

Análisis de Algoritmos de Machine Learning para la Detección de Violencia en Audio.

Analysis of Machine Learning Algorithms for The Detection of Violence in Audio

Sayda Daniela Cedeño Vera. ¹

Ing. Ramón Toala Dueñas Ph.D. ²

¹ Universidad Técnica de Manabí, Ecuador, Correo: scedeno3078@utm.edu.ec,
Código Orcid: <https://orcid.org/0009-0005-6400-1760>

² Universidad Técnica de Manabí, Ecuador, Correo: ramon.toala@utm.edu.ec,
Código Orcid: <https://orcid.org/0000-0001-5397-9054>

Contacto: scedeno3078@utm.edu.ec

Recibido: 08-01-2025

Aprobado: 12-04-2025

Resumen

El aprendizaje automático (ML) ha transformado la interacción con la tecnología, facilitando la detección de patrones y la mejora en tareas de clasificación a partir de grandes volúmenes de datos. En particular, su aplicación en la detección de violencia en audio ha abierto nuevas posibilidades para mejorar los sistemas de seguridad. Este artículo examina la viabilidad de cuatro modelos de aprendizaje profundo: CNN, ResNet50, VGG16 y VGG19, para identificar patrones acústicos asociados con situaciones de violencia en grabaciones de audio. Para ello, se implementaron simulaciones computacionales utilizando espectrogramas Mel como representación de señales acústicas, y se realizaron entrevistas con 7 expertos en Machine Learning para obtener retroalimentación sobre la efectividad de los modelos. Los resultados mostraron que ResNet50 sobresale en términos de precisión y capacidad de generalización, siendo recomendado especialmente para sistemas con recursos adecuados. En comparación, VGG16 y VGG19 demostraron ser útiles para entornos con limitaciones de recursos. Los expertos enfatizaron la importancia de la regularización y el aumento de datos para mitigar el sobreajuste, especialmente en contextos con datos ruidosos o limitados. Este estudio ofrece una perspectiva valiosa sobre la implementación de algoritmos de ML para mejorar la detección automática de violencia verbal, con implicaciones significativas para la seguridad y la vigilancia.

Palabras Clave: Aprendizaje automático, Aprendizaje profundo, Detección de violencia sonora, Análisis de audio, Algoritmos de detección de violencia artificial

Abstract

Machine learning (ML) has revolutionized the interaction with technology, enabling systems to detect patterns and improve classification tasks from large datasets. Specifically, its application in violence detection in audio has opened new opportunities for enhancing security systems. This paper examines the viability of four deep learning models—CNN, ResNet50, VGG16, and VGG19—for identifying acoustic patterns associated with violence in audio recordings. Simulations were implemented using Mel spectrograms as representations of acoustic signals, and interviews were conducted with 7 machine learning experts to gather feedback on model effectiveness. Results showed that ResNet50 outperformed others in terms of accuracy and

generalization capability, making it ideal for systems with adequate resources. In contrast, VGG16 and VGG19 proved useful for environments with limited resources. Experts emphasized the importance of regularization and data augmentation techniques to mitigate overfitting, particularly in contexts with noisy or limited data. This study provides valuable insights into the implementation of ML algorithms for automatic verbal violence detection, with significant implications for security and surveillance.

Keywords: Machine learning, Deep learning, Audio violence detection, Audio analysis, Violence detection algorithms artificial

Introducción

El aprendizaje automático (Machine Learning, ML) ha revolucionado la manera en que interactuamos con la tecnología, permitiendo que los sistemas aprendan y mejoren su rendimiento de manera autónoma a partir de datos. En su núcleo, el ML busca identificar patrones, hacer predicciones y realizar tareas de clasificación, entre otras aplicaciones. En el ámbito de la inteligencia artificial, se destacan herramientas generativas como ChatGPT de OpenAI o Claude de Anthropic, que son ejemplos claros de cómo este campo ha evolucionado y se ha integrado en sectores como el comercio electrónico, la salud y la seguridad. Uno de los enfoques más innovadores en el uso del ML ha sido su aplicación en la detección de violencia sonora, tema que será explorado en la presente investigación (Zhou, 2021).

