

Implementación y Evaluación de Arquitecturas de Deep Learning para el Reconocimiento de Actividades Humanas

Tatiana María Gaviria Sáenz
Facultad de Ingeniería, Universidad de Antioquia

Noviembre 19, 2023

Resumen

En este proyecto se exploró el análisis de series temporales obtenidas de sensores de smartphones para clasificar actividades humanas. Se emplearon técnicas de deep learning con arquitecturas como CNN, LSTM, GRU y sus variantes híbridas. Se investigó el impacto de transformar datos al dominio de frecuencia (FFT) y al de potencia (método de Welch) en la precisión de los modelos. Los resultados indicaron una mejora notable en la precisión con la FFT, especialmente en el modelo CNN+GRU. Este hallazgo subraya la relevancia de analizar las series temporales desde perspectivas de tiempo y frecuencia para tareas de clasificación.

1. Introducción

El reconocimiento de actividades humanas (HAR) ha emergido como un campo de estudio prominente adquiriendo una trascendencia particular en las últimas dos décadas debido a su aplicación en sectores críticos como la salud personalizada y la seguridad [1]. Este auge se ha visto catalizado por avances significativos en microelectrónica, tecnologías de sensores y ciencias computacionales, lo cual ha permitido una captura más eficiente y detallada de los movimientos humanos, posibilitando así que técnicas avanzadas de aprendizaje automático, transformen dicha información en inteligible y funcional para las máquinas [2,3].

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 [4]. En este contexto, las redes neuronales convolucionales (CNN) han demostrado ser de utilidad en el HAR, proporcionando un procesamiento eficaz de datos sensoriales de alta dimensionalidad [5]. No obstante, las CNN enfrentan limitaciones en el manejo de dependencias temporales, un aspecto crítico en el análisis de actividades humanas, que son inherentemente secuenciales y dinámicas [6]. Para abordar esta limitación, se ha explorado el desarrollo de modelos híbridos que integran las capacidades de análisis espacial de las CNN con arquitecturas especializadas en el procesamiento de secuencias temporales, como las Memorias de Corto y Largo Plazo (LSTM) y las Unidades Recurrentes con Puertas (GRU) [7,8]. Estos modelos híbridos buscan ofrecer

una comprensión más holística y precisa de las actividades humanas, combinando eficientemente el análisis espacial y temporal.

Este informe detalla la implementación y evaluación de cuatro arquitecturas de Deep Learning : CNN-LSTM, CNN-LSTM Mejorado, GRU y CNN-GRU en el contexto del reconocimiento de actividades humanas. Se presenta un análisis comparativo de estas arquitecturas, a partir de métricas como la precisión (accuracy), sensibilidad (recall), y la F1-Score. Además, se examina el impacto de transformaciones de las señales, como la Transformada de Fourier y de Welch, en el rendimiento de los modelos.

2. Estructura de los Notebooks Entregados

El proyecto se estructura en tres notebooks en los cuales se aborda a partir de la implementación de 4 arquitecturas (CNN-LSTM, CNN-LSTM Mejorado, GRU y CNN-GRU) el problema de clasificación de 6 actividades humanas : caminar (WALKING), subir escaleras (WALKING.UPSTAIRS), bajar escaleras (WALKING.DOWNSTAIRS), estar sentado (SITTING), estar de pie (STANDING) y estar acostado (LAYING) utilizando la base de datos de HAR tomada del repositorio de Machine Learning de la UCI [9].

2.1. 01_Proyecto_DL

2.1.1. Descarga del Conjunto de Datos

- **Importación de bibliotecas:** Se importan bibliotecas esenciales como os, numpy, pandas, random y tensorflow.
- **Inicialización de semillas:** Se establecen semillas para la reproducibilidad del experimento utilizando random.seed(), np.random.seed() y tf.random.set_seed().
- **Definición de rutas y descarga de datos:** Se especifican las rutas para la descarga del conjunto de datos UCI HAR desde un URL. Los datos se descargan y descomprimen utilizando comandos del sistema (wget y unzip).

