. *View the attendance list***.** This option allows us to display the results of the algorithm, as they are being acquired as shown in Figure 3. Being able to download the results obtained from the audio, voice, and/or assistance CNNs. In addition, the outline of the student's photograph when he attends turns green and red when he does not attend.

c. *Stop the attendance algorithm*. This option allows us to stop the algorithm, when the professor wishes and thus ends up consuming the resources required by the intelligent assistance pass.

8. ***End.***

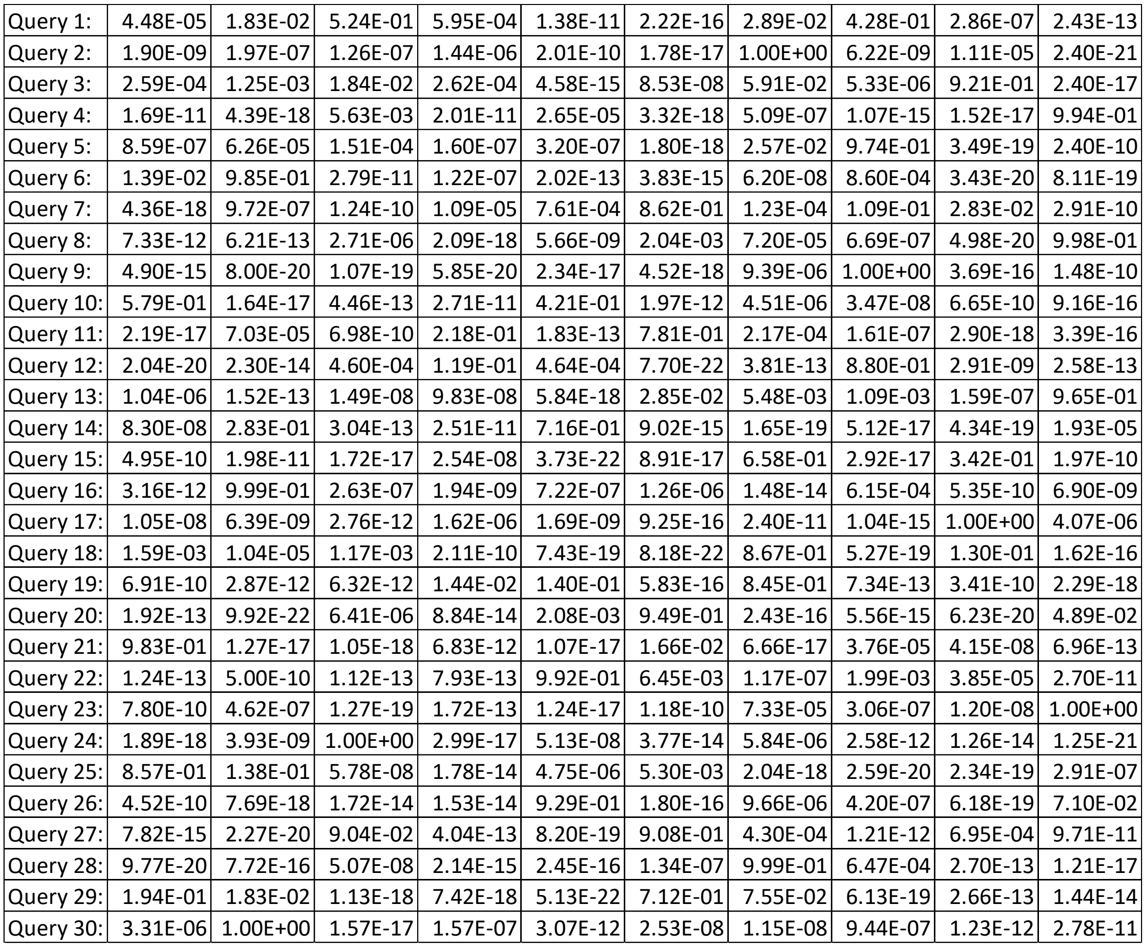


User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24
True	False	True	True	True	True	True	True	True	True	True	True	True	True	False	True	True	True	True	True	True	True	True	True

Figure 3. Attendance with the intelligent algorithm.

