

Identificación de instrumentos musicales de cuerdas pulsadas de la región andina colombiana en solo, mediante técnicas de aprendizaje de máquina

✉ INDIRA JULIANA TOBÓN-GONZALEZ¹
JIMMY ALEXANDER CORTÉS-OSORIO¹

Resumen

Son muchos los estudios propuestos sobre la identificación de instrumentos musicales, pero ninguno ha estado enfocado en instrumentos de cuerda pulsada de la región andina colombiana como lo son: tiple, tiple requinto, guitarra y bandola. Por ello, se propone la identificación de estos utilizando técnicas de aprendizaje de máquina tales como Análisis discriminante, Árbol de Decisión, kNN, SVM, ANNs y utilizando tres métodos de reducción de datos: Feature Selection; PCA con 1, 100 y 1000 componentes principales; y extrayendo las cinco primeras frecuencias parciales junto a sus amplitudes normalizadas. Esta investigación se realizó usando una base de datos de 1000 grabaciones de audio monofónicas, construida a partir del registro de las notas de la primera posición de cada instrumento en formato WAV. Se utilizó como Método de Validación Cruzada con un k igual a cinco para realizar las Matrices de Confusión y Curvas ROC. La mejor Exactitud se alcanzó con ANNs que tuvo un porcentaje de 99,8% en la identificación, además las curvas ROC mostraron un área bajo la curva muy cercana a uno para la guitarra.

Palabras clave: Aprendizaje de Máquina, Fourier, Identificación, Instrumentos Musicales, Matriz de Confusión, Región Andina Colombiana.

Artículo en prensa / Article in press

¹ Universidad Tecnológica de Pereira, Colombia.



Autor de correspondencia: Tobón-Gonzalez, I.J. (Indira Juliana): Carrera 31 # 15 -87 CDV Barrio San Luis, ParqueSoft Risaralda, Pereira, Colombia. Teléfono: 3128875217. Correo electrónico: ijtohon@utp.edu.co

Historia del artículo:

Artículo recibido: 18-V-2018/ Aprobado: 3-VI-2018
Disponible online: 3 de agosto 2018
Discusión abierta hasta abril de 2020

Identification of Musical Instruments of Plucked Strings of the Colombian Andean Region in Solo Using Machine Learning Techniques

Abstract

There are many studies on the identification of musical instruments, but none has focused on plucked string instrument of the Colombian Andean region such as the tiple, tiple requinto, guitar and bandola. Therefore, we propose to identify these instruments using machine-learning techniques such as: Discriminant Analysis, Decision Tree, k-Nearest Neighbors (kNN), Support Vector machines (SVM), Artificial Neural Network (ANNs) and three methods of data reduction: Feature Selection; Principal Components Analysis (PCA) with 10,100 and 1000 principal components, and extracting the first five partial frequencies along with their normalized amplitudes. We carried out this study using a database of 1000 digital monophonic audio recordings, built of the recordings of the first position of the notes played in solo for each string instrument in WAV format. Regarding the validity method, the Cross Validation Method was used with a k equal to five to perform the confusion matrices and the ROC Curves (Receiver Operating Characteristic). We reached the best results with ANNs that had an accuracy of 99.8%, besides the ROC curves showed an area under the curve very close to one for the guitar.

Keywords: Machine Learning, Fourier, Identification, Musical Instruments, Confusion Matrix, Colombian Andean region.

Identificação de Instrumentos Musicais de Cordas Pulsadas da Região Andina da Colômbia em Solo, por meio de Técnicas de Aprendizado de Máquinas

Resumo

Existem muitos estudos propostos sob a identificação dos instrumentos musicais, mas nenhum tem sido focado em instrumentos de cordas pulsadas da região andina colombiana, como o tiple, tiple requinto, violão e bandola. Portanto, estes instrumentos podem ser identificados usando técnicas de aprendizado de máquina, tais como Discriminant Analysis, Decision Tree, kNN, SVM, ANNs. Por meio do uso de três métodos de redução de dados a seguir: Seleção de Recursos; PCA com 1, 100 e 1000 componentes principais; e extraíndo as cinco primeiras frequências parciais, juntamente com suas amplitudes normalizadas. Para esta pesquisa foi utilizado um banco de dados de 1000 gravações de áudio monofônicas, a base de dados foi construída a partir da gravação das notas musicais da primeira posição de cada instrumento, no formato WAV. No método de validação utiliza-se o método de validação cruzada onde foi usada uma constante de trabalho k igual a cinco, para executar as matrizes de confusão e Curvas ROC do método. A melhor precisão foi alcançada com as RNAs que possuíam 99,8% de porcentagem de identificação, além das curvas ROC mostrarem uma área sob a curva muito próxima de uma para o violão.

Palavras-chave: Aprendizagem Mecânica, , Fourier, Identificação, Instrumentos Musicais, Matriz de Confusão, Região Andina Colombiana.

1. Introducción

La identificación de instrumentos musicales ha despertado gran interés en los últimos tiempos. Esto se debe al aumento de contenido multimedia que se ha venido presentando gracias al rápido desarrollo tecnológico en equipos para la captura de este tipo de datos, como es el caso de las tecnologías para la grabación, almacenamiento, edición y distribución de audio. Esto hace que la música digital tenga una participación importante en el contenido multimedia y sea unos de los tipos de datos más importantes distribuidos por internet (Essid et al. 2004; Xu et al. 2005).

Al aumentar el tamaño de las bases de datos de música disponible al público en general, se hace necesario el uso de sistemas que permitan una clasificación y reconocimiento automático de los contenidos de estos archivos. Por ejemplo, para la indexación automática de datos y para la recuperación de bases de datos (Chandwadkar y Sutaone, 2012), la identificación de instrumentos se vuelve necesaria, pues a partir de esto se pueden reconocer varios estilos musicales (Kitahara et al., 2004), editar audio, generar lista de reproducción, hacer recuperación de audio (Chandwadkar y Sutaone, 2012) y realizar transcripciones (Martin & Kim 1998; Wicaksana et al. 2006), entre muchas otras tareas que facilitan y disminuyen el tiempo de organización y análisis de datos relacionados con archivos musicales.

Por tal motivo, han sido muchos los estudios propuestos sobre la identificación de instrumentos musicales, sin embargo, en la revisión del tema hecha para esta investigación, no se evidenció ningún estudio relacionado con la identificación de los instrumentos de cuerdas típicos nacionales de Colombia, como lo son el tiple, la bandola, el tiple requinto y la guitarra. Por ende, en muchas ocasiones, los sonidos colombianos pueden pasar desapercibidos en las bases de datos de archivos musicales en el proceso de identificación.

Uno de los trabajos encontrados en la revisión de tema, es el descrito por Martin y Kim (1998), en el cual se aplicó una técnica estadística de reconocimiento de patrones a la clasificación de los tonos de instrumentos musicales dentro de una jerarquía taxonómica. En dicha método se obtuvo una Exactitud del 90% en la identificación de las familias de instrumentos, y un 70% en la identificación de instrumentos individuales.

Otro estudio hallado es el indicado por Essid, Richard, y David (2004), en el que se realizó la identificación de instrumentos de vientos- madera (saxofón alto, flauta, oboe, clarinete en Bb y fagot), utilizando dos clasificadores, Máquinas de Soporte Vectorial (SVM), y Modelos de Mezcla de Gaussiana (GMM). Teniendo en cuenta una base datos constituida por notas aisladas y fragmentos de solos, dando como resultado índices de reconocimiento que van desde 52% para el clarinete en Bb hasta el 96% para el oboe.

Por otra parte, se encontró la investigación de Xu, Maddage y Shao (2005), en el cual se demostró que el método de Máquina de Soporte Vectorial resulta tener mejor desempeño, en comparación con el método tradicional de distancia Euclidiana, para clasificar entre música instrumental y vocal. También se encuentra la investigación de Wicaksana, Hartono y Wei (2006), donde se propone el uso de varias características de la señales de audios y a las cuales se les agrupa en tres categorías, espectrales, temporales y cepstrales. Además, utilizaron para clasificar una Red Neuronal compuesta, en la que se obtuvo una Exactitud de 94%, y una estructura de Red Neuronal tradicional con la que se obtuvo una Exactitud del 88%.

