Clasificación de género basada en señales de voz mediante modelos difusos y algoritmos de optimización

Gender classification based on voice signals using fuzzy models and optimization algorithms

Luis Miguel Cortés-Martinez
Universidad Distrital Francisco José de Caldas
Bogotá, Colombia
lmcortesm@correo.udistrital.edu.co

Helbert Eduardo Espitia-Cuchango
Universidad Distrital Francisco José de Caldas
Bogotá, Colombia
heespitiac@udistrital.edu.co

Resumen– En este documento se describe un esquema de clasificación de género, basado en señales de voz, en el que se proponen y prueban 16 modelos difusos diferentes que son optimizados mediante cuatro algoritmos bioinspirados y el método cuasi-Newton. El esquema de clasificación considera cuatro conjuntos de datos y cinco características de voz diferentes para definir los valores de entrada de un algoritmo en el proceso de optimización. Los valores de entrada de cada modelo difuso definen la media y varianza de sus funciones de pertenencia gaussianas, y su desempeño se evalúa mediante los valores de entrada del algoritmo de optimización y el error cuadrático medio como función objetivo para minimizar. Se hace un análisis comparativo entre modelos, algoritmos y conjuntos de datos para obtener conclusiones de acuerdo con los resultados de cada modelo optimizado.

Palabras clave– Lógica difusa, optimización, algoritmos genéticos, búsqueda armónica, evolución diferencial, optimización con enjambre de partículas, método cuasi-Newton, clasificación de género.

Abstract– This paper describes a gender classification scheme based on voice signals in which 16 different fuzzy models are proposed and optimized using four bio-inspired optimization algorithms and the quasi-Newton method. The classification scheme considers four data sets and five different voice features to define the input values of an algorithm in the optimization process. The inputs of each fuzzy model define the mean and variance of their Gaussian membership functions, and their fitness is evaluated by the input values of the algorithm and mean squared error as objective function to be minimized. A comparative analysis between models, algorithms and data sets is made to obtain conclusions according to the results of each optimized model.

Keywords– Fuzzy logic, optimization, genetic algorithms, harmony search, differential evolution, particle swarm optimization, quasi Newton method, gender classification.

1. INTRODUCCIÓN

Las señales de voz han sido estudiadas con diversos propósitos de clasificación, detección o reconocimiento. Se han planteado e implementado sistemas de clasificación de género [1]-[6]. En [4] y [5] se mencionan posibles aplicaciones y motivaciones de este objetivo.

Otros propósitos del uso de señales de voz es la formulación e implementación de sistemas para la distinción del habla [7]-[9], reconocimiento del habla [10], reconocimiento de lenguaje [11], reconocimiento de emociones basados en habla y en género [12], [13], diagnóstico de enfermedades patológicas como la disfonía y la laringitis [14], diagnóstico de nódulos, edemas y parálisis unilateral de las cuerdas vocales [15], reconocimiento de disartria en sistemas de reconocimiento automático de habla [16] y la detección temprana de Parkinson mediante voz [17].

Además de señales de voz, se han propuesto alternativas de clasificación que utilizan otras características de los individuos para clasificar el género. Algunos autores han propuesto la clasificación mediante el procesamiento de características fisionómicas humanas como la forma de la

Recibido: 25/02/2019 Aceptado: 16/05/ 2019 ITECKNE Vol. 16 Número 2 • diciembre 2019 • ISSN 1692-1798 • ISSN Digital 2339-3483 • 126 - 143
https://doi.org/10.15332/iteckne.v16i2.2356

cara [18], las manos [19], las orejas [20] y los iris [21]. Otros autores han utilizado estas características para la estimación de la edad y su clasificación por rangos [22]-[25].

La aparición y evolución de enfoques de clasificación que utilizan modelos de lógica difusa, redes neuronales artificiales, redes adaptativas de inferencia neurodifusa o aprendizaje automático en sistemas de control ha generado numerosos estudios e implementaciones en sistemas de detección, reconocimiento o clasificación. Se han propuesto modelos de redes neuronales [10], [13], [17], algoritmos genéticos [2], sistemas de inferencia neurodifusa (ANFIS) [3], y máquinas de soporte vectorial (SVM) [6] con el fin de clasificar con un margen de error bajo el género de una señal de voz humana. En este artículo, la clasificación se realiza mediante la aplicación de algoritmos de optimización a modelos difusos.

En este trabajo se presenta un esquema implementado en Matlab®, en el que se prueban 16 diferentes modelos de inferencia difusa (modelos difusos) de clasificación de género, los cuales se diferencian principalmente en la cantidad de entradas y los tipos de entrada especificados.

El objetivo principal del esquema propuesto es encontrar modelos de clasificación de género, basados en señales de voz humanas con un bajo margen de error; se hace la comparación de resultados entre modelos difusos, algoritmos de optimización bioinspirados y conjuntos de datos de entrada probados.

Para las entradas de los modelos difusos se definen cinco características por cada señal de voz. Cada modelo difuso propuesto utiliza tres, cuatro o cinco de estas características como entradas. Esto se realiza principalmente para determinar las entradas con mayor capacidad de clasificar géneros con un menor margen de error.

Los modelos son optimizados mediante cuatro algoritmos bioinspirados, y posteriormente optimizados, utilizando el método cuasi-Newton. Se implementan tres configuraciones diferentes para cada algoritmo bioinspirado por cada modelo en cada uno de los cuatro conjuntos de datos obtenidos del preprocesamiento de señales de voz.

El análisis de los resultados de desempeño de este trabajo permite encontrar el modelo difuso con la mayor capacidad para clasificar el género de una señal de voz humana. Adicionalmente, la interpretabilidad de este modelo hace posible la descripción de la voz perteneciente a un género determinado en función de las diferentes características de voz con las que se relaciona.

Por otra parte, la variedad de resultados obtenidos es analizada con el fin de establecer los modelos, conjuntos de datos, algoritmos bioinspirados y configuraciones de parámetros de mayor rendimiento.

La estructura de este documento es la siguiente. En la Sección 2 se definen las bases de lógica difusa. En la Sección 3, los algoritmos de optimización. En la Sección 4, las características de las señales de voz utilizadas. En la Sección 5 se presenta el esquema propuesto de clasificación. En la Sección 6 se hace el análisis cuantitativo y cualitativo de los resultados obtenidos. En la Sección 7 se listan las conclusiones obtenidas con base en la implementación y los resultados.

2. LÓGICA DIFUSA

La lógica difusa nace de la teoría de conjuntos difusos, base para el desarrollo del enfoque lingüístico [26], y sus valores de verdad pueden ser intermedios entre los valores verdadero y falso. De este modo, la lógica difusa, a diferencia de la lógica proposicional, no es dicotómica, cuantifica valores de verdad y tiene la capacidad de lidiar con la inferencia causal aproximada [27].

La lógica difusa es la base del funcionamiento de los sistemas de inferencia difusa basados en reglas [28]. Estos están compuestos esencialmente de entradas, salidas, funciones de pertenencia y reglas de inferencia. A través de estos componentes y su configuración se obtienen los valores estimados de salida como valores aproximados de verdad para la clasificación, de acuerdo con los valores de entrada. El proceso de diseño de un modelo difuso involucra la definición del conjunto de entradas y salidas del sistema difuso, así como la posición, forma y dominio de sus funciones de pertenencia y las reglas de inferencia. La clasificación con bajo margen de error depende del diseño del sistema, lo que puede depender del criterio, pruebas y validación de expertos en el área particular de investigación [29].

A diferencia del enfoque de redes neuronales, los sistemas difusos pueden ser interpretables [28], lo que los hace útiles en problemas de control con alta interpretabilidad. Esto permite la descripción de las salidas en términos de las entradas, reglas de inferencia y funciones de pertenencia. La certeza en la descripción de las salidas dependerá del valor de desempeño del sistema de clasificación difuso. Sin embargo, la interpretación de un sistema difuso puede dificultarse si su estructura es de alta complejidad [30]. La complejidad estructural de un sistema difuso depende proporcionalmente de la cantidad de reglas de inferencia y funciones de pertenencia en entradas y salidas. Los modelos difusos propuestos en este artículo tienen una estructura de complejidad moderada para evitar la obstrucción de su interpretabilidad.

En las últimas décadas, la aplicación de conjuntos difusos se ha visto ignorada en el campo del procesamiento del lenguaje natural y del habla, y se ha cuestionado la utilidad y contribución de la lógica difusa en las aplicaciones relacionadas con el procesamiento de señales de voz [31]. No obstante, la evolución de la computación y contribuciones tecnológicas en los últimos años han favorecido al enfoque de lógica difusa para su utilización en diversos campos de estudio [32].

3. ALGORITMOS DE OPTIMIZACIÓN

En esta Sección se definen cuatro algoritmos bioinspirados y el método cuasi-Newton.

Los algoritmos bioinspirados tienen el objetivo de optimizar problemas que presentan múltiples soluciones locales, con el fin de converger a una solución óptima global. Simulan la forma en que la naturaleza se enfrenta a problemas de optimización mediante la evolución natural de especies [33]. Por esta razón, estos algoritmos tienen naturaleza aleatoria y dependiente de múltiples parámetros. También existe la posibilidad de converger prematuramente a soluciones sobresalientes.

Atributos comunes en los algoritmos bioinspirados son una cantidad de iteraciones, una población o número de individuos, variables aleatorias y un criterio de finalización que se define principalmente con un número máximo de iteraciones o generaciones, con opciones adicionales, como detenerse al lograr un valor mínimo de desempeño, o una desviación estándar pequeña en los valores de desempeño de la población en una iteración.

Los algoritmos genéticos (GA), de búsqueda armónica (HS), de evolución diferencial (DE) y de optimización con enjambre de partículas (PSO) son los algoritmos bioinspirados que han sido utilizados para la optimización de los modelos difusos propuestos.

