International Journal of Artificial Intelligence,
ISSN 0974-0635; Int. J. Artif. Intell.
2025, Volume 23, Issue Number 1

System for Learning Kinetic Communication Patterns in Patients with Cerebral Palsy Based on Deep Learning

P. A. Sánchez-Sánchez[1], **J. R. García-González**[2] and **L. A. Ávila Gutierrez**[3]

[1, 2, 3] Facultad de ingenierías, Universidad Simón Bolívar, Barranquilla, COLOMBIA
Email: paola.sanchezs@unisimon.edu.co, jose.garciag@unisimon.edu.co

ABSTRACT

This paper presents the design and development of a system aimed at improving communication for patients with cerebral palsy through the application of deep learning techniques. The system's components—including a video/image repository, a user interaction tool, and a kinesic language classifier—are meticulously defined and described. The system allows users to upload and evaluate videos that capture kinesic communication actions such as gestures, signs, and other forms of non-verbal communication. These actions are identified by pre-trained convolutional neural network models, providing a robust mechanism for recognizing and interpreting patient behavior. The study evaluates the performance of several deep learning models, including VGGNet, ResNet, and ResNetCRNN, within the kinesic language classifier component. Results indicate that the ResNetCRNN model outperforms the others in terms of accuracy and loss percentage, particularly due to its capability to assess sequences of related images rather than independent frames. The findings suggest that the proposed system offers a viable alternative for facilitating communication in cerebral palsy patients by analyzing and identifying their actions. The system's strength lies in its ability to learn from each patient, making it a constructive and adaptable communication tool.

Keywords: Cerebral palsy, deep learning, kinesic language recognition, non-verbal communication, convolutional neural networks, ResNetCRNN, gesture recognition.

Mathematics Subject Classification: 68T07, 68T45

1. INTRODUCTION

Cerebral palsy is a condition arising from a non-progressive, permanent brain injury that occurs during the brain's developmental phase, primarily impacting the central nervous system and resulting in alterations in muscle tone and motor response. This injury can occur prenatally (during pregnancy), perinatally (during childbirth), or postnatally (within the first six months of life). As a result, cerebral palsy is often classified as a form of intellectual disability (Camacho-Salas, et al., 2007); (Gulati & Vishal, 2018); (Evans, 1948).

Cerebral palsy is recognized as the most prevalent cause of disability among children in many countries (Aspace, 2014). Although there is currently no cure for cerebral palsy, it is important to understand that affected individuals are at a heightened risk for a broad spectrum of speech and language disorders. These disorders can vary significantly in severity, ranging from relatively straightforward speech issues such as dysarthria—a motor speech disorder—to more complex language pathologies, including agnosia, which involves the inability to process sensory information, and dysphasia, characterized by impaired word coordination (Busto Barcos, 1988); (Ayala, 2012); (Gulati & Vishal, 2018); (Pennington, 2008); (Kuban & Leviton, 1994).

Patients diagnosed with cerebral palsy often experience partial or complete disruption of the psychophysical mechanisms required for speech, significantly impairing their ability to communicate. This impairment is frequently due to motor dysfunctions that prevent the precise muscular movements necessary for producing intelligible speech (De Saussure, 1994; Riofrio & Carolina, 2015). Moreover, both orofacial and general motor disorders further limit their capacity to utilize these mechanisms effectively, thereby hindering the articulation of the complex combinations that form speech. As a result, the development of their linguistic communication competence is directly compromised (Novacheck, Trost, & Sohrweide, 2010); (Saville-Troike, 2008).

Cerebral palsy is a complex condition that necessitates a multidisciplinary approach, with each specialty contributing to the development and reeducation of the patient's communicative competencies, ultimately improving their quality of life. One critical aspect of these competencies is kinesics, defined by Ray Lee Birdwhistell (1952) as the branch of study focused on gestures, encompassing everything from facial expressions and body movements to signs, eye movements, head tilts, and other physical manifestations involved in communication (Baños, 2012); (Aripin & Noor, 2019).

In patients diagnosed with this condition, kinesic competence is of paramount importance, regardless of the individual's ability to speak. In many instances, these patients develop this competence in an abstract manner, often creating their own unique language code (Gulati & Vishal, 2018); (Gumperz & Hymes, 1972).

