Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 3 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i03.46765

PAPER

Design of a Sign Language-to-Natural Language Translator Using Artificial Intelligence

Hernando Gonzalez(✉),
Silvia Hernández,
Oscar Calderón

Universidad Autónoma de Bucaramanga (UNAB), Bucaramanga, Colombia

hgonzalez7@unab.edu.co

ABSTRACT

This paper describes the results obtained from the design and validation of translation gloves for Colombian sign language (LSC) to natural language. The MPU6050 sensors capture finger movements, and the TCA9548a card enables data multiplexing. Additionally, an Arduino Uno board preprocesses the data, and the Raspberry Pi interprets it using central tendency statistics, principal component analysis (PCA), and a neural network structure for pattern recognition. Finally, the sign is reproduced in audio format. The methodology developed below focuses on translating specific preselected words, achieving an average classification accuracy of 88.97%.

KEYWORDS

sign language, neural network, signal processing, pattern recognition

1 INTRODUCTION

According to the World Health Organization (WHO), an estimated 430 million individuals, representing over 5% of the global population, suffer from disabling hearing loss. Among them, approximately 500,000 are Colombians, constituting around 1% of Colombia's total population, which was estimated at 48 million in 2020 [1]. This significant demographic encounters communication challenges because they are unable to naturally articulate speech, leading them to depend on sign language as an alternative form of expression. Despite its significance, a comprehensive global implementation of sign language remains unrealized.

Various approaches have been proposed to bridge the gap between sign language and spoken language, with the aim of enhancing communication effectiveness and idea transmission. One common approach involves using image recognition systems. Researchers at Guilan University's electrical engineering department developed a method that uses nonparametric robust tracking algorithms, such as CAMSHIFT (Continuously Adaptive Mean-Shift), along with different classifiers. Impressively, this technique achieved a peak recognition accuracy of 95.56% for translating Persian words, with a swift recognition duration of 6.2 milliseconds [2].

Gonzalez, H., Hernández, S., Calderón, O. (2024). Design of a Sign Language-to-Natural Language Translator Using Artificial Intelligence. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(3), pp. 89–98. https://doi.org/10.3991/ijoe.v20i03.46765

Article submitted 2023-10-15. Revision uploaded 2023-11-18. Final acceptance 2023-11-21.

Furthermore, a novel technique was presented at the 2017 International Conference on Virtual Reality and Visualization. This approach utilized motion-capturing elements using leap motion equipment and Kth-nearest neighbor algorithms to interpret and translate Chinese sign language. The methodology involved detecting finger movements while disregarding their positional information. The experimental results, which included 100 test patterns consisting of digits and the Chinese alphabet, demonstrated an average classification accuracy of 96.8% [3].

Another avenue of exploration involves the integration of sensor-based gesture recognition gloves. A significant development occurred in the second edition of EIConCIT 2018, when researchers introduced gloves equipped with gyroscopes to measure hand orientation along with sensors tracking finger flexion [4], [17]. The employed classification algorithm, support vector machine (SVM), demonstrated outstanding performance. Remarkably, when using a training data-to-test data ratio of 70:30, this ensemble SVM technique achieved very high scores for accuracy (99.1%), sensitivity (99.6%), and specificity (98.7%) [5] [18] [19].

Similarly, students at Kochi University developed gloves equipped with flexion sensors and a palm-embedded gyroscope to record hand movement data. By employing heuristic algorithms based on class-specific threshold values, they attained an average classification rate of 51% for 20 Japanese words [6]. Furthermore, at the 2022 IEEE 4th Global Conference on Life Sciences and Technologies, scholars from Waseda University presented gloves equipped with finger flexion sensors and an inertial measurement unit to capture palm-acquired accelerations and angular velocities. Employing a variety of machine learning algorithms, the researchers achieved impressive accuracy percentages for the Japanese alphabet: SVM (99%); decision tree (94.25%); random forest (99.75%); and K-nearest neighbors (99.75%) [7].