3.1 Algoritmos genéticos (Genetic Algorithms - GA)

Un algoritmo genético es una forma de evolución que ocurre en un computador. Los algoritmos genéticos son un método de búsqueda útil para resolver problemas de optimización y modelar sistemas evolutivos [34].

La forma más simple de algoritmo genético involucra tres tipos de operadores: selección, cruce y mutación [35], [36]. Estos operadores aplican iterativamente una cantidad de generaciones.

En primer lugar, los desempeños de cada individuo en la población son evaluados. Luego, el proceso de selección genera valores de expectativa por individuo para influenciar su probabilidad de reproducción.

La reproducción es simulada mediante los algoritmos de cruce y mutación que utilizan los individuos seleccionados como padres para generar individuos nuevos para la próxima generación.

El cruce, también llamado recombinación, consiste en combinar el genotipo de dos padres en un solo individuo. La mutación, en cambio, es la alteración genética aleatoria en el genotipo de un único padre en la población.

Los operadores de cruce y mutación pueden verse como maneras de mover a la población en el paisaje definido por la función de aptitud [35]. Con estos operadores se busca la convergencia a un valor óptimo global y se logra una exploración mayor en el espacio de posibles soluciones.

Esta secuencia se repite por una cantidad de generaciones o hasta cumplir con el criterio de finalización. Tanto la mutación como la recombinación de individuos son procesos estocásticos, pues tienen un valor asociado de probabilidad.

El diagrama de flujo que describe el funcionamiento de los algoritmos genéticos se muestra en la Fig. 1.

Fig. 1. DIAGRAMA DE FLUJO DE ALGORITMO GA

Fuente: Los autores.

3.2 Algoritmo de búsqueda armónica (Harmony Search - HS)

La armonía musical es una combinación de sonidos considerada agradable desde un punto de vista estético. La armonía en la naturaleza es una relación especial entre varias ondas de sonido que tienen diferentes frecuencias [37].

El algoritmo de búsqueda armónica imita la improvisación musical en la que los músicos intentan encontrar mejores armonías basadas en la aleatoriedad o sus experiencias [38].

Se compone de una memoria armónica (HM) en la que se encuentran conjuntos de valores que representan las notas en una armonía. Esta memoria tiene un tamaño específico (HMS).

En el contexto de optimización, una nota es un valor de atributo de un individuo, y una armonía o conjunto de notas en la memoria armónica, es el individuo con todos sus atributos.

En cada iteración, las armonías existentes en la memoria armónica son improvisadas de acuerdo con una tasa de consideración de la memoria armónica (HMCR).

El proceso de improvisación de armonías puede utilizar una armonía de la memoria armónica (familiar) y alterar sus notas con un valor de probabilidad y una tasa de ajuste de tono (PAR), o generar una armonía completamente aleatoria dependiendo de un valor aleatorio que es comparado con el valor de HMCR como se define en la ecuación (1), donde A_i es una armonía y p_i un valor de probabilidad.

$$A_i = \begin{cases} A'_{HMSi}, & p_i < HMCR \\ A'_{rand}, & p_i \geq HMCR \end{cases} \quad (1)$$

Siendo A_i una armonía, p_i un valor aleatorio entre 0 y 1, A'_{HMS_i} una armonía en la memoria armónica con posibles modificaciones, y A_{rand} un acorde aleatorio.

Si alguna de las armonías producidas en una iteración tiene mejor desempeño que al menos una armonía en la memoria armónica, esta ocupará el espacio de la peor armonía en la memoria.

La tasa de ajuste de tono (PAR) es un valor opcional del algoritmo que imita el ajuste de tono de cada instrumento para afinar el conjunto [37]. El mecanismo de ajuste de tono se diseña como el desplazamiento de una nota a valores vecinos dentro de un rango de valores posibles. Esta tasa puede ser un valor constante, o un valor decreciente que inicia en un valor máximo y llega a un valor mínimo en la última iteración, con el fin de explorar en mayor profundidad la región del espacio de posibles soluciones donde se estima la ubicación del valor óptimo global.

El diagrama de flujo que describe el funcionamiento del algoritmo de búsqueda armónica se muestra en la Fig. 2.

3.3 Algoritmo de evolución diferencial (Differential Evolution – DE)

La idea crucial detrás de este algoritmo es un esquema para generar vectores de parámetros de prueba [39]. El algoritmo de evolución diferencial genera nuevos vectores de parámetros (individuos) al agregar un vector de diferencia ponderada entre dos miembros de la población a un tercer miembro. Si el vector resultante produce un valor de función objetivo más bajo que un miembro de población predeterminado, el vector recién generado reemplazará el vector con el que se comparó en la siguiente generación.

Para variar los parámetros en la población se define un operador de variación diferencial, que se realiza mediante el cambio en el valor de los parámetros de los vectores en una iteración, mediante la ecuación (2).

$$V_{r1,G} = X_{r1,G} + F(X_{r2,G} - X_{r3,G}) \quad (2)$$

Siendo el peso F un valor constante, y $X_{r1,G}$, $X_{r2,G}$, $X_{r3,G}$ vectores de la matriz de población en la generación G.

Luego se aplica un operador de cruce probabilístico, que funciona con una tasa de cruce y valores aleatorios por cada elemento de cada vector en la población. Esta es la primera estrategia definida para este algoritmo por Storn y Price [39].

Los elementos con un valor menor a la tasa de cruce (CR) toman el valor respectivo del vector V. Los elementos restantes, con un valor aleatorio mayor o igual a la tasa de cruce, permanecen sin cambios. El valor de cada parámetro j en cada vector i se muestra en la ecuación (3).

$$U_{i,j} = \begin{cases} V_{i,j}, & p_{i,j} < CR \\ X_{i,j}, & p_{i,j} \geq CR \end{cases} \quad (3)$$

Siendo $p_{i,j}$ un valor aleatorio entre 0 y 1.

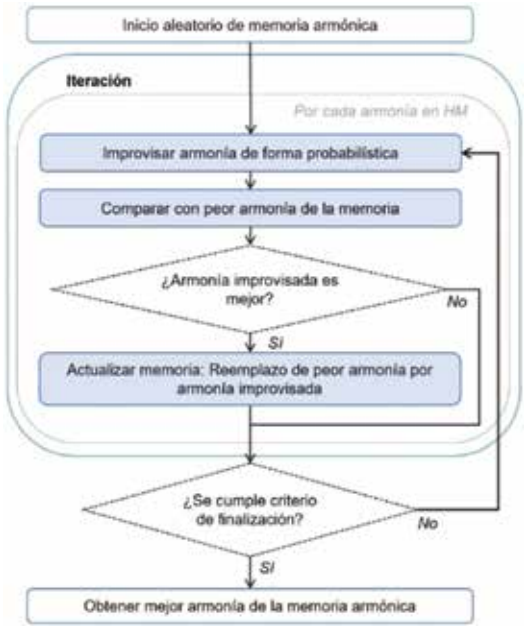
Una vez finalizado el proceso de cruce, se realiza el proceso de selección vector por vector de la matriz de población U. En este proceso, cada vector es comparado con los demás en la población. Un vector será miembro de la población de la siguiente generación si y solo si su desempeño es mejor que el desempeño de otro vector en la población de la generación actual; de lo contrario, la población no se verá alterada.

El diagrama de flujo que describe el funcionamiento del algoritmo de evolución diferencial se muestra en la Fig. 3.

3.4 Algoritmo de optimización con enjambre de partículas (Particle Swarm Optimization - PSO)

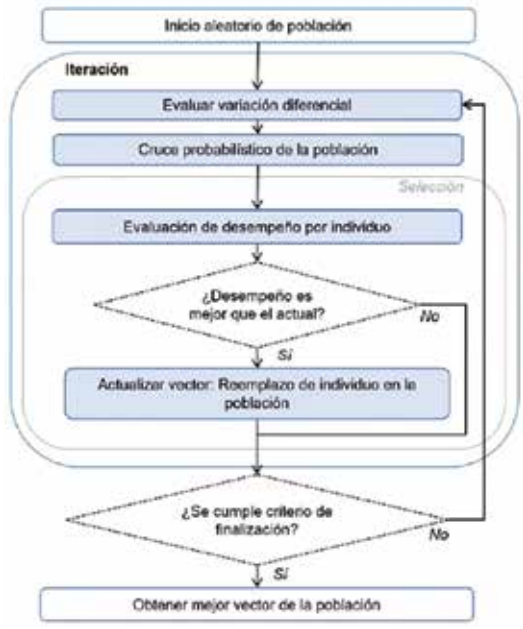
El algoritmo de optimización con enjambre de partículas se inspira en el comportamiento social emergente que puede observarse en bandadas de aves y cardúmenes de peces de algunas especies [40].

Fig. 2. DIAGRAMA DE FLUJO DE ALGORITMO HS



Fuente: Los autores.

Fig. 3. DIAGRAMA DE FLUJO DE ALGORITMO DE



Fuente: Los autores.

La población está compuesta de partículas, siendo la posición de cada una de ellas un punto solución en el espacio de soluciones posibles de un número de dimensiones igual al número de parámetros del problema.

En cada iteración se actualiza la mejor posición de cada partícula (mejor posición individual),

y la mejor posición del enjambre (mejor posición global). Con estos valores se calcula la velocidad de cada partícula mediante la ecuación (4).

$$\vec{v_1}(n+1) = w(n)\vec{v_1}(n) + \beta_{i(1)}(\vec{x_{pi}} - \vec{x_i}(n)) + \beta_{2(1)}(\vec{x_g} - \vec{x_i}(n)) \quad (4)$$

Donde:

i es el índice del individuo.

n es el índice de iteración.

$\vec{v_i}$ es la velocidad del i-ésimo individuo.

$\vec{x_i}$ es la posición del i-ésimo individuo.

$w(n)$ es la función de inercia.

$\vec{x_{pi}}$ es la mejor posición del i-ésimo individuo.

$\vec{x_g}$ es la mejor posición del enjambre.