To effectively support the development and reeducation of the patient's communicative competencies, specialized care in dedicated centers and assistance within educational institutions are essential. This care may take place in special education settings designed for cerebral palsy or in mainstream schools if the child can be integrated. The choice depends on the degree of impairment and the severity of the condition. Without proper treatment, there is a significant risk of regression in the development of daily life activities over time, as well as the potential for new impairments that could further impede the child's progress. (Busto Barcos, 1988); (Morales Vaquero, 2017).

Similarly, the severity of cerebral palsy can vary widely among patients. In some, the condition is barely noticeable, while in others, it can result in significant impairments, necessitating ongoing support for daily activities. Although cerebral palsy cannot be cured, patients diagnosed with this condition can improve their quality of life through appropriate interventions. These interventions aim to enhance motor skills, promote cognitive development, improve communication abilities, and stimulate social interaction. (Chew Ka, Iacono, & Tracy, 2009).

Treatment and resources for addressing language disorders in patients with cerebral palsy are diverse. They encompass various approaches, from language re-education methods such as Tardieu, Bobath, and PODD, to the use of augmentative and alternative communication (AAC) systems (Ayala, 2012). The primary aim of AAC is to enhance (augmentative) and/or support (alternative) communication and expression difficulties experienced by individuals with disabilities. AAC employs symbolic and graphic methods, including pictures, symbols, and letters; gestural methods, such as signs and manual signals; and supplementary support services (Román, 2017); (Ayala, 2012); (Baños, 2012).

However, these tools may not be universally applicable to all types of cerebral palsy due to the variability in associated conditions (Pennington, 2008). Additionally, current technological solutions are constrained by the requirement for patients to have a certain degree of motor control and/or cognitive abilities to associate images with their meanings. This limitation excludes individuals who lack these capabilities and overlooks the critical role of kinesic communication skills. Developing these skills is essential for effective communication, particularly for individuals with severe speech disorders within this population (Birdwhistell, 1952); (Busto Barcos, 1988); (Abril Abadín, Delgado Santos, & Vigara Cerrato, 2012); (Aripin & Noor, 2019).

The identification of non-verbal communication forms, classified under each patient's kinesic language, requires mechanisms capable of capturing and translating a variety of gestures, facial expressions, body movements, signs, and other non-verbal cues into a format comprehensible to the patient's interlocutors. In this context, artificial intelligence techniques, specifically deep learning methods using convolutional neural networks (CNNs), present a promising solution. CNNs are highly versatile for identifying features and behavior patterns in images and videos (Sánchez-Sánchez, García-González, & Rúa Ascar, 2020).

CNNs typically comprise numerous processing layers organized into blocks that iteratively combine and refine data to reveal hidden features and patterns. Their efficacy has been demonstrated across various healthcare applications. For instance, CNNs have been used to detect cancer by analyzing images from ultrasounds, magnetic resonance imaging (MRI), and nuclear medicine, thereby supporting medical diagnoses and identifying patterns or anomalies that might elude human observers (Hinton, 2018). Similarly, CNNs are employed to prevent heart attacks by training the network on echocardiogram images to detect potential issues, which aids in early diagnosis and prevention (Hinton, 2018; Miotto, Fei, Shuang, Xiaoqian, & Dudley, 2018).

Moreover, CNNs have been instrumental in the early detection of neurological disorders such as Alzheimer's, Parkinson's, epilepsy, and migraines by analyzing MRI scans to recognize disease-specific patterns, thus supporting diagnosis and management. They are also used to identify ocular diseases, including glaucoma, retinopathy, and macular edema, through the analysis of retinal images (Hinton, 2018; Miotto, Fei, Shuang, Xiaoqian, & Dudley, 2018). Recently, CNNs have been applied to detect COVID-19 cases by analyzing radiographic images, enhancing the ability to identify and diagnose the disease promptly (Panwar, Gupta, Mohammad, Morales-Menendez, & Vaishnavi, 2020).