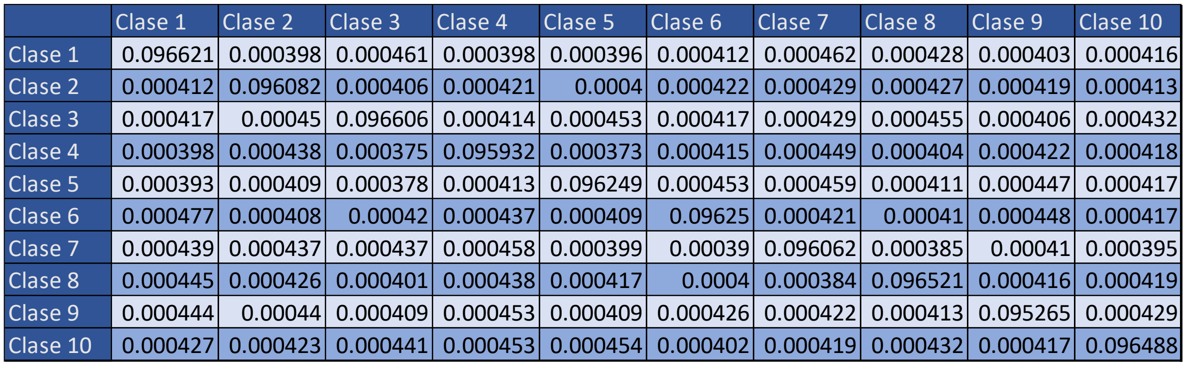
5 Results

In the CNN of images (as shown in Figure 4), an accuracy of 96% was obtained as can be seen in Figure 5, because images that do not correspond to the students or the professor are filtered, and this causes the network to generate predictions that are not adequate. For the audio network (as shown in Figure 6), an accuracy of 90% is achieved (as can be seen in Figure 7), mainly because users tend to have ambient noise, so the pre-processing is not able to eliminate it.