2.1.2. Exploración de datos

- **Definición de rutas a los datos:** Se establecen las rutas para acceder a los datos de series temporales y las etiquetas tanto para el conjunto de entrenamiento como para el de prueba.
- **Carga y visualización inicial:** Se carga una muestra de los datos de series temporales usando pandas y se visualiza para obtener una comprensión preliminar de su estructura.

2.1.3. Procesamiento de datos

Se define una función para formatear los datos que ingresarán a los modelos y se implementa para preparar los conjuntos de entrenamiento y prueba. Esta función carga y organiza los datos de las series temporales y etiquetas las cuales convierte a un formato categórico (one-hot encoding).

2.1.4. Definición de métricas

Con el fin de sistematizar la evaluación de los modelos, se establecen las siguientes funciones:

- **print_metrics:** Calcula e imprime métricas de evaluación como la precisión, matriz de confusión, precisión por clase, recall y puntuación F1. Esta función utiliza la biblioteca `sklearn.metrics`.
- **plot_accuracy_per_epoch:** Función útil para trazar la precisión de entrenamiento y prueba en cada época con el fin de entender el comportamiento del modelo a lo largo del entrenamiento.

2.1.5. Creación y evaluación de los modelos

En esta sección se sigue la siguiente estructura para las 4 arquitecturas (CNN-LSTM, CNN-LSTM Mejorado, GRU y CNN-GRU) implementadas:

- Definición del modelo.
- Construcción de las capas del modelo.
- Compilación y resumen del modelo.
- Entrenamiento del modelo.
- Predicciones y evaluación.

2.2. 02_Proyecto_DL Iteracion Frecuencia

Para este notebook se conserva la estructura anterior salvo que se realizan modificaciones en la sección del **procesamiento de datos** las cuales consisten en lo siguiente:

- **Transformación al dominio de la frecuencia:** En el primer notebook, los datos se manejaron directamente en el dominio del tiempo, manteniendo la forma original de las series temporales. En este segundo notebook, se aplica la FFT a cada serie temporal para convertirla al dominio de la frecuencia.
- **Visualización y análisis de frecuencia:** Se realiza un análisis exploratorio de la transformación al dominio de frecuencia. Esto incluye visualizar tanto la serie temporal original como su representación en el dominio de la frecuencia con el fin de verificar la transformación de los datos.
- **Preparación de los datos:** Tras aplicar la FFT, se ajusta la estructura de los datos. El número de pasos de tiempo se convierte en pasos de frecuencia y se muestra cómo la FFT altera la dimensión de los datos, donde el número de pasos de frecuencia (`n.frequency_steps`) reemplaza el número de pasos de tiempo.

2.3. 03_Proyecto_DL Iteracion Potencia

Para este tercer notebook se mantiene la estructura implementada en los dos anteriores, en esta iteración se realizan las siguientes modificaciones en la sección de **procesamiento de datos** para su transformación:

- **Transformación al dominio de espectros de frecuencia:** Se implementa la metodología de Welch con el fin de estimar la densidad espectral de potencia (PSD) de los datos para visualizar la manera en que se distribuye la potencia de la señal a lo largo de diferentes frecuencias.
- **Visualización y análisis del espectro de potencia:** De manera similar al segundo notebook, se realiza una visualización del espectro de potencia de las señales.
- **Preparación de los datos:** Con esta transformación cada muestra se describe por un conjunto de 'bins' de frecuencia modificando la segunda dimensión de los datos, pasando de 'pasos de tiempo' a 'bins' de frecuencia resaltando cómo la energía de la señal se distribuye en diferentes frecuencias para cada característica sensorial.

3. Descripción de la Solución

La solución propuesta tiene como finalidad el análisis y la clasificación de actividades humanas a partir de series temporales registradas mediante sensores integrados en un dispositivo móvil. El enfoque de la implementación es la exploración del efecto de transformar los datos al dominio de frecuencia y potencia, y la influencia de dichas transformaciones en el rendimiento y exactitud de los modelos.

Por otro lado, un punto de interés particular, basado en revisiones de literatura, es el análisis del desempeño de modelos híbridos en la tarea de clasificación multiclase de actividades humanas.