The objective of this article is to design a system that enhances communication for patients with cerebral palsy using convolutional neural network models to identify and learn kinesic communication patterns from videos. The system is designed with the following capabilities:

1. Input Processing: Accepts videos showcasing kinesic communication actions performed by patients.
2. Learning and Classification: Analyzes the non-verbal communication actions of each patient, learning from these interactions, and classifies the actions accordingly.
3. Action Notification: Provides feedback to users regarding the identification of actions performed by the patient.

The proposed system seeks to address the fundamental question: "How can we effectively promote the development of communication skills in children with cerebral palsy?"

2. PROPOSED METHODOLOGY

The proposed strategy aims to specify a system to enhance the kinesic communication of patients with cerebral palsy. This approach considers a systemic and systematic process, the initial definition and declaration of which align with the framework of a research project (García-González & Sánchez-Sánchez, 2020). This strategy is based on the design of a system that encompasses the interaction of three components: a repository of videos/images, a user interaction tool, and a kinesic language classifier.

The user interaction tool enables the recording of kinesic communication events and provides classification notifications to the interlocutor. Through this interaction tool, videos/images suitable for analysis are continuously collected, which in turn feed the repository of videos/images. The images are analyzed using the kinesic language classification component, which employs deep learning techniques based on convolutional neural networks to establish patterns that facilitate the determination of what the patient is communicating through gestures, body language, or signs. It also learns new patterns and updates the video repository with the classifications made.

2.1. Video/Image Repository

To operate the system, it's essential to have a repository of videos/images containing instances of kinesic communication events and their respective classifications (meanings). Starting from the premise that a video is a sequence of related images and sounds, this proposal utilizes the visual component by breaking down videos into multiple images, each of which represents kinesic communication events of patients with cerebral palsy.

Each video represents at least one event, which is associated with a classification. For example, a series of gestures and signs made by the patient to request food can be classified as hunger.

2.2. User Interaction Tool

The primary function of this component is user interaction, necessitating a user-friendly interface that facilitates user engagement and enables the capture of new videos for analysis. Additionally, real-time video analysis and device portability are required functionalities.

Considering the above, we propose the development of a mobile and portable app with the following core modules:

- Home, registration, and login module
- Video capture module
- Video exploration module
- Video analysis module
- Recommendation module

Access to the app requires prior user registration through the application and a login. The video capture module allows registered users to interact with the application by capturing videos directly within the app or uploading them from a local source. Users can access their video history and the respective classifications suggested by the system through the video exploration module. Furthermore, for new video captures, users can initiate an analysis, activating the kinesic language classifier, which provides the corresponding event classification and updates the classifier's video repository.

The recommendation module offers interactive and enjoyable tools based on the classified events, which can include activation/relaxation exercises, videos, songs, etc.

2.3. Kinesic Language Classifier

This component serves as the core of the system, as it is responsible for receiving new video captures, analyzing the event's classification, providing feedback to the user, and updating the video repository.

The fundamental goal of a classification system is to predict the class or label of a new instance based on past observations (Sánchez-Sánchez, et al. 2019). In the context of image processing, the objective is to recognize a visual pattern based on previous classifications of videos/images in a repository.

As previously mentioned, the literature has increasingly embraced knowledge of deep learning models based on convolutional neural networks, providing sufficiently robust and highly efficient pre-trained models for visual pattern recognition. However, these applications are often focused on static images rather than videos.

In this proposal, we employ pre-trained models of convolutional neural networks to process images generated from video fragments in the search for kinesic communication patterns (non-verbal).

The systemic model of the classifier is depicted in Figure 1. The source code of the model is written in Python and can be accessed in the public Git repository (Avila Gutierrez, 2020).

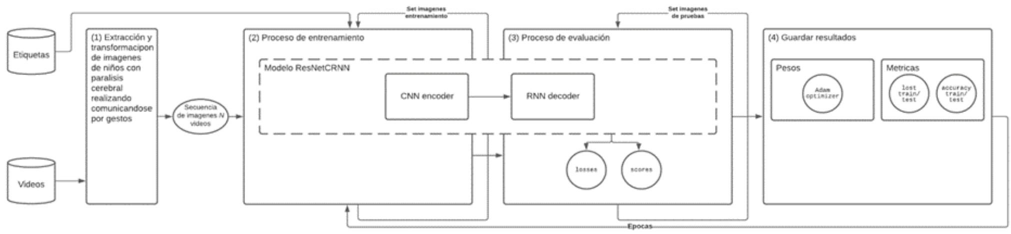

Figure 1. Diagram of the kinesic language classifier.

