

Implementación de una red neuronal MLP para la detección automática de Emisiones Otoacústicas Transientes

MLP neural network implementation for automatic detection of Transient Otoacoustic Emissions

Yosvani Pantoja Gómez^{1*}

[0000-0002-7197-4983](tel:0000-0002-7197-4983)

José Carlos Santos Ceballo¹

[0000-0003-3267-546X](tel:0000-0003-3267-546X)

Jorge Germán Pérez Blanco¹

[0000-0002-3631-4269](tel:0000-0002-3631-4269)

Ernesto Velarde Reyes¹

[0000-0002-5840-3537](tel:0000-0002-5840-3537)

¹ Centro de Neurociencias de Cuba (CNEURO), La Habana, Cuba.

*Autor para la correspondencia: ypantojag@gmail.com

RESUMEN

La hipoacusia presenta una incidencia de 6 por cada mil nacidos vivos, y afecta al 5% de la población mundial, ocupando el primer lugar entre los déficits sensoriales más prevalentes. A partir de los 6 meses de edad comienza un período crítico en el desarrollo del niño, donde se alcanza el mayor potencial de plasticidad neural, y el daño auditivo es irreversible. El diagnóstico e intervención temprana, permite lograr a largo plazo niveles de desarrollo intelectual adecuados, por lo que es necesaria la implementación de programas de cribado auditivo neonatal. Las técnicas fundamentales utilizadas en el cribado neonatal auditivo emplean las Emisiones Otoacústicas Transientes (EOAT) como indicador del funcionamiento integral de la cóclea, empleando herramientas de detección automática. En esta investigación se presenta la implementación de un método de detección automática de EOAT basado en una red neuronal multicapa. Se diseñó una *Wide Neural Network* de una sola capa intermedia con 100 neuronas, y se implementó en un microcontrolador sobre el sistema de cribado neonatal Infantix. El método basado en la red neuronal se validó mediante un estudio en 50 sujetos adultos. Los resultados de la validación muestran una mejora en el rendimiento de la prueba de EOAT en el sistema Infantix, al compararlo con un método basado en umbrales, implementado sobre el mismo sistema. De esta forma, el empleo de una red neuronal para clasificación de señales con patrones complejos dota al sistema Infantix de una herramienta fiable e inteligente en la detección automática de EOAT.

Palabras Clave: hipoacusia infantil; despistaje auditivo neonatal; detección automática del déficit auditivo; diagnóstico temprano de hipoacusia; redes neuronales; potenciales evocados auditivos (PEA); detección de bioseñales; aprendizaje automático; reconocimiento de patrones.



ABSTRACT

Hearing loss has an incidence of 6 per 1,000 live births and affects 5% of the global population, ranking as the most prevalent sensory deficit. At 6 months of age, a critical period begins in a child's development where the greatest potential for neural plasticity is reached, yet auditory damage becomes irreversible. Early diagnosis and intervention enable adequate levels of intellectual development over the long term, making newborn hearing screening programs essential. Key techniques in neonatal hearing screening utilize Transient Otoacoustic Emissions (TEOAE) as indicators of cochlear function, employing automated detection tools. This study presents the implementation of an automatic TEOAE detection method based on a multilayer neural network. A Wide Neural Network with a single hidden layer of 100 neurons was designed and implemented on the Infantix newborn screening system. The neural network-based method was validated through a study involving 50 adult subjects. Validation results show an improvement in TEOAE test performance in the Infantix system compared to a threshold-based method implemented on the same system. Thus, using a neural network for the classification of complex-pattern signals provides the Infantix system with a reliable and intelligent tool for automatic TEOAE detection.

Keywords: hearing loss in children; newborn hearing screening; automatic detection of hearing loss; early diagnosis of hearing loss; neural network; auditory evoked potentials (AEP); bio-signal detection; automatic learning; pattern recognition.

Recibido: 25/11/2024

Aprobado: 23/01/2025

Introducción

La Organización Mundial de la Salud (OMS) estima que alrededor del 5% de la población mundial padece algún tipo de hipoacusia. Esta pérdida de la capacidad auditiva ocupa el primer lugar entre los déficits sensoriales más prevalentes de acuerdo a estadísticas de la misma organización. Además, se estima que la incidencia de la hipoacusia en sus diferentes grados, es de 6 por cada mil nacidos vivos.⁽¹⁾ El anuario estadístico de salud en Cuba en el año 2009 cita a la sordera entre los primeros 40 padecimientos de la población cubana, ocupando un 2.6% del total de la población examinada.^{(2), (3)}

Este tipo de pérdida sensorial desde edades tempranas, tiene un impacto severo en el desarrollo del lenguaje. Los niños hipoacúsicos ven limitadas su capacidad para adquirir conocimientos, y su autonomía en el aprendizaje. Como consecuencia, se limita el desarrollo intelectual del niño y sus capacidades de comunicación, provocando un aislamiento social.⁽⁴⁾



Alrededor de los 6 meses de edad se establecen las conexiones cerebrales en las zonas destinadas al lenguaje dentro de la corteza cerebral, considerándose un período crítico en la neuromaduración.^{(5), (7)} La ausencia de estímulo en estas regiones, debido a la presencia de algún tipo de hipoacusia, provoca que no se realicen dichas conexiones o que se organicen de otra forma estos circuitos neuronales, siendo un proceso irreversible. El diagnóstico temprano de la hipoacusia, en niños menores de 6 meses de edad, permite una rápida intervención y el posterior desarrollo normal del niño. Por esta razón, es necesaria la implementación de programas de cribado o pesquisa auditiva neonatal que permitan una intervención temprana.

Los programas de cribado neonatal auditivo, en su mayoría, son implementados por los sistemas nacionales de salud, destacándose dos tipos fundamentales de acuerdo a la población de pacientes a los que se aplica: el cribado selectivo y el universal. En el cribado selectivo solo se aplica el examen a los recién nacidos con alguna condición de riesgo, mientras que en el cribado universal se realiza la prueba a todos los neonatos.

Más del 50% de los recién nacidos hipoacúsicos no presentan condición de riesgo conocida, por lo que se hace imposible su detección mediante un cribado selectivo.⁽⁸⁾ En este sentido, es necesario que los sistemas de salud cuenten con programas de cribado neonatal universal, para extender el examen a toda la población infantil.

En contraste, el alto costo en recursos humanos y tecnológicos que supone garantizar la evaluación de todos los infantes, conlleva a dividir estos programas de pesquisa en diferentes etapas que permitan racionalizar eficientemente los recursos disponibles. De esta forma, en la primera fase se utilizan tecnologías de cribado totalmente automáticas que no requieren de personal altamente especializado; mientras que en etapas posteriores se aplican exámenes más específicos e incrementa el grado de especialización.⁽⁴⁾

Para la implementación de cualquier programa de cribado neonatal se precisa contar con técnicas que permitan evaluar a los neonatos objetivamente. Entre las variables fisiológicas más empleadas para la evaluación objetiva de la audición, se encuentran los Potenciales Evocados Auditivos (PEA) y las Emisiones Otoacústicas (EOA).