In conclusion, the pressing issue of communication barriers faced by the deaf and hard-of-hearing population has led to the exploration of innovative techniques for translating sign language into natural language. These efforts span a spectrum of methodologies, ranging from image recognition systems to sensor-equipped gloves, each demonstrating remarkable successes in breaking down linguistic obstacles. Despite the advancements achieved thus far, the pursuit of seamless, universal sign language implementation remains a critical endeavor.

2 MATERIAL AND METHODS

The following subsections offers a description of the phases in which each of the subsystems implemented for the proper functioning of the glove prototype is determined.

2.1 Electronic circuit design

The placement of the sensors and the measurement variables is crucial, so they must be adjusted to the system's requirements. The correct selection of sensors, their placement, and the appropriate signal processing must be carried out in accordance with the necessary criteria in the LSC without disrupting the signal generation, thereby ensuring that the requirements are met.

Arduino Uno. The protocol implemented in the glove system corresponds to I2C communication. It is a synchronous protocol that uses only two wires, one for the clock (SCL) and one for the data (SDA). The code for preprocessing the data is implemented on this board. It maps the digital signals and scales them to the international system (m/s2 for acceleration and °/s for angular velocity). Finally, a low-pass filter is applied to eliminate the noise present in the signals.

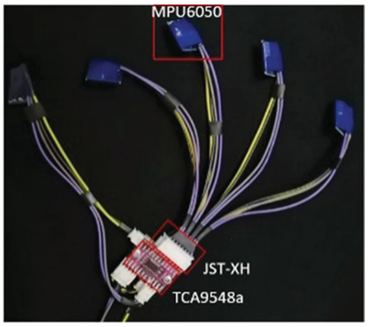
Raspberry Pi 3B. It is a small, single-board computer that functions as an embedded system. It stores algorithms developed in Python code for post-processing the data (normalized with z-scores), the neural network, and the corresponding classifier for each method. It also features a USB port that emulates serial communication for messages coming from the Arduino, eliminating the need to establish a configuration through code on the console. Finally, it is necessary to utilize the audio output available on the board to connect the peripheral (such as a speaker or baffle) by means of a 3.5mm plug cable in order to reproduce the sound of the words.

GPIO pins. GPIO stands for general purpose input/output, which refers to the general-purpose input/output pins that are part of the Raspberry Pi hardware. The prototype utilizes the GPIO pins to operate a LED, which serves as a display of the time window. During the sign-making process, the Arduino turns on the LED as soon as it acquires the data and keeps it on until the sign classification is complete. It then returns to its off state. This is done to indicate to individuals in the deaf-mute community when to make the sign.

MPU6050 sensor. The device is an inertial measurement unit (IMU) that combines a three-axis accelerometer and gyroscope. Communication is achieved through an I2C interface, which includes a 3.3V voltage regulator and pull-up resistors for direct use [8]. A sensor is placed in the distal phalanx of each finger of the prototype, enabling better information to be acquired during signing without interfering with movement.

TCA9548A module. This module contains 8 channels for the I2C bus, which help to multiplex the addresses of each sensor without conflict. By default, the TCA9548A module has the address 0x70 enabled. Since 2 modules are needed (1 for each hand), it is necessary to enable the next address (0x71).

Figure 1a depicts the construction of the prototype of the right glove, illustrating the connections of the 5 MPU6050 sensors to be positioned on each finger of the hand. Additionally, it shows the TCA4598a multiplexer for signal integration and the JST connectors for connecting the various elements in a generalized manner. Figure 1b depicts the complete prototype gloves (right and left) and also displays the LED (time window display) and kangaroo, which internally house the Arduino, Raspberry Pi, and audio system (speaker).

a) 

b) 

Fig. 1. Electronic circuit a) Right glove component assembly, b) LSC translator gloves

2.2 Signal processing

Once the data is acquired by the Arduino Uno, signal characterization must be performed through data processing.