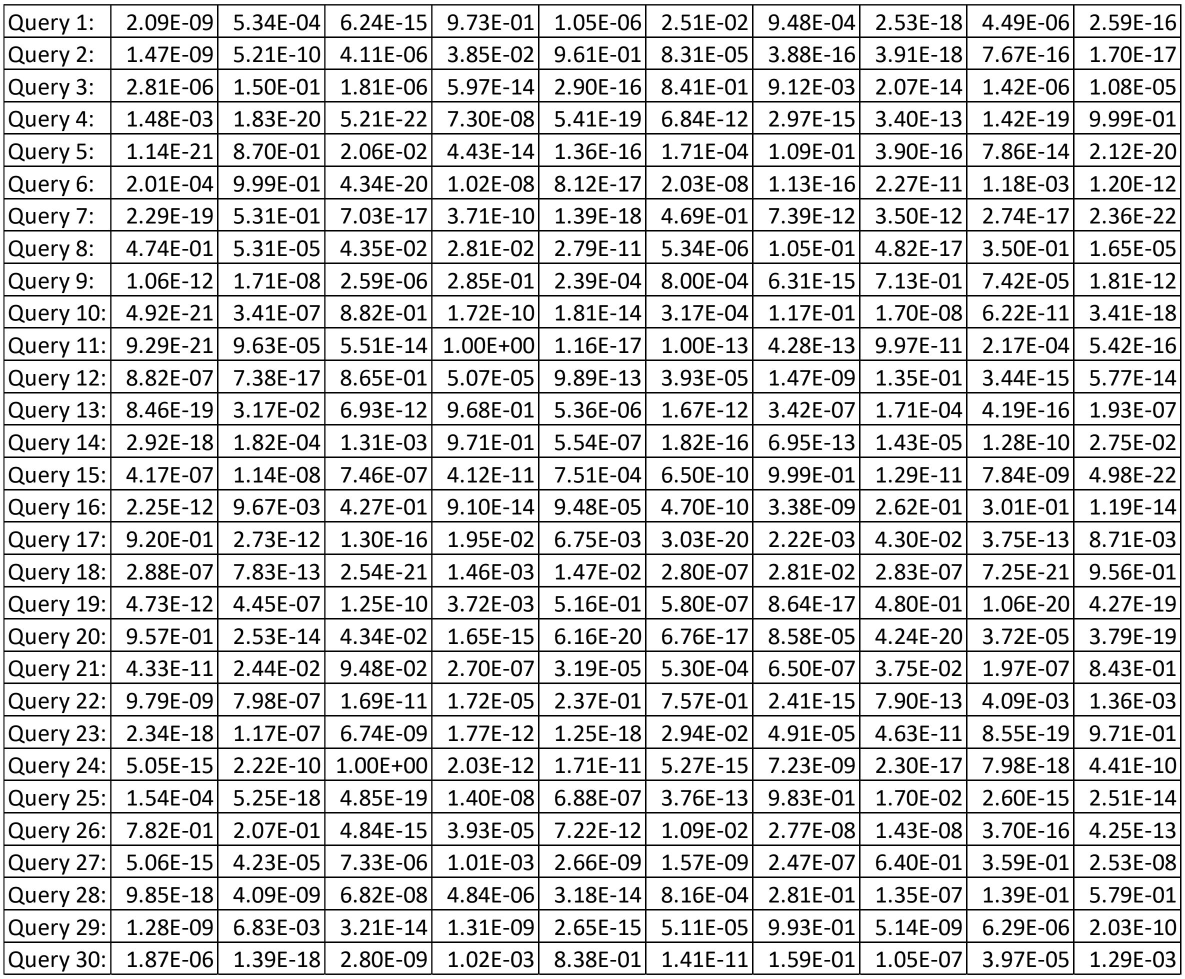
Query 1:	4.48E-05	1.83E-02	5.24E-01	5.95E-04	1.38E-11	2.22E-16	2.89E-02	4.28E-01	2.86E-07	2.43E-13
Query 2:	1.90E-09	1.97E-07	1.26E-07	1.44E-06	2.01E-10	1.78E-17	1.00E+00	6.22E-09	1.11E-05	2.40E-21
Query 3:	2.59E-04	1.25E-03	1.84E-02	2.62E-04	4.58E-15	8.53E-08	5.91E-02	5.33E-06	9.21E-01	2.40E-17
Query 4:	1.69E-11	4.39E-18	5.63E-03	2.01E-11	2.65E-05	3.32E-18	5.09E-07	1.07E-15	1.52E-17	9.94E-01
Query 5:	8.59E-07	6.26E-05	1.51E-04	1.60E-07	3.20E-07	1.80E-18	2.57E-02	9.74E-01	3.49E-19	2.40E-10
Query 6:	1.39E-02	9.85E-01	2.79E-11	1.22E-07	2.02E-13	3.83E-15	6.20E-08	8.60E-04	3.43E-20	8.11E-19
Query 7:	4.36E-18	9.72E-07	1.24E-10	1.09E-05	7.61E-04	8.62E-01	1.23E-04	1.09E-01	2.83E-02	2.91E-10
Query 8:	7.33E-12	6.21E-13	2.71E-06	2.09E-18	5.66E-09	2.04E-03	7.20E-05	6.69E-07	4.98E-20	9.98E-01
Query 9:	4.90E-15	8.00E-20	1.07E-19	5.85E-20	2.34E-17	4.52E-18	9.39E-06	1.00E+00	3.69E-16	1.48E-10
Query 10:	5.79E-01	1.64E-17	4.46E-13	2.71E-11	4.21E-01	1.97E-12	4.51E-06	3.47E-08	6.65E-10	9.16E-16
Query 11:	2.19E-17	7.03E-05	6.98E-10	2.18E-01	1.83E-13	7.81E-01	2.17E-04	1.61E-07	2.90E-18	3.39E-16
Query 12:	2.04E-20	2.30E-14	4.60E-04	1.19E-01	4.64E-04	7.70E-22	3.81E-13	8.80E-01	2.91E-09	2.58E-13
Query 13:	1.04E-06	1.52E-13	1.49E-08	9.83E-08	5.84E-18	2.85E-02	5.48E-03	1.09E-03	1.59E-07	9.65E-01
Query 14:	8.30E-08	2.83E-01	3.04E-13	2.51E-11	7.16E-01	9.02E-15	1.65E-19	5.12E-17	4.34E-19	1.93E-05
Query 15:	4.95E-10	1.98E-11	1.72E-17	2.54E-08	3.73E-22	8.91E-17	6.58E-01	2.92E-17	3.42E-01	1.97E-10