In the proposed convolutional neural network model, we begin with a set of videos obtained from public databases featuring humans performing actions, as well as some videos of cerebral palsy patients making communication gestures. These videos will be used throughout the training process.
In Step 1, the videos undergo an extraction and transformation process, resulting in images that identify the image blocks corresponding to the N videos in the dataset. These images will be distributed randomly in the subsequent steps, as depicted in Figure 2.

Process 2 involves training, where the neural network is provided with images in a supervised manner, indicating the actions performed in each image through labels. This enables the model to subsequently encode the image and learn from the supplied labels. As a result, it becomes capable of classifying the event depicted in the image, decoding it, and generating the image correctly, as illustrated in Figure 3.

In step 3, tests are conducted with new data that the model has not encountered during training, and without providing the labels that indicate the actions presented in these videos. The model's weights are updated in each iteration in the same manner as before.

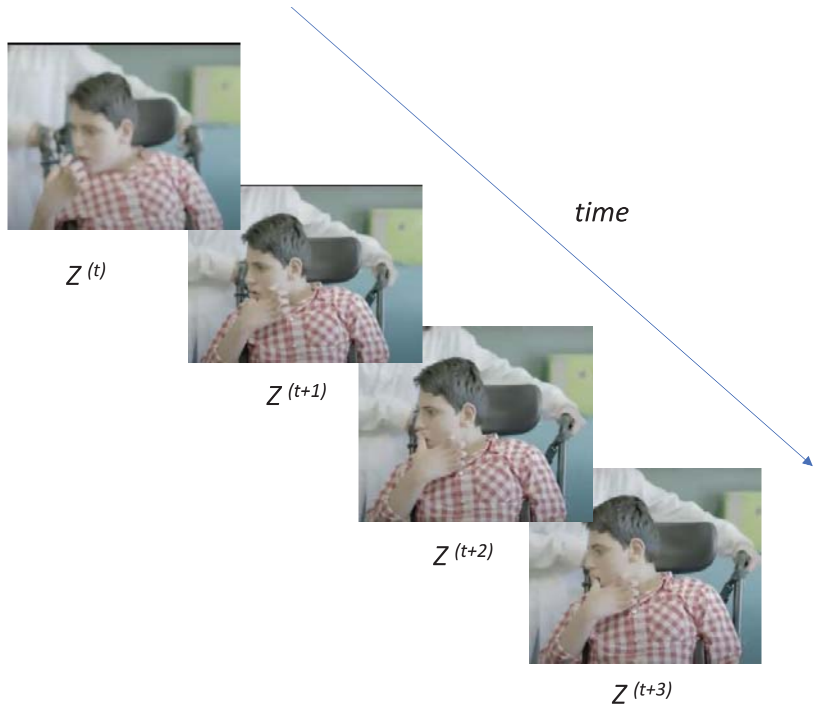

Figure 2. Sequential Image Flow in a Video. Source: Taken from the video on the website https://aspace.org/ Cerebral Palsy 168 Hours, Documentary (min 10:16)

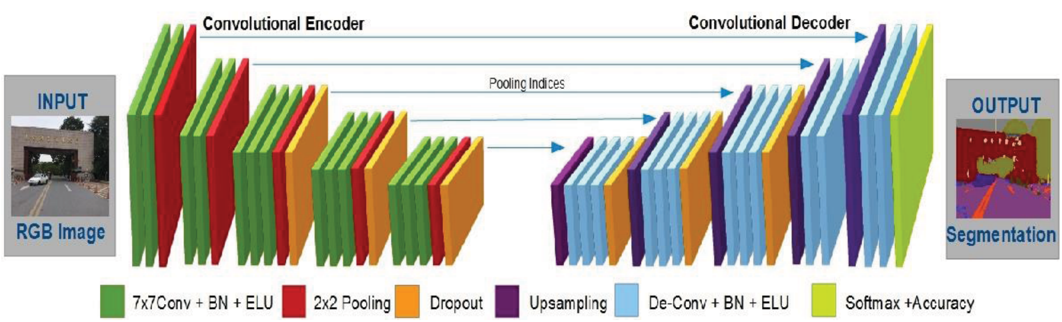

Figure 3. Example of encoding and decoding of inputs in an RNN. Source: (Ramírez & Castaño, 2007)