Otras investigaciones más recientes fueron las realizadas por Azarloo y Farokhi (2012) en la cual hacen la comparación del desempeño entre kNN (k-Nearest Neighbor) y el Perceptron Multicapa (MLP) para el reconocimiento automático de siete instrumentos musicales (guitarra, piano, percusión, voz, contrabajo, batería y trompeta) al ser tocados simultáneamente, teniendo MLP el mejor resultado; y

también se encuentra el estudio de Nagawade y Ratnaparkhe (2017) donde se utiliza MFCC como característica de las señales, y como clasificador un kNN, para la identificación del Piano, flauta, violín, trompeta y chelo, en este, la precisión de reconocimiento de violonchelo, piano y trompeta fue 91.66% y para flauta y violín fue 83.33%.

En esta investigación, se desea realizar la identificación del tiple, tiple requinto, bandola y guitarra, instrumentos típicos de la región andina colombiana, usando cinco técnicas de aprendizaje de máquina: kNN, Árbol de Decisión, Máquinas de Soporte Vectorial (SVM), las Redes Neuronales Artificiales (ANNs) y Análisis Discriminante. Además se proponen tres formas de selección y reducción de características: extrayendo las primeras 5 frecuencias parciales de las señales de audio, usando el método de Feature Selection para la selección, y utilizando Análisis de Componentes Principales (PCA) para la reducción.

A. Métodos de clasificación

El Clasificador kNN (k-Nearest Neighbor) es uno de los clasificadores más sencillos de emplear. Consiste en la clasificación de un elemento desconocido, determinando la distancia más corta entre el elemento a clasificar y los elementos de las clases conocidas (Mitchell 2009; Azarloo y Farokhi 2012). Para medir estas distancias, se utiliza métricas como la distancia Chebychev y la distancia Euclidiana.

La distancia Euclidiana es la más utilizada, y define la separación entre dos puntos a_i y b_i por medio del teorema de Pitágoras, mientras que la distancia Chebyshev, también llamada la métrica del valor máximo, presenta la separación de dos puntos como la mayor diferencia entre sus dimensiones (Rodríguez Jorge Enrique et al. 2007).

En cuanto a la clasificación a través de un Árbol de Decisión, se puede emplear el algoritmo llamado ID3, el cual realiza una evaluación estadística de cada atributo de entrenamiento, que consiste en la medición de la *ganancia de información*, definida como la reducción esperada de la entropía después de una di-

visión de datos, y elige el atributo que mayor ganancia tiene para realizar un nodo de la estructura del Árbol de Decisión en construcción (Mitchell 2009).

Las partes que componen un árbol de decisión son: raíz, nodos y ramas. Las ramas se representan con líneas y descienden de los nodos, los cuales representan una posible decisión, y la raíz es el nodo principal, por el cual se empieza a desplegar las ramas y los demás nodos (Mitchell 2009).

Por otro lado, la técnica de Máquinas de Soporte Vectorial (SVM), busca encontrar un hiperplano que permita separar las clases. Para ello, mapea los puntos de entrada en un espacio de características de una dimensión más alta, de este modo, si los puntos de entrada están en R^2 , la SVM los mapeará en R^3 (Bentacourt 2005; Mitchell 2009).

Asimismo, el método de Análisis Discriminante divide el espacio de entrenamiento en hiperplanos, cuyos límites están dados por funciones que dependiendo la separabilidad de los datos de entrenamiento, puede ser lineales o cuadráticas, es por ello que dentro de este método se encuentra dos tipos: Análisis Discriminante Lineal y Análisis Discriminante cuadrático (Bishop 2013).

En cambio, las Redes Neuronales Artificiales (ANNs), es un clasificador inspirado en las redes neuronales biológicas, por tanto un sistema de ANNs, está conformado por la unión de varios perceptrones, el cual actúa como una neurona artificial. Su trabajo es sumar los valores de entrada, que recibe normalmente de otros perceptrones, y compararlos con un umbral. Si los datos de entrada superan o igualan dicho umbral, envía una señal a su salida (José Montaña Moreno Director y Alfonso Palmer Pol 2002; Mitchell 2009). Un ejemplo de ello se puede apreciar en el trabajo de Gómez, Sánchez, Ocampo y Restrepo (2012), en el que se utiliza Redes Neuronales para la clasificación de arcillas.

Las características elegidas fueron las denominadas espectrales, que son las obtenidas a partir de la señal en el dominio de la frecuencia. Estas características cobran importancia debido a la influencia

del contenido frecuencial en la percepción auditiva (Wicaksana et al. 2006; Azarloo y Farokhi 2012). La herramienta utilizada en este grupo de características es la Transformada de Fourier (TF), pues ésta permite determinar el espectro más simple de una señal, en el cual se puede apreciar los componentes en frecuencia, amplitud y fase de dicha señal, al igual que permite evaluar características básicas de la forma de onda como el tono y el timbre (Cortés Osorio et al. 2012).

B. Métodos de reducción y selección de características

El Análisis de Componentes Principales, es un método que permite reducir la dimensión del espacio de características, transformando su sistema original de coordenadas, en un nuevo sistema, el cual se construye a partir de la varianza de los elementos que conforman el espacio. Estos ejes, que corresponden al nuevo sistema de coordenadas, son llamados Componentes Principales (Harrington 2012).

Por otra parte, la técnica de Feature Selection, busca elegir un subconjunto de características relevantes del espacio de características principal, según ciertos criterios. Un ejemplo es que la precisión de la clasificación no disminuya significativamente al reducir el número de características, y

también que la distribución de clases resultantes, sean lo más cercanas posible a la distribución de clases de los datos originales (Dash y Liu 1997; Tang et al. 2014).

C. Técnicas de evaluación

Para evaluar el comportamiento de cada uno de las técnicas de aprendizaje utilizadas en la identificación entre bandola, tiple, tiple requinto y guitarra, se utilizó la matriz de confusión y la curva ROC (Receiver Operating Characteristic).

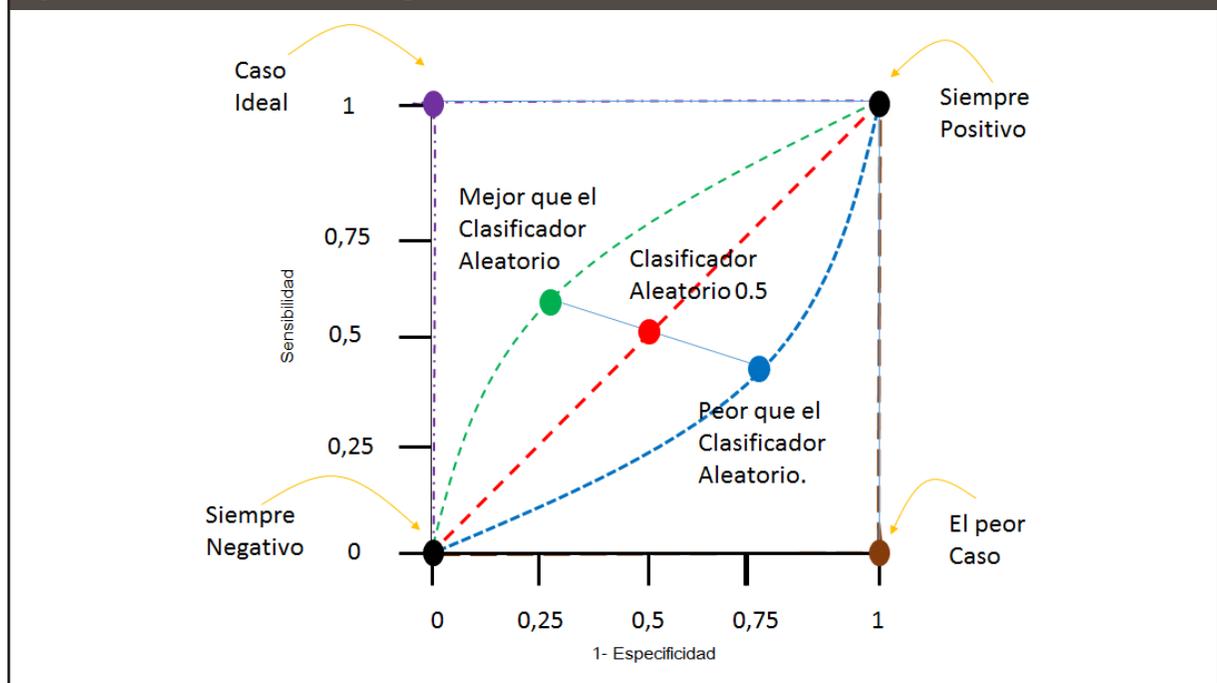
La Matriz de Confusión, mostrada en la **Figura 1**, también llamada Matriz de Error o de Contingencia, es una herramienta utilizada para mostrar el nivel de acierto entre las clases asignadas por el clasificador con respecto a una muestra de prueba no utilizada durante el proceso de entrenamiento. La Matriz de Confusión tiene un tamaño de $N \times N$, donde N es el número de clases asignado al clasificador (El-Habi y El-Ghareeb 2014; Cortés Osorio n.d.).