Query 16:	3.16E-12	9.99E-01	2.63E-07	1.94E-09	7.22E-07	1.26E-06	1.48E-14	6.15E-04	5.35E-10	6.90E-09
Query 17:	1.05E-08	6.39E-09	2.76E-12	1.62E-06	1.69E-09	9.25E-16	2.40E-11	1.04E-15	1.00E+00	4.07E-06
Query 18:	1.59E-03	1.04E-05	1.17E-03	2.11E-10	7.43E-19	8.18E-22	8.67E-01	5.27E-19	1.30E-01	1.62E-16
Query 19:	6.91E-10	2.87E-12	6.32E-12	1.44E-02	1.40E-01	5.83E-16	8.45E-01	7.34E-13	3.41E-10	2.29E-18
Query 20:	1.92E-13	9.92E-22	6.41E-06	8.84E-14	2.08E-03	9.49E-01	2.43E-16	5.56E-15	6.23E-20	4.89E-02
Query 21:	9.83E-01	1.27E-17	1.05E-18	6.83E-12	1.07E-17	1.66E-02	6.66E-17	3.76E-05	4.15E-08	6.96E-13
Query 22:	1.24E-13	5.00E-10	1.12E-13	7.93E-13	9.92E-01	6.45E-03	1.17E-07	1.99E-03	3.85E-05	2.70E-11
Query 23:	7.80E-10	4.62E-07	1.27E-19	1.72E-13	1.24E-17	1.18E-10	7.33E-05	3.06E-07	1.20E-08	1.00E+00
Query 24:	1.89E-18	3.93E-09	1.00E+00	2.99E-17	5.13E-08	3.77E-14	5.84E-06	2.58E-12	1.26E-14	1.25E-21
Query 25:	8.57E-01	1.38E-01	5.78E-08	1.78E-14	4.75E-06	5.30E-03	2.04E-18	2.59E-20	2.34E-19	2.91E-07
Query 26:	4.52E-10	7.69E-18	1.72E-14	1.53E-14	9.29E-01	1.80E-16	9.66E-06	4.20E-07	6.18E-19	7.10E-02
Query 27:	7.82E-15	2.27E-20	9.04E-02	4.04E-13	8.20E-19	9.08E-01	4.30E-04	1.21E-12	6.95E-04	9.71E-11
Query 28:	9.77E-20	7.72E-16	5.07E-08	2.14E-15	2.45E-16	1.34E-07	9.99E-01	6.47E-04	2.70E-13	1.21E-17
Query 29:	1.94E-01	1.83E-02	1.13E-18	7.42E-18	5.13E-22	7.12E-01	7.55E-02	6.13E-19	2.66E-13	1.44E-14
Query 30:	3.31E-06	1.00E+00	1.57E-17	1.57E-07	3.07E-12	2.53E-08	1.15E-08	9.44E-07	1.23E-12	2.78E-11