3.1. Preprocesamiento de los datos

- **Estructura y carga de los datos:** Los datos fueron leídos desde sus correspondientes archivos de texto, estos fueron previamente separados en conjuntos de prueba y entrenamiento. 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.
- **Etiquetado de Datos:** Se extrajeron las etiquetas correspondientes a cada actividad y se ajustaron a un formato categórico, esto es, se aplicó la técnica de one-hot encoding a las variables de salida, transformando las etiquetas de clase de cada muestra en vectores binarios.
- **Transformaciones :** Para las iteraciones correspondientes a la frecuencia y la potencia, se realizaron las transformaciones FFT y Welch respectivamente, reemplazando la

segunda dimensión del arreglo tridimensional según correspondía (pasos de frecuencia o bins).

-Al realizar la FFT y el método Welch, ocurre una reducción en la segunda dimensión del arreglo debido a que solo se toma el espectro de frecuencias positivas dada la simetría de los espectros de datos con valores de números reales.

3.2. Descripción de las arquitecturas

En las figuras 1-4 se presentan los esquemas correspondientes a las diferentes arquitecturas implementadas.

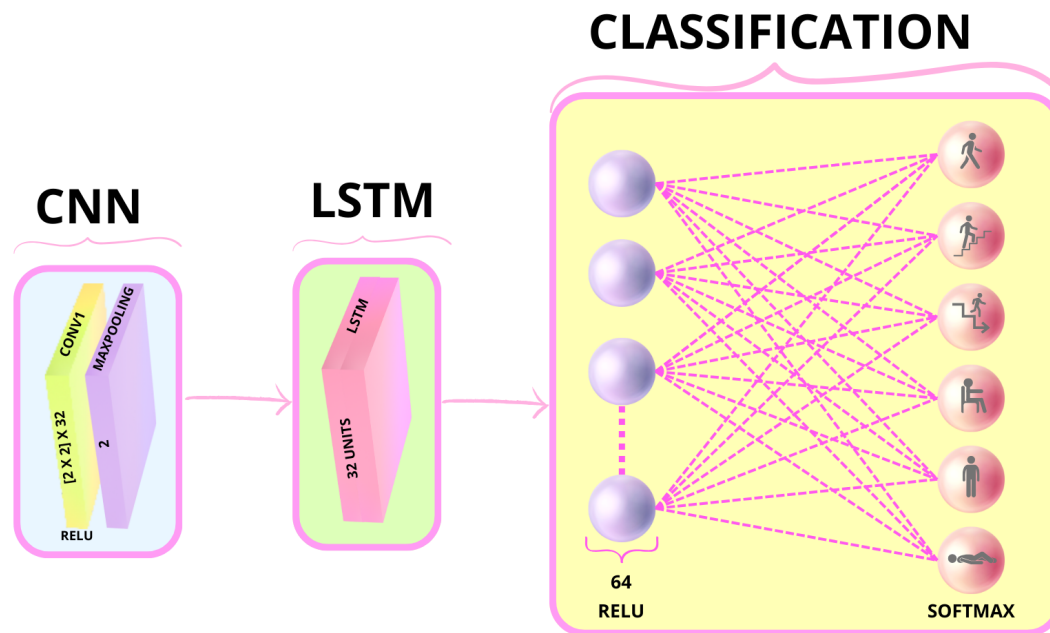


Figura 1: Esquema de la Arquitectura CNN-LSTM

1. CNN+LSTM

- **Estructura:** Combina una capa convolucional (Conv1D) para la extracción de características locales en los datos temporales, seguida de una capa LSTM para analizar dependencias temporales.
- **Capas:**
 - Conv1D con 32 filtros y kernel de tamaño 2, seguida de una capa de MaxPooling para reducir la dimensionalidad de los datos.
 - LSTM con 32 unidades.
 - Dos capas densas, una para clasificación y otra para activación softmax.
- **Compilación:** Usa el optimizador Adam y la pérdida de entropía cruzada categórica.

