

Análisis Comparativo de CNN-GRU y CNN-LSTM para la identificación de actividades humanas

Tatiana María Gaviria Sáenz¹

¹*Facultad de Ingeniería, Universidad de Antioquia, El Carmen de Viboral, Colombia.*

Resumen

El reconocimiento de actividades humanas (HAR) es fundamental en sectores críticos como la salud y la seguridad, donde la identificación precisa de acciones cotidianas potencia la efectividad de los sistemas de monitoreo. El avance en la tecnología de sensores y la informática ha reforzado este campo, mejorando significativamente la captura y el análisis de datos mediante dispositivos móviles y técnicas de inteligencia artificial. Este trabajo explora la aplicación de dos modelos de redes neuronales híbridas, CNN-LSTM y CNN-GRU, para la clasificación de actividades humanas utilizando datos de sensores de movimiento incorporados en un dispositivo móvil. Mediante el análisis del conjunto de datos de la UCI, se pone de manifiesto la eficiencia de estas arquitecturas híbridas en la interpretación de secuencias temporales complejas, con ambos modelos alcanzando una precisión superior al 90 %. La arquitectura CNN-LSTM demostró ser útil para la clasificación de actividades estáticas, mientras que el modelo CNN-GRU se destacó en la identificación precisa de actividades dinámicas. Estos resultados subrayan el potencial de las redes neuronales híbridas en la mejora de sistemas HAR marcando una dirección prometedora para futuras investigaciones en la clasificación precisa y eficiente de actividades humanas.

Palabras clave: HAR, CNN-LSTM, CNN-GRU, Sensores, Redes Neuronales Híbridas.

Abstract

Human Activity Recognition (HAR) is critical in essential sectors such as health and safety, where precise identification of daily activities enhances the effectiveness of monitoring systems. Advances in sensor technology and computing have strengthened this field, significantly improving data capture and analysis through mobile devices and artificial intelligence techniques. This work explores the application of two hybrid neural network models, CNN-LSTM and CNN-GRU, for human activity classification using motion sensor data embedded in a mobile device. The analysis of the UCI dataset reveals the efficiency of these hybrid architectures in interpreting complex temporal sequences, with both models achieving accuracy above 90 %. The CNN-LSTM architecture proved to be useful for classifying static activities, while the CNN-GRU model excelled in the precise identification of dynamic activities. These results underline the potential of hybrid neural networks in enhancing HAR systems, marking a promising direction for future research in the precise and efficient classification of human activities.

Keywords: HAR, CNN-LSTM, CNN-GRU, Sensors, Hybrid Neural Networks.

1. Introducción

El reconocimiento de actividades humanas (HAR) se ha establecido como un área de estudio relevante desde principios de la década de 1980, cobrando especial impulso en los últimos veinte años debido a su aplicación en campos críticos como la atención médica personalizada y la seguridad [1].

En este contexto, la identificación de actividades humanas cotidianas, tales como caminar, estar de pie o sentarse, ha adquirido una importancia fundamental para garantizar la funcionalidad y el rendimiento óptimo de los sistemas de monitoreo [2]. Avances en microelectrónica y tecnología de sensores, acompañados de progresos significativos en las ciencias computacionales, han permitido que la captura de datos sobre movimientos humanos sea más eficiente y detallada, posibilitando así que técnicas avanzadas de aprendizaje automático, transformen dicha información en inteligible y funcional para las máquinas [3, 4]. Este progreso ha desbloqueado nuevas posibilidades para el análisis de grandes conjuntos de datos provenientes de sensores, facilitando el desarrollo de modelos de HAR sofisticados y precisos [5].

Los datos necesarios para el desarrollo de modelos HAR, son adquiridos mediante sensores portables que incluyen giroscopios, acelerómetros, barómetros, magnetómetros, entre otros [6]. No obstante, en la actualidad los smartphones, que son omnipresentes y equipados con una variedad de sensores, se han convertido en una plataforma poderosa para adquirir datos de HAR ya que estos dispositivos móviles poseen la capacidad de recolectar información de movimientos en tiempo real permitiendo el registro continuo y no intrusivo de señales [7, 8].