Figure 4. CNN image result.



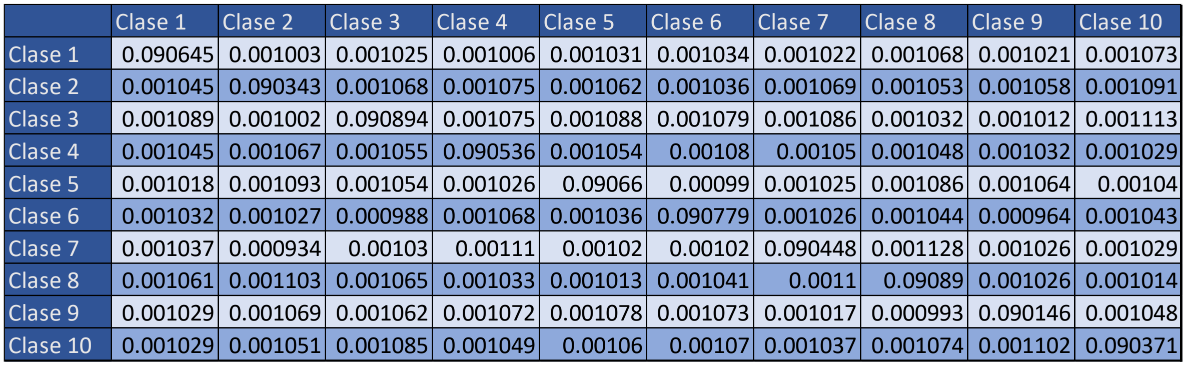
	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5	Clase 6	Clase 7	Clase 8	Clase 9	Clase 10
Clase 1	0.096621	0.000398	0.000461	0.000398	0.000396	0.000412	0.000462	0.000428	0.000403	0.000416
Clase 2	0.000412	0.096082	0.000406	0.000421	0.0004	0.000422	0.000429	0.000427	0.000419	0.000413
Clase 3	0.000417	0.00045	0.096606	0.000414	0.000453	0.000417	0.000429	0.000455	0.000406	0.000432
Clase 4	0.000398	0.000438	0.000375	0.095932	0.000373	0.000415	0.000449	0.000404	0.000422	0.000418
Clase 5	0.000393	0.000409	0.000378	0.000413	0.096249	0.000453	0.000459	0.000411	0.000447	0.000417
Clase 6	0.000477	0.000408	0.00042	0.000437	0.000409	0.09625	0.000421	0.00041	0.000448	0.000417
Clase 7	0.000439	0.000437	0.000437	0.000458	0.000399	0.00039	0.096062	0.000385	0.00041	0.000395
Clase 8	0.000445	0.000426	0.000401	0.000438	0.000417	0.0004	0.000384	0.096521	0.000416	0.000419
Clase 9	0.000444	0.00044	0.000409	0.000453	0.000409	0.000426	0.000422	0.000413	0.095265	0.000429
Clase 10	0.000427	0.000423	0.000441	0.000453	0.000454	0.000402	0.000419	0.000432	0.000417	0.096488

Figure 5. Image CNN confusion matrix.