En cuanto a la curva ROC mostrada en **Figura 2**, es una gráfica que representa la Sensibilidad o Tasa de Verdaderos Positivos (TPR True Positive Rate) frente a la Tasa de Falsos Positivos (FPR False Positive Rate). Es usada para visualizar el desempeño de los clasificadores (Burgueño et al. 1995; Zhu et al. 2010; Cortés Osorio n.d.).

Figura 1. Estructura de la matriz de confusión (Cortés Osorio n.d.).

		Predicción		
		Positivo	Negativo	
Valor Real	Positivo	TP	FN	Sensibilidad
	Negativo	FP	TN	Especificidad

Figura 2. Curva ROC (Receiver Operating Characteristic) (Cortés Osorio n.d.).



En la gráfica de la curva ROC expuesta en la **Figura 2**, la Sensibilidad es representada por el eje 'Y' y la Tasa de Falsos Positivos (FPR) por el eje 'X', sus coordenadas mínimas son (0,0) y las máximas son (1,1). Teniendo en cuenta lo anterior, se describen tres casos: El primero implica que si un clasificador posee una discriminación perfecta, con alta Sensibilidad y baja FPR, la curva de ROC pasa por la esquina superior izquierda. En el segundo caso si un clasificador tiene igual distribución, con igual Sensibilidad y FPR, la curva será una línea diagonal a 45° desde la esquina inferior izquierda hasta la esquina superior derecha y dicho clasificador tendrá un desempeño incierto. El caso tercero es cuando un clasificador tiene una alta FPR y baja Sensibilidad, en este, la curva cae por debajo de la línea diagonal de 45°, teniendo este clasificador un desempeño indeseado (Burgueño et al. 1995; Zhu et al. 2010; Cortés Osorio n.d.).

D. Descripción de los instrumentos musicales a identificar

La bandola andina (ver **Figura 3. A**), es un instrumento musical de cuerda pulsada de la familia de

la guitarra que se ejecuta con una pluma o plectro. Se le dice bandola andina ya que sus orígenes y uso se dan en la región andina colombiana. Este instrumento está constituido por seis órdenes pareados de afinación unísono. La distancia sonora entre un orden y otro es la misma (cuarta justa). Además se caracteriza por tener un cuerpo en forma de pera o gota de agua (Londoño y Tobón Restrepo 2004; Rincón Gómez n.d.).

El Tiple colombiano (ver **Figura 3. C**) es descendiente de la vihuela de cuatro órdenes y la guitarra. Consta de cuatro órdenes triples de cuerdas metálicas; el primero posee tres cuerdas de acero de calibre y afinación igual; por su lado los órdenes segundo, tercero y cuarto, constan de dos cuerdas de acero de calibre y afinación igual y una cuerda central entorchada de acero la cual es afinada a una octava abajo con respecto a las otras dos. Este instrumento se toca con los dedos y en la música colombiana hace su papel de acompañante, por lo que se pueden realizar rasgueos y ritmos andinos como el pasillo y el bambuco (Londoño y Tobón Restrepo 2004; Vanessa et al. 2009).

Figura 3. Instrumentos musicales a identificar. A. Bandola andina fabricada por Arvey Bastidas. B. Guitarra clásica Yamaha C-80. C. Tiple Colombiano elaborado por Tobías Bastidas. D. Tiple requinto fabricado por Javier Navarro.



El Tiple Requinto (ver **Figura 3. D**) es una variante del tiple, y en subregiones andinas orientales y centrales sustituye a la bandola como instrumento melódico. Es tocado con plectro o también con una cuchilla de afeitar. Al igual que el tiple, el tiple requinto posee cuatro órdenes triples de cuerdas metálicas y afinadas unísono, su diapasón es más largo y su caja es más pequeña con respecto al tiple, esto con el fin de obtener un sonido más brillante, además posee un sonido metálico y dulce (Londoño y Tobón Restrepo 2004).

La Guitarra Clásica (ver **Figura 3. B**), es un instrumento de cuerdas pulsadas de origen arábigo-asiático y de nombre greco-romano. Por su fácil ejecución se popularizó su uso en diferentes géneros musicales. Este instrumento musical posee seis órdenes de una sola cuerda; los tres primeros órdenes tienen cuerdas de nylon, y los siguientes tres poseen

cuerdas entorchadas de acero (Londoño y Tobón Restrepo 2004; von Vaseghi 2007).

2. Materiales y métodos

La base de datos utilizada en la identificación de los cuatro instrumentos musicales, fue construida para propósitos de esta investigación, la cual se encuentra disponible en [http:// academia.utp.edu.co/jacoper/music-id/](http://academia.utp.edu.co/jacoper/music-id/). Para esta se realizaron diez registros monofónicos, de cada una de las notas correspondientes a la primera posición de los instrumentos a identificar, en un formato WAV, en la **Figura 4** se presenta la ilustración de esta posición en el diapasón del instrumento. Estas grabaciones se hicieron con una frecuencia de muestreo de 44.100 Hz, y una duración de 0,5 segundos en condiciones de estudio de grabación. La guitarra aportó 300 sonidos de notas en la base de datos, al igual que la bandola, mientras que el tiple aportó 200 sonidos al igual que el tiple requinto, dando en total una base de datos de 1.000 sonidos.

En las **Tablas 1, 2, 3 y 4** se presentan las notas musicales grabadas equivalentes a la primera posición de cada instrumentos a identificar (bandola, tiple, tiple requinto y guitarra) y el número de muestras tomadas por cada una de ellas.

Por otro lado, la caracterización de la base de datos se hizo teniendo en cuenta la cualidad del timbre de los cuatro instrumentos musicales, por tanto se aplicó la FFT (Transformada Rápida de Fourier) a las señales de audio, para obtener los componentes en frecuencia de las muestras grabadas junto a sus respectivas amplitudes, las cuales fueron normalizadas para facilitar el análisis, luego se construyó las bases de datos de entrenamiento de tres maneras:

- Extrayendo las primeras cinco frecuencias parciales junto a sus amplitudes normalizadas (ver **Figura 5**).
- Utilizando el método de Análisis de Componentes Principales (PCA), eligiendo 1, 100 y 1000 Componentes Principales.
- Aplicando la técnica de Feature Selection.

Figura 4. Primera Posición de un instrumento de cuerda pulsada.

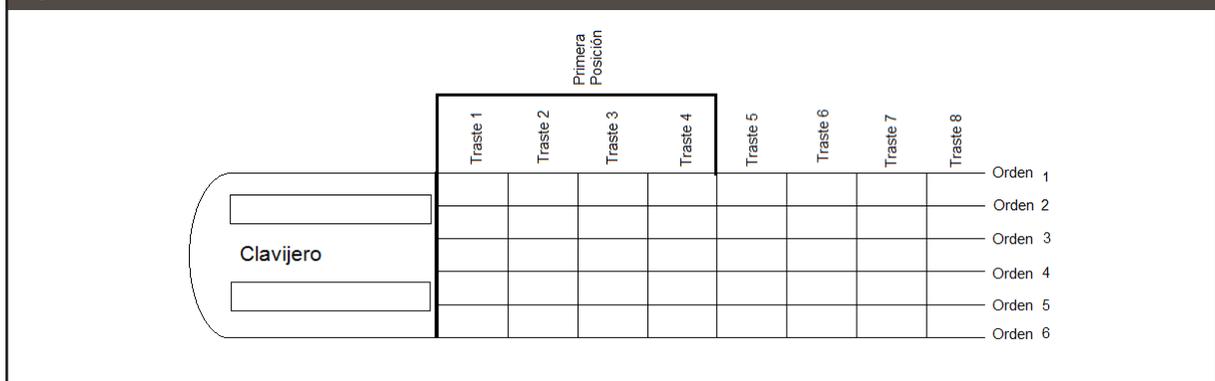


TABLA 1. NÚMERO DE MUESTRAS GRABADAS DE LAS NOTAS MUSICALES CORRESPONDIENTES A LA PRIMERA POSICIÓN DE LA GUITARRA.