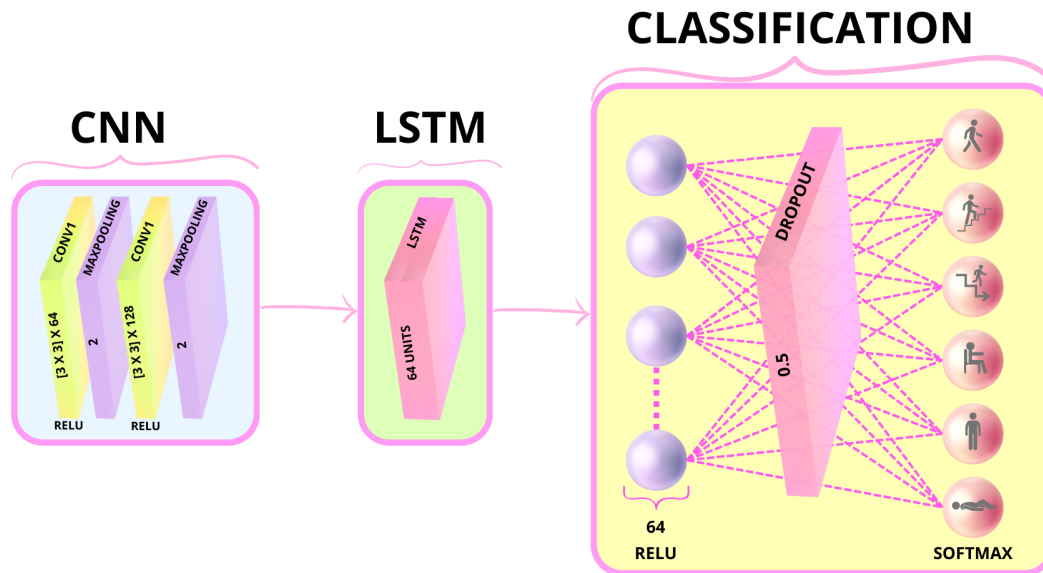


Figura 2: Esquema de la Arquitectura CNN-LSTM Mejorada

2. CNN+LSTM Mejorado

- **Estructura:** Versión modificada del primero, con capas adicionales para una mejor extracción de características y regularización.
- **Capas:**
 - Dos capas Conv1D (64 y 128 filtros) con MaxPooling.
 - LSTM con 64 unidades.
 - Capa densa con 64 unidades, Dropout para regularización, y capa de activación softmax.
- **Compilación:** Usa el optimizador Adam y la pérdida de entropía cruzada categórica.

3. GRU

- **Capas GRU:** Dos capas con 128 y 64 unidades respectivamente. La primera capa devuelve secuencias para mantener la estructura temporal para la segunda capa GRU.
- **Regularización:** Se aplica Dropout después de cada capa GRU para reducir el riesgo de sobreajuste.
- **Capas Densas:** Finaliza con una capa densa y una capa de salida con activación softmax para clasificación.
- **Compilación:** Usa el optimizador Adam y la pérdida de entropía cruzada categórica.

