

Implementación de una red neuronal profunda en tres etapas paralelas para el reconocimiento de actividades humanas e imágenes

Yafte A. Flores-Morales¹, Juan Villegas-Cortez²,
Graciela Román-Alonso¹, Arturo Zúñiga-López³,
César Benavides-Álvarez¹, Salomón Cordero-Sánchez⁴

¹ Universidad Autónoma Metropolitana,
Unidad Iztapalapa,
Departamento de Ingeniería Eléctrica,
México

² Universidad Autónoma Metropolitana,
Unidad Azcapotzalco,
Departamento de Sistemas,
México

³ Universidad Autónoma Metropolitana,
Unidad Azcapotzalco,
Departamento de Electrónica,
México

⁴ Universidad Autónoma Metropolitana,
Unidad Iztapalapa,
Departamento de Química,
México

{yafteaaron, scs, grac}@xanum.uam.mx,
cesarbenavides32@gmail.com, {juanvc, azl}@azc.uam.mx

Resumen. La importancia de la aplicación del reconocimiento de patrones para el reconocimiento de imágenes, se hace notar cuando el número y la complejidad de las imágenes aumenta. Las redes neuronales artificiales han demostrado su efectividad para atacar problemas de patrones complejos en los últimos quince años aproximadamente y han evolucionado hacia arquitecturas híbridas y profundas. En este artículo presentamos la implementación de una red neuronal profunda en tres capas paralelas, con una alimentación paralela de los patrones a tres niveles de granularidad: fina, mediana y gruesa; buscando conformar un análisis del patrón característico robusto aplicado a imágenes. La red profunda de la cual partimos la hemos modificado y orientado, de patrones de reconocimiento de actividades humanas (HAR), hacia atacar la complejidad de las imágenes, con resultados prometedores del 99 % de reconocimiento en las pruebas realizadas.

Palabras clave: RNC, aprendizaje profundo, clasificación, reconocimiento de patrones, HAR, reconocimiento de imágenes.

Implementation of Three Parallel Layer Deep Neural Network for Human Activity and Image Recognition

Abstract. The recognition of patterns for the recognition of faces becomes an issue of extreme importance when the number and complexity of images increases. For the last fifteen years, artificial neural networks have shown its effectiveness to address complex pattern problems at the same time that they have evolved to hybrid and deep architectures. This paper presents the implementation of a deep neural network developed in three parallel layers, and with parallel feeding of patterns with three levels of granularity: fine, medium and thick. This results in conforming an analysis of the characteristic robust pattern applied in the images. The initial deep network was modified and oriented to address the complexity of the images by integrating human activity recognition (HAR). The results are promising achieving 99 % of recognition of the performed tests.

Keywords: Deep neural networks, deep learning, pattern recognition, HAR, image recognition.

1. Introducción

El área de reconocimiento de patrones en la inteligencia artificial (IA) está tomando mucha importancia en los últimos años debido al tratamiento automático de grandes cantidades de datos complejos. El reconocimiento de patrones (RP) extrae de un conjunto de datos la información que permite establecer propiedades, vínculos y relaciones entre los datos, estas relaciones como patrones permiten una descripción estructural o cuantitativa de un objeto o de alguna otra entidad de interés en los datos [10, 14].

La aplicación del RP a diferentes estructuras de datos, como señales o imágenes, ha dado lugar al desarrollo de una gran variedad de modelos, técnicas y algoritmos en las áreas de aprendizaje automático (del inglés Machine Learning o ML) y aprendizaje profundo (del inglés Deep Learning o DL), que permiten extraer patrones de las estructuras de datos con cierta eficiencia y eficacia. Algunas aplicaciones del reconocimiento de patrones son, por ejemplo, algoritmos sofisticados para reconocimiento de voz en los dispositivos móviles, o los videojuegos basados en realidad aumentada donde pueden detectar los movimientos del jugador por medio de una cámara. Sin embargo, reconocer movimientos de una persona no es tarea fácil. La tarea de reconocer y clasificar acciones de un grupo de individuos en el área de la IA, se le llama Reconocimiento de Actividad Humana (del inglés Human Action Recognition o HAR). En este artículo se propone implementar una red neuronal profunda con diferentes niveles de extracción de características, basado en la red neuronal grueso-a-fino⁵ en [5] para la tarea HAR y el reconocimiento de patrones en imágenes.