In step 4, the results of the model's execution, which include the final weights, are saved, and performance measures such as the accuracy percentage and model loss are calculated. A lower loss indicates a better model. The loss is calculated during training and testing, and its interpretation relates to how well the model performs on these two sets. It is a sum of the errors made for each example.

Accuracy, on the other hand, refers to the proportion of correct classifications relative to the total number of examples in the training and testing sets.

The trained model can be used to predict classifications for new videos or retrained with new data.

3. RESULTS OF THE KINESIC LANGUAGE CLASSIFIER AND DISCUSSION

This article's focus is on the development of the kinesic language classifier, omitting the components of the video/image database and the user interaction tool.

To test the proposed kinesic language classifier scheme, the system is fed with a dataset of videos sourced from a public database (University of Central Florida, 2021). These videos feature humans performing actions, as well as videos of cerebral palsy patients making communication gestures, along with our own collected videos. The video/image database obtained is used throughout the training and testing processes.

During the training process, various pretrained convolutional neural network models were considered. After assessing the strengths and limitations of the models presented and considering that the primary challenge addressed in this work is the classification of human actions in videos, a selection of three models was made:

- VGGNet: Very Deep Convolutional Networks for Large-Scale Image Recognition. Karen Simonyan, Andrew Zisserman/USA/2015 (Simonyan, 2014).
- ResNet: Deep Residual Learning for Image Recognition. Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun/USA/2015 (He, 2016).
- ResNetCRNN: Stochastic Fusion for Multi-stream Neural Network in Video Classification. Yu-Min Huang. Huan-Hsin Tseng, Jen-Tzung Chien / Taiwan – USA / 2019 (Huang, Tseng, Chien, & Chien, 2019).

The selected models were tested for comparison with the video/image dataset. The common parameters for these experiments were established as shown in Table 1.

Table 1: Experiment Parameters

Parameters	Values
Input Videos	13,350
Epochs	150
Frame Rate	5
Target Image Size	224
Categories	101
Learning Algorithm	Adam
Metrics	Precision/Loss

Hidden Layers	VGGNet (19), ResNet (50), ResNetCRNN (512)
Output Neurons	101
GPU	1 Tesla K80

In all cases, the dataset was divided into two sets: training and testing. The training set comprised 80% of the videos, while the testing set contained the remaining 20%. This division was implemented to prevent overfitting, following the methodology of Sánchez-Sánchez and García-González in their work "Autoregressive moving average recurrent neural networks applied to the modeling of Colombian exchange rate" (2018). The number of hidden layers in the experiments varied depending on the model, following recommendations from relevant literature.

Table 2 presents the performance measures calculated for each model, enabling a comparison between them.

Table 2: Experiment Results

Model	Accuracy (%)	Loss (%)
VGGNet	44.8	64.99
ResNet	22.4	45.6
ResNetCRNN	84.6	10.4

Upon analyzing the results, the ResNetCRNN model demonstrates the highest performance, achieving an accuracy rate exceeding 80%. This result indicates that the proposed neural network model is proficient in accurately classifying 8 out of 10 human action videos and providing the action's name from a defined set of 101 categories in textual format.

The superior performance of the ResNetCRNN model can be attributed to its unique characteristics compared to other models. Originally developed for image classification, ResNetCRNN distinguishes itself by processing the entire sequence of images within a video, rather than treating each frame as isolated data. This capability allows it to effectively capture the temporal flow of images and understand the action performed across the entire sequence, rather than just within a single frame.

It is important to acknowledge that human action videos exhibit considerable variability. In this study, 101 categories were utilized, but this number may vary depending on the specific communication modes and events identified for each patient. As a result, periodic retraining of the model is necessary whenever substantial changes occur in the video/image dataset. Additionally, the ResNetCRNN model incurs higher computational costs compared to other models tested, due to the increased number of hidden layers required, leading to longer experiment runtimes.