β son valores aleatorios entre 0 y 1.

Una vez calculadas las velocidades se determina la posición de cada partícula para la siguiente iteración mediante la ecuación (5).

$$\vec{x_i}(n+1) = \vec{x}_i(n) + \vec{v_i}(n+1) \quad (5)$$

La inercia puede utilizarse como un valor constante o definirse con otros enfoques específicos [41]. Uno de estos enfoques define a la inercia como un valor linealmente dependiente de las iteraciones como se define en la ecuación (6) [42].

$$w(n) = w_{\max} - \frac{n(w_{\max} - w_{\min})}{N} \quad (6)$$

Donde $w_{\max}$ y $w_{\min}$ son los valores máximos y mínimos de inercia y N es el número total de iteraciones.

El diagrama de flujo que describe el funcionamiento del algoritmo de optimización con enjambre de partículas se muestra en la Fig. 4.

3.5 Método cuasi-Newton (Quasi-Newton method)

El método cuasi-Newton es un método de optimización basado en gradiente [43], útil en la optimización de problemas en los que se busca converger a una solución óptima local de una función objetivo a minimizar, partiendo de un punto inicial en el espacio de soluciones posibles.

Este se basa en el método de Newton, el cual establece que para un punto y una función objetivo , existe un punto dado por la ecuación (7) que se aproxima a un mínimo local en el espacio de soluciones.

$$\underline{X}_{k+1} = \underline{X}_k - (\nabla^2(f(\underline{X}_k)))^{-1}\nabla(f(\underline{X}_k)) \quad (7)$$

Donde ∇ es el gradiente de la función objetivo, ∇^2 es el Hessiano de la función objetivo, y $\underline{X}_k$ es un punto mínimo estimado de la función.

El cálculo del Hessiano de una función objetivo puede tener un alto costo computacional, especialmente cuando las funciones no son lineales y se definen con múltiples variables [43].

El método cuasi-Newton calcula un valor aproximado del Hessiano. Se han desarrollado métodos de actualización del Hessiano como la fórmula de Davidon, Fletcher y Powell (DFP), y la fórmula de Broyden, Fletcher, Goldfarb y Shanno (BFGS). Esta última se considera la más efectiva para el uso en un método de propósito general [44], [45].

El esquema general del algoritmo se muestra en la Fig. 5.

4. CARACTERÍSTICAS DE LAS SEÑALES DE AUDIO

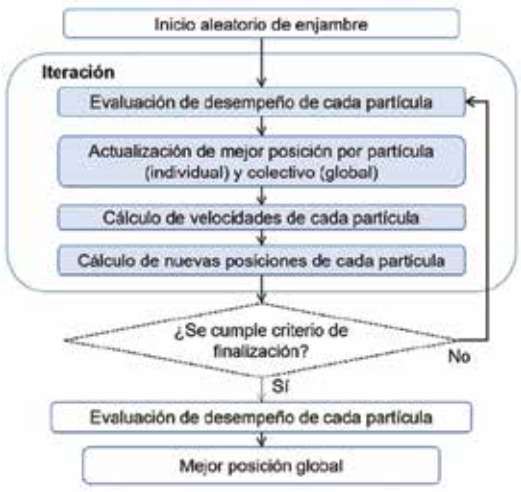
Esta Sección describe las cinco entradas utilizadas para los modelos difusos propuestos, abreviadas mediante las siglas EE, STE, ZCR, RMS y PSC.

Algunos artículos, con el propósito de clasificar el género utilizan las entradas EE, STE y ZCR como características fundamentales de las voces [1], [6]. En [46] se ha planteado que las entradas STE y ZCR pueden clasificar sonidos que se asocian a entornos específicos.

En [1] se ha planteado que las características STE y ZCR tienen un valor bajo en voces masculinas, y un valor alto y continuo en voces femeninas.

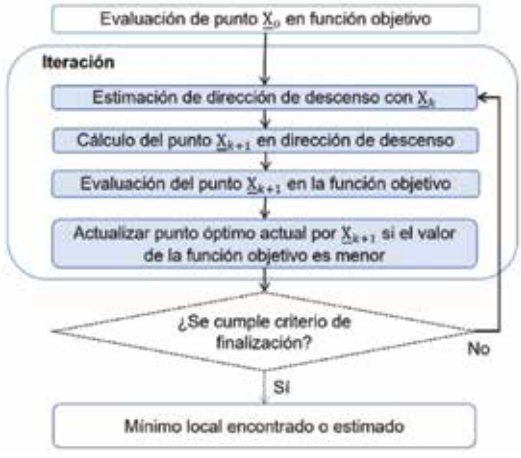
Otros artículos han trabajado con los coeficientes MFCC (Mel Frequency Cepstral Coefficients) para la clasificación de género [47], reconocimiento de oradores [10], y detección de patologías por medio de señales de voz [7], [15], [48].

Fig. 4. DIAGRAMA DE FLUJO DE ALGORITMO PSO



Fuente: Los autores.

Fig. 5. DIAGRAMA DE FLUJO MÉTODO CUASI-NEWTON



Fuente: Los autores.

4.1 EE (Energy Entropy)

La entropía en la señal de voz se mide con los cambios repentinos en el nivel de energía de una señal de voz [1]. Para calcularla, la señal de voz se divide inicialmente en k ventanas y luego se calcula la energía normalizada de la señal por cada ventana.

El valor de entropía es la suma de todos los valores calculados. La entropía de la señal está dada por la ecuación (8), donde es la energía de la señal normalizada.

$$EE = \sum_{i=0}^{k-1} \sigma^2 \log(\sigma^2) \quad (8)$$

Esta característica se ha utilizado para otros propósitos relacionados, como la detección del habla [9].

4.2 STE (Short Time Energy)

La característica STE de una señal de voz busca ver la tendencia que esta tiene de aumentar repentinamente en lapsos cortos de tiempo. El valor STE de una ventana de la señal se calcula mediante la ecuación (9).

$$STE = \sum_{r=-\infty}^{\infty} \frac{y(r)^2 h(s-r)}{S} \quad (9)$$

Donde y es una señal, s es el tamaño de una ventana de la señal, y $h(r)$ es la función de ventana definida en la ecuación (19).

$$h(r) = \begin{cases} 1, & 0 \le r \le s \\ 0, & en\ otro\ caso \end{cases} \quad (10)$$

Normalmente, la energía es alta si hay una voz en la señal [2]. Se ha planteado esta característica para la detección de frecuencia en el habla [49].

Esta entrada puede ser de utilidad para detectar los momentos en los que una señal tiene amplitud insuficiente facilitando la remoción de silencio de las señales de voz [50], [51]. Es posible detectar el silencio de una señal al establecer un umbral mínimo de STE de modo que los valores menores a este umbral pueden catalogarse como silencios en una señal.

4.3 ZCR (Zero Crossing Rate)

Para una señal descrita por un vector X, la tasa de cruces por cero (ZCR) se calcula por pares consecutivos de X como lo describe la ecuación (11).

$$ZCR = \frac{1}{2N} \sum_{r=1}^{N-1} sgn\{X(i)\} - sgn\{X(i-1)\} \quad (11)$$

Siendo sgn la función signo, definida en la ecuación (12).

$$sgn\{x(i)\} = \begin{cases} 1; X(i) > 0 \\ 0; X(i) = 0 \\ -1; X(i) < 0 \end{cases} \quad (12)$$

4.4 RMS (Root Mean Square)

Para una señal descrita por un vector , el valor cuadrático medio de una señal está definido por la ecuación (13).

$$RMS = \sqrt{\frac{\sum_{i=1}^{N} X^2(i)}{N}} \quad (13)$$

Se define como la raíz cuadrada del cuadrado medio, es decir, de la media aritmética de los cuadrados de un conjunto numérico. La diferencia entre esta entrada y STE es que esta característica no es calculada por ventanas, sino que resume el comportamiento de la señal total en un único valor. Además, como se trata de la raíz cuadrada tiene una sensibilidad menor que STE.

4.5 PSC (Perceptual Spectral Centroid)

El centroid espectral es una medida utilizada en el procesamiento digital de señales para caracterizar el espectro de frecuencia de una señal. Investigadores creen que la calidad del brillo del timbre se correlaciona con el aumento de la potencia en altas frecuencias [52]. Esta medida de frecuencia indica la posición en la que se encuentra el baricentro del espectro.

Perceptualmente, tiene una conexión robusta con la impresión de brillo del sonido [53]. Se calcula como la media ponderada de frecuencias, determinada mediante la transformada de Fourier, con sus magnitudes en el histograma de la señal. Se calcula mediante la ecuación (14).

$$PSC = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (14)$$

Siendo $x(n)$ el valor de frecuencia ponderada de una ubicación n en el histograma de la señal, y $f(n)$ la frecuencia central de esa ubicación [54].

Se ha utilizado esta medida para estudiar el brillo en el sonido de instrumentos musicales y clasificar las emociones que son capaces de evocar los sonidos [55].

5. ESQUEMA DE CLASIFICACIÓN PROPUESTO

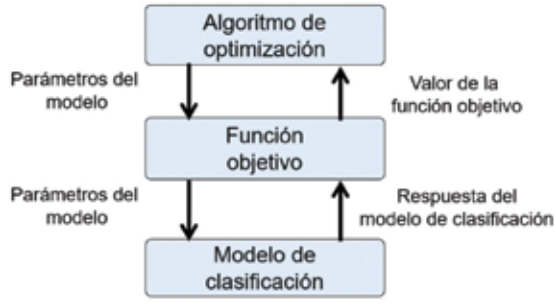
El esquema propuesto abarca 4 algoritmos bioinspirados, 4 conjuntos de datos, 3 configuraciones diferentes por algoritmo bioinspirado y 16 modelos difusos, dando lugar a 4x4x3x16=768 combinaciones posibles de clasificación probadas, es decir, 768 resultados de indicadores de desempeño.