Ordenes	Al Aire	1 Traste	2 Traste	3 Traste	4 Traste	Total Muestras
Orden 1	Mi4	Fa4	Fa4#	Sol4	Sol4#	50
Muestras por notas	10	10	10	10	10	
Orden 2	Si3	Do4	Do4#	Re4	Re4#	50
Muestras por notas	10	10	10	10	10	
Orden 3	Sol3	Sol3#	La3	La3#	Si3	50
Muestras por notas	10	10	10	10	10	
Orden 4	Re3	Re3#	Mi3	Fa3	Fa3#	50
Muestras por notas	10	10	10	10	10	
Orden 5	La2	La2#	Si2	Do3	Do3#	50
Muestras por notas	10	10	10	10	10	
Orden 6	Mi2	Fa2	Fa2#	Sol2	Sol2#	50
Muestras por notas	10	10	10	10	10	
Total Muestras Guitarra						300

TABLA 2. NÚMERO DE MUESTRAS GRABADAS DE LAS NOTAS MUSICALES CORRESPONDIENTES A LA PRIMERA POSICIÓN DE LA BANDOLA.

Ordenes	Al Aire	1 Traste	2 Traste	3 Traste	4 Traste	Total Muestras
Orden 1	Sol5	Sol5#	La5	La5#	Si5#	50
Muestras por notas	10	10	10	10	10	
Orden 2	D5	D5#	E5	F5	Fa5#	50
Muestras por notas	10	10	10	10	10	
Orden 3	La4	La4#	Si4	Do5	Do5#	50
Muestras por notas	10	10	10	10	10	
Orden 4	Mi4	Fa4	Fa4#	Sol4	Sol4#	50
Muestras por notas	10	10	10	10	10	
Orden 5	Si3	Do4	Do4#	Re4	Re4#	50
Muestras por notas	10	10	10	10	10	
Orden 6	Fa3#	Sol3	Sol3#	La3	La3#	50
Muestras por notas	10	10	10	10	10	
Total Muestras Bandola						300

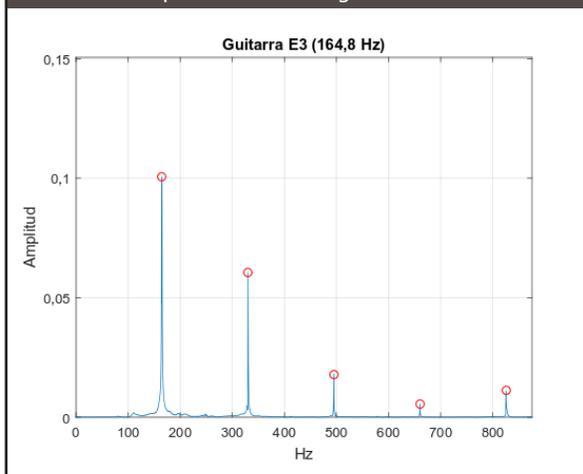
TABLA 3. NÚMERO DE MUESTRAS GRABADAS DE LAS NOTAS MUSICALES CORRESPONDIENTES A LA PRIMERA POSICIÓN DEL TIPLE.

Ordenes	Al aire	1 Traste	2 Traste	3 Traste	4 Traste	Total Muestras
Orden 1	Mi4	Fa4	Fa4#	Sol4	Sol4#	50
Muestras por notas	10	10	10	10	10	
Orden 2	Si3	Do4	Do4#	Re4	Re4#	50
Muestras por notas	10	10	10	10	10	
Orden 3	Sol3	Sol3#	La3	La3#	Si3	50
Muestras por notas	10	10	10	10	10	
Orden 4	Re3	Re3#	Mi3	Fa3	Fa3#	50
Muestras por notas	10	10	10	10	10	
Total Muestras Tiple						200

TABLA 4. NÚMERO DE MUESTRAS GRABADAS DE LAS NOTAS MUSICALES CORRESPONDIENTES A LA PRIMERA POSICIÓN DEL TIPLE REQUINTO.

Ordenes	Al Aire	1 Traste	2 Traste	3 Traste	4 Traste	Total Muestras
Orden 1	Mi4	Fa4	Fa4#	Sol4	Sol4#	50
Muestras por notas	10	10	10	10	10	
Orden 2	Si4	Do5	Do5#	Re5	Re5#	50
Muestras por notas	10	10	10	10	10	
Orden 3	Sol4	Sol4#	La4	La4#	Si4	50
Muestras por notas	10	10	10	10	10	
Cuerda 4	Re4	Re4#	Mi4	Fa4	Fa4#	50
Muestras por notas	10	10	10	10	10	
Total Muestras Tiple Requinto						200

Figura 5. Identificación de los 5 primeros armónica de una señal de audio proveniente de la guitarra.



Teniendo las cinco bases datos de entrenamiento, construidas a partir de los tres métodos anteriormente nombrados, se realizó el entrenamiento de las cinco Técnicas de Aprendizaje de Máquina con estas bases de datos, usando las especificaciones presentadas en la **Tabla 5**.

Finalmente, para realizar las Pruebas de validez de los clasificadores kNN, Árbol de Decisión, SVM y Análisis Discriminante, se utilizó el método denominado Validación Cruzada (Bishop 2013; Cortés Osorio n.d.), el cual consiste en dividir aleatoriamente los datos en k partes, procurando siempre que sea posible que estas tenga el mismo número de instancias, para usar una parte como entrenamiento

y la otra en la evaluación del desempeño (ver **Figura 6**). El resultado final de la Validación cruzada es un promedio de todos los resultados entregado por cada división k.

TABLA 5. ESPECIFICACIONES USADAS EN LOS CLASIFICADORES.

Técnica de Aprendizaje de Máquina	Especificaciones
kNN	Distancia Euclidiana con número de vecinos 1 y 3
	Distancia Chebyshev con número de vecinos 1 y 3
Árbol de Decisión	4 Divisiones
	100 Divisiones
SVM	Kernel Lineal
	Kernel Cúbica
Análisis Discriminante	Lineal
	Cuadrática
Redes Neuronales Artificiales	10 neuronas
	20 neuronas

En el caso de la validación cruzada, usada para evaluar el desempeño de la clasificación entre ban-

dola, tiple requinto, guitarra y tiple realizada por las técnicas kNN, Árbol de Decisión, SVM y Análisis Discriminante, se escogió un k igual a cinco, por tanto la base de datos de 1000 muestras, fue dividida en cinco partes iguales, utilizando 800 muestras para entrenar los clasificadores y 200 para realizar la evaluación, repitiendo el proceso cinco veces y promediando sus resultados.

En la evaluación del desempeño de la clasificación entre bandola, guitarra, tiple y tiple requinto, usando Redes Neuronales Artificiales, la base de datos se dividió de la siguiente manera: 700 muestras fueron utilizadas para el entrenamiento del clasificador, 150 muestras para realizar la validación y 150 muestras para probar el clasificador.

3. Resultados y discusión

Las Exactitudes alcanzadas por cada una de las cinco técnicas de aprendizaje de máquina, usadas para la identificación de la bandola, tiple, guitarra y tiple requinto se muestran en la **Tabla 6**, en ella se puede ver que el uso de PCA para la reducción de dimensionalidad genera diferentes resultados de Exactitudes en los clasificadores.

Figura 6. Ejemplo de Validación Cruzada para k=5, (Cortés Osorio n.d.).