⁵ Coarse-fine N. del T.

La contribución de este trabajo es dos puntos: *(i)*, se busca evaluar la red neuronal basada en [5] con dos diferentes bases de datos WISDM v2.0 y HAPT, mismas que no se habían considerado previamente; y *(ii)*, se prueba el rendimiento de la arquitectura de la red neuronal en el reconocimiento de imágenes.

El documento está organizado de la siguiente manera: en la Sección 2 presentamos el estado del arte del problema de la caracterización de HAR y el reconocimiento de imágenes en grandes cantidades desde la perspectiva del reconocimiento de patrones y de visión por computadora; en la Sección 3 presentamos la implementación y adecuación de la metodología de las bases de datos escogidas para ejemplificar los alcances. La Sección 4 comparte los resultados de los experimentos y una discusión de los mismos y por último, en la Sección 5, las conclusiones y perspectivas de trabajo futuro son compartidos.

2. Estado del arte

En la literatura existen dos enfoques principales para tratar la tarea HAR, el primer enfoque se basa en reconocer diferentes actividades por medio de cámaras de video, Jalal [16] obtiene buenos resultados con este enfoque, utilizando dispositivos de video para reconocer actividades realizadas por un individuo. Sin embargo, el costo de procesar video reduce la eficiencia de los algoritmos por el alto consumo de memoria y cómputo, debido a esto muchos autores utilizan diferentes técnicas capaces de procesar el gran volumen de datos producidos por sensores de video, Martinez [22] utiliza visión por computadora y otros enfoques para el reconocimiento de actividades obteniendo resultados arriba del 90 %. El segundo enfoque se basa en la utilización de sensores como giroscopio y acelerómetro colocados en diferentes partes del individuo, trabajando con señales obtenidas de estos sensores. En esta propuesta nos centraremos en este segundo enfoque.

El área de investigación de HAR ha recibido atención debido a la tendencia creciente de sus aplicaciones en diferentes áreas, la reducción en el precio de los sensores integrados en los dispositivos portátiles y la masificación de su producción [18]. Diferentes artículos realizan estudios sobre características importantes de los sensores; tipo, costo, capacidad, cantidad. sin embargo, en la mayoría de estos se utilizan teléfonos inteligentes debido a la relación calidad-precio y la accesibilidad [27].

El HAR es un problema tratado por diferentes técnicas de la IA, el ML y el DL. Existen métodos basados en ML como máquinas de soporte vectorial o árboles de decisión, estos métodos utilizan características estadísticas como media, mínimo, máximo, desviación estándar, asimetría, curtosis, ángulos, entropía, etc. [2, 29].

Un enfoque diferente para la tarea de extracción de características se basa en DL donde se centra en la utilización de Redes Neuronales Convolucionales (del inglés Convolution Neural Network o CNN), un tipo particular de las redes neuronales artificiales caracterizada por el uso de la convolución entre señales o sucesiones de valores discretos.

Existen diferentes tipos de redes neuronales artificiales, la CNN utiliza como operación principal la convolución para la extracción de características de los datos, estas características representan rasgos particulares de los datos.

La importancia de las CNN radica en el tratamiento de dos puntos principales: (i) La extracción de características más robustas a partir de la implementación de la operación de convolución, y (ii) la posibilidad de trabajar con una mayor cantidad de patrones complejos en longitud o cantidad de atributos, así como el manejo de valores en punto flotante.

Existen diversos conjuntos de datos para el HAR que ayudan a medir el rendimiento de los modelos. En este artículo se ocuparon cuatro conjuntos de datos: UCI HAR [3], WISDM v1.1 [17], WISDM v2 [17, 32] y HAPT [25].