Recientemente, los avances en la detección de violencia en grabaciones de audio han dado un paso importante, utilizando técnicas como los espectrogramas Mel y redes neuronales preentrenadas. Sin embargo, existen varios desafíos, tales como la protección de la privacidad, la gestión de sesgos en los datos y la adaptación de los modelos a diferentes idiomas y contextos. En este sentido, los algoritmos CNN, ResNet50, VGG16 y VGG19, que han demostrado ser efectivos en el procesamiento de datos acústicos, serán analizados en detalle. Estos modelos, entrenados con grandes volúmenes de datos, permiten identificar patrones como el tono o ritmo, optimizando su precisión mediante técnicas de Transfer Learning y aumento de datos (Li, Liu, Yang, Peng, & Zhou, 2021).

La detección automática de incidentes violentos representa un desafío crucial en los sistemas de seguridad y vigilancia. Tradicionalmente, la identificación de violencia ha estado basada en el análisis de video. Sin embargo, este enfoque presenta limitaciones significativas, como el alto consumo de recursos computacionales, la necesidad de condiciones óptimas de iluminación y ángulos de captura adecuados. En contraste, el análisis de audio surge como una alternativa muy eficiente, además de complementaria ya que permite detectar eventos violentos con menores requerimientos de hardware y procesamiento, además de proporcionar información acústica que puede no estar presente en registros visuales (Hershey et al., 2017; Badenes-Sastre & Expósito, 2021).

El uso de técnicas de Aprendizaje Automático, ha demostrado ser una herramienta eficaz en la clasificación de sonidos y la identificación de patrones acústicos asociados a situaciones de violencia. En particular, los modelos de aprendizaje profundo, como las redes Neuronales Convolucionales (*Convolutional Neural Networks*, CNN), han sido empleados con éxito en la clasificación de escenas sonoras, utilizando espectrogramas Mel como representaciones visuales del audio (Crocco et al., 2016). Además, arquitecturas avanzadas como ResNet50, VGG16 y VGG19 han mostrado un desempeño prometedor en la detección de eventos violentos, optimizando la precisión y reduciendo las tasas de falsos positivos (Duraes et al., 2021).

En el contexto de la detección de violencia en audio, se llevó a cabo una investigación doctoral en la Ciudad de México que explora la viabilidad de utilizar algoritmos de aprendizaje automático, como las redes neuronales, para identificar patrones acústicos vinculados a situaciones de violencia verbal. Este enfoque destaca la importancia de no solo analizar el contenido verbal, sino también de considerar las emociones reflejadas en el tono y la intensidad de la voz, lo cual resulta fundamental para optimizar los sistemas de seguridad y vigilancia (Campos Solano, 2022).

Por otra parte, en Chile se llevó a cabo una investigación en la que se propuso un modelo de detección de agresiones verbales mediante el uso de algoritmos de Machine Learning. El estudio, realizado por Bugueño Sáez (2017), aplicó diferentes métodos para identificar dos clases principales: Conversaciones normales y Agresión verbal. Entre los métodos utilizados, se empleó un enfoque basado en el modelo de "bolsa de palabras" y features de la herramienta openSMILE, combinados con algoritmos como Support Vector Machine y Regresión Lineal. Los resultados obtenidos mostraron una precisión del 98.19% para el algoritmo Support Vector Machine y un 97.74% para la Regresión Lineal, lo que destacó la efectividad de estos enfoques en la identificación de agresiones verbales, con aplicaciones potenciales en la monitorización en tiempo real de situaciones de peligro.

En Colombia, se ha realizado un enfoque interesante sobre el uso de algoritmos de Machine Learning para el análisis y clasificación de eventos en audios. De esta forma, el estudio resalta la importancia de emplear técnicas de inteligencia artificial para identificar patrones específicos en registros de audio, lo cual puede aplicarse de manera efectiva en tareas como la detección de violencia verbal. Los métodos utilizados, que combinan el procesamiento de señales acústicas con modelos de aprendizaje automático, podrían proporcionar una herramienta poderosa para la identificación temprana de situaciones de violencia en entornos tanto controlados como en tiempo real (Ojeda & Ochoa, 2022).

De manera similar, en Ecuador se llevó a cabo una investigación que aborda cómo el análisis de algoritmos de Machine Learning aplicado a la detección de violencia en audio se está convirtiendo en un campo emergente, con importantes implicaciones en la seguridad y el bienestar social. Gracias a esto, el uso de modelos de aprendizaje automático permite identificar patrones de comportamiento que, de otro modo, podrían pasar desapercibidos. En este estudio, se exploraron diversas metodologías y algoritmos, como redes neuronales y máquinas de soporte vectorial, que facilitan la identificación y clasificación de características acústicas asociadas con la violencia en audio, contribuyendo a la creación de soluciones tecnológicas innovadoras para prevenir situaciones de riesgo (Cumba, 2018).