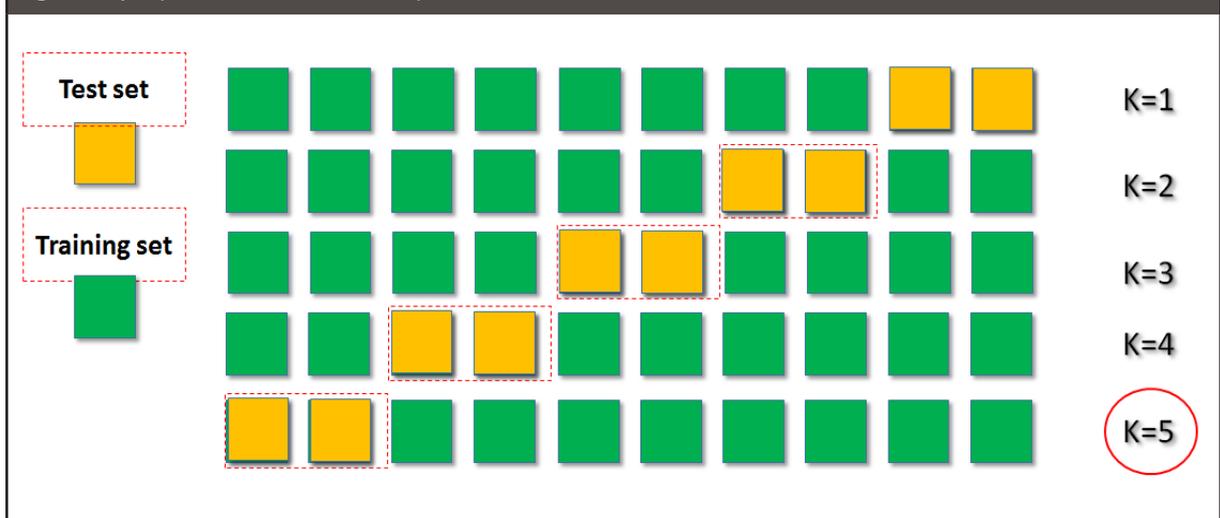


TABLA 6. PORCENTAJE DE EXACTITUD ALCANZADAS POR LAS TÉCNICAS DE APRENDIZAJE DE MÁQUINA UTILIZADAS EN LA IDENTIFICACIÓN DE LOS INSTRUMENTOS MUSICALES.

Exactitudes						
Técnica de Aprendizaje de Máquina	Especificaciones	las primeras cinco frecuencias parciales	Feature selection	PCA con 1	PCA con 100	PCAcon 1000
kNN	Distancia Euclidiana con número de vecinos 1	92,30%	98,80%	54,70%	99,10%	33,60%
	Distancia Euclidiana con número de vecinos 3	90,20%	98,9%	59,20%	97,60%	30,40%
	Distancia Chebyshev con número de vecinos 1	89,20%	97,80%	54,70%	98,10%	30,00%
	Distancia Chebyshev con número de vecinos 3	85,60%	97,20%	59,20%	95,3%	29,20%
Árbol de Decisión	4 Divisiones	60,90%	92,60%	61%	65%	65,30%
	100 Divisiones	91,90%	97,50%	59,20%	87,40%	83,20%
SVM	Kernel Lineal	68,40%	99,10%	56,50%	98,50%	46,60%
	Kernel Cúbica	90,80%	99,20%	44,00%	99,50%	32,40%
Análisis Discriminante	Lineal	56,40%	91,10%	56,40%	98,80%	97,40%
	Cuadrática	55,80%	97,90%	61,60%	75,50%	30,30%
ANNs	10 neuronas	85,60%	98,90%	62,80%	99,80%	91,90%
	20 neuronas	95,10%	99,60%	64,90%	99,70%	91,00%

A continuación, en la gráfica de la **Figura 7** se muestra las Exactitudes más altas de los cinco clasificadores, utilizando los métodos de reducción y selección de características. En ella se puede ver que la Exactitud menos favorable, entre las más altas, fue la alcanzada por Análisis Discriminante, aplicando como método reducción la extracción

de las primeras cinco frecuencias parciales y sus amplitudes normalizadas con un porcentaje de 56,4%. También se puede evidenciar que la técnica de ANNs presenta los resultados de las Exactitudes más altas y al mismo tiempo son los más uniformes en los tres métodos de reducción de características propuestos.

Figura 7. Exactitudes más altas presentadas por los 5 clasificadores, utilizando los tres métodos de reducción de características.

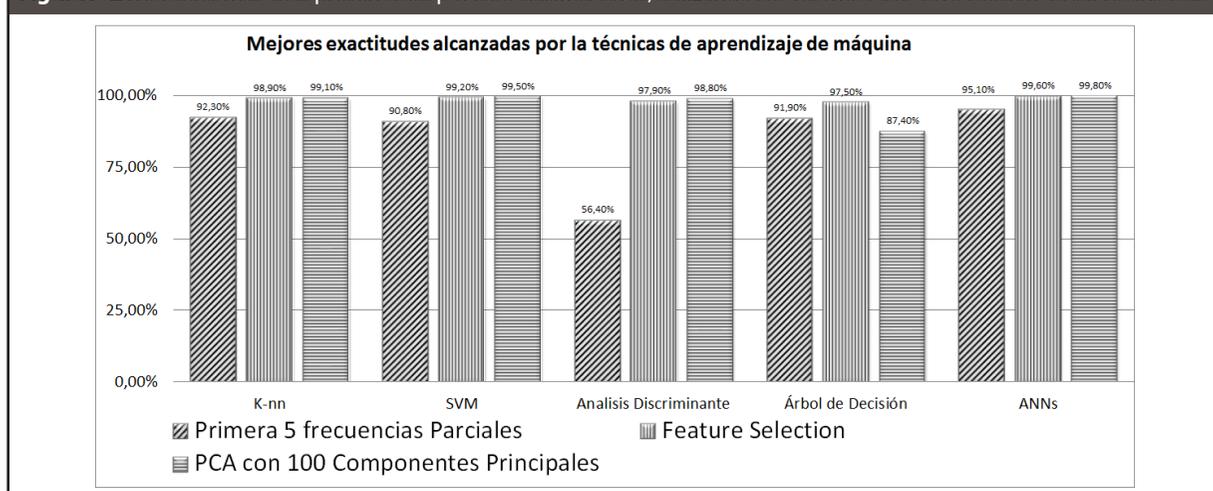
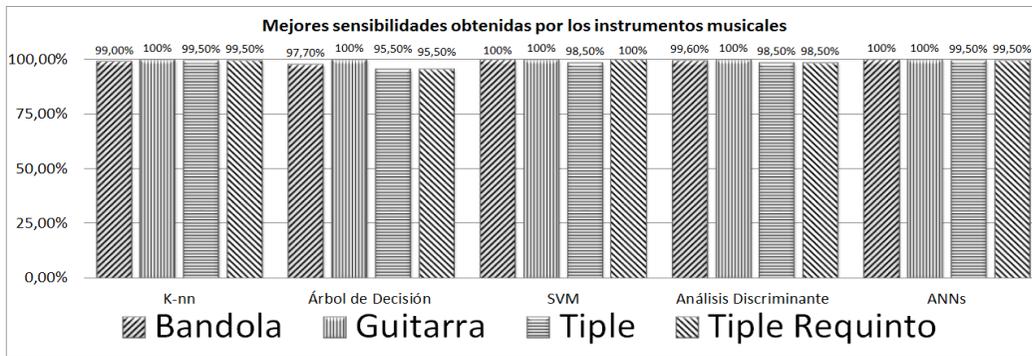


Figura 8. Sensibilidades más altas presentadas por los 4 Instrumentos Musicales en las 5 Técnicas de Aprendizaje de Máquina empleadas en su identificación.



En la **Figura 8**, se muestra las Sensibilidades más altas presentadas por la bandola, la guitarra, el tiple y el tiple requinto en cada una de las técnicas de aprendizaje utilizadas en su identificación. En ella se puede ver en forma global que las Sensibilidades alcanzadas fueron superiores al 95%, además también se puede apreciar que la Sensibilidad más alta es la obtenida por la guitarra con valor de 100% y la menos favorables, entre las mejores Sensibilidades, fueron las conseguidas por el tiple requinto y tiple con valor de 95,5%.

TABLA 7. RESULTADOS ARROJADOS POR LA TÉCNICA REDES NEURONALES ARTIFICIALES CON NÚMERO DE NEURONAS IGUAL A 10, APLICANDO PCA Y ELIJIENDO 100 COMPONENTES PRINCIPALES.

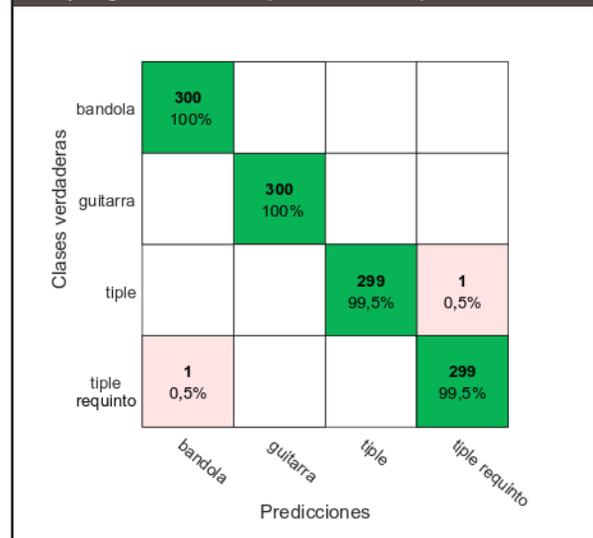
Medida	ANNs	
	No Neuronas	10
Sensibilidad		
FNR		
Bandola	100%	0%
Guitarra	100%	0%
Tiple	99,5%	0,5%
Tiple requinto	99,5%	0,5%
Exactitud	99,8%	
Error	0,2%	

Debido a lo extenso de la investigación, a continuación, solo se presentan los resultados más y menos destacados. En la **Tabla 7** se ilustran: lo valores en porcentaje de Exactitud y Error, junto a los porcentajes de la Sensibilidad y la Tasa de Falsos Negativos (FNR) presentada por la bandola, el tiple,

el tiple requinto y la guitarra, utilizando la técnica ANNs de 10 neuronas y aplicando a la matriz de entrenamiento el método PCA, eligiendo los primeros 100 componentes principales.