El PEA es la actividad eléctrica que se produce en la corteza cerebral en respuesta a algún tipo de estímulo auditivo. Para su registro, se colocan electrodos en la cabeza del sujeto empleando el sistema internacional 10-20, y se mide la diferencia de potencial eléctrico entre cada par de electrodos. El registro de estos potenciales constituye una prueba objetiva no invasiva que permite conocer el estado y la integridad de la vía auditiva, desde el oído externo hasta la corteza cerebral.⁽⁹⁾

Las EOA son señales sonoras de muy baja intensidad que se generan en la cóclea producto de la actividad de las células ciliadas externas, y se registran en el conducto auditivo externo. A diferencia de los PEA, en los que se puede evaluar la vía auditiva en su totalidad, las EOA no permiten evaluar pérdidas retrococleares, es decir, cuando está dañado o no existe el nervio auditivo.^{(10),(11)} Para la obtención de las EOA solo se



requiere la colocación de una sonda en el oído externo, con una pequeña bocina para la emisión de sonido, y un micrófono de alta sensibilidad para recoger la respuesta proveniente de la cóclea.⁽¹²⁾

Teniendo en cuenta la sencillez del examen, la técnica de EOA es la utilizada en el primer nivel de cribado auditivo neonatal, aportando bajo costo, rapidez, además de que requiere un menor nivel de especialización del personal médico que realiza la prueba.⁽¹³⁻¹⁶⁾ A pesar de ello, se requiere del empleo de los PEA para obtener un sistema de cribado auditivo neonatal más completo, con las ventajas que aportan cada una de estas técnicas.

Las EOA pueden ser generadas por la cóclea de forma espontánea o en respuesta a un estímulo sonoro externo (evocadas). Las aplicaciones clínicas de las EOA espontáneas para el cribado están limitadas por el hecho de no estar presentes en la totalidad de sujetos que no presentan patologías auditivas.⁽¹⁷⁻¹⁹⁾ Las EOA evocadas pueden dividirse en dos grupos fundamentales de acuerdo al tipo de estímulo que las provoca, las Emisiones Otoacústicas Transientes (EOAT) y las Emisiones Otoacústicas Producto de Distorsión (EOAPD).

Las EOAT se generan a partir de estímulos de muy corta duración, como el estímulo clic, con intensidades típicas entre los 60 y 90 dB SPL (Sound Pressure Level). Al emplearse un estímulo de banda ancha, se obtiene simultáneamente la respuesta de varias regiones de la cóclea, por lo que resulta más rápida la evaluación que en el caso de las EOAPD. Las EOAPD son más específicas en frecuencia, pero es necesario realizar un barrido para evaluar las distintas regiones de la cóclea, aumentando el tiempo de la prueba.⁽²⁰⁾ Esta investigación centra su análisis en el registro de las EOAT.

En la actualidad se han desarrollado diferentes metodologías que permiten emitir un resultado automático a partir del registro de las EOAT. El empleo de algoritmos automáticos de detección de EOAT es imprescindible en el contexto del cribado, ya que permite masificar su empleo, eliminando la necesidad de personal altamente especializado para la realización de las pruebas. Es por ello que se requieren métodos de detección de bioseñales que brinden resultados en muy corto tiempo, sin comprometer la veracidad en el diagnóstico.

El surgimiento de herramientas de inteligencia artificial y su uso en sistemas embebidos, marcan tendencia en el procesamiento digital de señales. Las redes neuronales multicapas MLP (Multilayer Perceptron) son un tipo de modelo de aprendizaje automático con un uso extendido en el reconocimiento de patrones y clasificación de señales. En los últimos años este tipo de herramienta ha tenido un auge en el campo de la neurofisiología, debido a la complejidad de las señales cerebrales y los patrones que describen su funcionamiento.⁽²¹⁾

En el Centro de Neurociencias de Cuba (CNEURO) se desarrolla el Sistema de Cribado Neonatal Infantix para la pesquisa auditiva y visual.⁽²²⁾ Este es un sistema modular, que incorpora módulos para el registro de Potenciales Evocados y Emisiones Otoacústicas.



El sistema Infantix permite la detección de bioseñales sin la fiscalización de un especialista, por lo que incorpora métodos automáticos de procesamiento digital de señales. El método de detección automática de EOAT implementado en el sistema Infantix ha presentado resultados favorables durante su uso, con valores de sensibilidad y especificidad aceptados para este tipo de examen fisiológico, pero con tiempos de prueba un poco elevados, alrededor de los 60 segundos.⁽²³⁾ Teniendo en cuenta esto, se hace necesario el empleo de técnicas de procesamiento de señales más potentes, con el objetivo de mejorar aún más la sensibilidad y especificidad, y reducir los tiempos de prueba. La presente investigación tiene como objetivos específicos, implementar y validar una red neuronal MLP en un microcontrolador, para la detección automática de Emisiones Otoacústicas Transientes.

Materiales y Métodos

Sistema de Cribado Neonatal Infantix

El Sistema Infantix está compuesto por un Módulo de Control que se comunica con un Módulo de Registro de Potenciales Evocados (PE) o un Módulo de Registro de Emisiones Otoacústicas (EOA), dependiendo de la prueba a realizar (Fig. 1). Este documento centrará el análisis en el último de los módulos.



Fig.1- Diagrama en bloques del Sistema Infantix.

El Módulo de Control es el encargado de establecer la interfaz con el usuario, y ejecutar los métodos de procesamiento digital de las bioseñales que se registran en tiempo real, para emitir el resultado automático de la prueba (“Pasó” o “Referir”). Además, permite el almacenamiento de los datos del paciente. En este módulo es donde se implementará el método automático de detección de EOAT propuesto.



Por otra parte, el Módulo de Registro de EOA tiene la función de adquirir y acondicionar las bioseñales registradas con la Sonda de EOA (transductor que recibe la respuesta del sujeto). Este registro está sincronizado con un estímulo auditivo, que genera el mismo módulo, y que es aplicado al sujeto a través de la sonda referida anteriormente.

Registro de Emisiones Otoacústicas

En el Sistema Infantix la prueba de EOA inicia con la evaluación automática del estado de la sonda (“dentro”, “fuera” y “ocluido”), lo que permite saber si está bien colocada en el conducto auditivo. De esta manera se garantiza que el registro se realice bajo las mejores condiciones, aportando validez al resultado. Una vez detectada la sonda “dentro”, comienza el registro del bioseñal.