La ampliación y sofisticación del HAR están intrínsecamente ligadas al progreso acelerado de la inteligencia artificial (IA), lo cual no solo extiende las capacidades y precisión de los sistemas HAR sino que también extiende su aplicabilidad [9]. En particular, las redes neuronales convolucionales (CNN) han demostrado su utilidad en el HAR ya que ofrecen un mecanismo eficaz pa-

ra el procesamiento y análisis automático de datos sensoriales de alta dimensionalidad además de captar sutilezas en los movimientos humanos que podrían ser imperceptibles para técnicas menos sofisticadas [5, 10, 11], sin embargo, mientras que las CNN son excelentes para capturar dependencias espaciales dentro de los datos de sensores, su capacidad para manejar dependencias temporales es limitada [12].

Lo anterior plantea un desafío dado que las actividades humanas son procesos dinámicos que se desarrollan en el tiempo. Por tanto, para comprender completamente la naturaleza secuencial de las actividades humanas, es esencial no solo analizar el aspecto espacial, sino también el temporal de los datos [13]. En la búsqueda de soluciones más eficientes, los investigadores han comenzado a explorar modelos híbridos que combinan la potencia de las CNN para el análisis espacial con modelos especializados en secuencias temporales, como la Memoria a Corto y Largo Plazo (LSTM) y las Unidades Recurrentes con Puertas (GRU), para mejorar la identificación y clasificación de las actividades humanas [14, 15, 16, 17].

Este estudio se enfoca en la implementación y evaluación comparativa de dos modelos híbridos avanzados (CNN-LSTM y CNN-GRU) mediante diversas métricas estadísticas. El objetivo principal es discernir las capacidades y eficiencia de cada arquitectura en la tarea de reconocimiento de actividades humanas, proporcionando así una comprensión más profunda de su aplicabilidad y rendimiento en escenarios de HAR.

2. Metodología

2.1. Adquisición y Procesamiento de Datos

Los datos para este estudio fueron extraídos del reconocido repositorio de aprendizaje automático de la Universidad de California, Irvine (UCI) [18], un referente en la comunidad de HAR. En las secciones subsiguientes, se proporcionará una descripción detallada de las características

del conjunto de datos, así como de los métodos empleados para su procesamiento y preparación para su uso en los modelos propuestos.

2.1.1. Conjunto de Datos UCI HAR

El conjunto de datos de HAR de la UCI contiene información de un experimento que incluyó a 30 voluntarios, con edades entre 19 y 48 años. Cada voluntario portaba un smartphone Samsung Galaxy S II, con el cual se registraron datos de aceleración lineal tridimensional y velocidad angular tridimensional. Estos datos se capturaron a una tasa de muestreo constante de 50 Hz, empleando el acelerómetro y el giroscopio incorporados en el dispositivo [18].

Las actividades realizadas por los participantes se dividieron en seis categorías distintas: caminar (WALKING), subir escaleras (WALKING_UPSTAIRS), bajar escaleras (WALKING_DOWNSTAIRS), estar sentado (SITTING), estar de pie (STANDING) y estar acostado (LAYING).

Antes de su inclusión en el conjunto de datos, las muestras se sometieron a un proceso de filtrado para la cancelación de ruido y un filtro Butterworth pasabajas de tercer orden con una frecuencia de corte de 20 Hz. El conjunto de datos consta de 10,299 muestras divididas en dos grupos: un conjunto de entrenamiento y un conjunto de prueba. El primero contiene 7,352 muestras que corresponden al 71.39 % del total, mientras que el segundo incluye las 2,947 muestras restantes 28.61 %. En la tabla 1 se presenta información adicional acerca de los elementos que componen el conjunto de datos, así como de las identificaciones de las clases y las proporciones muestrales.