Según lo expuesto en la **Tabla 7**, se demuestra que aplicándole a la matriz de entrenamiento PCA, eligiendo 100 Componentes Principales, y utilizando la técnica Red Neuronal Artificial con 10 neuronas, en la identificación de los cuatro instrumentos musicales de cuerdas pulsadas, se puede alcanzar una Exactitud de 99,8%, dando Sensibilidades de 100% en la bandola y guitarra, y de 95,5% para los dos tiples.

Figura 9. Matriz de Confusión arrojada por la Red Neuronal Artificial con número de neuronas igual a 10, aplicando PCA y eligiendo 100 Componentes Principales

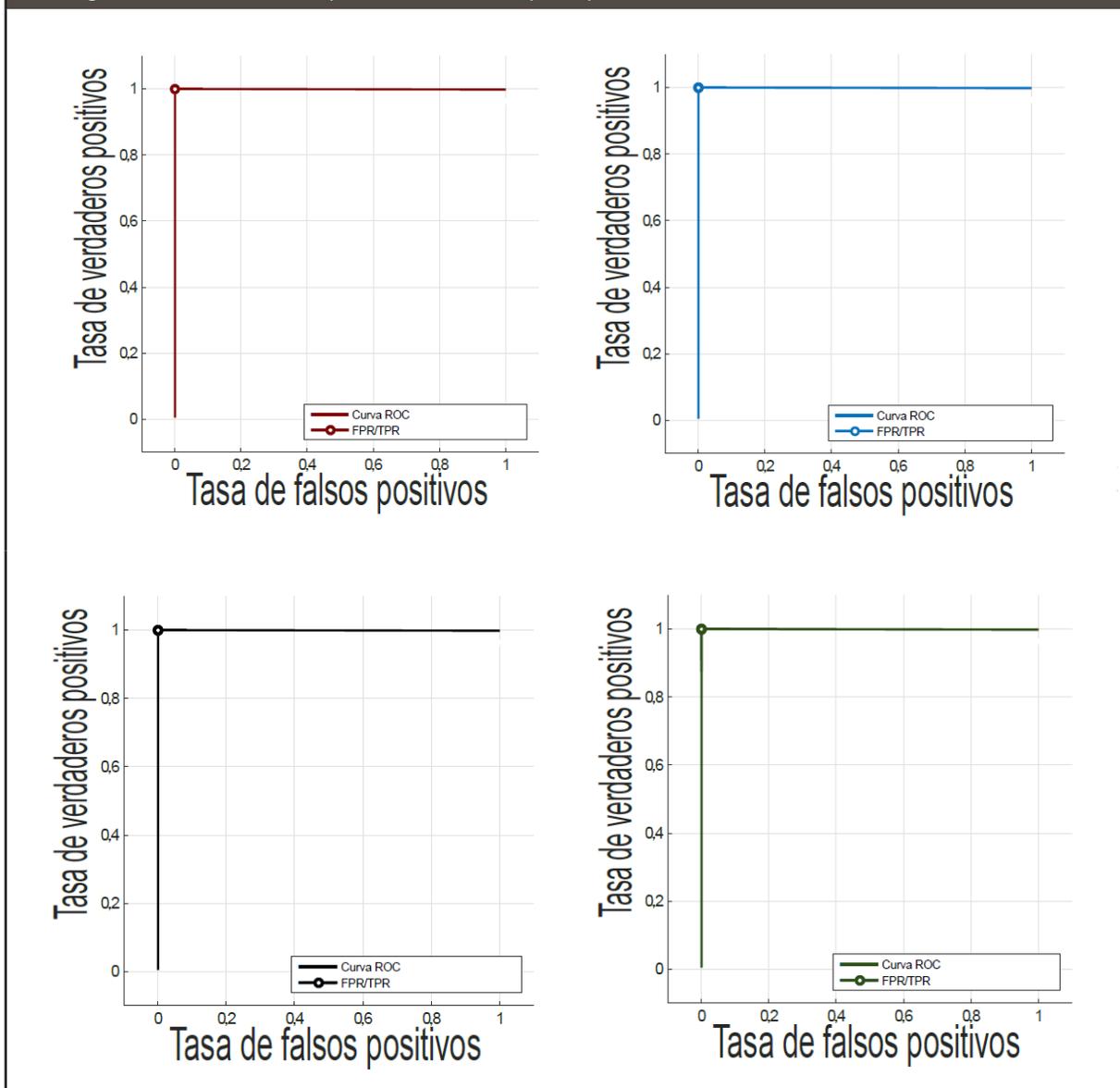


En la **Figura 9**, se presenta la matriz de confusión dada por el clasificador ANNs de 10 neuronas y aplicándole a la matriz de entrenamiento la técnica de reducción PCA, usando los 100 primeros componentes principales. En esta se puede observar que los instrumentos no poseen Tasas de Falsos Negativos relevantes (FNR), siendo el porcentaje más alto de 0,5%, además en esta matriz se observa que solo

una muestra de bandola fue identificada como un tiple requinto, y una muestra de tiple requinto fue identificado como tiple.

Después de exponer los resultados de la **Tabla 7**, en la **Figura 10** se muestran las curvas ROC que representan la Sensibilidad (TPR) frente a la Tasa de Falsos Positivos (FPR), de los cuatro instrumentos musicales identificados con ANNs.

Figura 10. Curvas ROC de la bandola, tiple, tiple requinto, y guitarra arrojadas por la técnica de Red Neuronal Artificial con número de Neuronas igual a 10, aplicando PCA y eligiendo 100 Componentes Principales. A. Curva ROC de bandola. B. Curva ROC de guitarra. C. Curva ROC de tiple. D. Curva ROC de tiple requinto.



Se puede apreciar, que las curvas ROC de los cuatro instrumentos musicales de cuerdas pulsadas, obtenidas en ANNs, se aproximan todas a la esquina superior izquierda lo que indica una adecuada clasificación.

En la **Tabla 8** se presentan: los valores en porcentaje de Exactitud y Error, junto a los porcentajes de la Sensibilidad y la Tasa de Falsos Negativos (FNR) presentada por la bandola, el tiple, el tiple requinto y la guitarra en la técnica kNN con distancia Chebyshev y número de vecinos igual a tres; aplicándole a la base de datos de entrenamiento el método de reducción de dimensionalidad PCA y eligiendo 1000 Componentes Principales.

TABLA 8. RESULTADOS ARROJADOS POR LA TÉCNICA KNN CON MÉTRICA DE DISTANCIA CHEBYSHEV, Y NÚMERO DE VECINOS IGUAL 3, APLICANDO PCA Y ELIGIENDO 1000 COMPONENTES PRINCIPALES.

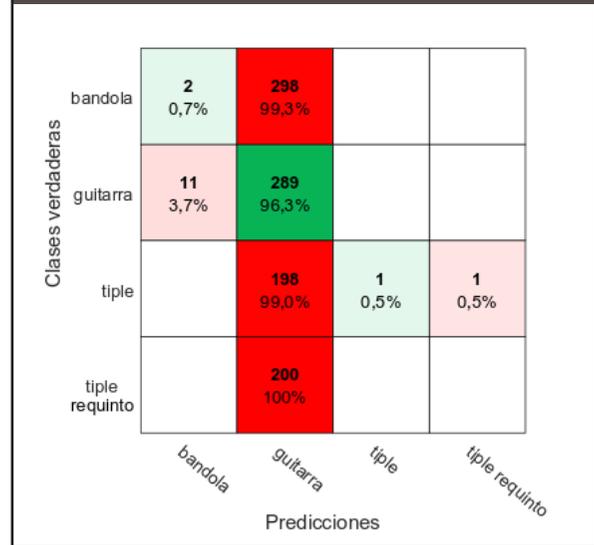
kNN		
Distancia	Chebyshev	
Número Vecinos	3	
Medida	Sensibilidad	FNR
Bandola	0,7%	99,3%
Guitarra	96,3%	3,7%
Tiple	0,5%	99,5%
Tiple requinto	0%	100%
Exactitud	29,2%	
Error	70,8%	

Conforme a la información dada en la **Tabla 8**, se puede observar que en la identificación de los cuatro instrumentos musicales, el método kNN con distancia Chebyshev y número de vecinos igual a tres; aplicándole a la base de datos de entrenamiento el método de reducción de dimensionalidad PCA y eligiendo 1000 Componentes Principales; arroja un Error de 70,8%, presentando los FNRs más altos en el tiple y tiple requinto con porcentaje de 99,5% y 100% respectivamente. También se puede ver que la guitarra posee la Sensibilidad más alta con valor de 96,3% en esta técnica.