El proceso de registro de EOAT comienza con la presentación de un estímulo auditivo tipo click no lineal, a través de la sonda de EOA, la cual registra por medio de un micrófono de alta sensibilidad, la respuesta proveniente de la cóclea sincronizada con el estímulo.

Para la obtención de la respuesta coclear se genera un estímulo a una intensidad de 83 dB peSPL y frecuencia interestímulo de 51 Hz. Esta frecuencia limita la ventana de registro a un máximo de 19.6 ms, que son suficientes para evaluar el rango de frecuencias comprendido entre 1 kHz y 5 Khz.⁽²⁴⁾ La frecuencia de muestreo de la señal es de 24 kHz y la ventana de análisis de 14 ms, lo que permite un registro de segmentos periódicos de señal de 336 muestras.

Esta señal es acondicionada y enviada al Módulo de Control donde se procesa y se realiza el análisis matemático-estadístico. Son evaluadas diversas características de la señal y se emite el resultado de la prueba (“Pasó” o “Referir”). El tiempo de la prueba está determinado por la rapidez con la que el método automático detecta una respuesta significativa.

Método de detección de Emisiones Otoacústicas basado en umbrales

El método automático de detección de EOAT en el Sistema Infantix está implementado en el microcontrolador (uC) del Módulo de Control. El uC que se emplea es un STM32F429 del fabricante STMicroelectronics que incorpora un procesador con una arquitectura ARM Cortex-M4. Este procesador cuenta con una Unidad de Punto Flotante (FPU, por sus siglas en inglés), que permite el empleo de instrucciones dedicadas a la aritmética de punto flotante y al procesamiento digital de señales en sentido general. El fabricante provee la biblioteca CMSIS DSP que implementa operaciones optimizadas para el procesamiento digital de señales. La implementación del método existente y el propuesto en este documento, hace un uso extensivo de dicha biblioteca para un mejor aprovechamiento de los recursos disponibles en el procesador.



La figura 2 muestra el diagrama de flujo del método automático ya implementado en el Sistema Infantix. Este método propone un análisis en el dominio del tiempo y la frecuencia, y evalúa la señal respecto a umbrales.

En este algoritmo, primeramente, se acondiciona la señal, eliminando la componente directa de cada uno de los registros obtenidos, y se realiza un filtrado digital de los mismos. Se emplea un filtro IIR paso banda, Butterworth, fase 0, de orden 6, con frecuencia de corte a las bajas de 500 Hz y a las altas de 8 kHz. Una vez que la señal registrada se encuentra libre de artefactos fuera de la banda de interés, se compara la amplitud de esta con un umbral por encima de los valores típicos de intensidades de las EOA. En caso de exceder el umbral, el registro es rechazado.

Posteriormente se promedian paquetes de 4 registros consecutivos, para implementar la técnica de estimulación no lineal. Luego se organizan estos promedios alternadamente en dos grupos, y utilizando la técnica de promediación coherente o sincrónica se obtienen los promedios de dichos grupos que denominaremos hemipromedios.

Mediante esta técnica, segmentos de señal de igual latencia respecto a la presentación del estímulo son promediados, manteniendo la respuesta evocada y reduciendo el ruido. De esta forma mejora la Relación Señal-Ruido (RSR o SNR por sus siglas en inglés) de la señal resultante.

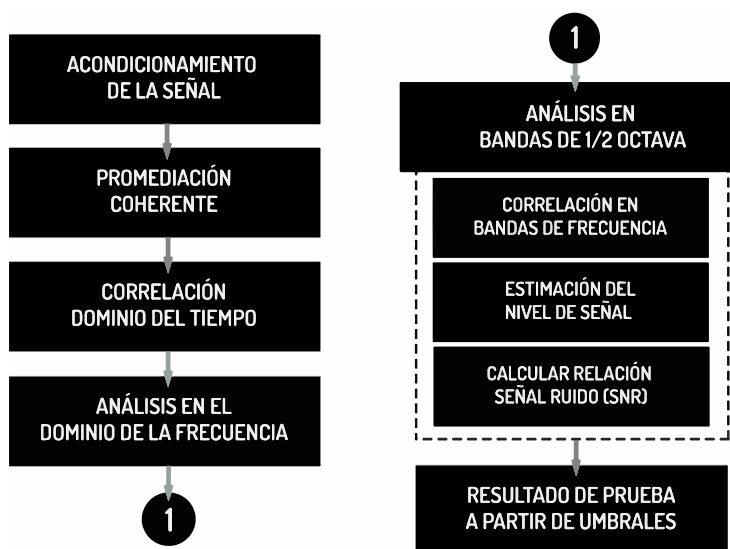


Fig. 2- Diagrama de flujo del método automático de EOAT basado en umbrales.

Como primer criterio de evaluación de la respuesta se calcula la correlación de los hemipromedios en el dominio del tiempo, lo que permite evaluar su reproducibilidad. Luego se estima el contenido espectral de los hemipromedios y se realiza un análisis en el dominio frecuencial, donde se evalúan las bandas de media octava centradas en 1.5, 2, 3 y 4 kHz respectivamente. Para cada una de estas bandas se analizan la correlación,



el nivel de la señal y la RSR, y se comparan con umbrales fijos. Este análisis se realiza cada 30 promediaciones, aproximadamente cada 2.4 segundos.

Finalmente se comprueba que el total de bandas en las que se encontró respuesta sea mayor o igual a 3. En caso de que se cumplan todos los criterios señalados, el método automático de detección de EOAT emite el resultado “Pasó”. Si al menos uno de estos criterios no se cumple, se emite el resultado “Referir”, siendo necesaria la remisión del paciente a estudios más profundos que permitan diagnosticar el tipo de pérdida auditiva que presenta. Este algoritmo de detección ha sido validado en estudios anteriores.⁽²³⁾

Método de detección de Emisiones Otoacústicas basado en redes neuronales

Teniendo en cuenta el flujo de análisis del método basado en umbrales, se propone una actualización del algoritmo para la detección de EOAT con el empleo de machine learning (aprendizaje automático) en la identificación de patrones para la clasificación de la bioseñal.

La figura 3 muestra el diagrama de flujo del método propuesto, el cual realiza un análisis en el dominio del tiempo y la frecuencia, al igual que el método basado en umbrales.

En una primera etapa de procesamiento, se acondiciona la señal, empleando los mismos recursos que el método basado en umbrales. Se utiliza un filtro digital IIR paso banda con las mismas características que el expuesto en el acápite anterior y se rechazan los registros más ruidosos. Se promedian paquetes de 4 registros consecutivos, para implementar la técnica de estimulación no lineal, y estos se organizan alternadamente en dos grupos.