Mapping. To visually understand and interpret the data, the signal values are converted to the International System of Units: °/sec for angular velocity and m/s2 for acceleration. This approach enables us to comprehend the signal peaks, their behavior, and whether the variations align with the expected movement, as well as identify errors in the data acquisition. This stage marks the beginning of data preprocessing and serves as the starting point for generating the database and training the neural network.

Emma filter. The MPU6050 sensor has internal filters for each variable, but they can still be affected by noise generated by external factors. Considering that the signals contain high-frequency noise, a low-pass filter is designed to obtain a filtered value from a measurement by applying equation (1), where the input signal M is represented by, A_n the output signal is represented by, and α is a constant between 0 and 1 [9]. This filter removes data spikes caused by noise that may interfere with data analysis.

$$A_n = \alpha M + (1 - \alpha)A_{n-1} \tag{1}$$

Z-score normalized. Signal normalization is achieved by applying a standardization method using z-scores. This is based on the analysis of the distance of each value from the mean (μ) by expressing it in units of the standard deviation (σ) [10].

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

2.3 Classification techniques

Once the signals have been properly processed, it is essential to extract descriptors that capture the most relevant characteristics of each signal in smaller and more representative values in order to facilitate the neural network algorithm's learning process.

Database. For the creation of the database, a member of the deaf-mute community assisted in accurately executing each sign. A database was created using the information collected by the sensors for the following five phrases: "Good morning," "How are you?" "Thank you," "Hello," and "Please." To gather information about the location, dispersion, and other behavioral patterns of the data, statistical measures such as mean, standard deviation, interquartile range (IQR), skewness, and kurtosis were utilized to capture the structure of the data [12] [13]. These are representative values of a collection of data and summarize the information of the entire dataset in a few key values. Statistical measures provide information about the central tendency, variability, and other patterns in the data, enabling a quick understanding of its structure. It is possible to interpret that the flat section in the accelerations, influenced by gravity, may affect the mean and make it unrepresentative. This statistic, while being a measure of central tendency, is not sensitive or unrepresentative when there are no extreme values in the data set. This is because the variations in acceleration when a movement is made are significantly greater than the influence of gravity on each axis. Therefore, an accurate measurement is obtained.

Principal component analysis. To reduce the computational cost and complexity of the neural network, as well as to eliminate irrelevant features that may negatively impact the model's performance, the principal component analysis (PCA) method is utilized. It should be noted that while this method simplifies the complexity of sample spaces with many dimensions, it does not imply that these complexities can be eliminated directly [14] [15]. This is because each principal component is calculated using linear combinations of all the variables involved in the system. Therefore, excluding a variable could result in a significant loss of information for the principal components. On the other hand, it is necessary to standardize the data set used with this method to prevent certain variables from carrying more weight than others. If this occurs, it is possible that the number of principal components describing the system may not be accurate. Finally, the acceptable number of principal components depends on the specific application. For descriptive purposes, it may be sufficient to explain 80% of the variance. However, if you intend to conduct further analyses with the data, it is advisable for the principal components to account for at least 90% of the variance. This is directly linked to the percentage of classification in the neural network.

Neural network. A neural network was trained for identification and classification [16]. This structure consists of two layers: a hidden layer with a hyperbolic tangent activation function, and an output layer with a softmax activation function. The training algorithms of the proposed neural network are based on back propagation, which enables the propagation of errors backward through the neural network, and a scaled conjugate gradient to accelerate convergence. For the neural network training, 140 samples were created for each class, resulting in a total of 700 patterns for the training process. With the development of the database and in consideration of the Pareto principle, which states that 70% of the consequences stem from 30% of the causes, a split of 70% for the training group and 30% for the validation set is defined. The validation set accumulates most of the variance in the system.

