Identificación de Instrumentos Musicales utilizando Análisis de Señales (MFCC) e Inteligencia Artificial

Identification of Musical Instruments Using Signal Analysis (MFCC) and Artificial Intelligence.

M. Alan Salomón Vázquez Robledo[1] Dr. Juan Manuel López Hernández [2] Dr. Francisco Javier García Ledesma[3]
[1] Departamento de Estudios Multidisciplinarios, Campus Irapuato-Salamanca (sede Yuriria), Universidad de Guanajuato. as.vazquezrobledo@ugto.mx
[2] Departamento de Estudios Multidisciplinarios, Campus Irapuato-Salamanca (sede Yuriria), Universidad de Guanajuato. jmlopez@ugto.mx
[3] Departamento de Música y Artes Escénicas, Campus Guanajuato, Universidad de Guanajuato. pacochelo22@gmail.com

Resumen

La música ayuda a crear una identidad cultural a cada persona y su aporte en cada cultura genera cualidades artísticas en cada región geográfica. En la actualidad, el desarrollo de la tecnología ha ayudado a tener una evolución constante de cómo se crea la música digital. Esta evolución ha permitido una mayor accesibilidad a diferentes contenidos de archivos musicales además presenta un campo emergente con nueva información, características distintivas y con ello nuevas problemáticas a resolver. La Identificación de Instrumentos Musicales ha despertado el interés de los investigadores en los campos de la inteligencia artificial y análisis de señales. En este trabajo, se propone un modelo de identificación de instrumentos musicales, utilizando el conjunto de audios Sound Samples, desarrollado por Philharmonia Orchestra (s. f.), mediante el uso de las técnicas de Aprendizaje de Máquina, como el Perceptrón Multicapa (MLP por sus siglas en inglés, Multi-layer Perceptron). Estas mismas serán utilizadas para el entrenamiento y posteriormente la evaluación del sistema mediante métricas como la Precisión, Exactitud, F1 score y Recall, contribuyendo al conocimiento en Inteligencia Artificial.

Palabras clave: Aprendizaje de Máquina (ML), Perceptrón Multicapa (MLP), Máquinas de Soporte Vectorial (SVM), Vecinos más Cercanos (KNN).

Introducción

A lo largo de la historia, se muestra una importante relación en la expresión artística del ser humano y la música. El gran valor cultural y su relación con nuevas áreas de investigación en la música, han tomado más importancia en los últimos años, relacionándose con nuevos desafíos tecnológicos. Entre estos desafíos podemos encontrar el concepto musical de digitalización y el uso de plataformas digitales que necesitan procesar un gran volumen de información

En este momento, los profesionales del área de la música tienen un gran desafío debido al crecimiento de datos disponibles en este campo y a causa de la complejidad de la organización de los mismos. Es necesario el desarrollo de una herramienta innovadora de análisis automático de datos que contenga un enfoque multidisciplinario. Que busque superar los desafíos de gestión de los datos y aprovechar las oportunidades proporcionadas por el entorno digital actual.

La tarea de los sistemas de Recuperación de Información Musical (MIR, por sus siglas en inglés, Music Information Retrieval systems) se define como el área encargada del análisis de la estructura y extracción de la información necesaria de una señal de audio como lo plantea (Lucena et al., 2020). Algunas de las tareas principales de los sistemas MIR se centran en funciones como identificación de artistas, clasificación de géneros, clasificación de estados de ánimo, notación musical y la identificación de Instrumentos Musicales. Estas tareas nos permiten obtener información musical de las señales de audio.

Estado del arte

Los sistemas de aprendizaje automático han demostrado tener éxito en tareas de identificación de instrumentos, como demuestran los estudios presentados en esta sección.

Una de las ventajas en el aprendizaje de máquina (Machine Learning), es el análisis específico donde los métodos necesitan un menor poder computacional y lo que permite reconocer los atributos de las señales de audio de las herramientas musicales de una forma óptima.

Ahora, considerando las técnicas de aprendizaje de máquina (Machine Learning), es importante destacar que los siguientes estudios se centran en el conjunto de datos Philharmonia Orchestra (s. f.). Las técnicas de aprendizaje de máquina (Machine Learning) son computacionalmente factibles debido a limitaciones de hardware o datos, que permiten trabajar sin la necesidad de contar con un hardware computacional muy exigente.