4. CNN+GRU

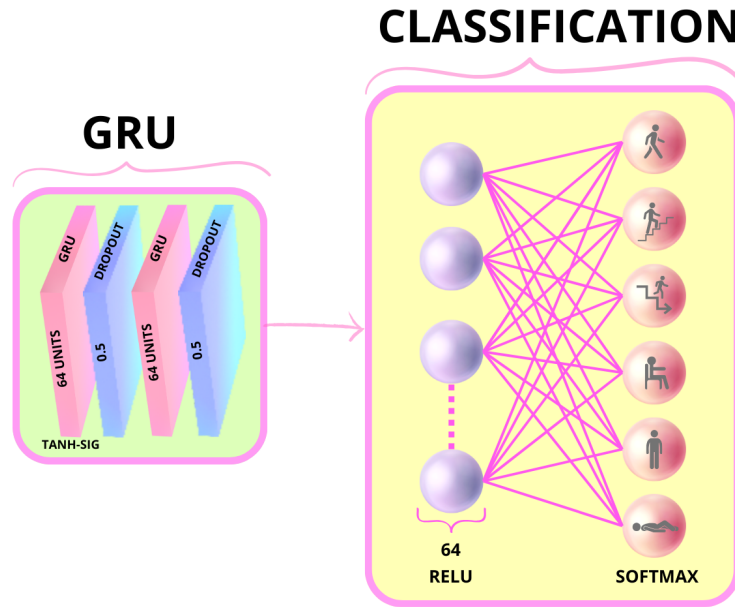


Figura 3: Esquema de la Arquitectura GRU

- **Capas Convolucionales:** Utiliza dos capas con MaxPooling para la extracción de características.
- **Capas GRU:** Dos capas GRU con 64 unidades cada una para capturar la dependencia temporal.
- **Regularización y Capas Densas:** Incorpora Dropout para evitar sobreajuste y termina con capas densas para la clasificación.
- **Compilación:** Usa el optimizador Adam y la pérdida de entropía cruzada categórica.

3.3. Descripción de las Iteraciones

- **Iteración 0:** Inicialmente se implementó una arquitectura básica de CNN+LSTM entrenándola directamente con los datos de series temporales obtenidos de los sensores con el fin de establecer un referente para las comparaciones entre los siguientes modelos.
- **Iteración 1:** En esta iteración se estableció una estructura mejorada para la arquitectura CNN+LSTM así como la configuración de las arquitecturas GRU y CNN+GRU las cuales se entrenaron con los datos en el dominio del tiempo y se evaluó su rendimiento.
- **Iteración 2:** Para esta iteración se realizó la transformada rápida de fourier FFT a los datos con los que se alimentaron los modelos previamente construidos para proceder a evaluarlos y comparar su rendimiento.

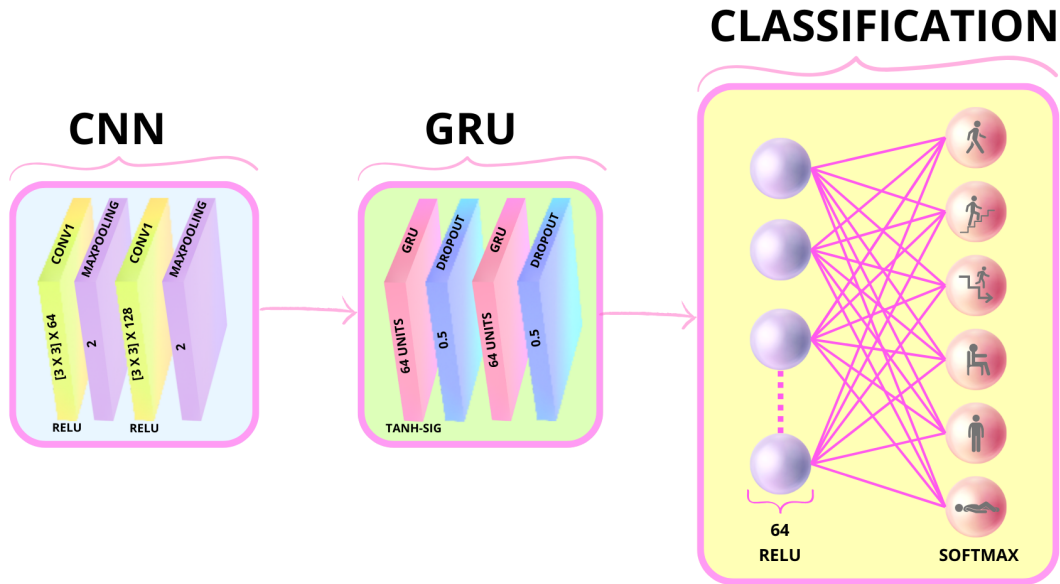


Figura 4: Esquema de la Arquitectura CNN-GRU

- **Iteración 3** Esta iteración final consistió en la transformación de los datos con la metodología Welch para alimentar los modelos construidos inicialmente para obtener sus métricas y determinar su rendimiento con respecto a las iteraciones anteriores.