Cada modelo difuso es optimizado por un algoritmo bioinspirado, y una vez el algoritmo bioinspirado termina, el modelo es optimizado localmente con el método cuasi-Newton.

5.1 Proceso de optimización

La evaluación de un individuo en el proceso de optimización con un algoritmo se hace mediante la construcción de un modelo difuso con los parámetros del individuo y la estructura definida en la función objetivo. Esta función es calculada con las respuestas de clasificación del modelo construido. La Fig. 6 muestra el esquema general que el algoritmo de optimización utiliza para la evaluación de los individuos, siendo el modelo de clasificación un modelo difuso de los 16 propuestos.

Fig. 6. ESQUEMA DE EVALUACIÓN DE MODELOS DIFUSOS CON UN ALGORITMO DE OPTIMIZACIÓN



Fuente: Los autores.

Un modelo difuso de clasificación es la representación funcional de un individuo en la población o un punto en el espacio de soluciones.

La función objetivo establecida es el error cuadrático medio (MSE). Con esta se evalúa el desempeño de todo modelo de clasificación difuso.

La Fig. 7 muestra los pasos que se realizan en un proceso de optimización con cualquier combinación posible probada de conjunto de datos, modelo y algoritmo bioinspirado.

Cada señal de voz se convierte en una serie de valores que se fija con un mecanismo de preprocesamiento de los cuatro definidos en la Subsección 5.3, y de cada serie se extraen las caracte-

rísticas definidas en la Sección 4; luego se realiza el proceso de optimización de un modelo difuso propuesto, el cual involucra evaluaciones iterativas del mismo. Si el modelo optimizado tiene resultados de desempeño altos, se hace un análisis de su interpretación.

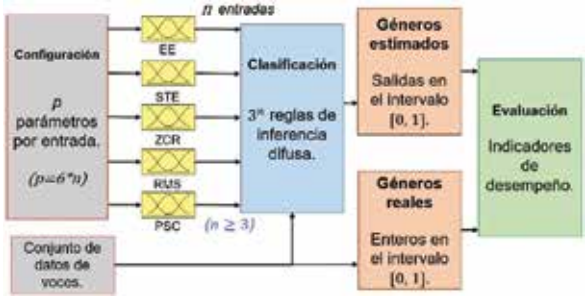
Fig. 7. PASOS EN EL PROCESO DE OPTIMIZACIÓN DE UN MODELO DIFUSO

Fuente: Los autores.

5.2 Estructura de los modelos difusos propuestos

La estructura general de los modelos difusos propuestos se muestra en la Fig. 8.

Fig. 8. ESTRUCTURA DE MODELOS DIFUSOS PROPUESTOS



Fuente: Los autores.

Los rangos de asignación factible de las funciones de pertenencia tienen en cuenta el valor mínimo y máximo por cada una de las entradas para el modelo difuso definidas en la Sección 4. Por ejemplo, para un modelo que utilice las primeras tres entradas, los rangos son los definidos en la ecuación (15).

$$\{[EE_{\min}, EE_{\max}], [STE_{\min}, STE_{\max}], [ZCR_{\min}, ZCR_{\max}]\} \quad (15)$$

Con la normalización de cada uno de estos rangos entre los valores 0 (mínimo) y 1 (máximo), se establece un rango general para todas las entradas y se promueve la exploración factible en los algoritmos de optimización en un espacio fijo de soluciones de acuerdo con los datos procesados.

En este documento se trabaja con modelos difusos tipo Sugeno. Este tipo de modelo tiene ventaja sobre los modelos tipo Mamdani, dado que requiere menos memoria y posee mayor velocidad [56]. Con este modelo se reduce la complejidad de evaluación por señal de voz, sin perjudicar el propósito de clasificación de género ni la utilidad de la información suministrada por las entradas.

Cada modelo de inferencia difusa produce una única salida compuesta de tres funciones de pertenencia de tipo constante entre 0 y 1. Esta indica el género estimado para una señal de voz, donde 0 representa al género femenino, 0.5 representa un género indefinido, y 1 representa al género masculino.

La forma gaussiana de las funciones de pertenencia es considerada más flexible que la forma triangular, y se aproxima más a los mecanismos de inferencia humanos [57], [58]. A diferencia de las funciones de pertenencia triangulares, estas se definen con dos parámetros: media y varianza.

Se definen tres funciones de pertenencia gaussianas en cada entrada del modelo difuso. Tres se considera el número suficiente de funciones de pertenencia para inducir la interpretabilidad del modelo [28], mantener bajo control la complejidad de este y evitar la redundancia en las funciones de pertenencia.

El número de parámetros que configuran estas funciones de pertenencia es de dos parámetros por cada una de las tres funciones de pertenencia en cada entrada considerada. Así, el valor de parámetros de un sistema difuso, en función de sus entradas, está dado por la ecuación (16).

$$p = 6n \quad (16)$$

Estos parámetros se encuentran en el intervalo para definir la media, y en el intervalo para definir la varianza de cada función de pertenencia. Fuera de estos intervalos se aumenta la probabilidad de obtener funciones de pertenencia redundantes o de poca influencia para las salidas en el modelo optimizado.

El número de reglas de un modelo difuso en función del número de entradas está dado por la ecuación (17).

$$R = 3^n \quad (17)$$

Siendo el número de entradas del modelo difuso a optimizar. Estas reglas utilizan el operador AND, para evaluar toda combinación posible de funciones de pertenencia por entrada para su ajuste en el proceso de optimización.

Los modelos difusos propuestos se diferencian principalmente en el conjunto de entradas y la cantidad de estas. La Tabla I muestra los modelos propuestos probados.

Tabla I.
MODELOS DIFUSOS PROPUESTOS

Sistema	No. Entradas	Entradas				
		EE	STE	ZCR	RMS	PSC
1	3	X	X	X		
2	3	X	X		X	
3	3	X	X			X
4	3	X		X	X	
5	3	X		X		X
6	3	X			X	X
7	3		X	X	X	
8	3		X	X		X
9	3		X		X	X
10	3			X	X	X
11	4	X	X	X	X	
12	4	X	X	X		X
13	4	X	X		X	X
14	4	X		X	X	X
15	4		X	X	X	X
16	5	X	X	X	X	X

Fuente: Los autores.

5.3 Conjuntos de datos propuestos

Se han implementado alteraciones sobre archivos de audio con el fin de atenuar [59], segmentar [60], detectar silencios [50], [51], y comprimir señales de voz mediante cuantización vectorial [61]. En este artículo, los fragmentos de cada señal normalizada en los que el valor de STE es menor al 1% fueron removidos para evitar el procesamiento innecesario de silencios en el cálculo de sus características.

Para la obtención de los datos se ha establecido un conjunto de 50 archivos de audio con una señal de voz masculina o femenina. El 78% de los archivos de audio son obtenidos de una base de datos [62], y el 22% restante fue creado en un entorno arbitrario. Todos los archivos de audio pronuncian una palabra o frase legible. El 50% de los archivos pertenecen a voces masculinas, y el 50% restante a voces femeninas.

Se prueba la optimización con cuatro conjuntos de datos obtenidos de una base de datos de señales de voz, y cuatro mecanismos de preprocesamiento de las señales de audio, con el propósito de variar los valores de entrada sin alterar la base de datos de voces y determinar cuál de estos es clasificado con mayor exactitud. Los cuatro mecanismos de preprocesamiento de señales son:

- Procesamiento puro de la señal (sin filtro promediador ni transformada rápida de Fourier).
- Procesamiento de la señal aplicando transformada rápida de Fourier.
- Procesamiento de la señal con filtro promediador de la señal en 10 períodos anteriores y sin transformada rápida de Fourier.
- Procesamiento de la señal con filtro promediador de la señal en 10 períodos anteriores y transformada rápida de Fourier.

Por cada mecanismo se produce un conjunto de datos diferente. Los conjuntos producidos serán identificados como conjuntos 1, 2, 3 y 4, respectivamente.

5.4 Indicadores de desempeño

Los indicadores de desempeño son calculados en cada modelo difuso propuesto y género, y son el medio por el que se determinan los mejores resultados. Estos indicadores son la exactitud pura, la exactitud de acierto y el error cuadrático medio.

Para el cálculo de estos indicadores, toda señal de voz evaluada en un modelo difuso tiene tres valores: género real, género estimado y género definido.

Los valores de género real y género definido son enteros, donde 1 representa el género masculino, y 0 representa el género femenino. El valor de género estimado tiene un valor real que oscila entre estos valores.

El primer indicador, la exactitud pura (EP), es el valor que describe el éxito de clasificación del sistema sin un umbral definido. Mientras mayor sea la diferencia entre las salidas obtenidas (género estimado), y las esperadas (género real), la exactitud pura es menor. Se define mediante la ecuación (18).

$$EP = \frac{1}{N}\sum_{i=1}^{N} 1 - \left| G_{Estimado_i} - G_{Real_i} \right| \quad (18)$$

Donde N es el número de señales de voz procesadas y clasificadas por el modelo difuso.

El segundo indicador, la exactitud aproximada o de acierto (EA), a diferencia de la exactitud pura, utiliza el género definido que utiliza un umbral para clasificar la salida como masculina o femenina. El género definido se calcula mediante la ecuación (19).

$$G_{Definido_i} = \begin{cases} 1\, si\, G_{Estimado_i} > U \\ 0\, si\, G_{Estimado_i} \leq U \end{cases} \quad (19)$$

Donde el umbral U para determinar el género es de 0,5. Si la salida final es mayor al umbral, el género será clasificado como masculino, y de otro modo, femenino. La exactitud de acierto se define mediante la ecuación (20).

$$EA = \frac{1}{N}\sum_{i=1}^{N} 1 - \left| G_{Definido_i} - G_{Real_i} \right| \quad (20)$$

El propósito de la EA es concretar un género de acuerdo con la exactitud pura, siendo un indicador dicotómico de acierto o fallo por señal de voz.