Figura 1. Orquesta Philharmonia grabando en los estudios AIR.
Fuente: Fotografía por M. Vidor, 2020, Philharmonia Orchestra (https://philharmonia-assets.s3.eu-west-1.amazonaws.com/uploads/2020/02/05110834/Philharmonia_Air-Studios_screen-shot_3_credit-Marina-Vidor_medium-res-960x561.jpg).

(Toghiani-Rizi y Windmark, 2017), utilizaron una red neuronal artificial (RNA) entrenada con muestras de audio transformadas al dominio de la frecuencia. Para verificar el potencial informativo de los datos proporcionados, los autores experimentaron con un conjunto de características tanto del dominio temporal como del dominio de la frecuencia. Sus resultados experimentales registraron una precisión del 93,5 %, lo que verifica que la RNA puede reconocer patrones de audio complejos cuando se utilizan diferentes características.

(Uruthiran y Ranathunga, 2019), derivaron muchas características de los dominios de tiempo y frecuencia (centroide espectral, raíz cuadrática media, tasa de cruce por cero, MFCC, etc.) para entrenar los clasificadores de árboles de decisión, k vecinos más cercanos (KNN) y máquinas de vectores de soporte (SVM). De estos clasificadores, la SVM tuvo la máxima precisión media del 93,37%, lo que indica su competencia para manejar conjuntos de características mixtas en tareas de clasificación de instrumentos.

En el estudio de (Tu y Li, 2023), se utilizaron modelos de perceptrón multicapa (MLP) y máquina de vectores de soporte (SVM) para la clasificación de notas orquestales aisladas. Los resultados indicaron que la MLP superó a la SVM, con una tasa de precisión del 87%. En base a esta observación, se puede afirmar que la MLP ofrece un rendimiento óptimo al procesar muestras de audio aisladas.

Además, (Su, 2023) realizó un análisis comparativo de seis modelos para la clasificación de instrumentos, incluyendo k vecinos más cercanos (KNN), máquina de vectores de soporte (SVM), modelado de mezcla gaussiana (GMM), redes neuronales artificiales (ANN), redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN). El modelo de aprendizaje profundo CNN alcanzó la máxima precisión del 96,82%. Entre los modelos de aprendizaje automático, Los algoritmos KNN, SVM y GMM alcanzaron precisiones del 93, 34%, 73, 35% y 64,85 %, respectivamente. Además, SVM logró el mejor tiempo de cálculo, de 0,53 segundos. Estos resultados indican un desequilibrio entre la precisión de la clasificación y el tiempo de procesamiento en los distintos tipos de modelos.

Además, (Bhagyalakshmi y Anandaraju, 2025), propusieron un modelo KNN que utiliza características MFCC optimizadas y ajustadas, y lo compararon con la SVM. Se observó que la SVM alcanzó una precisión de clasificación de instrumentos del 90%, mientras que el método propuesto con KNN obtuvo una precisión superior, del 94%. Estos hallazgos demuestran la eficacia de utilizar características MFCC junto con KNN para lograr un alto rendimiento en la clasificación.

(Ezirmik y Çiloğlugil, 2026), propusieron un enfoque basado en múltiples características para la clasificación de instrumentos musicales mediante algoritmos de aprendizaje automático. El mejor rendimiento en términos de precisión se obtuvo fue de 91% con Random Forest y 93% con XGBoost. Estos algoritmos de aprendizaje automático se adaptaron particularmente bien para distinguir diferencias acústicas en la música.

Metodología

La metodología que se implemento es de carácter cuantitativo experimental, ya que se hará uso de datos numéricos y métricas para evaluar los resultados obtenidos.

Selección de Herramientas y Tecnologías

Uno de los primeros pasos a desarrollar en este trabajo, y como primera propuesta, se trabajó con la herramienta Python que es un entorno de computación numérica y un lenguaje de programación de alto nivel. Es ampliamente utilizado en ingeniería, matemáticas, ciencias de la computación y otras disciplinas técnicas y científicas. Se utilizaron audios del conjunto de muestras de audio Sound Samples, desarrollado por Philharmonia Orchestra (s. f.), el cual contiene grabaciones de alta calidad de diversos instrumentos musicales en diferentes notas, dinámicas e interpretaciones.