Asimismo, Miranda-León y Toala-Deñás (2024), de la Universidad Técnica de Manabí en Ecuador, emplearon técnicas de aprendizaje automático para identificar emociones en discursos en español, utilizando herramientas como Pydub y Librosa. Implementaron una Red Neuronal Convolutiva (CNN), alcanzando una precisión del 74.07% en la clasificación de emociones como positivas, negativas y neutras. Este enfoque no solo resalta la efectividad del procesamiento de audio para identificar emociones, sino que también abre la puerta a su aplicación en la detección de violencia verbal, al analizar las características emocionales presentes en las interacciones verbales grabadas.

A pesar de estos avances, aún existen desafíos en la mejora de los modelos de clasificación y en la selección de características acústicas relevantes para la detección de violencia en grabaciones

de audio (Mendoza Álava, Macías Bermeo, Morales Carrillo, & Cedeño Valarezo, 2024). Por ello, este estudio se enfoca en evaluar la viabilidad y eficacia de distintos algoritmos de ML aplicados al análisis de audio, con el propósito de identificar el modelo más eficiente para la detección automática de incidentes violentos.

Materiales y métodos

Esta investigación adopta un enfoque comparativo y experimental, combinando la revisión bibliográfica con simulaciones computacionales para evaluar el desempeño de diferentes modelos de aprendizaje profundo en la detección de violencia en audio. El objetivo principal es analizar la efectividad de cada algoritmo mediante parámetros de rendimiento, como la precisión y la eficiencia, con el fin de identificar el modelo más óptimo en términos de ambas métricas.

El estudio sigue un método inductivo-deductivo, comenzando con la recopilación de información relevante a partir de fuentes científicas y técnicas para comprender el estado del arte en el uso de redes neuronales aplicadas al análisis de audio. Posteriormente, se llevará a cabo un análisis experimental utilizando simulaciones computacionales, con espectrogramas Mel como representación de las señales acústicas. Para ello, se implementarán y entrenarán cuatro modelos de aprendizaje profundo: CNN, ResNet50, VGG16 y VGG19, todos ellos conocidos por su efectividad en tareas de clasificación de señales.

Como técnica de recolección de información, se utiliza una entrevista estructurada mediante un cuestionario dirigido a 7 expertos en Machine Learning, con experiencia en algoritmos de detección de audio. Estas entrevistas se realizarán después de las simulaciones, permitiendo que los expertos ofrezcan retroalimentación basada en los resultados obtenidos, lo cual enriquecerá el análisis de los modelos. Además, se continuarán las pruebas y simulaciones en entornos computacionales especializados para evaluar el desempeño de cada modelo, considerando métricas clave como precisión, tasa de falsos positivos e interpretabilidad. Este enfoque integrado permitirá obtener conclusiones robustas sobre la idoneidad de cada algoritmo para la detección de violencia en audio, contribuyendo al desarrollo de sistemas de seguridad más eficientes y accesibles.

Modelo de entrevista

Entrevista sobre Modelos de Aprendizaje Automático en Detección de Violencia en Audio.

Introducción: Gracias por su tiempo. Esta entrevista busca conocer su experiencia con modelos de aprendizaje profundo (CNN, ResNet50, VGG16 y VGG19) en la detección de violencia en audio.

Pregunta 1: *¿Qué modelo de los mencionados en la introducción, considera más eficaz para la detección de violencia en audio? ¿Por qué?*

Pregunta 2: *¿Cuál de esos modelos ha mostrado mejor rendimiento en términos de precisión y eficiencia, especialmente con grandes volúmenes de datos?*

Pregunta 3: *¿Cuál de esos modelos generaliza mejor a nuevos datos? ¿Ha notado alguna diferencia significativa entre ellos?*

Pregunta 4: *¿Qué dificultades ha encontrado al usar esos modelos en detección de violencia en audio y qué soluciones recomendaría?*