El tercer indicador, el error cuadrático medio (MSE), es la media aritmética de la suma de las diferencias entre las salidas esperadas y las salidas obtenidas al cuadrado:

$$WSE = \frac{1}{N}\sum_{i=1}^{N} \left(G_{Estimado_i} - G_{Real_i} \right)^2 \quad (21)$$

El MSE es el parámetro más usado en el propósito de prueba de modelos [63], y es usado en este artículo como indicador de desempeño (fitness) en los algoritmos de optimización.

Se busca que el MSE sea el mínimo posible. Si este valor es menor, los indicadores de exactitud tienden a ser mayores y los resultados obtenidos (géneros estimados) se aproximan a los deseados (géneros reales).

5.5 Configuración de los algoritmos de optimización

Para hacer la comparación de resultados entre algoritmos bioinspirados y sus configuraciones, se probó un número de iteraciones de 500, y una población de 20 individuos en todas las configuraciones.

La Tabla II muestra las tres diferentes configuraciones por cada algoritmo bioinspirado.

Tabla II.
CONFIGURACIONES DE PARÁMETROS EN ALGORITMOS BIOINSPIRADOS

Opc. GA	Población	Generaciones	P. cruce	P. mutación
A	20	500	0.60	0.005
B			0.90	0.01
C			0.50	0.04
Opc. HS	HMS	Iteraciones	HMCR	PAR
A	20	500	0.90	0.50
B			0.95	0.50
C			0.95	0.35
Opc. DE	Vectores	Generaciones	P. cruce	Peso (F)
A	20	500	0.98	0.68
B			0.95	0.58
C			1.00	0.75
Opc. PSO	Partículas	Iteraciones	β	Inercia (w)
A	20	500	1.70	0.60
B			1.49	0.729
C			2.05	1.00

Fuente: Los autores.

Para el método cuasi-Newton se establece una única configuración con un máximo de 100 iteraciones, y un máximo de 10.000 evaluaciones de la función objetivo.

6. ANÁLISIS DE RESULTADOS

En esta Sección se lista una serie de tablas y gráficas con los resultados de indicadores de desempeño, en las que se establecen los mejores conjuntos de datos y modelos por algoritmo bioinspirado.

6.1 Análisis cuantitativo

Las tablas III a VI resumen los resultados de desempeño en cada opción por cada algoritmo bioinspirado, independientemente del modelo difuso de clasificación.

Los resultados de los mejores modelos difusos de clasificación optimizados encontrados por opción en cada algoritmo son especificados en las tablas VII a X, y se determina la mejor opción y modelo con base en sus indicadores de desempeño.

Los mejores resultados por modelo se muestran en la Tabla XI, con el algoritmo bioinspirado y conjunto de datos utilizado para lograr cada uno de sus indicadores. La Tabla XII muestra los resultados promedio por número de entradas y la Tabla XIII muestra los resultados promedio de los modelos con una entrada en común.

Tabla III.

RESULTADOS DE ALGORITMO GA POR OPCIÓN EN CADA CONJUNTO DE DATOS

GA	Conjunto	Veces mejor en opción			Promedio de indicadores por opción		
		EP	EA	MSE	EP	EA	MSE
Opción A	1	43,75%	52,94%	50,00%	62,39%	72,25%	18,07%
	2	31,25%	41,18%	37,50%	62,71%	72,50%	17,81%
	3	12,50%	0,00%	6,25%	58,38%	65,38%	20,36%
	4	12,50%	5,88%	6,25%	59,48%	68,88%	19,87%
Opción B	1	37,50%	25,00%	37,50%	63,27%	71,13%	17,36%
	2	50,00%	50,00%	56,25%	63,90%	74,25%	17,47%
	3	12,50%	10,00%	6,25%	59,61%	64,63%	19,78%
	4	0,00%	15,00%	0,00%	59,52%	67,88%	19,74%
Opción C	1	43,75%	57,89%	50,00%	68,80%	76,13%	15,18%
	2	25,00%	26,32%	31,25%	64,69%	73,50%	16,70%
	3	12,50%	0,00%	12,50%	62,11%	68,38%	18,63%
	4	18,75%	15,79%	6,25%	63,01%	70,25%	18,25%
Total veces mejor	1	41,67%	44,64%	45,83%	Promedio final		
	2	35,42%	39,29%	41,67%	62,32%	70,43%	18,27%
	3	12,50%	3,57%	8,33%			
	4	10,42%	12,50%	4,17%			
Mejor conjunto de datos		1	1	1			

Fuente: Los autores.

Tabla IV.

RESULTADOS DE ALGORITMO HS POR OPCIÓN EN CADA CONJUNTO DE DATOS

HS	Conjunto	Veces mejor en opción			Promedio de indicadores por opción		
		EP	EA	MSE	EP	EA	MSE
Opción A	1	37,50%	25,00%	43,75%	58,33%	66,25%	19,90%
	2	43,75%	25,00%	25,00%	59,05%	67,25%	20,10%
	3	0,00%	6,25%	6,25%	55,82%	60,88%	21,81%
	4	18,75%	43,75%	25,00%	58,31%	67,00%	20,44%
Opción B	1	37,50%	36,84%	43,75%	58,82%	69,25%	19,56%
	2	37,50%	31,58%	37,50%	59,08%	69,13%	20,01%
	3	0,00%	5,26%	0,00%	55,14%	60,00%	22,15%
	4	25,00%	26,32%	18,75%	58,60%	66,50%	20,57%
Opción C	1	50,00%	38,89%	50,00%	60,44%	68,63%	19,03%
	2	12,50%	44,44%	18,75%	58,73%	69,63%	19,89%
	3	18,75%	11,11%	25,00%	56,84%	61,50%	21,28%
	4	18,75%	5,56%	6,25%	57,67%	64,25%	21,22%
Total veces mejor	1	41,67%	33,96%	45,83%	Promedio final		
	2	31,25%	33,96%	27,08%	58,07%	65,85%	20,50%
	3	6,25%	7,55%	10,42%			
	4	20,83%	24,53%	16,67%			
Mejor conjunto de datos		1	1 y 2	1			

Fuente: Los autores.

Tabla V.

RESULTADOS DE ALGORITMO DE POR OPCIÓN EN CADA CONJUNTO DE DATOS

DE	Conjunto	Veces mejor en opción			Promedio de indicadores por opción		
		EP	EA	MSE	EP	EA	MSE
Opción A	1	50,00%	50,00%	56,25%	65,75%	74,50%	16,43%
	2	25,00%	22,73%	18,75%	64,86%	72,63%	17,06%
	3	6,25%	4,55%	6,25%	61,93%	68,25%	18,51%
	4	18,75%	22,73%	18,75%	63,29%	71,75%	17,94%
Opción B	1	37,50%	44,44%	50,00%	65,86%	73,38%	16,53%
	2	25,00%	27,78%	31,25%	63,70%	71,50%	17,69%
	3	0,00%	0,00%	0,00%	59,05%	66,63%	20,00%
	4	37,50%	27,78%	18,75%	63,30%	71,75%	17,91%
Opción C	1	31,25%	50,00%	37,50%	63,43%	72,75%	17,29%
	2	37,50%	25,00%	25,00%	62,63%	72,38%	18,04%
	3	6,25%	0,00%	6,25%	59,31%	66,25%	19,89%
	4	25,00%	25,00%	31,25%	62,36%	69,38%	18,65%
Total veces mejor	1	39,58%	48,21%	47,92%	Promedio final		
	2	29,17%	25,00%	25,00%	62,96%	70,93%	18,00%
	3	4,17%	1,79%	4,17%			
	4	27,08%	25,00%	22,92%			
Mejor conjunto de datos		1	1	1			

Fuente: Los autores.

Tabla VI.

RESULTADOS DE PSO POR OPCIÓN EN CADA CONJUNTO DE DATOS

PSO	Conjunto	Veces mejor en opción			Promedio de indicadores por opción		
		EP	EA	MSE	EP	EA	MSE
Opción A	1	31,25%	25,00%	43,75%	64,19%	71,13%	17,36%
	2	31,25%	35,00%	31,25%	62,61%	71,75%	18,18%
	3	6,25%	0,00%	6,25%	59,80%	63,63%	19,72%
	4	31,25%	40,00%	18,75%	62,96%	70,50%	18,69%
Opción B	1	37,50%	44,44%	56,25%	63,19%	71,75%	17,50%
	2	37,50%	27,78%	31,25%	62,54%	71,63%	18,17%
	3	6,25%	5,56%	12,50%	58,83%	64,63%	20,13%
	4	18,75%	22,22%	0,00%	61,92%	69,75%	19,52%
Opción C	1	43,75%	38,89%	50,00%	64,83%	72,13%	17,00%
	2	25,00%	33,33%	31,25%	63,25%	72,75%	17,87%
	3	0,00%	0,00%	0,00%	58,58%	66,88%	20,24%
	4	31,25%	27,78%	18,75%	63,48%	73,75%	17,92%
Total veces mejor	1	37,50%	35,71%	50,00%	Promedio final		
	2	31,25%	32,14%	31,25%	62,18%	70,02%	18,53%
	3	4,17%	1,79%	6,25%			
	4	27,08%	30,36%	12,50%			
Mejor conjunto de datos		1	1	1			

Fuente: Los autores.

Tabla VII.

RESULTADOS DE MEJOR MODELO POR OPCIÓN EN GA

GA		EP	EA	MSE
Opción A	Valor	72,41%	82,00%	12,39%
	Modelo	15	13 y 15	15
Opción B	Valor	73,51%	84,00%	13,19%
	Modelo	1	9	9
Opción C	Valor	87,97%	86,00%	10,00%
	Modelo	1	1 y 8	1
Mejor modelo	Opción	C	C	C
	Modelo	1	9	1
Mejor opción		C		

Fuente: Los autores.

Tabla VIII.