Preprocesamiento de Datos

Como primera actividad del desarrollo técnico de la metodología, se analizaron las señales disponibles en la base de datos. A partir de esto, y dependiendo de la necesidad, podremos aplicar procesamientos de adecuación de la señal como: filtrado, limpieza, normalización, etc.

Extracción de Características

En esta propuesta, una etapa importante es la extracción de características relevantes de las señales de audio, como espectrogramas, coeficientes cepstrales de frecuencia Mel (MFCC, por sus siglas en inglés, Mel-Frequency Cepstral Coefficients). Los MFCC son características ampliamente utilizadas en el procesamiento de señales de audio y en aplicaciones de reconocimiento de voz. Están inspiradas en la percepción auditiva humana y se derivan del análisis cepstral de la señal de audio. Los MFCC representan la amplitud del espectro del sonido de manera compacta, esto los ha vuelto la técnica de extracción de características más usada en reconocimiento del sonido.

Selección y desarrollo del Modelo

(Shinde, 2018), plantea que el Machine Learning, o aprendizaje automático es uno de los campos del mundo de la informática moderna. Se han realizado muchas investigaciones para hacer que las máquinas sean inteligentes. El aprendizaje es un comportamiento humano natural que también se ha convertido en un aspecto esencial de las máquinas. Es una disciplina de la inteligencia artificial que se centra en el desarrollo

de algoritmos y modelos que permiten a las computadoras aprender a partir de datos y realizar tareas específicas sin una programación explícita. En lugar de ser programadas para llevar a cabo una tarea particular de manera específica, las máquinas aprenden de forma autónoma a través de la experiencia.

(Lezama Gutierrez *et al.,* 2019) mencionan que una de las figuras más destacadas en el desarrollo temprano de los perceptrones multicapa fueron Marvin Minsky, junto con Seymour Papert. En su libro "Perceptrones", publicado en 1969, Minsky y Papert investigaron las limitaciones de los perceptrones de una sola capa y señalaron la necesidad de arquitecturas más complejas para abordar problemas más complejos de aprendizaje.

La propuesta para el modelo de clasificación inicia con la selección de la aproximación a utilizar. Se introducirá la viabilidad de utilizar las herramientas de Machine Learning como el MLP.La arquitectura de la red está formada de 13 neuronas en la capa de entrada correspondientes a los 13 (MFCC) principales de cada archivo de audio, después cuenta con dos capas ocultas de 64 y 32 neuronas, implementando una función de activación ReLU y una capa de salida correspondiente a las 20 clases de instrumentos musicales. La capa de salida utiliza una función Softmax para convertir las salidas en probabilidades, con tantas neuronas como clases a predecir (por ejemplo, guitarra, bajo y batería).

Métricas de Evaluación

El modelo se evaluó utilizando métricas estándar como la exactitud, precisión, recall y F1-score. La exactitud mide el porcentaje total de predicciones correctas, proporcionando una visión general del rendimiento del modelo. La precisión indica qué proporción de las predicciones positivas son correctas, mientras que el recall mide cuántos de los casos positivos reales fueron correctamente identificados. El F1-score, es la media armónica entre precisión y recall, las fórmulas de estas métricas se presentan a continuación:

$$Exactitud = \frac{Número\ de\ predicciones\ correctas}{Número\ total\ de\ predicciones}, \quad (1)$$

$$Precisión = \frac{Verdaderos\ positivos}{Verdaderos\ positivos\ +\ Falsos\ positivos}, \quad (2)$$

$$Recall = \frac{Verdaderos\ positivos}{Verdaderos\ positivos\ +\ Falsos\ negativos}, \quad (3)$$

$$F1 = 2 \times \frac{Precision\ x\ Recall}{Precision\ +\ Recall}, \quad (4)$$

Resultados

A continuación, se muestran los resultados obtenidos al realizar las pruebas con el concentrado de audios de cada instrumento musical: acordeón, guitarra acústica, bajo, violonchelo, clarinete, contrabajo, flauta, guitarra, arpa, oboe, órgano, piano, saxofón, trompeta, trombón, tuba, viola, violín, xilófono, percusión, los cuales fueron procesados para extraer sus 13 Coeficientes Cepstrales de Frecuencia Mel (MFCC por sus siglas en inglés, Mel Frequency Cepstral Coefficients) principales para realizar la posterior clasificación por medio del algoritmo de Perceptrón Multicapa (MLP por sus siglas en inglés, Multilayer Perceptron). Del total de la data set: 13,681 archivos de audio, se utilizó el conjunto de entrenamiento (80%), aproximadamente de 10,944 archivos, además el conjunto de prueba (20%), consta de 2,737 archivos.