Fig. 3- Diagrama de flujo del método automático de EOAT basado en red neuronal MLP.



Estos dos grupos son promediados empleando la técnica de promediación bayesiana, obteniéndose los hemipromedios A y B. Mediante esta técnica, segmentos de señal de igual latencia son promediados, teniendo en cuenta el peso de cada segmento, el cual es determinado por su nivel de ruido. Se estima el ruido de cada segmento a partir de su varianza, y el inverso de esta para determinar su respectivo peso. La promediación bayesiana (ecuación 1) se adapta a la variabilidad de la señal, ponderando cada segmento de acuerdo con su calidad, utilizando el inverso de la varianza de cada segmento como peso (ecuación 2). Esto significa que los segmentos con mayor ruido (mayor varianza) tienen menor impacto en el promedio final, acelerando la reducción de ruido y mejorando la RSR de la señal recuperada.

$$\bar{X}_{bayes} = \frac{\sum_{i=1}^N x_i * w_i}{\sum_{i=1}^N w_i} \quad (1)$$

$$w_i = \frac{1}{\sigma_i^2} \quad (2)$$

Donde x_i es el segmento analizado y w_i es el peso del mismo, obtenido a partir del inverso de su varianza σ_i^2 . La varianza de la señal en el μC se calcula empleando la función *arm_var_f32* de la biblioteca CMSIS DSP. Este esquema ponderado garantiza una reducción de varianza más rápida y eficiente.

En contraste, la promediación sincrónica clásica basada en la media aritmética, empleada en el método de umbrales, aunque es efectiva en la reducción de ruido para señales con varianza uniforme, presenta limitaciones cuando el ruido varía significativamente entre segmentos. Al asignar el mismo peso a todos los segmentos de la señal, la media aritmética no considera las diferencias de ruido en cada segmento. Esto conlleva a una convergencia subóptima y a una reducción de ruido menos eficiente, especialmente cuando los segmentos contienen ruido aleatorio de alta varianza.

Posteriormente se determina la correlación en el dominio del tiempo entre los promedios A y B, lo cual permite evaluar su reproducibilidad temporal. Para ello se emplea la función *arm_correlate_f32* de la biblioteca CMSIS DSP.

Luego la señal es transformada al dominio de la frecuencia para analizar el contenido espectral de los promedios A y B, mediante la Transformada Rápida de Fourier (FFT por sus siglas en inglés), empleando la función *arm_rfft_f32* de CMSIS DSP. Se calcula la RSR y el nivel de correlación en bandas de media octava centradas en 1.5, 2, 3 y 4 kHz.

El nivel de la señal se estima mediante el cálculo de la raíz cuadrática media (ecuación 3), y a través de la ecuación 4 se estima el nivel de ruido, mediante el cálculo de la desviación cuadrática media. Con estos valores se obtiene la RSR (ecuación 5).



$$S = 20 * \log \sqrt{\frac{1}{N} \sum_1^N FFT(A) * FFT(B)} \quad (3)$$

$$R = 20 * \log \sqrt{\frac{1}{N} \sum_1^N (FFT(A) - FFT(B))^2} \quad (4)$$

$$RSR = S - R \quad (5)$$

Los valores calculados de correlación temporal y características en bandas de frecuencia, son pasados como parámetros de entrada a una red neuronal MLP. El empleo de esta red tiene como objetivo clasificar la señal de acuerdo a dos clases que determinan el resultado de la prueba (“Pasó” o “Referir”).

Como modelo de aprendizaje automático se determinó el uso de redes neuronales, debido a la naturaleza compleja de los datos (bioseñales acústicas de baja amplitud), y la necesidad de procesamiento en tiempo real sobre un μC con limitaciones de memoria y cómputo. Se escoge una red neuronal MLP, teniendo en cuenta el balance entre su capacidad para clasificar señales complejas y su eficiencia computacional, además de que puede ser optimizada para su implementación en el μC (STM32F429).

Se utilizó la herramienta Google Colab para diseñar la red y generar el modelo haciendo uso de la biblioteca TensorFlow Lite. Esta es una versión ligera de la librería TensorFlow de Google, diseñada para ejecutar modelos de aprendizaje automático en dispositivos de recursos limitados como los microcontroladores. Luego de generado el modelo, este se exportó en formato .tflite para su implementación en el μC , empleando la herramienta STM32 X-CUBE-AI para optimizarlo. STM32 X-CUBE-AI es un set de librerías y plugins incorporado en el asistente de configuración STM32CubeMx para la implementación de modelos de inteligencia artificial en microcontroladores de la familia STM32, garantizando eficiencia en el uso de la memoria y velocidad en el procesamiento.

Fueron evaluados 25 modelos de redes neuronales MLP en Google Colab, y se escogió el que mejor porcentaje de exactitud y sensibilidad arrojó en el entrenamiento. Para el diseño de cada modelo se variaron características claves como la cantidad de neuronas en las capas ocultas, la función de activación, el costo de clasificación errónea, y el número de parámetros o predictores de entrada (según el algoritmo de peso estadístico utilizado). Los resultados de este análisis se muestran en la sección Resultados y Discusión de este documento.

Para la ejecución de la red neuronal en el μC se utilizó una función de inferencia de la librería de STM32Cube-AI, que recibe como entrada 13 características de la señal (predictores de entrada) y devuelve un valor en el rango de 0 a 1 (Fig. 4). Este valor de salida indica la probabilidad de que los valores de entrada pertenezcan a una clase específica, permitiendo así la clasificación de la señal. Se estableció un umbral de 0.5 para el valor de salida, teniendo en cuenta que se trata de una clasificación binaria, “Referir” (< 0.5) y “Pasó” (> 0.5).