En la **Figura 11**, se presenta la matriz de confusión dada por la técnica kNN con distancia Che-

byshev y número de vecinos igual a 3, aplicando PCA y usando 1000 componentes principales, en ella se puede observar que los falsos negativos presentes en la bandola, tiple y tiple requinto fueron identificados como guitarra.

Figura 11. Matriz de Confusión arrojada por la técnica kNN con métrica de distancia Chebyshev, y número de vecinos igual tres, aplicando PCA y eligiendo 1000 Componentes Principales.



Después de presentar en la **Tabla 8**, los valores de Exactitud y Error, junto a la Sensibilidad y FNR de cada instrumento musical, usando la técnica kNN, a continuación en la **Figura 12** se muestran las curvas ROC que representan la Sensibilidad (TPR) frente a la Tasa de Falsos Positivos (FPR), de los cuatro instrumentos de cuerdas pulsadas identificados con esta técnica.

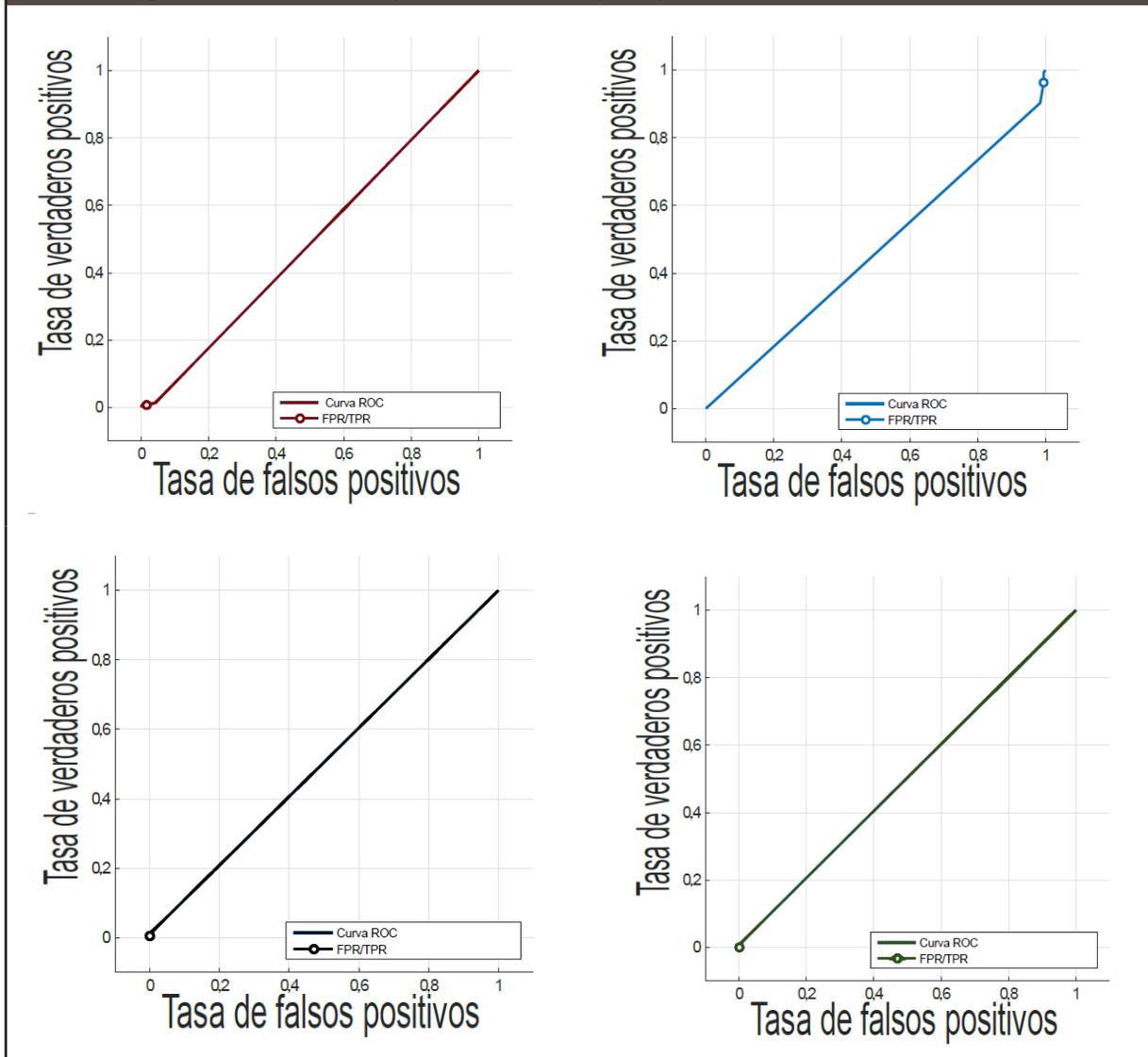
En la **Figura 12**, se puede apreciar que las curvas ROC obtenidas en el clasificador kNN, se aproximan a la diagonal de 45° que va desde la esquina inferior izquierda hasta la esquina superior derecha, lo que indica una clasificación incierta.

En esta investigación se evidenció, al igual como en la realizada por Essid, Richard, y David (2004), que la aplicación de Análisis de Componentes Principales (PCA) a la base de datos de entrenamiento, logra incrementar de forma significativa las exactitudes de las identificaciones, por otro lado,

los resultados también demuestran que la identificación realizada por medio de Redes Neuronales, da mejores Exactitudes que la realizada por el algoritmo kNN tal como lo mencionan Azarloo y Farokhi (2012), los cuales propusieron dos métodos de clasificación, redes neuronales de perceptrón multicapa y kNN, para clasificar siete instrumentos musicales diferentes (guitarra, piano, trompeta, percusión, voz, tambores, y contrabajo) dando mejores resultados las redes neuronales de perceptrón

trón multicapa, de igual modo, Wicaksana, Hartono y Wei (2006), realizaron la identificación de piano, violin, violonchelo, flauta, piccolo, y xilofono proponiendo dos estructuras de Redes Neuronales diferentes, con las cuales se obtuvieron Exactitudes de 94% y 88%, valores que se asemejan a las Exactitudes alcanzadas en este trabajo en la identificación de la bandola, guitarra, tiple y tiple requinto por el método de redes neuronales, valores que van desde 95,10% hasta 99,8%.

Figura 12. Curvas ROC de la bandola, tiple, tiple requinto, y guitarra arrojadas por el clasificador kNN con distancia Chebyshev y número de vecinos igual a tres, aplicando PCA y eligiendo 1000 Componentes Principales. A. Curva ROC de bandola. B. Curva ROC de guitarra. C. Curva ROC de tiple. D. Curva ROC de tiple requinto.



4. Conclusiones

En esta investigación se presenta la identificación de la bandola, la guitarra, el tiple y el tiple requinto en solo, instrumentos musicales de cuerdas pulsadas de la región andina colombiana, usando cinco técnicas de aprendizaje de máquina, las cuales fueron Késimo Vecino más Cercano (kNN), Árbol de Decisión, Máquina de Soporte Vectorial (SVM), Red Neuronal Artificial (ANNs) y Análisis Discriminante. Adicionalmente, a la matriz de características se le redujeron los datos de tres maneras: encontrando las 5 primeras frecuencias parciales y amplitudes normalizadas de la señales de audio, aplicando PCA y empleando Feature Selection. De este modo el mejor resultado fue arrojado por la técnica de Red Neuronal Artificial de 10 neuronas y aplicándole a la matriz de características PCA, utilizando 100 Componentes Principales, dando una Exactitud de 99,8%, Sensibilidades en guitarra y tiple del 100%, y 99,5% para bandola y tiple requinto.