The computational cost of the ResNetCRNN model is a significant consideration. Although it is currently manageable, exploring Transfer Learning techniques is essential to mitigate these costs and prevent an exponential increase in computational demands.

4. CONCLUSIONS

Individuals with cerebral palsy deserve equal opportunities for the development of their communicative competencies as those without disabilities. It is crucial to advance research that fosters the development of kinesic communication skills in these patients, particularly given that existing technological tools often fail to address non-verbal communication as a viable alternative for this population.

Addressing these needs can help overcome the economic and accessibility barriers that current technologies impose on individuals with cerebral palsy. This includes those with limited physical capabilities and those unable to afford services designed to enhance communicative skills, ultimately improving their quality of life (Castellanos & Castellanos, 2007; Canale & Swain, 1980; Savignon, 2018).

Family members and caregivers are integral to the rehabilitation process, often serving as interpreters and providing motor support, which contributes to the patient's social inclusion. During school years, communicative deficits become more pronounced as patients attempt to integrate socially, often revealing gaps in their ability to communicate independently. Thus, establishing a communication code that is accessible to all individuals in the patient's environment—beyond just primary caregivers and professionals—is essential (Aspace, 2014; Canale & Swain, 1980; Pennington, 2008; Savignon, 2018).

The pursuit of new technological solutions to assist in the treatment and reeducation of patients with cerebral palsy is of paramount importance. This is vital for enhancing communicative competence, improving quality of life, increasing autonomy, and reducing barriers to social inclusion, regardless of the severity of the condition (Ayala, 2012).

The complexity and variability of communication processes in individuals with cerebral palsy highlight the need for designing and implementing systems that can supplement or replace kinesic communication competencies using emerging technological tools, as proposed in this study.

Our holistic methodological approach allowed for an in-depth exploration of the communication process in patients with cerebral palsy. This comprehensive perspective addressed developmental aspects and facilitated a thorough examination of each component's functioning and its interactions with the environment. The insights gained from this study pave the way for an integrative framework that could benefit future research and advancements in this field.

5. REFERENCES

Abril Abadín, D., Delgado Santos, C. I., & Vigara Cerrato, Á., 2012, *Comunicación aumentativa y alternativa*, Guía de referencia (3 ed.). Madrid: CEAPAT.

Aripin, N., & Rahmat, N., 2019, Non-verbal communication and writing: exploring writing anxiety in the writing process through kinesics. International Innovation in Teaching and Learning & Language Education Conference, *E-Proceeding LEC 2019.*

Aspace, C., 2014, *Descubriendo la parálisis cerebral.* Confederation Aspace, In : https://www.guiadisc.com/wp-content/uploads/Descubriendo-la-paralisis-cerebral.pdf

Avila Gutierrez, L., 2020. *Github Video-Classification,* In: https://github.com/zluis1992/video-classification

Ayala, J., 2012, La Competencia Comunicativa en Personas que Presentan Parálisis Cerebral, *ReiDoCrea: Revista electrónica de investigación y docencia creativa* **1**, 158-163.

Baños, R., 2012, Los aspectos no verbales en la comunicación intercultural, *Ra Ximhai*, **8** (2), 223-239.

Birdwhistell, R., 1952, *Introduction to kinesics : an annotation system for analysis of body motion and gesture,* University of Michigan Library.

Busto Barcos, M.D., 1988, *Reeducación del Habla y del Lenguaje en el Paralítico Cerebra*l, Cepe.

Camacho-Salas, A., Pallás-Alonso, C., de la Cruz Bértolo, J., Simón de las Heras, R., & Mateos-Beato, F., 2007, Parálisis cérébral: concepto y registros de base poblacional. *Revista de Neurologia*, **45**(8), 503-508.

Canale, M., & Swain, M., 1980, Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, **1**(1), 1-47.

Castellanos, G., & Castellanos, R., 2007, Evaluación diagnóstica del niño con parálisis cerebral, *Revista Cubana de Pediatría*, **79**(2).

Chew Ka, L., Iacono, T., & Tracy, J., 2009, Overcoming communication barriers: Working with patients with intellectual disabilities, *Australian family physician*, **38**(1/2), 10-14.