Query 1:	2.09E-09	5.34E-04	6.24E-15	9.73E-01	1.05E-06	2.51E-02	9.48E-04	2.53E-18	4.49E-06	2.59E-16
Query 2:	1.47E-09	5.21E-10	4.11E-06	3.85E-02	9.61E-01	8.31E-05	3.88E-16	3.91E-18	7.67E-16	1.70E-17
Query 3:	2.81E-06	1.50E-01	1.81E-06	5.97E-14	2.90E-16	8.41E-01	9.12E-03	2.07E-14	1.42E-06	1.08E-05
Query 4:	1.48E-03	1.83E-20	5.21E-22	7.30E-08	5.41E-19	6.84E-12	2.97E-15	3.40E-13	1.42E-19	9.99E-01
Query 5:	1.14E-21	8.70E-01	2.06E-02	4.43E-14	1.36E-16	1.71E-04	1.09E-01	3.90E-16	7.86E-14	2.12E-20
Query 6:	2.01E-04	9.99E-01	4.34E-20	1.02E-08	8.12E-17	2.03E-08	1.13E-16	2.27E-11	1.18E-03	1.20E-12
Query 7:	2.29E-19	5.31E-01	7.03E-17	3.71E-10	1.39E-18	4.69E-01	7.39E-12	3.50E-12	2.74E-17	2.36E-22
Query 8:	4.74E-01	5.31E-05	4.35E-02	2.81E-02	2.79E-11	5.34E-06	1.05E-01	4.82E-17	3.50E-01	1.65E-05
Query 9:	1.06E-12	1.71E-08	2.59E-06	2.85E-01	2.39E-04	8.00E-04	6.31E-15	7.13E-01	7.42E-05	1.81E-12
Query 10:	4.92E-21	3.41E-07	8.82E-01	1.72E-10	1.81E-14	3.17E-04	1.17E-01	1.70E-08	6.22E-11	3.41E-18
Query 11:	9.29E-21	9.63E-05	5.51E-14	1.00E+00	1.16E-17	1.00E-13	4.28E-13	9.97E-11	2.17E-04	5.42E-16
Query 12:	8.82E-07	7.38E-17	8.65E-01	5.07E-05	9.89E-13	3.93E-05	1.47E-09	1.35E-01	3.44E-15	5.77E-14
Query 13:	8.46E-19	3.17E-02	6.93E-12	9.68E-01	5.36E-06	1.67E-12	3.42E-07	1.71E-04	4.19E-16	1.93E-07
Query 14:	2.92E-18	1.82E-04	1.31E-03	9.71E-01	5.54E-07	1.82E-16	6.95E-13	1.43E-05	1.28E-10	2.75E-02
Query 15:	4.17E-07	1.14E-08	7.46E-07	4.12E-11	7.51E-04	6.50E-10	9.99E-01	1.29E-11	7.84E-09	4.98E-22
Query 16:	2.25E-12	9.67E-03	4.27E-01	9.10E-14	9.48E-05	4.70E-10	3.38E-09	2.62E-01	3.01E-01	1.19E-14
Query 17:	9.20E-01	2.73E-12	1.30E-16	1.95E-02	6.75E-03	3.03E-20	2.22E-03	4.30E-02	3.75E-13	8.71E-03
Query 18:	2.88E-07	7.83E-13	2.54E-21	1.46E-03	1.47E-02	2.80E-07	2.81E-02	2.83E-07	7.25E-21	9.56E-01
Query 19:	4.73E-12	4.45E-07	1.25E-10	3.72E-03	5.16E-01	5.80E-07	8.64E-17	4.80E-01	1.06E-20	4.27E-19
Query 20:	9.57E-01	2.53E-14	4.34E-02	1.65E-15	6.16E-20	6.76E-17	8.58E-05	4.24E-20	3.72E-05	3.79E-19
Query 21:	4.33E-11	2.44E-02	9.48E-02	2.70E-07	3.19E-05	5.30E-04	6.50E-07	3.75E-02	1.97E-07	8.43E-01
Query 22:	9.79E-09	7.98E-07	1.69E-11	1.72E-05	2.37E-01	7.57E-01	2.41E-15	7.90E-13	4.09E-03	1.36E-03
Query 23:	2.34E-18	1.17E-07	6.74E-09	1.77E-12	1.25E-18	2.94E-02	4.91E-05	4.63E-11	8.55E-19	9.71E-01
Query 24:	5.05E-15	2.22E-10	1.00E+00	2.03E-12	1.71E-11	5.27E-15	7.23E-09	2.30E-17	7.98E-18	4.41E-10
Query 25:	1.54E-04	5.25E-18	4.85E-19	1.40E-08	6.88E-07	3.76E-13	9.83E-01	1.70E-02	2.60E-15	2.51E-14
Query 26:	7.82E-01	2.07E-01	4.84E-15	3.93E-05	7.22E-12	1.09E-02	2.77E-08	1.43E-08	3.70E-16	4.25E-13
Query 27:	5.06E-15	4.23E-05	7.33E-06	1.01E-03	2.66E-09	1.57E-09	2.47E-07	6.40E-01	3.59E-01	2.53E-08
Query 28:	9.85E-18	4.09E-09	6.82E-08	4.84E-06	3.18E-14	8.16E-04	2.81E-01	1.35E-07	1.39E-01	5.79E-01
Query 29:	1.28E-09	6.83E-03	3.21E-14	1.31E-09	2.65E-15	5.11E-05	9.93E-01	5.14E-09	6.29E-06	2.03E-10
Query 30:	1.87E-06	1.39E-18	2.80E-09	1.02E-03	8.38E-01	1.41E-11	1.59E-01	1.05E-07	3.97E-05	1.29E-03

Figure 6. ANN audio result.



	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5	Clase 6	Clase 7	Clase 8	Clase 9	Clase 10
Clase 1	0.090645	0.001003	0.001025	0.001006	0.001031	0.001034	0.001022	0.001068	0.001021	0.001073
Clase 2	0.001045	0.090343	0.001068	0.001075	0.001062	0.001036	0.001069	0.001053	0.001058	0.001091
Clase 3	0.001089	0.001002	0.090894	0.001075	0.001088	0.001079	0.001086	0.001032	0.001012	0.001113
Clase 4	0.001045	0.001067	0.001055	0.090536	0.001054	0.00108	0.00105	0.001048	0.001032	0.001029
Clase 5	0.001018	0.001093	0.001054	0.001026	0.09066	0.00099	0.001025	0.001086	0.001064	0.00104
Clase 6	0.001032	0.001027	0.000988	0.001068	0.001036	0.090779	0.001026	0.001044	0.000964	0.001043
Clase 7	0.001037	0.000934	0.00103	0.00111	0.00102	0.00102	0.090448	0.001128	0.001026	0.001029
Clase 8	0.001061	0.001103	0.001065	0.001033	0.001013	0.001041	0.0011	0.09089	0.001026	0.001014
Clase 9	0.001029	0.001069	0.001062	0.001072	0.001078	0.001073	0.001017	0.000993	0.090146	0.001048
Clase 10	0.001029	0.001051	0.001085	0.001049	0.00106	0.00107	0.001037	0.001074	0.001102	0.090371

Figure 7. Audio ANN confusion matrix.