3 RESULTS AND DISCUSSION

Signal mapping and filtering are programmed and implemented on the Arduino Uno board, while normalization with Z-scores is programmed in Python and implemented on the Raspberry Pi. It is important to keep in mind that, due to the variety of phrases and words in LSC, it is necessary to use both hands to differentiate between them. Figure 2 illustrates the transformation of the data through each process by consistently performing the Good Morning, Thank You, and Hello signs, respectively.

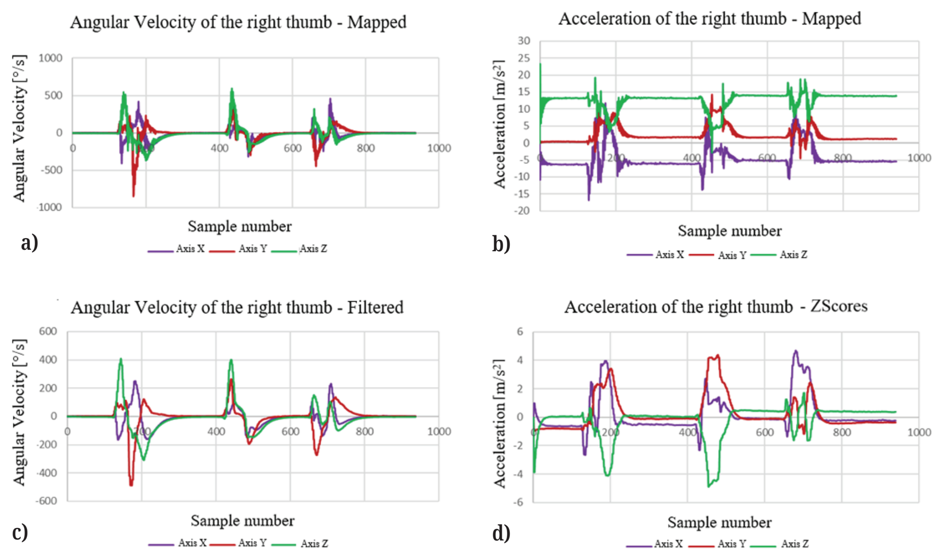


Fig. 2. *(Continued)*

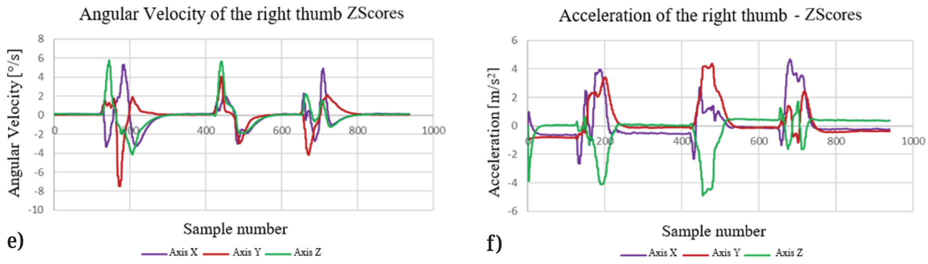


Fig. 2. Result of the processing stage of the information recorded by the right thumb: speed signal and acceleration signal

The signals indicate that gravitational forces inherently affect at least one of the three axes in the acceleration variable. This phenomenon is derived from the use of MPU6050 sensor filaments, which detect acceleration or angular velocity caused by rotational motion along any of the three axes. The resulting excitation and interaction of the filaments are converted into digital signals. Mitigation strategies exist to counteract the effects of gravity on the sensors, thereby facilitating the extraction of net movement values. One such approach involves calculating the tilt angle. By utilizing this approach and the initial axis-specific gravity measurements, corrective adjustments and offsets can be implemented for each variable. While this approach offers a partial solution, it introduces an inherent challenge. Since integrating angular velocity is necessary for determining tilt angle, any signal noise leads to increased error propagation through integration—a phenomenon termed drift [11]. Consequently, as time passes, the recorded data diverges further from the actual data. Considering these implications and pragmatic project execution, the decision has been made to utilize signals that include gravitational influence. This indicates that when the MPU6050 sensor is at rest, gravity-driven acceleration affects at least one of the axes. However, during active signal transmission or movement, recorded accelerations correspond exclusively to the intended movement.