2.1. Métodos basados en ML

UCI HAR, Anguita [4] mediante máquinas de soporte vectorial logra atacar el problema HAR obteniendo buenos resultados, Xiangbin Zhu [35] utiliza cadenas de markov. UCI HAPT, Taufeeq [28] obtiene excelentes resultados utilizando árboles de decisión. Zheng [34] utiliza máquinas de soporte vectorial combinado un modelo de agrupación dispersa de dos capas. WISDM v1.1 Kishor Walse [30] utiliza árboles de decisión. Cagatay Catal [6] realiza una combinación de diferentes técnicas, árboles de decisión, regresión logística y perceptrón multicapa. WISDM v2, Majid Ali Khan Quaid [23] por medio de un algoritmo genético logra alcanzar buenos resultados.

2.2. Métodos basados en DL

HAR, Avilés Cruz *et al.* [5] obtienen el 100% utilizando una CNN con 3 diferentes niveles de alimentación paralela. Cho *et al.* [7] utilizaron el paradigma de divide y vencerás y una CNN para identificar acciones realizadas por humanos. Yong Zhang *et al.* [33] construyeron una CNN capaz de adaptarse a diferentes longitudes de datos. En HAPT, Yong Zhang *et al.* [33], el único trabajo encontrado con enfoque DL, construyeron una CNN capaz de adaptarse a diferentes longitudes de datos.

WISDM v1.1 Avilés Cruz *et al.* [5] obtienen el 100% utilizando una CNN con 3 diferentes niveles de alimentación paralela. Xinxin Han [13] utiliza una CNN con tres niveles de alimentación. Daniele Ravi [24] crea una CNN tradicional probando diferentes capas de convolución. WISDM v2 Girmaw Abebe [1] obtiene un buen resultado combinando una CNN con memorias de largo plazo (LSTM). Ignatov [15] utiliza una CNN tradicional con métodos de preprocesamiento. MNIST, Li Wan [31] por medio de una CNN y un enfoque Dropconnect obtiene un buen resultado. Dan Claudiu Cireşan [8] construye una red neuronal multicolumna. Ming Liang [21] por medio de una red neuronal recurrente logra mejores resultados que una CNN clásica.

En las Tablas 1 a la 5 se presentan los trabajos mencionados de las diferentes técnicas de ML y DL.

Tabla 1. Comparación de exactitud de UCI HAR [3].

Nombre	Año	Exactitud %	Método
Avilés Cruz [5]	2019	100.0	CNN
Yong Zhang [33]	2018	98.4	U-CNN
San-Segundo [26]	2016	98.0	Marcov+ Estadística
Andrey Ignatov [15]	2018	97.63	CNN

Tabla 2. Comparación de exactitud de HAPT [25].

Nombre	Año	Exactitud %	Método
Taufeeq [28]	2016	100.0	Random Forrest
Zheng [34]	2018	96.26	TASG+SVM
Zheng [34]	2018	95.83	TASG+RNN
Yong Zhang [33]	2018	93.1	U-CNN

Tabla 3. Comparación de exactitud de WISDM 1.1v [17].

Nombre	Año	Exactitud %	Método
Avilés Cruz [5]	2019	100	CNN
Xinxin Han [13]	2020	98.83	CNN+FusionTrial
Daniel Ravi [24]	2017	98.2	CNN
kishor walse [30]	2016	98.09	Random forrest
Yong Zhang [33]	2018	97.0	U-CNN

Tabla 4. Comparación de exactitud de WISDM 2.0v [17] [32].

Nombre	Año	Exactitud %	Método
Girmaw Abebe [1]	2017	97.9	CNN+LSTM+Handcrafted
Majid Ali Khan Quaid [23]	2019	94.02	Genético
andrey Ignatov [15]	2018	93.32	CNN + tamaño 200
Daniele Ravi [24]	2017	92.7	CNN

3. Propuesta

La propuesta es implementar una CNN con diferentes niveles de extracción paralela de características, estos niveles de extracción están divididos en grueso, medio y fino. En la Figura 1 se muestra la arquitectura de la red neuronal [5], la idea principal está basada en extraer diferentes niveles de características, donde el nivel fino extrae características más sutiles de los datos, el nivel medio extrae

Tabla 5. Comparación de exactitud de MNIST [19].