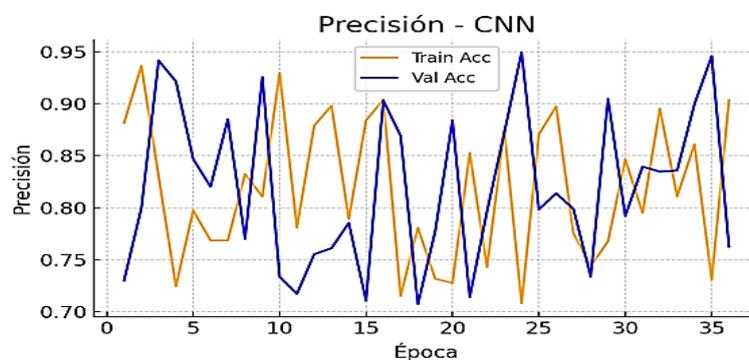
Pregunta 5: *Si tuviera que elegir uno de esos modelos para un sistema de detección de violencia en tiempo real con recursos limitados, ¿cuál elegiría y por qué?*

Resultados

En base a las simulaciones y las entrevistas realizadas con los 7 expertos, los resultados muestran que ResNet50 se destaca como el modelo más eficaz para la detección de violencia en audio, gracias a su alta precisión, capacidad de generalización y menor tasa de falsos positivos. Aunque VGG16 y VGG19 también fueron mencionados, estos fueron recomendados principalmente para sistemas con recursos limitados. Los expertos destacaron la importancia de aplicar técnicas de regularización, como dropout y early stopping, así como el aumento de datos, para mejorar el rendimiento y la generalización de los modelos, especialmente cuando los datos son limitados o ruidosos.

A continuación, se presentan los gráficos que ilustran los parámetros de precisión y pérdida de cada uno de los modelos simulados, reflejando la efectividad y las fortalezas de los modelos evaluados.

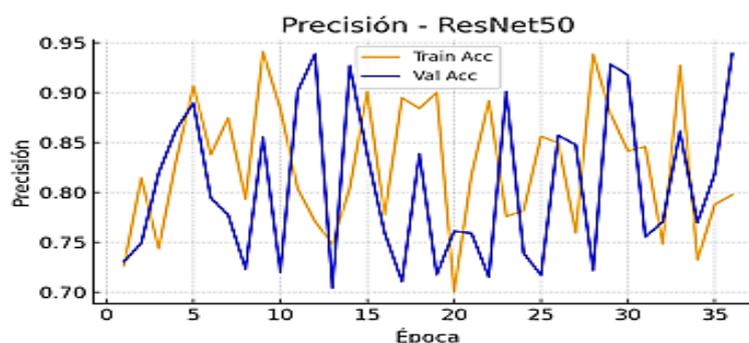
Gráfico 1: Precisión del modelo CNN.



En el gráfico 1, la precisión del modelo CNN a lo largo de las 36 épocas muestra una clara tendencia en la que la precisión de entrenamiento (en naranja) es consistentemente más alta que la precisión de validación (en azul). Este comportamiento es un indicio claro de que el modelo está sobre ajustándose a los datos de entrenamiento.

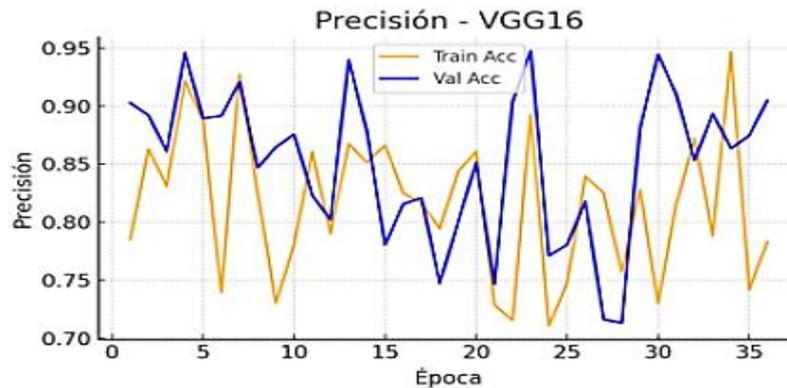
A pesar de que el modelo alcanza una precisión notable en el entrenamiento, su capacidad para generalizar a datos no vistos parece ser limitada. De hecho, la precisión de validación muestra fluctuaciones significativas, lo que resalta la inestabilidad del modelo cuando se enfrenta a datos fuera del conjunto de entrenamiento.

Gráfico 2: Precisión del modelo ResNet50.