RESULTADOS DE MEJOR MODELO POR OPCIÓN EN HS

HS		EP	EA	MSE
Opción A	Valor	65,92%	78,00%	16,59%
	Modelo	3	2 y 13	3
Opción B	Valor	66,20%	82,00%	16,07%
	Modelo	16	8	13
Opción C	Valor	65,77%	80,00%	15,81%
	Modelo	9	8	8
Mejor modelo	Opción	B	B	C
	Modelo	16	8	8
Mejor opción		B		

Fuente: Los autores.

Tabla IX.

RESULTADOS DE MEJOR MODELO POR OPCIÓN EN DE

DE		EP	EA	MSE
Opción A	Valor	72,93%	86,00%	12,43%
	Modelo	8	8	8
Opción B	Valor	77,61%	86,00%	11,43%
	Modelo	2	7	2
Opción C	Valor	73,36%	84,00%	13,33%
	Modelo	5	5	5
Mejor modelo	Opción	B	A y B	B
	Modelo	2	8 y 7	2
Mejor opción		B		

Fuente: Los autores.

Tabla X.

RESULTADOS DE MEJOR MODELO POR OPCIÓN EN PSO

PSO		EP	EA	MSE
Opción A	Valor	78,18%	90,00%	10,90%
	Modelo	6	2	2
Opción B	Valor	70,51%	82,00%	14,23%
	Modelo	10	12	9
Opción C	Valor	74,82%	84,00%	11,79%
	Modelo	7	8	7
Mejor modelo	Opción	A	A	A
	Modelo	6	2	2
Mejor opción		A		

Fuente: Los autores.

Tabla XI.

RESULTADOS DE MEJORES MODELOS

Modelo	Mejor resultado			Algoritmo			Conjunto de datos		
	EP	EA	MSE	EP	EA	MSE	EP	EA	MSE
1	87,97%	86,00%	10,00%	GA	GA	GA	1	1	1
2	77,61%	90,00%	10,90%	DE	PSO	PSO	1	1	1
3	75,18%	82,00%	12,04%	GA	PSO	GA	1	1	1
4	70,52%	78,00%	15,04%	GA	PSO	GA	4	4	4
5	73,36%	84,00%	13,33%	DE	DE	DE	1	1	1
6	78,18%	82,00%	13,24%	PSO	PSO	PSO	1	1	1
7	74,82%	86,00%	11,79%	PSO	DE	PSO	2	1	2
8	73,63%	86,00%	12,43%	DE	DE	DE	2	1	1
9	72,06%	84,00%	12,20%	GA	GA	GA	1	1	1
10	70,96%	78,00%	14,47%	DE	GA	DE	3	2	3
11	78,84%	82,00%	12,13%	GA	DE	GA	1	1	1
12	71,05%	82,00%	14,03%	DE	PSO	GA	4	1	1
13	70,33%	82,00%	13,77%	PSO	DE	GA	4	1	2
14	71,87%	76,00%	15,61%	PSO	PSO	PSO	2	2	2
15	72,41%	82,00%	12,39%	GA	GA	GA	2	2	2
16	73,93%	82,00%	13,31%	DE	DE	DE	1	1	1

Fuente: Los autores.

Tabla XII.

PROMEDIO DE RESULTADOS DE MEJORES MODELOS POR NÚMERO DE ENTRADAS

Entradas del mejor modelo	Promedio			Mejor valor		
	EP	EA	MSE	EP	EA	MSE
n=3	75,43%	83,60%	12,54%	87,97%	90,00%	10,00%
n=4	72,90%	80,80%	13,58%	78,84%	82,00%	12,13%
n=5	73,93%	82,00%	13,31%	73,93%	82,00%	13,31%

Fuente: Los autores.

Tabla XIII.

RESULTADOS PROMEDIO DE MSE POR CARACTERÍSTICA EN MEJORES MODELOS

MSE promedio				
EE	STE	ZCR	RMS	PSC
13,03%	12,27%	13,14%	13,17%	13,35%

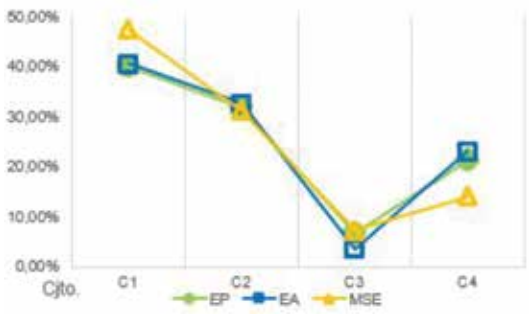
Fuente: Los autores.

6.2 Análisis cualitativo

En esta Subsección se resumen gráficamente los resultados de la Subsección anterior.

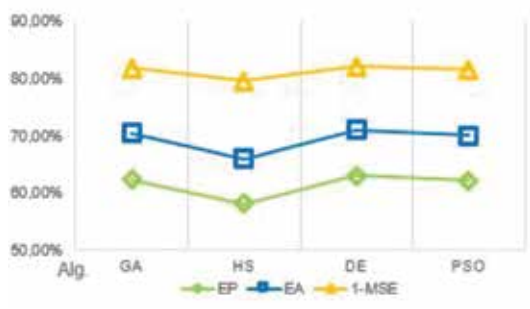
La Fig. 9 muestra el promedio de veces en que el resultado de desempeño de cada conjunto de datos fue mejor que los demás. La Fig. 10 muestra el desempeño por algoritmo bioinspirado. La Fig. 11 muestra el mejor desempeño por opción de cada algoritmo. La Fig. 12 compara los mejores resultados en cada modelo difuso. La Fig. 13 compara el promedio de estos resultados de acuerdo con el número de entradas de los modelos. La Fig. 14 compara el desempeño de las entradas de acuerdo con los mejores resultados de cada modelo difuso.

Fig. 9. PROMEDIO DE VECES EN QUE CADA CONJUNTO DE DATOS FUE MEJOR



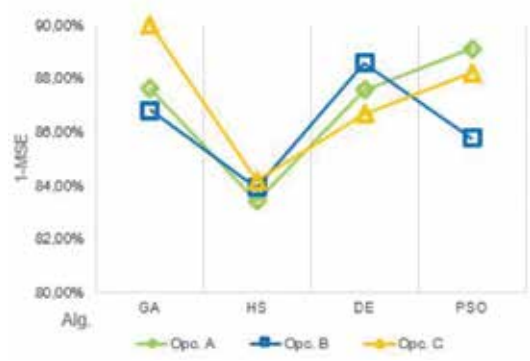
Fuente: Los autores.

Fig. 10. PROMEDIO DE DESEMPEÑO POR ALGORITMO



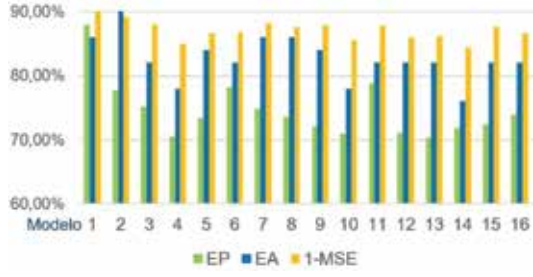
Fuente: Los autores.

Fig. 11. COMPARACIÓN DE 1-MSE POR OPCIÓN DE ALGORITMO BIOINSPIRADO



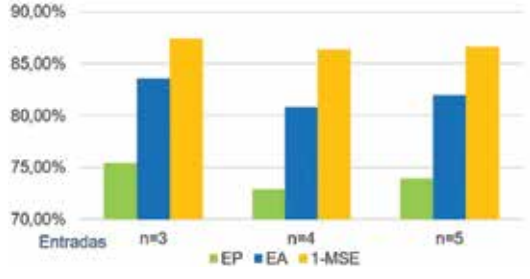
Fuente: Los autores.

Fig. 12. COMPARACIÓN DE MEJORES RESULTADOS DE CADA MODELO DIFUSO



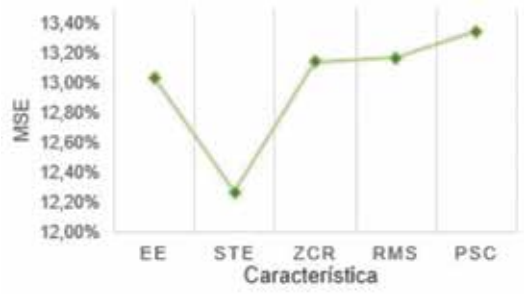
Fuente: Los autores.

Fig. 13. PROMEDIO DE INDICADORES POR NÚMERO DE ENTRADAS DE MODELOS DIFUSOS



Fuente: Los autores.

Fig. 14. PROMEDIO DE MSE EN MODELOS CON CARACTERÍSTICA



Fuente: Los autores.

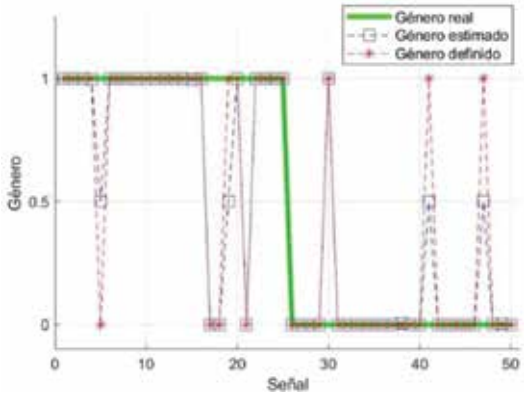
6.3 Análisis de modelos difusos de mayor desempeño

La Fig. 15 muestra los resultados de clasificación de género por voz del primer modelo difuso optimizado propuesto con el que se obtuvo el menor valor de MSE.

Se hace gráficamente evidente que el género estimado se aproxima con precisión al género real, con algunas excepciones en las que se aproxima al valor de umbral establecido (0,5) o al valor del género opuesto.