El Error más alto presentado por los clasificadores, fue de 70,8% arrojado por la técnica kNN con distancia Chebyshev y número de vecinos igual a tres; aplicándole a la base de datos de entrenamiento el método de reducción de dimensionalidad PCA y eligiendo 1000 Componentes Principales.

Por otro lado, analizando la Sensibilidad mostrada por los instrumentos musicales en cada una de las técnicas y con los tres modos de reducción de datos, la guitarra resulta tener la mejor Sensibilidad con respecto a la Sensibilidad de los tres instrumentos restantes, pues cuando se reduce los datos extrayendo las primeras cinco frecuencias parciales junto a sus amplitudes normalizadas, la guitarra presenta el porcentaje más alto de Sensibilidad en las técnicas kNN, Máquina de Soporte Vectorial y Redes Neuronales Artificiales; aplicándole Feature Selection a la matriz de entrenamiento, la guitarra presenta la Sensibilidad más alta en las técnicas kNN, Árbol de Decisión, Análisis Discriminante Cuadrático, SVM, y Red Neuronales Artificiales. Cuando se le aplica PCA a los datos de entrenamiento y se eligen 1, 100 y

1000 Componentes Principales la guitarra presenta la más alta Sensibilidad en todos los clasificadores.

En cuanto, la Sensibilidad menos favorables fueron las presentadas por el tiple y el tiple requinto con valores del 0%, obtenidos cuando se le aplicó PCA a los datos de entrenamiento y se eligieron 1000 Componentes Principales, en las técnicas kNN y Análisis Discriminante Cuadrático.

Respecto a la calidad del sonido usada para realizar la caracterización de las señales de audio provenientes de la bandola, guitarra, tiple y tiple requinto fue el timbre, la cual se puede analizar cuando la señales de audio se encuentran en el dominio de la frecuencia, por ende el análisis de Fourier resultó una herramienta muy útil en la construcción de la matriz de entrenamiento de los clasificadores utilizados en la identificación de estos instrumentos musicales.

Se recomienda en futuros trabajos tomar muestras de las notas musicales teniendo en cuenta los diferentes recursos tímbricos que ofrecen la bandola, el tiple, el tiple requinto y la guitarra (sul ponticello, sul tasto y natural), también explorar otro tipo de características del sonido para incluirlas en la construcción de la base de datos de entrenamiento.

Referencias

- Azarloo, A. & Farokhi, F., 2012. Automatic Musical Instrument Recognition Using K-NN and MLP Neural Networks. In *2012 Fourth International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, pp. 289–294. Available at: <http://ieeexplore.ieee.org/document/6274357/> [Accessed February 27, 2017].
- Bentacourt, G.A., 2005. Las Máquinas de Soporte Vectorial (SVMs). *Scientia et Technica*, 1(27). Available at: <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/6895/4139> [Accessed March 14, 2017].
- Bishop, C.M., 2013. *Pattern Recognition and Machine Learning*.
- Burgueño, M.J., García-Bastos, J.L. & González-Buitrago, J.M., 1995. Las curvas ROC en la evaluación de las pruebas diagnósticas. *Medicina Clínica*, 104(17), pp.661–670. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7623495>.

- Chandwadkar, D.M. & Sutaone, M.S., 2012. Role of features and classifiers on accuracy of identification of musical instruments. *Proceedings - 2012 2nd National Conference on Computational Intelligence and Signal Processing, CISP 2012*, pp.66–70.
- Cortés Osorio, J.A., Evaluación del clasificador. Available at: [https://drive.google.com/drive/folders/0B_Gjz\]cocbgtZWk5Y2gyUlc2TEE](https://drive.google.com/drive/folders/0B_Gjz]cocbgtZWk5Y2gyUlc2TEE) [Accessed March 29, 2017a].
- Cortés Osorio, J.A., Machine Learning UTP. Available at: [https://drive.google.com/drive/folders/0B_Gjz\]cocbgtZWk5Y2gyUlc2TEE](https://drive.google.com/drive/folders/0B_Gjz]cocbgtZWk5Y2gyUlc2TEE) [Accessed September 21, 2017b].
- Cortés Osorio, J.A., Knott, A.M. & Chaves Osorio, J.A., 2012. Aproximación a la síntesis de la música a través del análisis de fourier. , (52), pp.129–135.
- Dash, M. & Liu, H., 1997. Feature selection for classification. *NeuroImage*, 57(3), pp.1003–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21984758>.
- El-Habi, A.M. & El-Ghareeb, M., 2014. Evaluation of Data Mining Classification Models. , 22(1), pp.151–165.
- Essid, S., Richard, G. & David, B., 2004. Musical instrument recognition on solo performances. In *European Signal Processing Conference*. pp. 1289–1292.
- Gómez, J. et al., 2012. Aplicación de redes neuronales en la clasificación de arcillas. *Revista EIA*, 17, pp.183–191.
- Harrington, P., 2012. *Machine Learning in Action*, Available at: <https://account.manning.com/dashboard%5Cnpapers3://publication/uuid/3F648CE7-406F-4663-BB0B-52E5D557D0F0>.
- Kitahara, T., Goto, M. & Okuno, H.G., 2004. Category-level identification of non-registered musical instrument sounds. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4.
- Londoño, M.E. & Tobón Restrepo, A., 2004. Bandola tiple y guitarra: de las fiestas populares a la música de cámara. *Artes, la revista*, 4(7), pp.44–68. Available at: <http://dialnet.unirioja.es/servlet/articulo?codigo=1213852>.
- Martin, K.D. & Kim, Y.E., 1998. Musical instrument identification: A pattern-recognition approach. *The Journal of the Acoustical Society of America*, 104, p.1768.
- Mitchell, T.M., 2009. *Machine learning*,
- Montaño Moreno, J. José & Palmer Pol, A., 2002. Redes Neuronales Artificiales aplicadas al Análisis de Datos. Available at: <http://www.tdx.cat/bitstream/handle/10803/9441/tjmm1de1.pdf> [Accessed March 10, 2017].
- Nagawade, M.S. & Ratnaparkhe, V.R., 2017. Musical Instrument Identification using MFCC. , pp.2198–2202.
- Rincón Gómez, J.R., Una escuela de bandola. *Para sonar mejor*, pp.68–82.
- Rodriguez Jorge Enrique, Rojas, E.A. & Franco, R.O., 2007. Clasificación de datos usando el método k-nn. *Vinculos*, 4. Available at: <http://revistavinculos.udistrital.edu.co/files/2012/12/Clasificación-de-datos-ED7.pdf> [Accessed March 8, 2017].
- Tang, J., Alelyani, S. & Liu, H., 2014. Feature Selection for Classification: A Review. *Data Classification: Algorithms and Applications*, pp.37–64.
- Vanessa, L. et al., 2009. Análisis acústico de un tiple colombiano por medio de la obtención de la respuesta al impulso con método directo. *Primeras jornadas regionales de acústica*. Available at: <https://www.fceia.unr.edu.ar/acustica/biblio/A009> (Acero) Analisis acustico de un tiple Colombiano.pdf [Accessed March 26, 2017].
- von Vaseghi, S., 2007. Multimedia signal processing: theory and applications in speech, music and communications. In pp. 417–481. Available at: <http://books.google.com/books?id=80EekFWZY5UYC&printsec=frontcover>.
- Wicaksana, H., Hartono, S. & Wei, F.S., 2006. Recognition of musical instruments. In *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS*. pp. 327–330.
- Xu, C., Maddage, N.C. & Shao, X., 2005. Automatic music classification and summarization. *IEEE Transactions on Speech and Audio Processing*, 13(3), pp.441–450.
- Zhu, W., Zeng, N. & Wang, N., 2010. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. *Northeast SAS Users Group 2010: Health Care and Life Sciences*, pp.1–9.

**PARA CITAR ESTE ARTÍCULO /
TO REFERENCE THIS ARTICLE /
PARA CITAR ESTE ARTIGO /**

Tobón-Gonzalez, I.J.; Cortés-Osorio, J.A. (2018). Identificación de instrumentos musicales de cuerdas pulsadas de la región andina colombiana en solo, mediante técnicas de aprendizaje de máquina. *Revista EIA*, 15(30), julio-diciembre, pp. 177-193. [Online]. Disponible en: <https://doi.org/10.24050/reia.v15i30.1245>