En la figura 1 se presenta la incrustación de vecinos estocásticos t-distribuidos (t-SNE) con el fin de ilustrar la distribución de las muestras de diferentes clases sobre un plano bidimensional, en la cual es evidente que aunque las muestras corresponden a categorías distintas, hay una agrupación densa de los datos [4].

Tabla 1: Descripción del conjunto de datos

Actividades	Entrenamiento	Pruebas
Caminar	1226 (16.67 %)	496 (16.83 %)
Subir escaleras	1073 (14.60 %)	471 (15.99 %)
Bajar escaleras	986 (13.41 %)	420 (14.25 %)
Estar sentado	1286 (17.49 %)	491 (16.66 %)
Estar de pie	1374 (18.69 %)	532 (18.05 %)
Estar acostado	1407 (19.14 %)	537 (18.22 %)

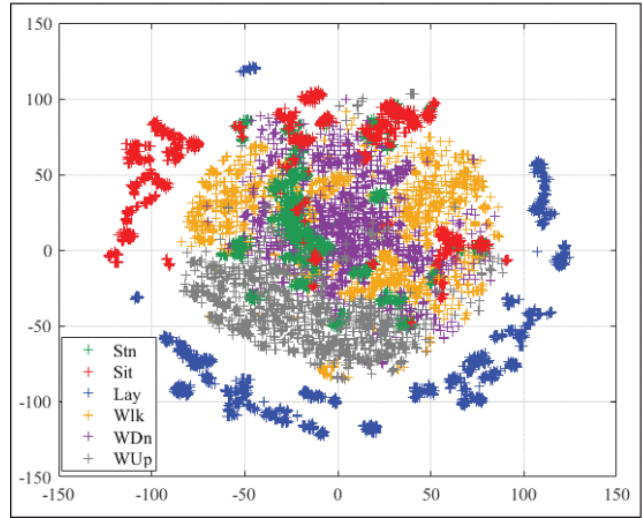


Figura 1: t-SNE Con las etiquetas correspondientes a Stn: estar de pie, Sit: estar sentado, Lay: estar acostado, Wlk: caminar, WDn: bajar escaleras y WUp: subir escaleras [4].

2.1.2. Procesamiento

Los datos fueron procesados y analizados utilizando el software Python (3.10.12) implementando bibliotecas especializadas como TensorFlow para la construcción y entrenamiento de los mo-

delos de redes neuronales, Pandas para la gestión eficiente de datos, y Numpy para operaciones matemáticas y de matrices. Los datos se ordenaron en arreglos tridimensionales de la forma: (muestras, pasos de tiempo, características) por lo que se tienen muestras correspondientes a series temporales de 128 elementos/pasos de tiempo con valores de 9 características, las cuales corresponden a la velocidad angular triaxial proveniente del giroscopio, la aceleración triaxial proveniente del acelerómetro (aceleración total) y la correspondiente aceleración triaxial corporal estimada. Adicionalmente, se aplicó la técnica de one-hot encoding a las variables de salida, transformando las etiquetas de clase de cada muestra en vectores binarios para facilitar su procesamiento por parte de los algoritmos de aprendizaje automático.

2.2. Construcción de los Modelos

Se desarrollaron modelos híbridos CNN-LSTM y CNN-GRU para evaluar su rendimiento en la clasificación de las actividades registradas. Los modelos consistieron en capas convolucionales para la extracción de características de las señales de tiempo, seguidas de capas LSTM o GRU para interpretar las dependencias temporales. Cada modelo fue compilado con el optimizador Adam, adecuado para el rápido descenso de gradientes en problemas de múltiples clases, junto con la función de pérdida 'categorical_crossentropy', que es estándar para la clasificación categórica. La precisión se selecciona como la métrica de rendimiento durante las fases de entrenamiento y validación.

2.2.1. Arquitectura del Modelo CNN-LSTM

Capa Convolucional: El modelo comienza con una capa convolucional (Conv1D), configurada con 64 filtros y un tamaño de kernel de 3, encargada de recibir las series temporales como entrada y extraer características espaciales relevantes. La función de activación ReLU se emplea para introducir no linealidades beneficiosas para el aprendizaje de patrones complejos en los datos.