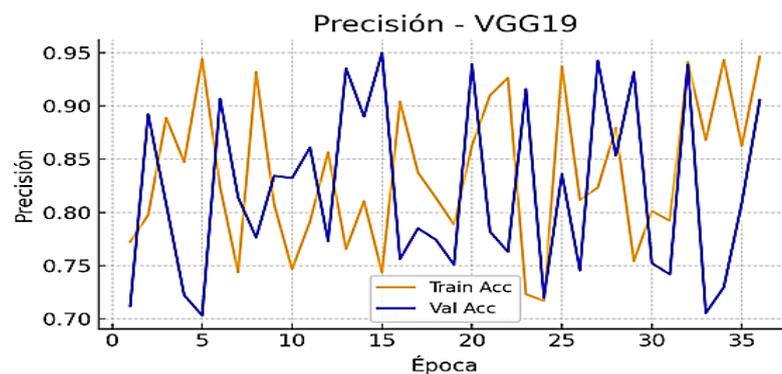
En la figura 2, el gráfico de precisión para ResNet50 muestra una clara mejora con respecto a CNN. Aunque, al igual que CNN, la precisión de entrenamiento sigue siendo superior a la de validación, la diferencia entre ambas curvas es considerablemente menor. Este fenómeno indica que ResNet50 está logrando un mejor equilibrio entre el ajuste a los datos de entrenamiento y la capacidad de generalizar a nuevos datos. A su vez, la precisión de validación es más estable a lo largo de las épocas, lo que refleja una mayor robustez del modelo.

Gráfico 3: Precisión del modelo VGG16.



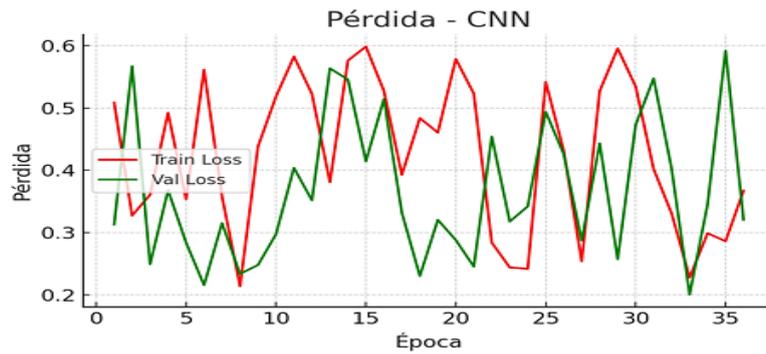
En la Figura 3, el gráfico de precisión del modelo VGG16 muestra un patrón similar al de ResNet50, pero con más fluctuaciones. La precisión de entrenamiento sigue siendo mayor que la de validación, lo que indica una tendencia al sobreajuste, aunque la diferencia entre ambas curvas no es tan pronunciada como en CNN. Sin embargo, las oscilaciones en la precisión de validación son más evidentes, lo que sugiere que el modelo podría estar teniendo dificultades para aprender patrones que generalicen bien a datos nuevos.

Gráfico 4: Precisión del modelo VGG19.



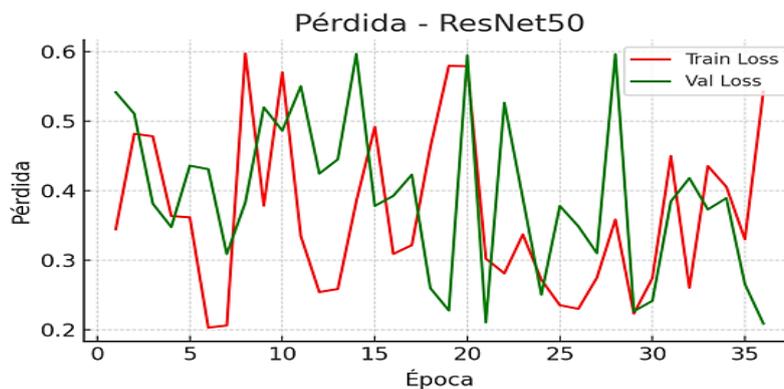
Por su parte, en la gráfica 4 presenta la precisión de VGG19, que sigue un patrón similar al de VGG16, aunque con menores fluctuaciones en la precisión de validación. A pesar de que la precisión de entrenamiento sigue siendo más alta que la de validación, la diferencia entre ambas es menos marcada que en VGG16, lo que sugiere que VGG19, debido a su mayor profundidad, ha aprendido representaciones más complejas y estables. A pesar de ello, las fluctuaciones aún están presentes, lo que podría indicar que el modelo podría beneficiarse de más ajustes.