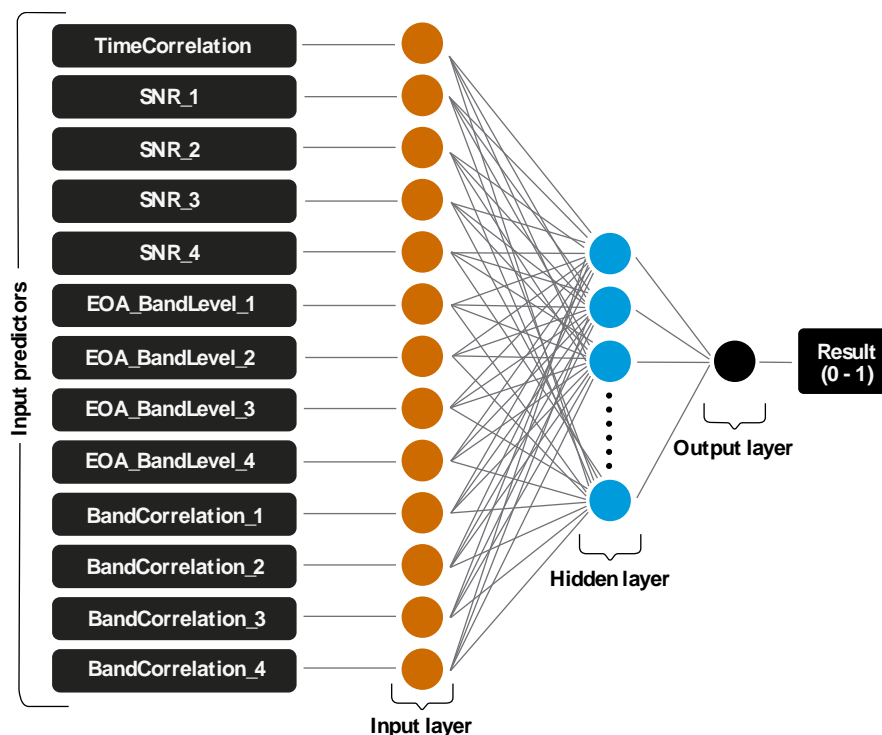


Fig. 4- Diagrama de la red neuronal MLP propuesta, con sus señales de entrada (Input predictors) y su señal de salida (Result (0 - 1)).

Hiperparámetros de la red neuronal MLP

Los hiperparámetros de una red neuronal son parámetros configurados manualmente antes del entrenamiento, que controlan aspectos clave de su estructura y aprendizaje, y que no se ajustan directamente durante el proceso de entrenamiento. En esta investigación, los hiperparámetros ajustados fueron la cantidad de neuronas en las capas ocultas o intermedias, la función de activación, y el costo de clasificación errónea.

Cantidad de neuronas y funciones de activación

En una red neuronal, la cantidad de neuronas se refiere al número de unidades o nodos presentes en cada capa de la red. Cada neurona procesa y transmite información a la siguiente capa mediante conexiones ponderadas.

Las funciones de activación, por otro lado, son funciones matemáticas aplicadas a la salida de cada neurona para introducir no linealidades en la red y permitir que aprenda relaciones complejas en los datos. Estas funciones determinan si la neurona se activa (produce una salida significativa) o no, influenciando así la capacidad de la red para capturar patrones complejos y realizar predicciones precisas.

En los modelos diseñados en esta investigación, se empleó una sola capa intermedia, y se variaron la cantidad de neuronas y funciones de activación en dicha capa.



Costo de clasificación errónea

El costo de clasificación errónea en un modelo de aprendizaje automático se refiere a la penalización o "costo" asociado a los errores que comete el modelo cuando clasifica incorrectamente una instancia.

En los modelos diseñados en esta investigación, solo se varió el costo asociado a la clasificación incorrecta de los sujetos con discapacidad auditiva (Falsos Negativos).

Diseño del entrenamiento de las redes neuronales MLP

Para el entrenamiento de las redes neuronales, se confeccionó una base de datos compuesta por registros de EOAT obtenidos en sujetos adultos. Los registros fueron clasificados en dos categorías: "Pasó" (normo-oyente) y "Referir" (pérdida auditiva de algún grado: leve, moderada o severa). La clasificación de las señales fue realizada por un especialista en audiología, quien se basó en el análisis de una audiometría, y la morfología y características de las señales evaluadas.

Los registros de las EOAT se realizaron con el Sistema Infantix y los datos fueron guardados para el posterior análisis por parte del especialista. Se obtuvieron registros de 1200 promediaciones, sin tener en cuenta los segmentos de señal rechazados, con un tiempo total de registro entre 60 y 90 seg.

Durante la evaluación se presentó al especialista una visualización de la señal, en diversos instantes de tiempo durante la prueba. Con base en el criterio complementario de la audiometría, el especialista determinó si cada instante de tiempo debía ser clasificado como "Pasó" o "Referir", asumiéndose como el criterio de verdad. Se empleó esta estrategia, para lograr un modelo que permita clasificar rápidamente, con la intención de mejorar el tiempo de la prueba de EOAT.

La muestra utilizada para el entrenamiento incluyó a 88 sujetos adultos de ambos sexos, lo que equivale a 176 oídos en total. De estos, 105 oídos fueron clasificados como "Pasó" y 71 oídos como "Referir". Antes de las pruebas de EOAT, todos los sujetos brindaron su consentimiento informado, y fueron sometidos a una otoscopia y una audiometría tonal. La otoscopia fue correcta en todos los casos.

Cada red neuronal fue entrenada empleando el mismo juego de datos, los cuales fueron normalizados antes de pasarlos a la red. Se utilizaron 4389 instantes de tiempo, garantizando una clasificación balanceada entre las clases "Pasó" y "Referir". Aunque el número de sujetos que refieren la prueba es menor, se incluyeron instantes de tiempo iniciales de todas las pruebas, en los que el especialista considera que los sujetos aún no pasan la prueba.

Se utilizaron hasta 13 predictores de entrada para el entrenamiento, y la clasificación final fue binaria ("Pasó" o "Referir"). Para evaluar el rendimiento del modelo, se empleó un esquema de validación basado en una cross-validación de 5 etapas, además de un set de prueba que representó el 10 % de los datos suministrados. Con el esquema propuesto se disminuye la probabilidad de sobreajuste durante el



entrenamiento. Los resultados de este análisis se exponen en la sección Resultados y Discusión.

Diseño del estudio para evaluación del método de detección automática de EOAT basado en una red neuronal MLP

Para evaluar el desempeño del método basado en una red neuronal MLP (en lo adelante método_RN), se compararon los resultados de este frente al método basado en umbrales (en lo adelante método_Umbral), en un estudio en sujetos adultos. Ambos métodos fueron implementados en el Sistema Infantix, y evaluados en tiempo real cada uno por separado.

Descripción de la muestra

La muestra para el estudio incluyó a 55 sujetos adultos de ambos sexos, con una edad promedio de 39 años, para un total de 110 oídos. Antes de las pruebas, todos los sujetos brindaron su consentimiento informado, y fueron sometidos a una otoscopia. En un solo caso, la otoscopia reveló una condición que llevó a la exclusión de un sujeto del estudio, por lo que se evaluaron finalmente, 108 oídos. Se realizó una audiometría tonal a cada sujeto, por parte de un especialista en audiología, como criterio de verdad, donde 77 oídos fueron clasificados como “Pasó” y 31 oídos como “Referir”.

Diseño del estudio en sujetos adultos

En una primera etapa, se realizaron registros de EOAT con el Sistema Infantix, sin implementación de ninguno de los dos métodos. Los registros de EOAT se fijaron en 1200 promediaciones, sin tener en cuenta los segmentos de señal rechazados, con una duración entre 60 seg y 90 seg de tiempo total de registro. Los registros fueron guardados en ficheros individuales, en la memoria extraíble del sistema (microSD), para su posterior análisis con cada método.