When distinguishing between the signs for “Good morning” and “Thank you,” a clear pattern emerges. Both acceleration and angular velocity variables show fluctuations for both the left and right hands. This trend is consistent with the nature of these signs, which involve bilateral hand movements. Conversely, when evaluating the gestures associated with the word “Hello,” significant differences are only evident in the signals made by the right hand. This difference highlights the importance of the specific movements used to convey the “Hello” sign, a gesture that involves only the right hand. The relatively subtle fluctuations in the data from the left hand, although appearing to indicate a neutral zone, remain analytically relevant. They contribute to identifying and differentiating signs by relying on unilateral hand movements, thereby facilitating more effective interpretation and classification of these gestures.

Once the database has been generated, the next step is to train the neural network. In order to achieve acceptable classification percentages, 30 neurons are defined in the hidden layer. Numerous neurons can cause the network to memorize the patterns used during training, leading to a loss of the ability to generalize and classify patterns that are not part of the dataset. However, if the number of neurons is too low, the capacity for generalization increases, but the ability to differentiate between different patterns decreases. The classification system includes a SoftMax layer, which produces probabilities for each of the five words in the algorithm’s training set.

The expression with the highest likelihood is then processed using text-to-speech conversion algorithms, allowing it to be transformed into understandable speech. The converted speech is then played through the designated audio output device.

To facilitate this validation process, a specific time window for data acquisition was carefully determined. This section considers the average duration required for executing a sign gesture as well as the temporal resources consumed during processing and classification within the neural network (refer to Table 1). Spanning a duration of seven seconds, this time window encompassed data acquisition, signal processing, and eventual audio translation. Notably, an LED indicator visually marked the beginning and end of this specific time window. This feature aims to allow users to synchronize their sign gestures with the precise data acquisition window, ensuring that the data entering the neural network contains comprehensive information necessary for accurate translation.

Table 2 presents the results of the experimental validation stage. One person used sign language to replicate the five words, repeating each one 20 times. The variation in the results obtained for each word using the neural network-based classification system can be attributed to several interconnected factors. First, the inherent complexity of each sign language word may affect the neural network's ability to distinguish distinct patterns. Words with distinct and easily detectable movements may achieve higher accuracy rates, while words with similar movements may pose challenges for the neural network in terms of discrimination. Another crucial factor is the quantity and quality of the training dataset. Words that appear more frequently in the training dataset will have a more pronounced representation in the neural network, thereby enhancing its ability to accurately classify them. Conversely, if a word has limited representation in the training dataset, the neural network may struggle to generalize accurately and achieve a high level of accuracy. The architecture and parameters of the neural network also play a crucial role. A deeper neural network or one with more neurons in specific layers may capture more subtle and complex features, potentially improving accuracy on more challenging words. Moreover, the hyperparameters of the network, including learning rates and activation functions, impact its capacity to accurately fit the data and generalize effectively. Ultimately, the intrinsic nature of machine learning leads to some variability in the results. Input data may contain noise or uncertainty, which can impact accuracy. Furthermore, random initialization of the network weights can result in various optimal solutions during different training sessions, leading to discrepancies in the observed results. To enhance and standardize the results, it is crucial to consider expanding and diversifying the training dataset, adjusting the network parameters, applying regularization techniques, and conducting thorough analyses to identify and address specific constraints that impact the accuracy of each word.