De Saussure, F., 1994, *Curso de Linguistica General*, Losada .

Evans, P. R., 1948, Problems in the classification of cerebral palsy in childhood. *Archives of Disease in Childhood*, **23**, 213.

García-González, J., & Sánchez-Sánchez, P., 2020, Theoretical design of research: methodological instructions for the development of scientific research proposals and projects, *Información tecnológica*, **31**(6), 159-170.

Gulati, S., & Vishal, S., 2018, Cerebral palsy: an overview. *The Indian Journal of Pediatrics*, **85**, 1006-1016.

Gumperz, J., & Hymes, D., 1972, *Directions in Sociolinguistics: The Ethnography of Communication*, John Wiley & Sons.

He, K. Z., 2016, Deep residual learning for image recognition, *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, NV, USA, 770-778.

Hinton, G., 2018, Deep learning—a technology with the potential to transform health care, *JAMA*, **320**(11), 1101-1102.

Huang, Y.-M., Tseng, H.-H., Chien, J.-T., & Chien, J.-T., 2019, Stochastic Fusion for Multi-stream Neural Network in Video Classification, *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (APSIPA ASC), 69-74.

Kuban, K. C., & Leviton, A., 1994, Cerebral palsy. *New England Journal of Medicine*, **330**, 188--195.

Saville-Troike, M., 2008, *The ethnography of communication: An Introduction*, 3 Ed, Wiley-Blackwell.

Miotto, R., Fei, W., Shuang, W., Xiaoqian, J., & Dudley, J. T., 2018, Deep learning for healthcare: review, opportunities and challenges, *Briefings in bioinformatics*, **19** (6), 1236-1246.

Morales Vaquero, R., 2017, *Terapias logopédicas emergentes aplicadas a la parálisis cerebral infantil: Propuesta de intervención en la Educación Primaria,* Universidad de Valladolid.

Novacheck, T., Trost, J., & Sohrweide, S., 2010, Examination of the child with cerebral palsy, *Orthopedic Clinics of North America*, **41**(4), 469-488.

Panwar, H., Gupta, P. K., Mohammad, K. S., Morales-Menendez, R., & Vaishnavi, S., 2020, Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet, *Chaos, Solitons & Fractals*, **138**, 109944.

Pennington, L., 2008, Cerebral palsy and communication, *Paediatrics and Child Health*, **18**(9), 405-409.

Perez, F.O., & Fernàndez, H., 2007, Las redes neuronales y la evaluación del riesgo de crédito, *Revista Ingenierías Universidad de Medellín*, **6**(10), 77-91.

Riofrio, R., & Carolina, D., 2015, *Habilitación del habla y lenguaje a través del método Bobath en los niños con parálisis cerebral que asisten al centro de rehabilitación e hipoterapia "Angelitos de Luz" de la ciudad de Loja año 2013,* Universidad Nacional de Loja.

Román, A., 2017, *Los GIF como recurso para la comunicación alternativa y aumentativa. Diseño de un SAAC,* Universidad de Valladolid.

Sánchez-Sánchez, P. A., García-González, J. R., & Rúa Ascar, J. M., 2020, Automatic migraine classification using artificial neural networks, *F1000 Research,* **9**, 618

Sánchez-Sánchez, P., & García-González, J., 2018, Autoregressive moving average recurrent neural networks applied to the modelling of Colombian exchange rate, *International Journal of Artificial Intelligence,* **16**(2), 194-207.

Sánchez-Sánchez, P., Cano, J., Garcia, D., Pinzon, A., Rodriguez, G., Garcia-Gonzalez, J., & Perez-Coronell, L., 2019, Knowledge Discovery in Musical Databases for Moods Detection, I*EEE Latin America Transactions*, **17**(12), 2061–2068.

Savignon, S., 2018, *Communicative competence*, The TESOL encyclopedia of English language teaching.

Simonyan, K. &. Zisserman, A., 2014, Very deep convolutional networks for large-scale image recognition, *arXiv*, 1409.1556.

University of Central Florida, 2021, *UCF101 - Action Recognition Data Set*, In: https://www.crcv.ucf.edu/data/UCF101.php