Con los 108 registros guardados, se procedió a la implementación de cada método en el Sistema Infantix, siguiendo en cada caso el diagrama de flujo de la prueba, presentado anteriormente. Cada registro fue cargado desde la memoria extraíble, y analizado por cada método en tiempo real, simulando la realización de una prueba de EOAT. En cada caso, una vez el método determina la presencia de EOAT, se detiene la prueba y se marca el registro como “Pasó”, en caso contrario se marca como “Referir”. El resultado y el tiempo de prueba se guardaron al instante que se detuvo la prueba. Estos valores fueron analizados y comparados entre los dos métodos, evaluando métricas como la sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), y tiempo medio de prueba. Los resultados de este análisis se exponen en la sección Resultados y Discusión.

Medidas de análisis estadístico

Sensibilidad

La sensibilidad indica la capacidad de nuestro método automático para dar como casos positivos los casos que realmente presentan alguna discapacidad auditiva. Es decir, la



sensibilidad caracteriza la capacidad de la prueba para detectar la hipoacusia en sujetos con discapacidad auditiva. La sensibilidad se calcula por la siguiente ecuación:

$$\text{Sensibilidad} = \frac{VP}{VP+FN} \quad (5)$$

Siendo VP los Verdaderos Positivos y FN los Falsos Negativos.

Especificidad

La especificidad indica la capacidad de nuestro método automático para dar como casos negativos, los casos realmente sanos. Es decir, la especificidad caracteriza la capacidad de la prueba para detectar la ausencia de hipoacusia en sujetos sanos. La especificidad se calcula por la siguiente ecuación:

$$\text{Especificidad} = \frac{VN}{VN+FP} \quad (6)$$

Siendo VN los Verdaderos Negativos y FP los Falsos Positivos.

Valor Predictivo Positivo

El Valor Predictivo Positivo (*VPP*) o *Positive Predictive Value (PPV)* es la probabilidad de tener discapacidad auditiva si el resultado de la prueba es positivo. Indica qué tan confiable es un resultado positivo de la prueba. El *VPP* se calcula por la siguiente ecuación:

$$VPP = \frac{VP}{VP+FP} \quad (7)$$

Siendo VP los Verdaderos Positivos y FP los Falsos Positivos.

Valor Predictivo Negativo

El Valor Predictivo Negativo (*VPN*) es la probabilidad de estar sano si el resultado de la prueba es negativo. Indica qué tan confiable es un resultado negativo de la prueba. El *VPN* se calcula por la siguiente ecuación:

$$VPN = \frac{VN}{VN+FN} \quad (8)$$

Siendo VN los Verdaderos Negativos y FN los Falsos Negativos.

Resultados y Discusión

Análisis de los modelos de redes neuronales entrenados

Se presentan los resultados obtenidos a partir del entrenamiento de los diversos modelos de redes neuronales diseñados. El objetivo de este análisis es identificar cómo las variaciones en los parámetros de diseño afectan el rendimiento de cada modelo, y de esta forma determinar la configuración óptima para la clasificación de la Emisión Otoacústica Transiente.

Fueron evaluados 25 modelos de redes neuronales MLP, en función de métricas de rendimiento como la exactitud (accuracy), sensibilidad (recall) y precisión (PPV, Positive Predictive Value). En la figura 5 solo se muestran los 12 modelos que arrojaron más de un 94 % de exactitud en la validación (validation) y la generalización (test).



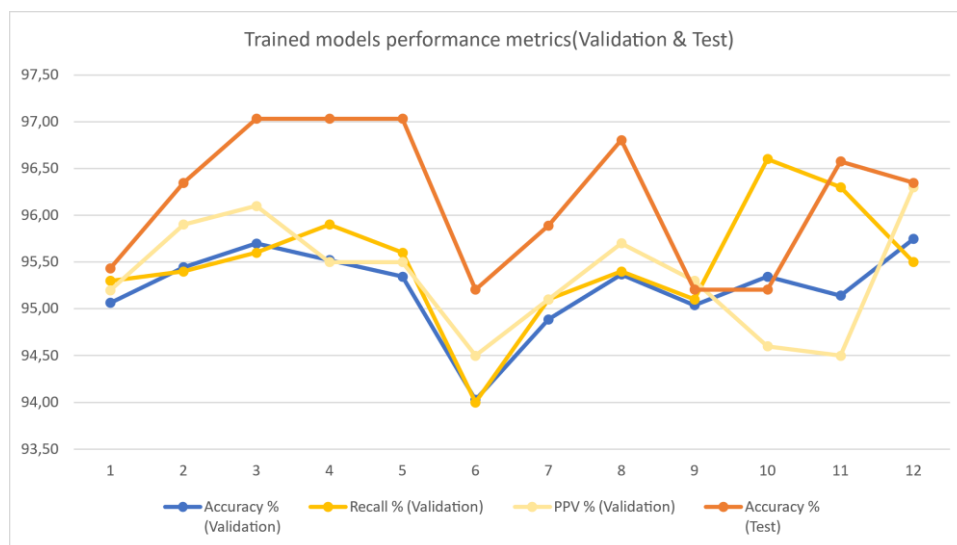


Fig. 5- Métricas de rendimiento (Exactitud (Accuracy), Sensibilidad (Recall), Precisión (PPV)) de 12 modelos de entrenamiento de redes neuronales MLP durante la validación (Validation) y la generalización (Test).

Variaciones en la cantidad de neuronas y funciones de activación

Se configuró para todos los casos una sola capa oculta, para disminuir el costo de procesamiento. La cantidad de neuronas en la capa oculta influyó significativamente en el rendimiento, donde se observó que un número medio-alto de neuronas (50 - 100) proporcionó un mejor resultado durante el aprendizaje y la generalización (modelos 3, 4, 8 y 11), aunque el modelo 5 con 25 neuronas, obtuvo muy buen rendimiento. A partir de 100 neuronas se observó una pequeña disminución en la exactitud del modelo, con un costo computacional más elevado (modelo 12 -> 150 neuronas). El modelo 6 con 25 neuronas fue el modelo con más bajo rendimiento.

Se analizaron las funciones de activación ReLU, Sigmoide y Tanh, en todos los casos efectivas en clasificación binaria. Los modelos que utilizaron Sigmoide y Tanh mostraron un rendimiento superior (modelos 3, 4, 5 y 11), con altos niveles de exactitud durante el aprendizaje y la generalización. Sin embargo, Tanh mostró una ligera mejora en la sensibilidad (modelos 4 y 11). Esto puede deberse a la naturaleza de las señales de EOAT que tienen características positivas y negativas, y a la aparición de no linealidades en las relaciones de los datos.