Table 1. Execution time of each of the algorithm stages

Algorithm	Processing Time (s)
Arduino readout	1.5764
Neural network	0.0159
Text to speech	0.6013
Player	0.1865
Total	2.3801

Table 2. Accuracy of the classification system

Classes	% Classification
Good morning	96.23%
How are you	70.11%
Thank you	96.51%
Hello	85.48%
Please	96.53%
Average	88.97%

4 CONCLUSION

This paper presents a comprehensive system designed for LSC recognition, processing, and translation. The core of this system is a neural network structured as a pattern recognition model. The neural network was trained using statistics from the velocity and acceleration signals of hand finger movement, achieving an average classification rate of 88.97%. This result is primarily due to the inherent nature of sign language development, where the articulation of each word involves the use of both hands. This intervention provides information that enables the extraction of a greater number of variables, thereby facilitating the derivation of essential descriptive distinctions necessary for pattern identification. The presented results demonstrate performance similar to that of methods that incorporate image recognition techniques. However, these methods often lack accessibility and portability. In contrast, the prototype developed in this study yields favorable results by prioritizing ease of use and simple sensor distribution. Similarly, while glove-based methods may produce high levels of noise in the captured signals, it is evident that the primary focus should be on developing algorithms for signal processing, descriptor extraction, and data analysis. The project is a crucial step in the development of LSC translation systems, as it combines the measurement of motion variables, signal processing, and artificial intelligence algorithms. The preliminary results indicate positive progress towards achieving this general objective.

5 REFERENCES

[1] R. Kushalnagar, "Deafness and hearing loss," in *Web Accessibility. Human–Computer Interaction Series*, Y. Yesilada and S. Harper, Eds., Springer, London, 2019, pp. 35–47. https://doi.org/10.1007/978-1-4471-7440-0_3

[2] M. Korban and M. Nahvi, "An algorithm on sign words extraction and recognition of continuous Persian sign language based on motion and shape features of hands," *Pattern Analysis and Applications*, vol. 21, pp. 323–335, 2018. https://doi.org/10.1007/s10044-016-0579-2

[3] Y. Xue S. Gao, H. Sun, and W. Qin, "A Chinese sign language recognition system using leap motion," in *International Conference on Virtual Reality and Visualization (ICVRV)*, Zhengzhou, China, 2017, pp. 180–185. https://doi.org/10.1109/ICVRV.2017.00044

[4] S. Suharjito, H. Gunawan, N. Thiracitta, and A. Nugroho, "Sign language recognition using modified convolutional neural network model," in *Indonesian Association for Pattern Recognition International Conference (INAPR)*, Jakarta, Indonesia, 2018, pp. 1–5. https://doi.org/10.1109/INAPR.2018.8627014

[5] N. Hardiyanti, A. Lawi, Diaraya, and F. Aziz, "Classification of human activity based on sensor accelerometer and gyroscope using ensemble SVM method," in *2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, Makassar, Indonesia, 2018, pp. 304–307. https://doi.org/10.1109/EIConCIT.2018.8878627

[6] Y. Mori and M. Toyonaga, "Data-glove for Japanese sign language training system with gyro-sensor," in *Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, Toyama, Japan, 2018, pp. 1354–1357. https://doi.org/10.1109/SCIS-ISIS.2018.00211

[7] L. Ji, J. Liu, and S. Shimamoto, "Recognition of Japanese sign language by sensor-based data glove employing machine learning," in *IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech)*, IEEE, 2022, pp. 256–258. https://doi.org/10.1109/LifeTech53646.2022.9754851

[8] V. T. Dau, T. Otake, T. X. Dinh, and S. Sugiyama, "Design and fabrication of convective inertial sensor consisting of 3DOF gyroscope and 2DOF accelerometer," in *TRANSDUCERS 2009 – 2009 International Solid-State Sensors, Actuators and Microsystems Conference*, Denver, CO, USA, 2009, pp. 1170–1173. https://doi.org/10.1109/SENSOR.2009.5285911