***Tabla 1**. Métricas de evaluación de resultados.*

Métrica	**Resultado**
Exactitud (Accuracy)	**94.51 %**
Precisión (Precision)	94.61 %
Sensibilidad (Recall)	94.51 %
F1-Score (Weighted)	94.44 %

La exactitud del modelo clasificó correctamente el 94.51% de los datos de prueba de las tres clases. Por lo cual la precisión nos muestra que, de todas las predicciones positivas hechas por el modelo el 94.61% fueron correctas. El recall identificó correctamente el 94.51% de todos los datos que en realidad pertenecían a las clases correctas. El F1-score es del 94.44%. Esta métrica es una combinación de la precisión y el recall, proporcionando un equilibrio entre ambas métricas. El promedio macro y promedio ponderado nos muestra un resultado similar en las tres diferentes métricas del 94% que utilizamos, finalmente se implementó el uso de la validación cruzada en 5 pruebas promediando los resultados las métricas para dicho número de pruebas, podemos visualizar dichos resultados en la siguiente tabla que tiene el nombre de Tabla 1. Métricas de evaluación de resultados.

***Tabla 2**. Comparación con trabajos previos.*

Estudio	**Método(s)**	**Características**	**Exactitud**
(Toghiani-Rizi y Windmark, 2017)	RNA	Características del dominio de la frecuencia	**93.5 %**
(Uruthiran y Ranathunga, 2019)	SVM, KNN	Características del dominio del tiempo y de la frecuencia	**93.37 %**
(Su, 2023)	Diversas características	Diversas características	**93.34 %**
(Tu y Li, 2023)	MLP, SVM	Notas orquestales aisladas	**87 %**
(Bhagyalakshmi y Anandaraju, 2025)	KNN, SVM	MFCC, RMS, ZCR, centroide espectral	**94 %, 90 %**
(Ezirmik y Çiloğlugil, 2026)	Bosque aleatorio, XGBoost	Características de MFCC	**93 %**
Este trabajo	**MLP**	**Características de MFCC**	**94.51 %**

Conclusiones

Se llevo a cabo la clasificación de 13,681 archivos de audio correspondientes a 20 clases de instrumentos musicales, mediante la extracción de características de los 13 principales (MFCC) y el uso del Perceptrón Multicapa (MLP). Después de aplicar las métricas de evaluación a los resultados obtenidos muestran un desempeño sobresaliente en la tarea de clasificación de instrumentos musicales, alcanzando una exactitud superior al 94.51 %. La similitud entre los valores de Accuracy, Recall y F1-Score indica un comportamiento equilibrado del clasificador y una adecuada capacidad de generalización sobre datos no observados durante el entrenamiento, podemos notar una mejora significativa del modelo propuesto en comparación con trabajos previos.

Algunos aspectos para mejorar son la implementación de otras técnicas de Aprendizaje Automático y probar ajustar la arquitectura para el perceptrón para clase que tuvieron un bajo desempeño al momento de llevar a cabo el experimento esto podría llevar a un mejorar el resultado.

Referencias

Bhagyalakshmi, R., y Anandaraju, M. B. (2025). Identification of specific musical instruments using machine learning models. *Journal of Integrated Science and Technology, 13(5)*, 1108. https://doi.org/10.62110/sciencein.jist.2025.v13.1108

Bello Rivera, M. A., Reyes García, C. A., Talavera Rojas, T. C., Quintero Flores, M., y Pérez Loaiza, R. E. (2023). Identificación automática de disfonías: Uso de algoritmos de aprendizaje automático. Ciencias de la Computación Aplicadas, 19(4), 14–25. https://doi.org/10.35784/acs202332

Constantinescu, C., y Brad, R. (2023). An overview on sound features in time and frequency domain. *International Journal of Advanced Statistics and ITyC for Economics and Life Sciences,* 13(1), 45–58. https://doi.org/10.2478/ijasitels-2023-0006

Cortes, C., y Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20(3),* 273–297. https://doi.org/10.1007/BF00994018