Variaciones en el costo de clasificación errónea

Se varió el costo de clasificación errónea, de manera tal que la tasa de Falsos Negativos (FN) disminuyera, buscando una mejora en la sensibilidad. Estas variaciones provocaron mejora en la sensibilidad de algunos modelos (modelos 10 y 11), pero en contraste afectaron en mayor medida la precisión. Esto sugiere que el conjunto de datos y las características de las clases permiten una óptima separación, sin necesidad de un costo elevado para los FN. En este caso el costo por defecto proporcionó el



mejor balance entre la sensibilidad y precisión, permitiendo al modelo clasificar correctamente sin comprometer su rendimiento en ninguna clase.

Variaciones en el número de predictores

Se exploraron diferentes algoritmos de pesos estadísticos, para la selección del número de parámetros de entrada, entre ellos MRMR, Chi2, RelieF, ANOVA y Kruskal Wallis. Los resultados indicaron que el algoritmo ANOVA (modelo 8), en particular, logró converger más rápidamente y presentó mejores métricas de rendimiento en comparación con los otros métodos que desestimaron al menos uno de los 13 predictores de entrada (modelo 7 y 9).

Sin embargo, al variar el número de parámetros de entrada, eliminando al menos uno, se observó que disminuyó la exactitud y sensibilidad de salida, observándose un máximo con el empleo de los 13 predictores (modelo 4), con una diferencia de al menos 1 punto porcentual por encima de los modelos que desestiman parámetros. Esto se debe a que cada uno de los parámetros de entrenamiento tiene un peso importante, y son incapaces de clasificar individualmente, pero aportan información relevante al combinarse con el resto.

En resumen, este análisis de los resultados de los modelos entrenados revela que tanto la selección de parámetros de entrada del modelo, como los hiperparámetros de entrenamiento, son cruciales para maximizar el rendimiento en la clasificación de las EOAT. Teniendo en cuenta estos resultados, el modelo con mejores métricas de rendimiento fue el modelo 4, correspondiente a una Wide Neural Network de una sola capa oculta con 100 neuronas, y una función de activación Tanh.

Este modelo se implementó en el Sistema Infantix y se evaluó en sujetos adultos como parte del método de detección automática de EOAT. Los resultados del estudio se muestran a continuación.

Análisis del estudio en sujetos adultos

La tabla 1 muestra los resultados del método_Umbral y el método_RN respectivamente, teniendo en cuenta que los sujetos marcados como "Referir" tanto en la audiometría como en el método, son los considerados como verdaderos positivos (VP).



Tabla 1- Resultados alcanzados con los métodos basado en umbrales (método_Umbral) y método basado en red neuronal (método_RN). Se muestran el tamaño de la muestra, el tiempo medio de prueba, la cantidad de Verdaderos Positivos (VP), Falsos Negativos (FN), Verdaderos Negativos (VN), Falsos Positivos (FP), valores de Sensibilidad, Especificidad, Valor Predictivo Positivo (VPP) y Valor Predictivo Negativo (VPN).

Métricas Método estadísticas	Basado en umbrales	Basado en red neuronal MLP
Muestra (oídos)	108	108
Tiempo medio de prueba (seg)	24.67	14.88
VP (oídos)	28	27
FN (oídos)	3	4
VN (oídos)	66	74
FP (oídos)	11	3
Sensibilidad (%)	90.32	87.09
Especificidad (%)	85.71	96.1
VPP (%)	71.79	90.0
VPN (%)	95.65	94.87

Especificidad

El método_RN tiene una especificidad del 96.1%, superior al 85.71% del método_Umbral. Esto significa que el clasificador de red neuronal es mucho más efectivo en identificar correctamente los oídos sanos (verdaderos negativos), lo cual reduce drásticamente los falsos positivos.

Valor Predictivo Positivo (VPP)

El método_RN también logra un aumento del VPP (90.0%), en comparación con el 71.79% del método_Umbral. Esto sugiere que cuando el modelo de red neuronal identifica un oído como “Referir”, tiene una alta probabilidad de ser correcto. Esto es relevante en el contexto del cribado, donde los sujetos marcados como “Referir” son sometidos a exámenes más exhaustivos y costosos, con la intervención de un especialista. De esta forma, se garantiza que en un primer nivel de cribado se identifiquen realmente los sujetos que presentan alguna discapacidad auditiva, sin llegar a saturar el segundo nivel de cribado.



Sensibilidad

Aunque la sensibilidad es ligeramente mayor en el método_Umbral (90.32% frente a 87.09%), esta diferencia es leve y se ve compensada por las mejoras en especificidad y VPP en el método_RN. En términos de la cantidad de sujetos en la muestra total, esta diferencia representa la discrepancia en la clasificación de un único sujeto, viéndose afectado el resultado por el tamaño de la muestra, y el número de sujetos marcados como “Referir” dentro de esta.

Valor Predictivo Negativo (VPN)

Ambos métodos tienen un alto VPN, pero el método_Umbral mantiene una ligera superioridad (95.65% frente a 94.87%), lo que indica que ambos son confiables para descartar correctamente los oídos sanos.

Tiempo Medio de Prueba

El método_RN también ofrece una ventaja significativa en términos de tiempo de prueba, con una reducción del tiempo medio de prueba a 14.88 segundos en comparación con 24.67 segundos en el método_Umbral. Esta es una variable a destacar, debido a las características del cribado, donde se hace necesario pesquisar la mayor cantidad de sujetos en el menor tiempo posible.

Conclusiones

La red neuronal multicapa se presenta como una herramienta potente en la clasificación de señales de Emisiones Otoacústicas Transientes (EOAT). El diseño de diversos modelos de redes neuronales multicapa permitió la selección de un modelo con resultados de entrenamiento muy favorables. El modelo escogido responde a una Wide Neural Network de una sola capa intermedia con 100 neuronas, y una función de activación Tanh. La red neuronal seleccionada se integró a un método de detección de EOAT como clasificador automático, y se implementó sobre el sistema de cribado auditivo neonatal Infantix. Se realizó un estudio en sujetos adultos para validar dicho método y se comparó con otro método automático basado en umbrales. El método basado en la red neuronal demostró un mejor rendimiento en términos de especificidad, valor predictivo positivo, y tiempo de prueba, haciendo que sea un algoritmo más confiable y eficiente. A pesar de una ligera reducción en la sensibilidad, la red neuronal proporciona una clasificación mucho más precisa al reducir los falsos positivos y ofrecer resultados en menor tiempo, lo cual es imprescindible en el contexto del cribado auditivo donde se requiere una alta fiabilidad y rapidez de la prueba.