6 Conclusions and Future Work

An innovative algorithm that uses both image and voice recognition to automatically take attendance of students in class has been developed and applied. This algorithm was applied to create an intelligent system that automatically takes attendance based on information available (image and voice) from users on the Microsoft Teams platform. The algorithm utilizes convolutional neural networks to carry out image recognition based on the profile images of the users, and voice recognition based on speech. This algorithm is unique because it combines both image and voice recognition technologies to automate the attendance process. Furthermore, the profile images of the users were taken to carry out an image recognition process based on artificial neural networks. In addition, a voice recognition system was developed to recognize users through speech, also using this type of network. Three CSV files were generated, which were analyzed, obtaining a percentage of 98% accuracy in the results of the combined attendance of both artificial neural networks. Finally, this algorithm can be used in any kind of meeting. Future work will consist of scaling the Google Chrome extension to increase its compatibility with the different environments used to hold meetings remotely. In addition, it will seek to improve the precision in the results of both networks, as well as improve their performance in terms of training time for data sets of considerable size.

References

1. Ellis, C. A., Gibbs, S. J., & Rein, G. L. (1991). Groupware: Some issues and experiences. *Communications of the ACM, 34*(1), 39–58.
2. Anzures-García, M., Sanchez-Gálvez, L. A., Hornos, M. J., & Paderewski-Rodríguez, P. (2018). Tutorial function groupware based on a workflow ontology and a directed acyclic graph. *IEEE Latin American Transactions, 16*(1), 294–300.
3. Anzures-García, M., & Sanchez-Gálvez, L. A. (2020). PROMISE: Proposing an ontological model for developing collaborative systems. *Journal of Intelligent & Fuzzy Systems, 39*, 2545–2557. https://doi.org/10.3233/JIFS-179913
4. Slavin, R. E. (1983). *Cooperative learning*. New York: Longman.
5. Ouadoud, M., Nejjari, A., Chkouri, M. Y., & El Kadiri, K. E. (2018). Educational modeling of a learning management system. *Proceedings of the International Conference on Electrical and Information Technologies*, 1–6.
6. Al-Busaidi, K. A., & Al-Shihi, H. (2009). A framework for evaluating instructors' acceptance of learning management systems. *Knowledge Management Innovation Advances in Economic Analysis Solutions: Proceedings of the 13th International Business Information Management Association Conference (IBIMA)*, 3, 1199–1207.
7. Medina-Flores, R., & Morales-Gamboa, R. (2015). Usability evaluation by experts of a learning management system. *Revista Iberoamericana de Tecnología del Aprendizaje, 10*(4), 197–203.
8. Bao, S., & Meng, F. (2015). The design of a massive open online course platform for English learning based on Moodle. *Proceedings of the Conference on Communication Systems and Network Technologies*, 1365–1368.
9. Dobre, I. (2015). Learning management systems for higher education – An overview of available options for higher education organizations. *Procedia - Social and Behavioral Sciences*.
10. Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386–408. https://doi.org/10.1037/h0042519