Nombre	Año	Exactitud %	Método
Li Wan [31]	2013	99.79	CNN Dropconnect
Dan Claudiu Cireşan [8]	2012	99.77	CNN Multi-Column
Benjamin Graham [12]	2014	99.68	CNN Fraccional
Dan Claudiu Cireşan [9]	2010	99.65	Big Deep CNN

características más robustas y el nivel grueso extrae características más burdas como líneas, círculos, colores, etc.

La arquitectura presentada en la Figura 1 consta de tres niveles:

1. **Fino:** Está compuesto por cuatro capas de convolución, cada capa de convolución es seguida por una capa de agrupación máxima, los mapas de características son reducidos en cada capa. Las primeras dos capas de convolución cuentan con 18 filtros y las últimas dos capas cuentan con 36 filtros, el tamaño del filtro es de (1×2) para todas las capas de convolución. La ventana de agrupación máxima es de (1×2) y el paso es de 2.
2. **Medio:** Está compuesto por dos capas de convolución, cada capa de convolución es seguida de una capa de agrupación máxima, los mapas de características son reducidos en cada capa. La primera capa de convolución cuenta con 18 filtros y la última cuenta con 36 filtros, el tamaño del filtro es de (1×2) para todas las capas de convolución. La ventana de agrupación máxima es de (1×4) y el paso es de 4.
3. **Grueso:** Está compuesto por una capa de convolución, la capa de convolución es seguida por una capa de agrupación máxima. La capa de convolución cuenta con 36 filtros de tamaño (1×2) . La ventana de agrupación máxima es de (1×16) y el paso es de 16.

La salida de los tres niveles se agrupan y se aplanan, formando un vector que pasa a una etapa de clasificación por medio de una red neuronal totalmente conectada. Finalmente la salida de la capa totalmente conectada se pasa a una capa softmax que calcula la distribución de probabilidad de las clases [11]. En la Tabla 6 se muestran los parámetros utilizados por la red neuronal propuesta.

4. Implementación

La implementación se realizó en una computadora tipo Workstation con las siguientes características: sistema operativo Linux, 8 GB RAM, procesador Intel Core i7, GPU NVIDIA GTX-1050. Esta sección describe y muestra las pruebas realizadas de la red neuronal profunda gruesa-fina usando diferentes conjuntos de datos, como se detallan a continuación en su generalidad.