Referencias

1. Olusanya BO, Wirz SL, Luxon LM. Community-based infant hearing screening for early detection of permanent hearing loss in Lagos, Nigeria: a cross-sectional study. *Bulletin of the World Health Organization*. 2008; 86 (12): 956-63.
2. Álvarez Amador HE, Vega Ulloa N, Castillo Toledo L, Santana Álvarez J, Betancourt Camargo MÁ, Miranda Ramos MI, Comportamiento de la hipoacusia neurosensorial en niños. *Revista Archivo Médico de Camagüey*. 2011;15, (5): 826-38.
3. Prensa Latina. Cuba consolida con éxito programas de cuidado del oído y la audición [Internet]. La Habana: Agencia Informativa Latinoamericana; 2024 [citado 19 ago 2024]. Disponible en: <https://archivo.prensa-latina.cu/2023/03/03/cuba-consolida-con-exito-programas-de-cuidado-del-oido-y-la-audicion>
4. Monsalve González A, Núñez Batalla F. La importancia del diagnóstico e intervención temprana para el desarrollo de los niños sordos: Los programas de detección precoz de la hipoacusia. *Psychosocial Intervention*, 2006; 15(1): 7-28.
5. Medina Alva DP, Kahn IC, Muñoz Huerta P, Leyva Sánchez J, Moreno Calixto J, Vega Sánchez SM. Neurodesarrollo infantil: características normales y signos de alarma en el niño menor de cinco años. *Revista Peruana de medicina experimental y salud pública*. 2015; 32: 565-73.
6. Moore JK. Maturation of human auditory cortex: implications for speech perception. *Annals of Otology, Rhinology & Laryngology*. 2002; 111(5): 7-10.
7. Martínez-Beneyto P, Morant A, Pitarch MI, Latorre E, Platero A, Marco J. La implantación coclear pediátrica en el periodo crítico de la vía auditiva, nuestra experiencia. *Acta Otorrinolaringológica Española*. 2009; 60 (5): 311-7.
8. U. P. S. T. Force, Universal screening for hearing loss in newborns: US Preventive Services Task Force recommendation statement. *Pediatrics*. 2008; 122 (1): 143-8.
9. Mayor LC, Burneo J, Ochoa J. Manual de electroencefalografía: Handbook of Electroencephalography. Bogotá: Ediciones Uniandes-Universidad de los Andes; 2013.
10. Botelho FA, Bouzada CF, Resende LM, Silva CF, Oliveira EA. Prevalence of hearing impairment in children at risk. *Brazilian Journal of Otorhinolaryngology*. 2010; 76: 739-44.
11. Ngo RY, Tan HK, Balakrishnan A, Lim SB, Lazaroo DT. Auditory neuropathy/auditory dys-synchrony detected by universal newborn hearing screening. *International journal of pediatric otorhinolaryngology*. 2006; 70 (7): 1299-306.
12. Bray P, Kemp D. An advanced cochlear echo technique suitable for infant screening. *British Journal of Audiology*. 1987; 21 (3): 191-204.
13. Akinpelu OV, Peleva E, Funnell WR, Daniel SJ. Otoacoustic emissions in newborn hearing screening: a systematic review of the effects of different protocols on test outcomes. *International journal of pediatric otorhinolaryngology*. 2014; 78 (5): 711-7.



14. Gorga MP, Preissler K, Simmons J, Walker L, Hoover B. Some issues relevant to establishing a universal newborn hearing screening program. *Journal of the American Academy of Audiology*. 2001; 12 (2): 101-12.
15. Ong KM, Rivera AS, Chan AL, Chiong CM. Determining concordance and cost impact of otoacoustic emission and automated auditory brainstem response in newborn hearing screening in a tertiary hospital. *International journal of pediatric otorhinolaryngology*. 2020; 128: 109704.
16. Levit Y, Himmelfarb M, Dollberg S. Sensitivity of the automated auditory brainstem response in neonatal hearing screening. *Pediatrics*. 2015; 136 (3): e641-e647.
17. Bonfils P. Spontaneous otoacoustic emissions: clinical interest. *The Laryngoscope*. 1989; 99 (7): 752-6.
18. Penner M, Zhang T. Prevalence of spontaneous otoacoustic emissions in adults revisited. *Hearing Research*. 1997; 103 (1-2): 28-34.
19. Morlet T. Spontaneous otoacoustic emissions in preterm neonates: prevalence and gender effects. *Hearing research*. 1995; 90 (1-2): 44-54
20. Infomed. Emisiones Otoacústicas (EOA) [Internet]. La Habana: Sociedad Cubana de Otorrinolaringología y Cirugía de Cabeza y Cuello; 2024 [citado 19 agos 2024]. Disponible en: <https://especialidades.sld.cu/otorrinolaringologia/2016/04/02/emisiones-otoacusticas-eoa-2/>
21. Delgado K, Ledesma S, Rostro H. Análisis de electroencefalograma usando redes neuronales artificiales. *Acta universitaria*. 2019; 29.
22. Pantoja-Gómez Y, Martín-González F, Torres-Fortuny A, Eimil-Suarez E, Velarde-Reyes E. Implementation of the NEURONIC INFANTIX Newborn Hearing Screening System. En: VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering: Proceedings of CLAIB-CNIB 2019, October 2-5, 2019, Cancún, México, 2019. Philadelphia: Springer Nature; 2019.pp. 453.
23. Pérez Blanco JG, González Fernández RI, Pantoja Gómez Y, Santos Ceballo JC, Velarde Reyes E. Registro de emisiones otoacústicas transientes basado en microcontroladores de alto rendimiento. *Revista Cubana de Informática Médica*. 2020; 12 (2).
24. Maxon AB, White KR, Behrens TR, Vohr BR. Referral rates and cost efficiency in a universal newborn hearing screening program using transient evoked otoacoustic emissions. *Journal of the American Academy of Audiology*. 1995; 6 (4): 271-7.



Conflicto de interés

Los autores declaran no tener conflictos de intereses que puedan comprometer la integridad científica del presente estudio.

Declaración de autoría

Yosvani Pantoja Gómez: Conceptualización integral del proyecto, desarrollo metodológico completo, diseño e implementación de la red neuronal MLP, obtención de los datos, procesamiento de los datos, análisis estadístico, redacción y edición del manuscrito en todas sus etapas.

José Carlos Santos Ceballo: Colaboración en el diseño de la arquitectura de la red neuronal y su implementación, revisión técnica del manuscrito y validación de la metodología computacional.

Jorge Germán Pérez Blanco: Contribución en la revisión crítica del manuscrito, validación de la metodología de procesamiento de los datos.

Ernesto Velarde Reyes: Supervisión general del proyecto, revisión final del manuscrito, análisis crítico de resultados y validación científica del estudio.