Capa de Agrupación Máxima: Se aplica una capa de agrupación máxima (MaxPooling1D) con un tamaño de ventana de 2 para reducir la dimensionalidad de los datos y resaltar las características más significativas, lo que mejora tanto la eficiencia computacional como la capacidad de generalización del modelo.

Segunda Capa Convolucional: Se implementa una segunda capa convolucional con 128 filtros y un tamaño de kernel de 3 para una extracción aún más profunda de características espaciales de las señales de entrada.

Segunda Capa de Agrupación Máxima: Se incluye en la arquitectura con el fin de proporcionar otra reducción dimensional para destacar las características significativas de los datos.

Capa LSTM: Para interpretar las características extraídas por la capa convolucional en el contexto temporal, se introduce una capa LSTM con 64 unidades. Esta capa permite al modelo aprender dependencias a largo plazo y es crucial para entender la secuencialidad inherente a los datos de HAR.

Capas Conectadas y Regularización: La red continúa con una capa densa (Dense) de 64 unidades con activación ReLU para la interpretación de los datos procesados. Para mitigar el sobreajuste, se incluye una capa de (Dropout) con una tasa de 0.5, con el fin de regularizar el modelo descartando aleatoriamente un conjunto de activaciones durante el entrenamiento.

Capa de Salida: Finalmente, se añade otra capa densa con una unidad por cada clase de actividad y una función de activación softmax para la clasificación final.

La siguiente figura (2) proporciona una representación visual esquemática de la arquitectura del modelo CNN+LSTM utilizado en el estudio.

Capa (tipo)	Dimensión de salida	# de parámetros
conv1d_1 (Conv1D)	(None, 126, 64)	1792
max_pooling1d_1 (MaxPooling 1D)	(None, 63, 64)	0
conv1d_2 (Conv1D)	(None, 61, 128)	24704
max_pooling1d_2 (MaxPooling 1D)	(None, 30, 128)	0
lstm_1 (LSTM)	(None, 64)	49408
dense_2 (Dense)	(None, 64)	4160
dropout (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 6)	390
Parámetros totales: 80,454		
Parámetros entrenables: 80,454		

Figura 2: Resumen del modelo CNN-LSTM

2.2.2. Arquitectura del Modelo CNN-GRU

la arquitectura de la sección de CNN para este modelo guarda las mismas características que las implementadas en la construcción del modelo CNN-LSTM en cuanto a la estructura correspondiente a las dos capas convolucionales y de agrupación máxima, una vez establecida esta configuración, se añade la siguiente sección:

Capas GRU: El modelo incorpora dos capas de GRU. La primera capa GRU está configurada para devolver la secuencia completa lo que permite que la información temporal se pase a través de la capa. La segunda capa GRU, en cambio, solo devuelve el último paso de tiempo, condensando así la información temporal para la clasificación. Ambas capas GRU tienen 64 unidades diseñadas para aprender las dependencias a largo plazo entre los pasos de tiempo.

Regularización con Dropout: Entre las capas GRU se introduce una capa de (Dropout) con una tasa del 50 %, al igual que la establecida en la arquitectura CNN-LSTM.

Capas Densas para la Clasificación: Al final del modelo hay dos capas densas (Dense), una con 64 unidades y activación ReLU y la otra con una unidad por cada clase y activación softmax para la salida de clasificación final.

El resumen de la arquitectura del modelo se presenta en la figura 3.