Cover, T. M., y Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13(1),* 21–27. https://doi.org/10.1109/TIT.1967.1053964

Dash, S. K., Solanki, S. S., y Chakraborty, S. (2023). A comprehensive review on audio-based musical instrument recognition: Human–machine interaction toward Industry 4.0. *Journal of Scientific and Industrial Research, 82(1),* 26–37. https://doi.org/10.56042/jsir.v82i1.70251

Ezirmik, A. H., y Çiloğlugil, B. (2026). A multi-feature approach for musical instrument classification using machine learning. *Journal of the Balıkesir University Institute of Science and Technology (BAUN Fen Bilimleri Enstitüsü Dergisi), 28(1),* 299–312. https://doi.org/10.25092/baunfbed.1706872

Ghisingh, S. V. M. (2016, December 16–18). Classifying musical instruments using speech signal processing methods. En 2016 IEEE Annual India Conference (INDICON). IEEE. 10.1109/INDICON.2016.7839034

Kostrzewa, B. K. (2022). Designing a training set for musical instruments identification. En Lecture Notes in Computer Science (Vol. 13350, pp. 599–610). *Springer*. https://doi.org/10.1007/978-3-031-08751-6_43

Lezama Gutiérrez, A., Suárez Carreño, F., y Rosales, L. (2019). *Computación inteligente y estados emocionales (1era. Ed)*. AutanaBooks.

Lucena, A., Pires, C., Nose-Filho, K., y Suyama, R. (2020, October 26). Musical instruments recognition using machine learning Techniques: MLP and SVM. *Brazilian Technology Symposium.*

Mahanta, S. K., Khilji, A. F. U. R., y Pakray, P. (2021). Deep neural network for musical instrument recognition using MFCCs. *Computación y Sistemas,* 25(2), 351–360. https://doi.org/10.13053/cys-25-2-3946

Martínez, G., y Aguilar, G. (2013). Reconocimiento de voz basado en MFCC, SBC y espectrogramas. I*ngenius. Revista de Ciencia y Tecnología, (10)*, 12-20. https://www.redalyc.org/articulo.oa?id=505554816003

Müller, M., Ellis, D. P. W., Klapuri, A., Richard, G., y Sein, J.-P. (2011, November). Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing, 5(6),* 1088–1110. https://doi.org/10.1109/JSTSP.2011.2112333

Philharmonia Orchestra. (s. f.). Sound samples. https://philharmonia.co.uk/resources/sound-samples/

Shah, D., Narayanan, A., y Espinosa-Ramos, J.I. (2022). Utilizing the Neuronal Behavior of Spiking Neurons to Recognize Music Signals Based on Time Coding Features. I*EEE Access,* 10, 37317-37329.

Shinde, P. P., y Shah, S. (2018). A review of machine learning and deep learning applications. En 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1–6). IEEE. https://doi.org/10.1109/ICCUBEA.2018.8697857

Su, Y. (2023). Instrument classification using different machine learning and deep learning methods. Highlights in Science, *Engineering and Technology,* 34, 136–142. https://doi.org/10.54097/hset.v34i.5435

Toghiani-Rizi, B., y Windmark, M. (2017). Musical instrument recognition using their distinctive characteristics in artificial neural networks (arXiv:1705.04971). *arXiv.* https://doi.org/10.48550/arXiv.1705.04971

Tu, H., y Li, Y. (2023). Neural network for music instrument identification [Proyecto final del curso CS229: Machine Learning]. Stanford University.

Uruthiran, P., y Ranathunga, L. (2019). Optimization of feature selection and classification of oriental music instruments identification. En 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 120–125). IEEE. https://doi.org/10.1109/AiDAS47888.2019.8970974

Vidor, M. (2020, 5 de febrero). Orquesta Philharmonia grabando en los estudios AIR [Fotografía]. Philharmonia Orchestra. https://philharmonia-assets.s3.eu-west-1.amazonaws.com/uploads/2020/02/05110834/Philharmonia_Air-Studios_screen-shot_3_credit-Marina-Vidor_medium-res-960x561.jpg

Vieira, J. T. R. (2021). Automatic classification of instruments from supervised methods of machine learning. En Simpósio Brasileiro de Computação Musical (p. 7). *Departamento de Ciências da Computação, UFSJ.* https://doi.org/10.5753/sbcm.2021.19418