11. Melamane, S., Rai, M., Khan, S., & Belgamwar, A. (2023). Artificial neural network–based inference of drug-target interactions. In *Nanotechnology Principles in Drug Targeting and Diagnosis* (pp. 35–62). Elsevier. https://doi.org/10.1016/B978-0-323-91763-6.00015-1
12. AbdelRaouf, H. (2023). Efficient convolutional neural network-based keystroke dynamics for boosting user authentication. *Sensors, 3*(10), 4898. https://doi.org/10.3390/s23104898
13. Devi, S. K., et al. (2023). Intelligent deep convolutional neural network-based object detection model for visually challenged people. *Computer Systems Science & Engineering, 46*(3), 3191–3207. https://doi.org/10.32604/csse.2023.036980
14. Baradaran, F., et al. (2023). Customized 2D CNN model for automatic emotion recognition based on EEG signals. *Electronics, 12*, 2232. https://doi.org/10.3390/electronics12102232
15. Yamashita, R., et al. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging, 9*, 611–629. https://doi.org/10.1007/s13244-018-0639-9
16. LeCun, Y., et al. (1989). Backpropagation applied to handwritten ZIP code recognition. *Neural Computation, 1*, 541–551.
17. Zhang, L., et al. (2023). A transfer residual neural network based on ResNet-50 for detection of steel surface defects. *Applied Sciences, 13*(9). https://doi.org/10.3390/app13095260
18. Hossain, B., et al. (2022). Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. *Informatics in Medicine Unlocked, 30*. https://doi.org/10.1016/j.imu.2022.100983
19. Liu, J., et al. (2021). Rock image intelligent classification and recognition based on ResNet-50 model. *Journal of Physics: Conference Series, 2076*. https://doi.org/10.1088/1742-6596/2076/1/012011
20. Abed, M., et al. (2021). Application of long short-term memory neural network technique for predicting monthly pan evaporation. *Scientific Reports, 11*(20742). https://doi.org/10.1038/s41598-021-99999-y
21. Elsaraiti, M., & Merabet, A. (2021). Application of long short-term memory recurrent neural networks to forecast wind speed. *Applied Sciences*, *11*(5). https://doi.org/10.3390/app11052387
22. Zahn, R., et al. (2021). Application of a long short-term memory neural network for modeling transonic buffet aerodynamics. *Aerospace Science and Technology*, *113*. https://doi.org/10.1016/j.ast.2021.106632
23. AindraLabs. (n.d.). Retrieved October 20, 2014, from https://www.aindralabs.com/company
24. Kandjimi, H., et al. (n.d.). Attendance system with emotion detection: A case study with CNN and OpenCV.
25. Zhuang, J., & Huang, W. (2021). Design and implementation of intelligent teaching attendance. *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*, 516–520. https://doi.org/10.1145/3501409.3501503
26. Kamel, P., et al. (2021). Conference attendance tracking and evaluation in the era of virtual conferences. *Academic Radiology*, *29*(5), 576–581. https://doi.org/10.1016/j.acra.2021.12.012
27. Li, T. (2021). Research on intelligent classroom attendance management. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-021-03042-x
28. Jelil, S. (2019). Speechmarker: A voice-based multi-level attendance application. Presented at *INTERSPEECH 2019: Show & Tell Contribution*, 15–19.
29. Uddin, N., et al. (2016). Development of voice recognition for student attendance. *Global Journal of Human-Social Science, 1*(6).
30. Vignesh, K., et al. (2023). Smart attendance system using deep learning. *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 1081–1087. https://doi.org/10.1109/ICOEI56765.2023.10126022
31. Kushwaha, K., et al. (2023). A CNN-based attendance management system using face recognition. *Proceedings of the Fourth International Conference on Smart Electronics and Communication (ICOSEC-2023)*.
32. Kapse, A., et al. (2022). Face recognition attendance system using HOG and CNN algorithm. *ITM Web of Conferences, 44*, 03028. https://doi.org/10.1051/itmconf/20224403028
33. Natesan, P., et al. (2021). Smart staff attendance system using convolutional neural networks. *International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 1–5. https://doi.org/10.1109/ICCCI50826.2021.9402589
34. Kakarla, S., et al. (2020). Smart attendance management system based on face recognition using CNN. *IEEE HYDCON*, Hyderabad, India, 1–5. https://doi.org/10.1109/HYDCON48903.2020.9242847

35. Muthunagai, R., Muruganandhan, D., & Rajasekaran, P. (2020). Classroom attendance monitoring using CCTV. *International Conference on System, Computation, Automation, and Networking (ICSCAN 2020)*, 1–4. https://api.semanticscholar.org/CorpusID:227276921
36. Khan, M. Z., et al. (2019). Deep unified model for face recognition based on convolutional neural network and edge computing. *IEEE Access*, *7*, 72622–72633. https://doi.org/10.1109/ACCESS.2019.2918275
37. Gupta, S. K., et al. (2018). CVUCAMS: Computer vision-based unobtrusive classroom attendance management system. *IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, Mumbai, India, 101–102. https://doi.org/10.1109/ICALT.2018.00131
38. Arya, S., Mesariya, H., & Parekh, V. (2020). Smart attendance system using CNN. *ArXiv*.