Capa (tipo)	Dimensión de salida	# de parámetros
conv1d_3 (Conv1D)	(None, 126, 64)	1792
max_pooling1d_3 (MaxPooling 1D)	(None, 63, 64)	0
conv1d_4 (Conv1D)	(None, 61, 128)	24704
max_pooling1d_4 (MaxPooling 1D)	(None, 30, 128)	0
gru_2 (GRU)	(None, 30, 64)	37248
dropout_3 (Dropout)	(None, 30, 64)	0
gru_3 (GRU)	(None, 64)	24960
dropout_4 (Dropout)	(None, 64)	0
dense_6 (Dense)	(None, 64)	4160
dense_7 (Dense)	(None, 6)	390
Parámetros totales : 93,254		
Parámetros entrenables : 93,254		

Figura 3: Resumen del modelo CNN-GRU

2.3. Análisis estadístico

para una evaluación integral de las arquitecturas CNN-LSTM y CNN-GRU en la clasificación de actividades humanas, se aplicó un conjunto diversificado de métricas analíticas . Se incluyeron métricas como el accuracy, la precisión, sensibilidad (Recall) por clase, y la medida F1, junto con su promedio macro, para brindar una valoración comprensiva del desempeño de cada modelo. Además, se utilizó la pérdida de entropía cruzada media (CE-L) como un indicativo de la calidad del ajuste de los modelos a los datos. Para complementar, se realizó un análisis de la curva ROC (Receiver Operating Characteristic) y se calculó el área bajo la curva (AUC), pro-

porcionando una medida de la capacidad de los modelos para diferenciar entre clases.

A continuación se presentan las ecuaciones implementadas para el cálculo de las métricas evaluadas.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

Donde, TP = Verdadero Positivo, TN = Verdadero Negativo, FP = Falso Positivo, FN = Falso Negativo.

$$Precisión = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = \frac{2X \text{ Precision } X \text{ Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Pérdida de entropía cruzada media

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(\hat{y}_{ic}) \quad (5)$$

Donde,

N es el número total de ejemplos en el conjunto de datos.

C es el número de clases.

y_{ic} es una variable binaria que indica si la clase real del i -ésimo ejemplo es la clase c .

\hat{y}_{ic} es la probabilidad predicha de que el i -ésimo ejemplo pertenezca a la clase c .

3. Resultados y análisis

En esta sección se detallan los resultados obtenidos en el estudio, donde se desarrollaron y evaluaron comparativamente dos modelos híbridos de clasificación (CNN-LSTM y CNN-GRU) conforme al método descrito previamente. A continuación se presentan las figuras correspondientes a las matrices de confusión y las curvas ROC-AUC para cada modelo implementado así como sus respectivas métricas.

Tabla 2: Métricas por clase para el modelo CNN-LSTM

Métrica/ Clase	Caminar	Subir escaleras	Bajar escaleras	Estar sentado	Estar de pie	Estar acostado
Precisión	1.00	0.91	0.86	0.83	0.85	0.99
Recall	0.92	0.95	1.00	0.78	0.84	0.97
F1-Score	0.95	0.93	0.93	0.81	0.84	0.98

Tabla 3: Métricas por clase para el modelo CNN-GRU

Métrica/ Clase	Caminar	Subir escaleras	Bajar escaleras	Estar sentado	Estar de pie	Estar acostado
Precisión	0.99	0.91	0.88	0.83	0.80	1.00
Recall	0.95	0.92	0.98	0.76	0.85	0.96
F1-Score	0.97	0.91	0.93	0.79	0.83	0.98

Tabla 4: Métricas de los modelos implementados.

Métrica/ Modelo	CNN-LSTM	CNN-GRU
Precisión	0.91	0.90
Recall	0.91	0.90
F1-Score	0.91	0.90
CE-L	0.49	0.40