4. Descripción de los Resultados

4.1. Primera Iteración

Los modelos CNN+LSTM, CNN+LSTM Mejorado, GRU y CNN+GRU mostraron una alta eficacia, con un rendimiento en términos de precisión (Accuracy) de entre el 86 % y el 91 %. Estos modelos combinan características de aprendizaje temporal y espacial, lo que parece ser efectivo para este tipo de datos, sin embargo, el modelo CNN+LSTM Mejorado obtuvo la mayor precisión (91 %), reflejando la importancia de una arquitectura más compleja. Por otro lado, el modelo GRU y CNN+GRU demostraron un rendimiento similar, lo que sugiere que la combinación de CNN y GRU es tan efectiva como GRU solo, pero con una estructura de red más compleja, finalmente se evidencia que la Precisión por clase y la F1-Score varían significativamente entre las clases, lo que indica diferencias en la facilidad de clasificación de diferentes tipos de actividades para los modelos.

4.2. Iteración Frecuencia

La mayoría de los modelos presentaron una mejora en la precisión con la transformación FFT, destacando la eficacia de analizar los datos en el dominio de la frecuencia para este tipo de tareas, entre los modelos destaca el CNN+GRU con una precisión del 94 %, este modelo se benefició considerablemente de la transformación FFT, lo que indica una mejor capacidad de este modelo para capturar características frecuenciales. A diferencia de los mo-

delos CNN, el modelo GRU disminuyó su rendimiento con la FFT, lo que puede deberse a su enfoque en dependencias temporales más que en características frecuenciales, lo anterior indica que la sección CNN le aporta al GRU la capacidad de captar de manera más efectiva las características frecuenciales, un aspecto que el GRU por sí solo no maneja tan eficientemente, lo cual resalta la sinergia entre ambas arquitecturas para esta aplicación.

4.3. Iteración Potencia

Se observó una disminución significativa en el rendimiento de los modelos con el método de Welch, especialmente con el modelo CNN+LSTM, que solo alcanzó un 35 % de precisión. Esta disminución drástica podría deberse a la pérdida de información temporal crítica durante la transformación de Welch, lo que afecta negativamente la capacidad de los modelos para clasificar las actividades. Aunque también disminuyó, el modelo CNN+GRU Mejorado fue más resistente a la transformación de Welch, con una precisión del 64 %.

5. Conclusiones

La implementación de la Transformada Rápida de Fourier (FFT) reveló una mejora significativa en la eficacia de los modelos lo cual da cuenta de la utilidad de examinar las series temporales no solo en su representación temporal original, sino también en el espectro de frecuencias para una comprensión más holística de las características subyacentes en este tipo de datos.

Los resultados indican que las arquitecturas híbridas, particularmente el modelo CNN+GRU, mostraron un desempeño superior en la tarea de clasificación. La combinación de CNN y GRU permitió aprovechar las fortalezas de ambas técnicas: la capacidad de las CNN para extraer características espaciales y de las GRU para modelar dependencias temporales lo cual subraya la importancia de seleccionar y combinar arquitecturas de redes neuronales de manera estratégica, especialmente en tareas que requieren un análisis integral.

El modelo GRU experimentó una disminución en su rendimiento tras la aplicación de la FFT, lo que subraya la relevancia de la selección adecuada de la arquitectura en función de las características de los datos. Por otro lado, la tercera iteración, que implicó el uso del método de Welch, disminuyó el rendimiento de los modelos lo que sugiere que esta técnica de transformación puede no ser la más adecuada para este tipo de datos o tarea.

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] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698–703, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050919310166>

-
- [3] 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)*, 2019, pp. 560–565.
 - [4] 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*. Springer Netherlands, 2022, vol. 55, no. 6. [Online]. Available: <https://doi.org/10.1007/s10462-021-10116-x>
 - [5] 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)*, 2023, pp. 1–4.
 - [6] 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 (PerCom Workshops)*, 2022, pp. 124–126.
 - [7] P. Rojanavas, A. Jitpattanakul, and S. Mekruksavanich, "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*, 2021, pp. 269–272.
 - [8] 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)*, 2021, pp. 1–4.
 - [9] 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 People and Smart City Innovation, SmartWorld/UIC/ATC/SCALCOM/IOP/SCI 2019*, no. April, 2013.