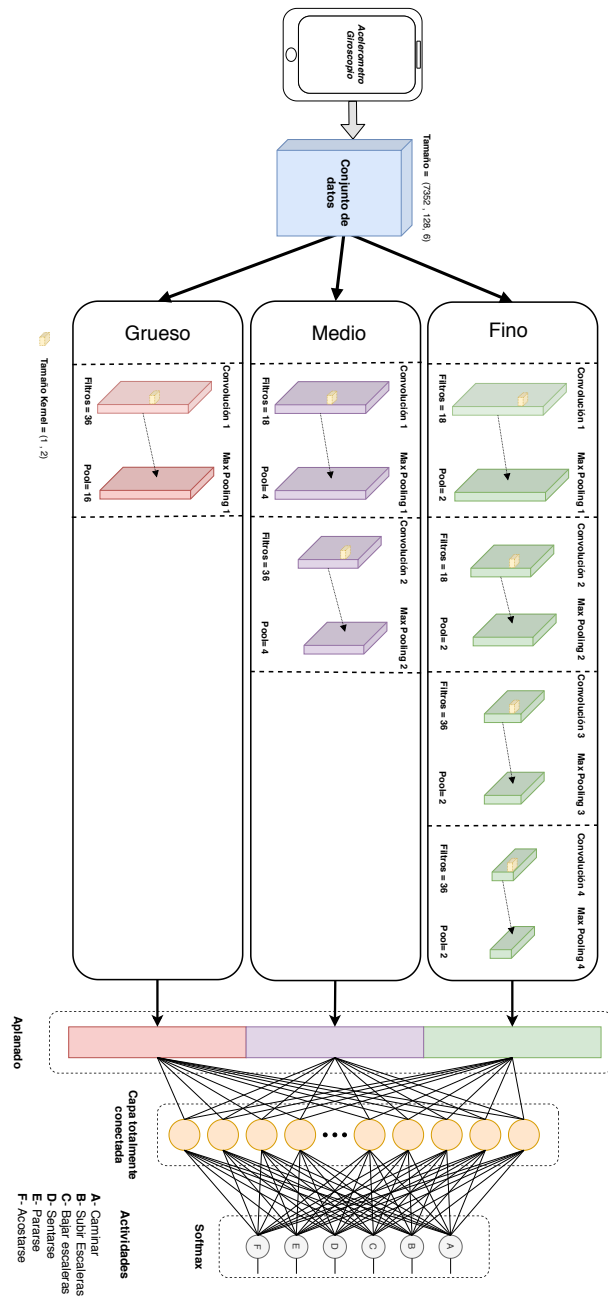


Fig. 1. Arquitectura de la red neuronal propuesta aplicada al reconocimiento de actividad humana [5].

Tabla 6. Parámetros establecidos para la red neuronal profunda gruesa-fina.

Parámetros	Valor	Tipo
Número de épocas	500	Entero
Tamaño de lote	300	Entero
Dropout	0.8	Flotante
Función activación	RELU	Función
Función perdida	Entropía cruzada	Función
Numero de filtros	18,36	Entero
Dimensión filtro (fila)	1	Entero
Dimensión filtro (columna)	2	Entero
Dimensión maxpooling (fila)	1	Entero
Dimensión maxpooling (columna)	2,4,16	Entero
Paso de maxpooling	2,4,16	Entero
Optimizador	Adadelata	Función

4.1. Conjuntos de datos

En los experimentos se utilizaron dos tipos de conjunto de datos: 1. Reconocimiento de actividad humana; UCI HAR [3], WISDM v1.1 [17], WISDM v2 [17] [32] y HAPT [25]. 2. Reconocimiento de imágenes; MNIST [19].

UCI HAR es un conjunto de señales preprocesadas de seis distintas actividades humanas: caminar, subir escaleras, bajar escaleras, sentarse, pararse, acostarse. Estas señales fueron tomadas con un teléfono inteligente colocado en la cintura de treinta voluntarios, se tomaron muestras del giroscopio y acelerómetro integrados en el teléfono con una velocidad de 50 Hz. Las señales del giroscopio y del acelerómetro están compuestas por tres ejes, (X, Y, Z). En la Tabla 7 se muestra la cantidad de instancias por clase.

Tabla 7. Número de instancias por clase del conjunto de datos HAR.

Actividad	Entrenamiento	Porcentaje	Prueba	Porcentaje
Bajar escaleras	986	13.4 %	420	14.25 %
Subir escaleras	1,073	14.6 %	471	15.98 %
Caminar	1,226	16.7 %	496	16.83 %
Sentarse	1,286	17.5 %	491	16.66 %
Pararse	1,374	18.7 %	532	18.05 %
Acostarse	1,407	19.1 %	537	18.22 %
Total	7,352	100 %	2,947	100 %

Human Activities and Postural Transitions (HAPT) es una actualización de HAR, tiene doce clases: caminar, subir escaleras, bajar escaleras, sentarse, pararse, acostarse, pararse a sentarse, sentarse a pararse, sentarse a acostarse, acostarse a sentarse, pararse a acostarse, acostarse a pararse. HAPT proporciona un conjunto de datos no procesados, se realizó un preprocesamiento aplicando el método de ventana deslizante con empalme de 50 % y tamaño de ventana de 128

muestras equivalente a 6.4 segundos, se aplicó un filtro Butterworth de tercer orden con frecuencia de corte de 10 *Hz*. En la Tabla 8 se muestra la cantidad de instancias por clase del conjunto de entrenamiento y del conjunto prueba.

Tabla 8. Número de instancias por clase del conjunto de datos HAPT procesada y particionada.