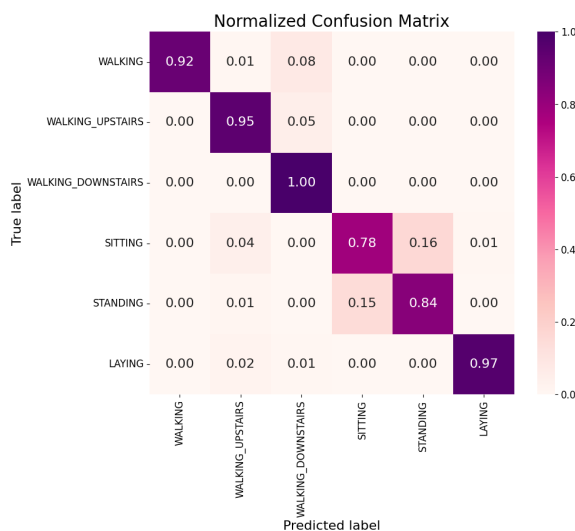


Figura 4: Matriz de confusión para el modelo CNN-LSTM

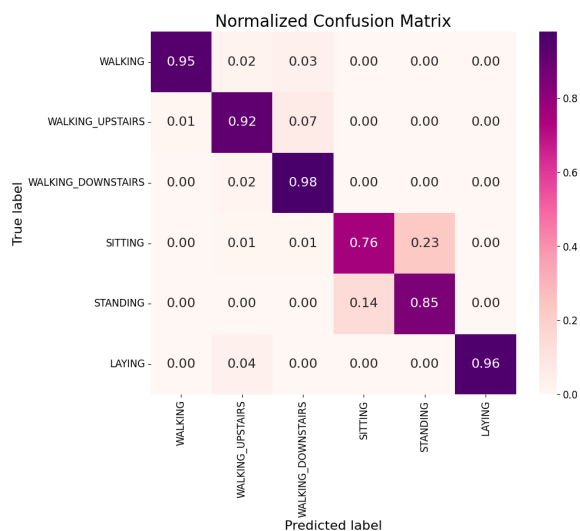


Figura 6: Matriz de confusión para el modelo CNN-GRU

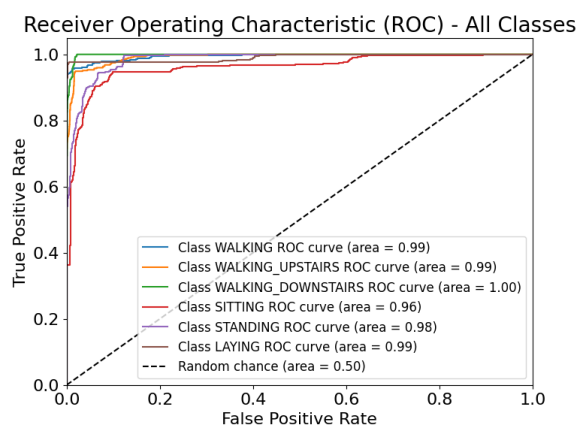


Figura 5: Curva ROC-AUC para el modelo CNN-LSTM

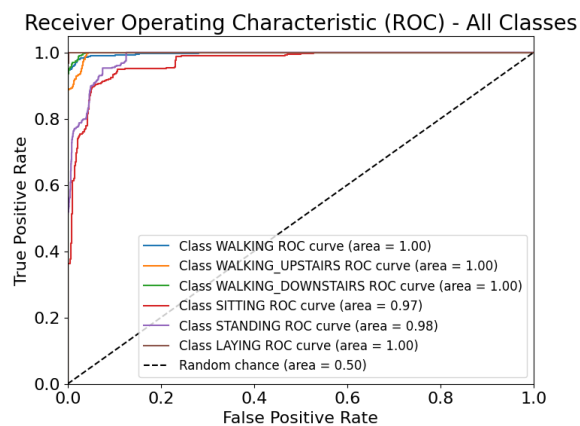


Figura 7: Curva ROC-AUC para el modelo CNN-GRU

El modelo CNN-LSTM alcanzó una precisión global del 91 %, reflejando una alta capacidad general para clasificar correctamente las actividades humanas. Al examinar la precisión por clase, el modelo demostró una capacidad perfecta para identificar la actividad 'caminar' con una precisión del 100 %, y mantuvo un desempeño sólido en las demás clases con la más baja siendo 'estar sentado' con un 83 %. La sensibilidad del mode-

lo, medida por el recall, fue alta en la mayoría de las clases, destacando en 'bajar escaleras' con un 97 % y mostrando áreas de mejora en 'estar sentado' con un 78 %.