[9] B. Alexander, T. Ivan, and B. Denis, "Analysis of noisy signal restoration quality with exponential moving average filter," in *International Siberian Conference on Control and Communications (SIBCON)*, Moscow, Russia, 2016, pp. 1–4. https://doi.org/10.1109/SIBCON.2016.7491750

[10] N. Costadopoulos, M. Z. Islam, and D. Tien, "Using Z-score to extract human readable logic rules from physiological data," in *11th International Conference on Knowledge and Systems Engineering (KSE)*, Da Nang, Vietnam, 2019, pp. 1–6. https://doi.org/10.1109/KSE.2019.8919473

[11] H. J. Luinge, P. H. Veltink, and C. T. M. Baten, "Estimation of orientation with gyroscopes and accelerometers," in *Proceedings of the First Joint BMES/EMBS Conference. 1999 IEEE Engineering in Medicine and Biology 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society*, Cat. N, Atlanta, GA, USA, vol. 2, 1999, p. 844. https://doi.org/10.1109/IEMBS.1999.803999

[12] G. A. Susto, A. Cenedesec, and M. Terzi, "Time-series classification methods: Review and applications to power systems data," *Big Data Application in Power Systems*, 2018, pp. 179–220. https://doi.org/10.1016/B978-0-12-811968-6.00009-7

[13] Y. Umeda, "Time series classification via topological data analysis," *Information and Media Technologies*, vol. 12, pp. 228–239, 2017. https://doi.org/10.1527/tjsai.D-G72

[14] B. M. S. Hasan and A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, 2021. https://doi.org/10.30880/jscdm.2021.02.01.003

[15] V. Bruni, M. L. Cardinali, and D. Vitulano, "A short review on minimum description length: An application to dimension reduction in PCA," *Entropy*, vol. 24, no. 2, p. 269, 2022. https://doi.org/10.3390/e24020269

[16] V. H. Benitez, "Pattern classification and its applications to control of biomechatronic systems," in *Artificial Neural Networks for Engineering Applications*, 2019, pp. 139–154. https://doi.org/10.1016/B978-0-12-818247-5.00020-4

[17] Y. Liang, "Intelligent emotion evaluation method of classroom teaching based on expression recognition," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 4, pp. 127–141, 2019. https://doi.org/10.3991/ijet.v14i04.10130

[18] A. Y. Saleh and L. H. Chern, "Autism spectrum disorder classification using deep learning," *International Journal of Online and Biomedical Engineering*, vol. 17, no. 8, pp. 103–114, 2021. https://doi.org/10.3991/ijoe.v17i08.24603

[19] P. G. Veronica, R. K. Mokkapati, L. P. Jagupilla, and C. Santhosh, "Static hand gesture recognition using novel convolutional neural network and support vector machine," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 9, pp. 131–141, 2023. https://doi.org/10.3991/ijoe.v19i09.39927

6 AUTHORS

Hernando Gonzalez received the B.E., degree in Electronic Engineering from UIS, Master in Electronic Engineering from UIS, student of Doctorate in Engineering from UNAB, currently is an Associate Professor of the Mechatronics Engineering program at UNAB and member of GICYM, His research interests include machine learning, control systems, signal treatments, robotics, electronics and mecatronics devices (E-mail: hgonzalez7@unab.edu.co).

Silvia Hernández received the B.E., degree in Mechatronics engineering from UNAB. She is an engineer with ability to design and implement automatic control systems, automate process, develop applications on the signal processing, machine learning, and data mining (E-mail: shernandez865@unab.edu.co).

Oscar Calderón received the B.E., degree in Mechatronics engineering from UNAB. Oscar is an engineer with ability to design and implement automatic control systems, automate process, develop applications on the signal processing, machine learning and data mining field. With knowledge in pneumatics, hydraulics, robotics, power electronics, and artificial intelligence. He has experience in variable speed driver maintenance used in oil and gas industry for controlling the speed of installed motors in artificial lift units (E-mail: ocalderon337@unab.edu.co).