El criterio del género estimado no logró inferir el género del 8% de las señales de voz, y se equivocó de género en otro 8% de señales de voz. No obstante, el 84% restante de señales de voz fue clasificada con precisión, lo que puede interpretarse como la seguridad intrínseca del modelo difuso y su similitud con la inferencia de un ser humano para la clasificación de género.

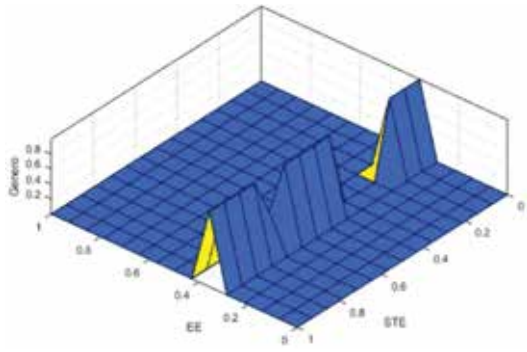
Fig. 15. RESULTADOS DE CLASIFICACIÓN POR VOZ CON MEJOR MODELO OPTIMIZADO



Fuente: Los autores.

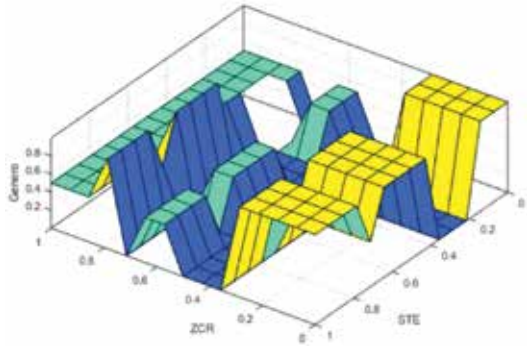
El primer modelo utiliza las tres primeras entradas definidas en la Sección 4. Las figuras 16, 17 y 18 muestran los valores que conforman la superficie de salida en función del valor de cada par posible de entradas.

Fig. 16. SALIDA EN FUNCIÓN DE EE Y STE



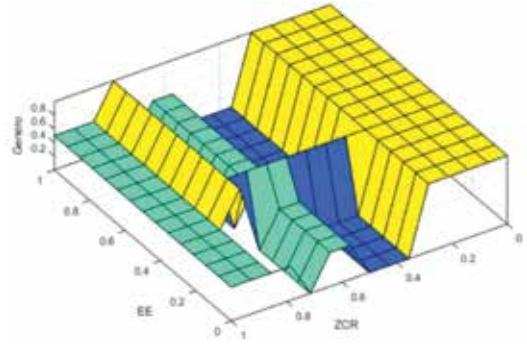
Fuente: Los autores.

Fig. 17. SALIDA EN FUNCIÓN DE ZCR Y STE



Fuente: Los autores.

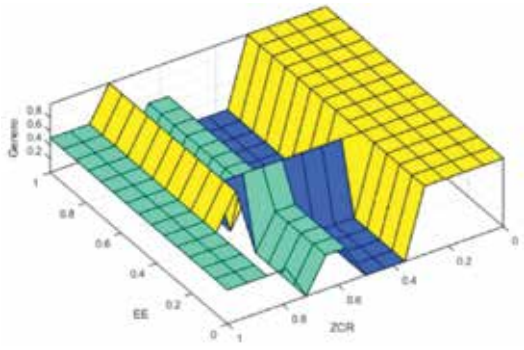
Fig. 18. SALIDA EN FUNCIÓN DE EE Y ZCR



Fuente: Los autores.

La Fig. 19 muestra los resultados de clasificación de género por voz del segundo mejor modelo difuso optimizado con el que se obtuvo un MSE del 10,9%.

Fig. 19. RESULTADOS DE CLASIFICACIÓN POR VOZ CON SEGUNDO MODELO OPTIMIZADO



Fuente: Los autores.

En este modelo la exactitud de acierto es del 90%, pero el criterio del género estimado tiene menor precisión, lo que lo hace menos confiable para la clasificación.

CONCLUSIONES

De acuerdo con los resultados se han obtenido las siguientes conclusiones:

Los conjuntos de datos de mayor a menor desempeño son 1, 2, 4 y 3. La Fig. 9 permite inferir que el procesamiento de la señal pura es sin duda la mejor opción, y que una señal promediada en 10 períodos anteriores no tiene la información que se requiere en el proceso de clasificación.

El modelo propuesto de mayor desempeño es el número 1, seguido de los modelos 2, 7 y 3. Todos estos usan tres entradas y los conjuntos de datos 1 y 2.

El modelo propuesto de menor desempeño es el número 14, seguido de los modelos 4, 10 y 12. Estos modelos tuvieron mejores resultados con los conjuntos de datos 3 y 4.

Los modelos de más de tres entradas no logran superar a los primeros cuatro mejores modelos. Esto puede atribuirse al tamaño del espacio de búsqueda, debido a que este es proporcional a la cantidad de entradas de un modelo difuso, a pesar de tener una cantidad de datos mayor para la clasificación.

El modelo obtenido de mejor desempeño se obtuvo con algoritmos genéticos en el modelo número 1, tuvo un MSE del 10% y logró una exactitud pura cercana al 88%. Las entradas de mayor influencia para el éxito de clasificación son EE, STE y ZCR.

La exactitud de acierto es un indicador de desempeño del que no se puede asegurar un modelo difuso optimizado con precisión; para este fin se utiliza la exactitud pura.

La entrada que más logra caracterizar el género de una señal de voz es STE. La entrada que menos logra caracterizar el género de una señal de voz es PSC.

Los algoritmos GA, DE y PSO tuvieron desempeños promedio cercanos. DE obtuvo el mejor promedio de indicadores de desempeño.

El algoritmo HS obtuvo el menor desempeño. Esto significa que el número de iteraciones que este requiere para lograr convergencia a mínimos globales debe ser mayor que la de los demás algoritmos.

Las mejores opciones para los algoritmos GA, HS, DE y PSO son las opciones C, C, B y A, respectivamente.

De acuerdo con la interpretación del mejor modelo optimizado, el valor de ZCR tiende a ser menor en voces masculinas.

De acuerdo con la interpretación del mejor modelo optimizado, un valor alto de ZCR junto a un valor bajo de STE describe mejor a las voces femeninas.

Este esquema puede ser utilizado para la clasificación de salidas diferentes al género y su descripción en función de las entradas de los modelos como ventaja de su interpretabilidad.

REFERENCIAS

[1] K. Meena, K. Subramaniam and M. Gomathy, "Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network," *The International Arab Journal of Information Technology,* vol. 10 (5), Sept. 2013.

[2] T. Jayasankar, K. Vinothkumar and A. Vijayaselvi, "Automatic gender identification in speech recognition by genetic algorithm," *Appl. Math. Inf. Sci,* vol. 11 (3), pp. 907-913, 2017.

[3] S. Lakra, J. Singh and A. K. Singh, "Automated pitch-based gender recognition using an adaptive neuro-fuzzy inference system," *in IEEE International Conference on Intelligent Systems and Signal Processing (ISSP*), 2013.

[4] P. Gupta, S. Goel and A. Purwar, "A stacked technique for gender recognition through voice," *in IEEE Eleventh International Conference on Contemporary Computing (IC3)*, Noida, India, 2018.

[5] P. Kumar, P. Baheti, R. K. Jha, P. Sarmah and K. Sathish, "Voice gender detection using gaussian mixture model," *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 8 (4), pp. 132-136, Apr. 2018.

[6] M. Gomathy, K. Meena and K. Subramaniam, "Gender clustering and classification algorithms in speech processing: a comprehensive performance analysis," *International Journal of Computer Applications*, vol. 51 (20), pp. 9-17, 2012.

[7] M. P. Gual, "Voice gender identification using deep neural networks running on FPGA," B.S. thesis, Fac. d'Inf. de Barcelona (FIB), Univ. Politècnica de Catalunya (UPC), 2016.

[8] M. Algabri, M. Alsulaiman, G. Muhammad, M. Zakariah, M. Bencherif and Z. Ali, "Voice and unvoiced classification using fuzzy logic," *Int'l Conf. IP, Comp. Vision, and Pattern Recognition (IPCV'15)*, pp. 416-420, 2015.

[9] G. Sun, Z. Fan, N. E. Mastorakis, S. D. Kaminaris and X. Zhuang, "The complexity analysis of voiced and unvoiced speech signal based on sample entropy," *in IEEE Fourth International Conference on Mathematics and Computers in Sciences and in Industry*, 2017.

[10] S. Jain, P. Jha and R. Suresh, "Design and implementation of an automatic speaker recognition system using neural and fuzzy logic in matlab," *in 2013 Int. Conf. on Signal Processing and Communication (ICSC)*, Noida, India, 2013.

[11] R. Kiran, K. Nivedha, S. Pavithra Devi and T. Subha, "Voice and speech recognition in Tamil language," *in 2017 Second International Conference on Computing and Communications Technologies (ICCCT'17)*, 2017.

[12] A. Austermann, N. Esau, L. Kleinjohann and B. Kleinjohann, "Fuzzy emotion recognition in natural speech dialogue," de *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, 2005.

[13] D. Gharavian, M. Sheikhan, A. Nazerieh and S. Garoucy, "Speech emotion recognition using FCBF feature selection method and GA-optimized fuzzy ARTMAP neural network," *Neural Computing and Applications*, vol. 21 (8), pp. 2115-2126, May. 2011.

[14] D. Panek, A. Skalski and J. Gajda, "Voice pathology detection by fuzzy logic," *in 2015 IEEE Int. Instrumentation and Measurement Technology Conf. (I2MTC) Proc.*, Pisa, Italy, 2015.

[15] H. Cordeiro, C. Meneses and J. Fonseca, "Continuous speech classification systems for voice pathologies identification," *in IFIP AICT*, vol. 450, L. Camarinha-Matos *et al*, 2015, pp. 217-224.