El equilibrio entre la precisión y el recall se captura en las puntuaciones F1 por clase las cuales se mantuvieron robustas en todas las categorías, especialmente en 'estar acostado' con un 98 %. Las métricas macro para precisión, recall y F1 Score presentaron un valor consistente de 0.91, sugiriendo un rendimiento balanceado en la clasificación de las actividades. Sin embargo, la pérdida de entropía cruzada media fue de 0.49, indicando que hay margen para mejorar la eficacia del modelo en la generalización de los datos.

En comparación, el modelo CNN-GRU registró una precisión global del 90 %, mostrando una competencia similar en la tarea de clasificación de actividades humanas. La precisión por clase reveló una habilidad casi perfecta para identificar 'estar acostado' y 'caminar' con precisiones del 100 % y 99 % respectivamente. Sin embargo, se notó una menor precisión en 'estar de pie' con un 80 %, lo que indica una oportunidad para optimizar el modelo en esta clase específica.

El recall por clase fue alto en la mayoría de las categorías, con un mínimo de 76 % en 'estar sentado', lo que sugiere que el modelo puede ser menos sensible en la detección de esta actividad en particular. Las puntuaciones de F1 por clase fueron consistentemente altas, con un mínimo de 0.79 en 'estar sentado', lo que refleja un buen balance entre precisión y sensibilidad en la mayoría de las clases.

Las métricas macro mostraron un rendimiento adecuado con valores de 0.90 para precisión, recall y F1 Score, lo que indica un rendimiento general equilibrado y confiable del modelo. La pérdida de entropía cruzada media fue notablemente baja (0.40), lo que señala una afinidad y ajuste superior del modelo CNN-GRU con los datos.

Ambos modelos exhibieron una alta eficiencia

en la clasificación de actividades humanas, con métricas macro consistentes y precisión global superior al 90 %. Mientras que el modelo CNN-LSTM mostró una ligera ventaja en términos de precisión global, el modelo CNN-GRU tuvo un mejor rendimiento en la clasificación de algunas actividades clave y registró una menor pérdida de entropía cruzada media. Esto podría indicar una mejor capacidad del modelo CNN-GRU para ajustarse a los datos y una posible ventaja en la generalización a nuevos conjuntos de datos.

Estos resultados se manifiestan igualmente en las curvas ROC, donde se observó una cercanía a las coordenadas ideales de (0,1). La aproximación a este punto de referencia ideal indica que los modelos tienen una alta tasa de verdaderos positivos (sensibilidad) manteniendo al mismo tiempo una baja tasa de falsos positivos (1 - especificidad), confirmando así la robustez del modelo CNN-GRU y la competencia del CNN-LSTM en la identificación de actividades humanas.

Los hallazgos de este estudio se alinean coherente y favorablemente con los resultados previamente reportados en la literatura científica [4, 14, 16], corroborando la eficacia de las metodologías empleadas. Específicamente, estos resultados muestran una notable correspondencia con los obtenidos por Verma y colaboradores, quienes adoptaron un enfoque CNN-GRU empleando el mismo conjunto de datos y obtuvieron una precisión del 92.53 %.[16] . Este paralelismo en los resultados subraya la consistencia y reproducibilidad de las técnicas de CNN en la identificación de actividades humanas.

4. Conclusiones

En el presente estudio, se examinaron y compararon dos arquitecturas híbridas de aprendizaje profundo, CNN-LSTM y CNN-GRU, para la clasificación de actividades humanas a partir de datos de series temporales. Los modelos demostraron un rendimiento sobresaliente, con precisión global superior al 90 %, evidenciando la eficacia de las combinaciones de CNN con RNN pa-

ra el manejo de datos complejos. Se destacó la precisión del modelo CNN-LSTM en la detección de la actividad 'estar acostado' y la alta sensibilidad del modelo CNN-GRU en la mayoría de las categorías. No obstante, se identificaron oportunidades de mejora en ciertas actividades más desafiantes como 'estar sentado' y 'estar de pie', señalando la necesidad de refinamientos futuros.