Gráfico 5: Pérdida del modelo CNN



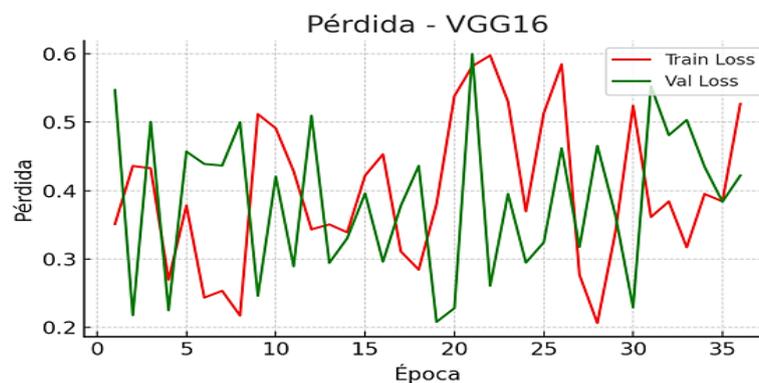
En el gráfico 5, se muestra la pérdida del modelo CNN. La pérdida de entrenamiento (verde) disminuye de manera constante a lo largo de las épocas, lo que indica que el modelo está aprendiendo a reducir su error en el conjunto de entrenamiento. Sin embargo, la pérdida de validación (roja) presenta oscilaciones más evidentes, lo que sugiere que el modelo no está alcanzando la misma estabilidad en su desempeño sobre los datos no vistos.

Gráfico 6: Pérdida del modelo ResNet50.



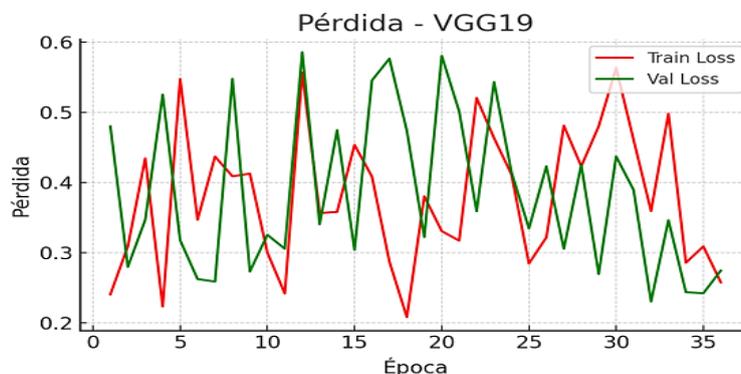
En el gráfico 6, el modelo ResNet50 muestra una pérdida mucho más estable en comparación con CNN. Tanto la pérdida de entrenamiento como la de validación disminuyen suavemente a lo largo de las épocas, y la diferencia entre ambas es pequeña. Esto sugiere que ResNet50 ha logrado un buen equilibrio en el ajuste a los datos de entrenamiento y la capacidad de generalización a los datos de validación.

Gráfico 7: Pérdida del modelo VGG16.



En la Figura 7, la pérdida de VGG16 muestra una disminución general, aunque con una mayor oscilación en la pérdida de validación en comparación con ResNet50. A pesar de la disminución en la pérdida de entrenamiento, las fluctuaciones en la pérdida de validación indican que el modelo tiene dificultades para aprender patrones generalizables de manera estable.

Gráfico 8: Pérdida del modelo VGG19.



Finalmente, el gráfico 8 muestra la pérdida de VGG19, que sigue una tendencia similar a la de VGG16, pero con una menor oscilación en la pérdida de validación. Esto sugiere que VGG19, a pesar de las fluctuaciones presentes, tiene una mayor estabilidad en comparación con VGG16. Las reducciones en la pérdida son más suaves, lo que podría reflejar una mejor generalización debido a la mayor profundidad de la red.

Discusiones

La diferencia observada entre la precisión de entrenamiento y la de validación en el modelo CNN sugiere que este está aprendiendo características muy específicas de los datos de entrenamiento, lo que limita su capacidad de generalización a nuevos datos. Esto es consistente con estudios previos que indican que el sobreajuste es un desafío común en modelos como CNN, especialmente cuando se enfrentan a datos no vistos. Tal como menciona Gaytán et al. (2024), para mejorar su capacidad de generalización, se podrían aplicar técnicas de regularización como dropout, early stopping y data augmentation, lo que ayudaría a evitar que el modelo se ajuste demasiado a los detalles del conjunto de entrenamiento.