[16] A. Asemi, S. S. B. Salim, S. R. Shahamiri, A. Asemi and N. Houshangi, "Adaptive neuro-fuzzy inference system for evaluating dysarthric automatic speech recognition (ASR) systems: a case study on MVML-based ASR," *Springer-Verlag GmbH Germany, part of Springer Nature,* Feb. 2018.

[17] F. T. Putri, M. Ariyanto, W. Caesarendra, R. Ismail, K. A. Pambudi and E. D. Pasmanasari, "Low cost parkinson's disease early detection and classification based on voice and electromyography signal," *in Computational Intelligence for Pattern Recognition. Studies in Computational Intelligence*, vol. 777, W. Pedrycz and S. Chen, Ed. 2018, pp. 397-426.

[18] J. Chen, S. Liu and Z. Chen, "Gender classification in live videos," *in IEEE International Conference on Image Processing*, Beijing, China, 2017.

[19] G. Amayeh, G. Bebis and M. Nicolescu, "Gender classification from hand shape," *in 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops,* Anchorage, AK, USA, 2008.

[20] J. Lei, J. Zhou and M. Abdel-Mottaleb, "Gender classification using automatically detected and aligned 3D ear range data," *in 2013 IEEE International Conference on Biometrics (ICB)*, Madrid, Spain, 2013.

[21] A. Bansal, R. Agarwal and R. Sharma, "SVM based gender classification using iris images," *in 2012 Fourth International Conference on Computational Intelligence and Communication Networks,* Mathura, India, 2012.

[22] S. S. Lee, H. G. Kim, K. Kim and Y. M. Ro, "Adversarial spatial frequency domain critic learning for age and gender classification," *in 2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 2018.

[23] S. E. Bekhouche, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed and A. Hadid, "Facial age estimation and gender classification using multi-level local phase quantization," *in 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, May. 2015.

[24] M. Shin, J.-H. Seo and D.-S. Kwon, "Face image-based age and gender estimation with consideration of ethnic difference," *in 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Lisbon, Portugal, 2017.

[25] D. Yaman, F. I. Eyiokur, N. Sezgin and H. K. Ekenel, "Age and gender classification from ear images" *in 2018 International Workshop on Biometrics and Forensics (IWBF)*, Sassari, Italy, 2018.

[26] P. P. Bonissone, "A fuzzy sets based linguistic approach: theory and applications," *in Proc. of the 1980 Winter Simulation Conf.*, California, 1980.

[27] T. Haider and M. Yusuf, "A fuzzy approach to energy optimized routing for wireless sensor networks," *The Int. Arab Journal of Information Technology (IAJIT),* vol. 6 (2), pp. 179-185, 2009.

[28] M. Gacto, R. Alcalá and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures," *Information Sciences*, vol. 181, pp. 4340-4360, 2011.

[29] C. Carlsson, "On the relevance of fuzzy sets in analytics," *in On fuzziness, Studies in fuzziness and soft computing 298*, vol. 1, pp. 83-89, 2013.

[30] T. R. Razak, J. M. Garibaldi, C. Wagner, A. Pourabdollah and D. Soria, "Interpretability and complexity of design in the creation of fuzzy logic systems — a user study," *in IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 420-426, 2018.

[31] J. P. Carvalho, F. Batista and L. Coheur, "A critical survey on the use of fuzzy sets in speech and natural language processing," in *2012 IEEE World Congress on Computational Intelligence (WCCI)*, Brisbane, Australia, Jun. 2012.

[32] H.-N. L. Teodorescu, "A retrospective assessment of fuzzy logic applications in voice communications and speech analytics," *International Journal of Computers*

Communications & Control (IJCCC), pp. 865-872, Dec. 2015.

[33] C. E. Borges and J. L. Montaña, "Algoritmos bioinspirados,", 2011 [Online]. Available: https://docplayer.es/2484561-Algoritmos-bioinspirad os.html. [Accessed: 1- Oct- 2018].

[34] S. Forrest, "Genetic algorithms: Principles of natural selection applied to computation," *in Science*, vol. 261(5123), pp. 872-878, Aug.1993.

[35] M. Mitchell, *An introduction to genetic algorithms*, Cambridge, Massachusetts. London, England: A Bradford Book The MIT Press, 1996.

[36] T. Weise, *Global optimization algorithms. Theory and application,* 2009.

[37] Z. W. Geem, J. H. Kim and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," *Simulation*, vol. 76 (2), pp. 60-68, 2001.

[38] Z. W. Geem, "Global optimization using harmony search: Theoretical foundations and applications," *Studies in computacional intelligence*, vol. 203, pp. 57-73, 2009.

[39] R. Storn and K. Price, "Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11 (4), pp. 341-359, 1997.

[40] J. Kennedy and R. Eberhart, "Particle swarm optimization," *in Proc. of the IEEE Int. Conf. on Neural Networks,* vol. 8 (3), pp. 1943-1948, 1995.

[41] J. C. Bansal, P. K. Singh, M. Saraswat, A. Verma, S. S. Jadon and A. Abraham, "Inertia weight strategies in particle swarm optimization," *in 2011 Third World Congress on Nature and Biologically Inspired Computing (NaBIC)*, pp. 633-640, 2011.

[42] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proc. of IEEE Int. Conf. on Evolutionary Computation.*, vol. 3, pp. 1945-1950, 1999.

[43] J. Nocedal and S. J. Wright, Numerical optimization, 2 ed., Berlin, Nueva York: Springer Verlag, 2006.

[44] D. F. Shanno and K. H. Phua, "Effective comparison of unconstrained optimization techniques," *Management science*, vol. 22 (3), pp. 321-330, Nov. 1975.

[45] M. Contreras and R. A. Tapia, "Sizing the BFGS and DFP updates: A numerical study," *Optim. Theory Appl*., vol. 78, pp. 93-108, 1993.

[46] F. Rong, "Audio classification method based on machine learning," *in 2016 International Conference on Intelligent Transportation, Big Data & Smart City*, 2017.

[47] A. DeMarco and S. J. Cox, "An accurate and robust gender identification algorithm," *Journal of Neuroscience Methods,* vol. 172 (1), pp. 122-130, 2008.

[48] D. E. Rey Lancheros, H. J. Gavilán Acosta y H. E. Espitia Cuchango, "Implementación de un algoritmo para la identificación de usuarios considerando problemas fisiológicos que afectan el habla," *Revista ITECKNE*, vol. 14 (2), pp. 131-139, 2017. https://doi.org/10.15332/iteckne.v14i2.1767

[49] T. T. Swee, S. H. S. Salleh and M. R. Jamaludin, "Speech pitch detection using short-time energy," *in International Conference on Computer and Communication Engineering (ICCCE),* Kuala Lumpur, Malasia, 2010.

[50] G. Saha, S. Chakroborty and S. Senapati, "A new silence removal and endpoint detection algorithm for speech and speaker recognition applications," *Indian Institute of Technology Kharagpur,* Kharagpur, India, 2005.

[51] D. Ortiz P, L. F. Villa, C. Salazar and O. L. Quintero, "A simple but efficient voice activity detection algorithm through Hilbert transform and dynamic threshold for speech pathologies," *in 20th Argentinean Bioengineering Society Congress (SABI 2015)*, 2016.

[52] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid," *Acta Acustica united with Acustica*, vol. 92, pp. 820-825, 2006.

[53] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *The Journal of the Acoustical Society of America (JASA)*, vol. 63, pp. 1493-1500, 1978.

[54] R. Thiruvengatanadhan, P. Dhanalakshmi and S. Palanivel, "GMM based indexing and retrieval of music using MFCC and MPEG-7 features," *in Proceedings of the 49th Annual Conference of the Computer Society of India (CSI)*, Chidambaram, Tamil Nadu, India, 2015.

[55] B. Wu, A. Horner and C. Lee, "Musical timbre and emotion: The identification of salient timbral features in sustained musical instrument tones equalized in attack time and spectral centroid," *in Proc. of 40th International Computer Music Conference (ICMC) 2014 and 11th Sound and Music Computing Conference (SMC)*, Athens, Greece, 2014.

[56] T. Uzunović, S. Konjicija and I. Turković, "Adjustment of fuzzy reasoning for implementation on microcontroller," *in 2011 18th International Conference on Systems, Signals and Image Processing*, Sarajevo, Bosnia-Herzegovina, Jan. 2011.

[57] I. A. Hameed, "Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems," *Expert Systems with Applications*, vol. 38, p. 7135–7142, 2011.

[58] Liu, Xiang-Jie; Zhou, Xiao-Xin, "Structural analysis of fuzzy controller with gaussian membership function," *in 14th World Congress of International Federation of Automatic Control (IFAC)*, Beijing, China, 1999.

[59] W. Meiniar, F. A. Afrida, A. Irmasari, A. Mukti and D. Astharini, "Human voice filtering with band-stop filter design in MATLAB," *in 2017 International Conference on Broadband Communication, Wireless Sensors and Powering (BCWSP),* Jakarta, Indonesia, 2017.

[60] G. K. Berdibaeva, O. N. Bodin, V. V. Kozlov, D. I. Nefed'ev, K. A. Ozhikenov and Y. A. Pizhonkov, "Pre-processing voice signals for voice recognition systems," *in 18th Int. Conf. of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM)*, 2017.

[61] M. Suell Dutra, C. H. Valencia Niño, S. García y Rodolfo, "Codificación y compresión de señales de voz con cuantización vectorial no determinística," *Revista ITECKNE*, vol. 6 (1), pp. 14-19, Jun. 2009. https://doi.org/10.15332/iteckne.v6i1.291

[62] "Wavsource," [Online]. Available: wavsource.com/people/people.htm. [Accessed: 15-Sep-2018].

[63] F. R. Jimenez López, C. E. Pardo Beainy and E. A. Gutiérrez Cáceres, "Adaptive filtering implemented over TMS320c6713 DSP platform for system identification," *Revista ITECKNE,* vol. 11 (2), pp. 157-171, Dec. 2014. https://doi.org/10.15332/iteckne.v11i2.726