El desempeño optimizado del CNN-GRU, reflejado en una menor pérdida de entropía cruzada, sugiere una calibración más precisa y una posible mejora en la generalización del modelo. Para trabajos futuros, se enfatiza la importancia de refinar hiperparámetros, ampliar conjuntos de datos y aplicar métodos de regularización más avanzados para potenciar la exactitud y la generalización de estas arquitecturas híbridas, abriendo camino a aplicaciones más precisas y eficientes en el campo del reconocimiento de patrones de movimiento humano.

Referencias

- [1] X. Zhou, W. Liang, K. I.-K. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for internet of healthcare things," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6429–6438, 2020.
- [2] W. Gerych, H. Kim, J. DeOliveira, M. Martin, L. Buquicchio, K. Chandrasekaran, A. Alajaji, H. Mansoor, E. Rundensteiner, and E. Agu, "Gan for generating user-specific human activity data from an incomplete training corpus," in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 4705–4714, 2021.
- [3] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698–703, 2019.
- [4] N. Sikder, M. S. Chowdhury, A. S. M. Arif, and A.-A. Nahid, "Human activity recognition using multichannel convolutional neural network," in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 560–565, 2019.
- [5] S. C. Gupta, D. Kumar, and V. Athavale, "A review on human action recognition approaches," in *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 338–344, 2021.
- [6] S. Mekruksavanich and A. Jitpattanakul, "Smartwatch-based human activity recognition using hybrid lstm network," in *2020 IEEE SENSORS*, pp. 1–4, 2020.
- [7] S. Jain, S. Alam, and K. S. Prabhu, "Human activity recognition using smartphone sensor data," in *Social Networking and Computational Intelligence* (R. K. Shukla, J. Agrawal, S. Sharma, N. S. Chaudhari, and K. K. Shukla, eds.), (Singapore), pp. 533–541, Springer Singapore, 2020.
- [8] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Trends in human activity recognition using smartphones," *Journal of Reliable Intelligent Environments*, vol. 7, no. 3, pp. 189–213, 2021.
- [9] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, *Human activity recognition in artificial intelligence framework: a narrative review*, vol. 55. Springer Netherlands, 2022.
- [10] N. Zehra, S. H. Azeem, and M. Farhan, "Human activity recognition through ensemble learning of multiple convolutional neural networks," in *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–5, 2021.
- [11] A. Singh, T. Margaria, and F. Demrozi, "Cnn-based human activity recognition on edge computing devices," in *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, pp. 1–4, 2023.
- [12] A. Mohamed, F. Lejarza, S. Cahail, C. Claudel, and E. Thomaz, "Har-gcnn: Deep graph cnns for human activity recognition from

highly unlabeled mobile sensor data,” in *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (Per-Com Workshops)*, pp. 124–126, 2022.

People and Smart City Innovation, Smart-World/UIC/ATC/SCALCOM/IOP/SCI 2019, no. April, 2013.

- [13] Y. Hao, Z. Shi, and Y. Liu, “Wifi-based spatiotemporal human action perception,” in *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 3581–3585, 2022.
- [14] P. Rojanavas, A. Jitpattanakul, and S. Me-krusavanich, “Comparative analysis of lstm-based deep learning models for har using smartphone sensor,” in *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, pp. 269–272, 2021.
- [15] R. Mutegeki and D. S. Han, “A cnn-lstm approach to human activity recognition,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 362–366, 2020.
- [16] U. Verma, P. Tyagi, and M. Kaur, “Smartphone-based cnn-gru framework for human activity recognition,” in *2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES)*, pp. 1–4, 2021.
- [17] M. A. Khatun, M. A. Yousuf, and M. A. Mo-ni, “Deep cnn-gru based human activity recognition with automatic feature extraction using smartphone and wearable sensors,” in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6, 2023.
- [18] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. .Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” *Proceedings - 2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of*