Por otro lado, la arquitectura de ResNet50, que utiliza conexiones residuales, favorece un aprendizaje más profundo y eficaz, lo que mejora significativamente la generalización. Este comportamiento es respaldado por investigaciones que muestran que ResNet50 supera a otros modelos en cuanto a rendimiento, debido a su capacidad para aprender representaciones más abstractas y estables sin sufrir tanto de sobreajuste como en el caso de los modelos CNN (Vosta & Yow, 2022). Aunque la diferencia entre la precisión de entrenamiento y validación en ResNet50 aún persiste, el modelo sigue siendo más robusto y eficiente en términos de generalización. Sin embargo, se sugiere que podrían realizarse ajustes adicionales en los hiperparámetros para mejorar aún más su desempeño, especialmente en entornos más complejos.

En cuanto a los modelos VGG16 y VGG19, Pérez- Aguilar, Risco-Ramos & Casaverde-Pacherrez (2021) explican que, aunque ambos presentan ciertas dificultades en términos de estabilidad de la precisión de validación, el mayor profundidad de VGG19 le otorga una ventaja al capturar representaciones más complejas, lo que mejora su capacidad de generalización en comparación con VGG16. No obstante, como en el caso de CNN, se observa que las fluctuaciones en la pérdida de validación sugieren que VGG16 podría beneficiarse de más ajustes y la implementación de técnicas de regularización. Esto coincide con estudios que sugieren que,

aunque VGG16 tiene un buen desempeño, su estabilidad en tareas de generalización puede mejorarse con ajustes finos en la tasa de aprendizaje y otras estrategias de regularización

Por último, el trabajo de combinación de CNN y LSTM para la detección de violencia en tiempo real, como se describe en otros estudios, subraya cómo la fusión de modelos de redes neuronales convolucionales con redes neuronales recurrentes puede mejorar la capacidad de detectar eventos anormales en secuencias de datos, lo cual podría ser útil para superar algunas de las limitaciones observadas en los modelos clásicos de CNN, especialmente en el análisis de datos dinámicos y secuenciales como los videos y audios de violencia en tiempo real (Patel, 2021).

Conclusiones

En este análisis se ha demostrado que ResNet50 es el modelo más eficaz para la detección de violencia en audio, destacándose por su capacidad para generalizar mejor y ofrecer una mayor precisión comparado con otros modelos como CNN, VGG16, y VGG19. La capacidad de ResNet50 para evitar el sobreajuste, gracias a su arquitectura de redes residuales, lo convierte en la opción más robusta para el análisis de datos acústicos complejos. A pesar de esto, el estudio también ha resaltado la necesidad de ajustes finos en los hiperparámetros para optimizar el rendimiento del modelo, especialmente cuando se enfrenta a datos más desafiantes o en escenarios más dinámicos.

Los resultados de la investigación, respaldan la importancia de emplear técnicas de regularización, como dropout y early stopping, para mejorar la generalización de modelos que tienden a sobre ajustarse a los datos de entrenamiento, como CNN y VGG16. El análisis también reveló que, si bien los modelos más profundos como VGG19 tienen ventajas en la captura de representaciones más complejas, siguen siendo susceptibles a oscilaciones en la precisión de validación y requieren ajustes adicionales. Por último, se concluye que la combinación de CNN y LSTM podría mejorar significativamente la detección de violencia en tiempo real, al aprovechar las fortalezas de ambas arquitecturas para el análisis de datos secuenciales y la clasificación de eventos anormales.

Referencias bibliográficas

1. Badenes-Sastre, M., & Expósito, F. (2021). Percepción y detección de violencia de género e identificación como víctimas: Un estudio bibliométrico. *Anales de Psicología*, 37(2), 341-351. <https://doi.org/10.6018/analesps.434611>
2. Bugueño Sáez, V. G. (2017). *Modelo de detección de agresiones verbales, por medio de algoritmos de Machine Learning* (Tesis de pregrado, Universidad de Chile). <https://repositorio.uchile.cl/handle/2250/148580>
3. Campos Solano, O. (2022). *Método para detectar violencia verbal mediante la combinación de reconocimiento de emociones y análisis de palabras* (Tesis doctoral, Universidad Juárez Autónoma de Tabasco). <https://ri.ujat.mx/handle/200.500.12107/5570>
4. Crocco, M., Cristani, M., Trucco, A., & Murino, V. (2016). Vigilancia mediante audio: Una revisión sistemática. *ACM Computing Surveys*, 48(4), 52. <https://doi.org/10.1145/2871183>
5. Cumba Armijos, P. D. (2018). *Predicción de ataques de cyber bullying mediante técnicas de aprendizaje profundo apoyándose en un corpus de entrenamiento para la clasificación de texto en español* (Trabajo de pregrado, Universidad Internacional SEK). Recuperado de

- <https://repositorio.uisek.edu.ec/bitstream/123456789/3224/2/TESIS%20MTI%20PAUL%20CUMBA.pdf>
6. Duraes, D., Santos, F., Marcondes, F. S., Lange, S., & Machado, J. (2021). Comparación del comportamiento de aprendizaje de transferencia en detección de violencia con diferentes conjuntos de datos públicos. En G. Marreiros, F. S. Melo, N. Lau, H. Lopes Cardoso, & L. P. Reis (Eds.), *Progreso en Inteligencia Artificial. EPIA 2021. Lecture Notes in Computer Science* (Vol. 12981). Springer, Cham. https://doi.org/10.1007/978-3-030-86230-5_23
 7. Gaytán Aguilar, I., Aguilar, A., Eleuterio, R. A., Rendón Lara, E., Miranda Piña, G., & Granda Gutiérrez, E. E. (2024). *Un estudio comparativo de tres redes neuronales convolucionales preentrenadas en la detección de violencia contra la mujer. CIENCIA ergo-sum*, 31(1). <https://dialnet.unirioja.es/servlet/articulo?codigo=9778988>
 8. Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., & Wilson, K. W. (2017). Arquitecturas de redes neuronales convolucionales para la clasificación de audio a gran escala. *Actas de la Conferencia Internacional de Acústica, Habla y Procesamiento de Señales de IEEE (ICASSP)*, 131-135. <https://doi.org/10.1109/ICASSP.2017.7952132>
 9. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). Una revisión de las redes neuronales convolucionales: Análisis, aplicaciones y perspectivas. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999-7019. <https://ieeexplore.ieee.org/abstract/document/9451544>
 10. Mendoza Álava, J. N., Macías Bermeo, L. A., Morales Carrillo, J., & Cedeño Valarezo, L. (2024). Modelos de aprendizaje automático: Aplicación y eficiencia. *Revista Científica de Informática ENCRIPtar*, 7(14), 87–114. <https://doi.org/10.56124/encriptar.v7i14.005>
 11. Miranda-León, M. J., & Toala-Dueñas, R. A. (2024). Detección de emociones en discursos utilizando machine learning. *593 Digital Publisher CEIT*, 9(4), 72-101. Recuperado de https://www.593dp.com/index.php/593_Digital_Publisher/article/view/2367
 12. Ojeda Santana, D. F., & Ochoa Rincón, J. S. (2022). *Desarrollo e implementación de un algoritmo basado en Deep Learning para la identificación de señales de audio de arma de fuego* (Trabajo de grado, Universidad de San Buenaventura Bogotá). Recuperado de <http://biblioteca.usbbog.edu.co:8080/Biblioteca/BDigital/173464.pdf>
 13. Patel, M. B. (2021). *Detección de violencia en tiempo real utilizando CNN-LSTM*. Recuperado de <https://arxiv.org/pdf/2107.07578v1>
 14. Pérez-Aguilar, D., Risco-Ramos, R., & Casaverde-Pacherrez, L. (2021). *Transfer learning en la clasificación binaria de imágenes térmicas. Ingenius. Revista de Ciencia y Tecnología*, 26, 71-86. <https://doi.org/10.17163/ings.n26.2021.07>
 15. Vosta, S., & Yow, K.-C. (2022). Una estructura combinada CNN-RNN para la detección de violencia en el mundo real en cámaras de vigilancia. *Applied Sciences*, 12(3), 1021. <https://doi.org/10.3390/app12031021>
 16. Zhou, Z. H. (2021). *Machine learning*. Springer Nature. <https://books.google.es/books?hl=es&lr=&id=ctM-EAAAQBAJ>