Actividad	Procesada	Entrenamiento	Prueba
Sentarse-Pararse	123	93	30
Pararse-Sentarse	155	108	47
Acostarse-Sentarse	169	111	59
Acostarse-Pararse	170	118	51
Sentarse-Acostarse	195	142	53
Pararse-Acostarse	223	157	66
Bajar escaleras	1,691	1,202	489
Subir escaleras	1,817	1,227	590
Caminar	1,905	1,314	591
Sentarse	1,983	1,421	562
Acostarse	2,144	1,503	641
Pararse	2,167	1,523	644
Total	12,742	8,919	3,823

Wireless Sensor Data Mining (WISDM) v1.1, contiene 1,098,204 señales no preprocesadas de seis diferentes actividades humanas: caminar, subir escaleras, bajar escaleras, sentarse, pararse, trotar. Estas señales fueron tomadas con un teléfono inteligente que midió la aceleración en tres ejes (X,Y,Z) con una frecuencia de 20 *Hz*. Para el preprocesamiento de las señales se utilizó el método de ventana deslizante con empalme de 50% y un tamaño de ventana de 128 muestras equivalente a 6.4s, se aplicó un filtro de Butterworth de tercer orden con frecuencia de corte de 10 *Hz*. La Tabla 9 muestra la cantidad de instancias.

Tabla 9. Número de instancias por clase del conjunto de datos WISDM v1.1.

Actividad	Sin procesamiento	Procesada	Entrenamiento	Prueba
Pararse	48,395	757	544	213
Sentarse	59,939	936	650	286
Bajar escaleras	100,427	1,565	1,089	476
Subir escaleras	122,869	1,927	1,327	600
Trotar	342,176	5,346	3,737	1,609
Caminar	424,398	6,627	4,663	1,964
Total	1,098,204	17,158	12,010	5,148

WISDM v2.0 es una actualización del conjunto de datos WISDM v1.1, es un conjunto de 2,980,765 señales no preprocesadas de seis diferentes actividades humanas: caminar, escaleras, sentarse, pararse, trotar, acostarse. Estas señales fueron tomadas con un teléfono inteligente el cual midió la aceleración en tres ejes (X,Y,Z) con una frecuencia de 20 Hz. Se utilizó el mismo preprocesamiento de las señales aplicado en WISDM v1.1. La Tabla 10 muestra la distribución de estas instancias por clase. La Tabla 11 compara las clases contenidas en cada conjunto de datos.

Tabla 10. Número de instancias por clase del conjunto de datos WISDM v2.0 procesada y particionada.

Actividad	Sin procesar	Procesada	Entrenamiento	Prueba
Escaleras	57,425	896	644	252
Acostarse	275,967	4,315	2,972	1,343
Pararse	288,873	4,512	3,141	1,371
Trotar	438,871	6,859	4,833	2,026
Sentarse	663,706	10,373	7,336	3,037
Caminar	1,255,92	19,618	13,675	5,943
Total	2,980,765	46,573	32,601	13,972

Tabla 11. Clases contenidas en los conjuntos de datos.

Dataset	HAPT	UCI	HAR	WISDM v1.1	WISDM v2.0
Pararse-Sentarse	*				
Sentarse-Pararse	*				
Sentarse-Acostarse	*				
Acostarse-Sentarse	*				
Pararse-Acostarse	*				
Acostarse-Pararse	*				
Subir escaleras	*	*	*		
Bajar escaleras	*	*	*		
Acostarse	*	*			*
Caminar	*	*	*		*
Sentarse	*	*	*		*
Pararse	*	*	*		*
Trotar			*		*
Escaleras					*
Total clase	12	6	6		6

Modified National Institute of Standards and Technology (MNIST) [20], es una base de datos de números escritos a mano, consta de 10 clases que representan los números del 0 al 9. Los dígitos están representadas por medio de imágenes en escala de grises con un tamaño de 28×28 píxeles. En la Tabla 12 se muestra la cantidad de instancias por clase.

Tabla 12. Número de instancias por clase de la base de datos MNIST.

Dígito	Entrenamiento	Prueba
Cero	5,923	980
Uno	6,742	1,135
Dos	5,958	1,032
Tres	6,131	1,010
Cuatro	5,842	982
Cinco	5,421	892
Seis	5,918	958
Siete	6,265	1,028
Ocho	5,851	974
Nueve	5,949	1,009
Total	60,000	10,000

4.2. Análisis de resultados

El análisis de resultados se llevó a cabo por medio de dos etapas; entrenamiento y prueba. La primera consiste en entrenar la red neuronal propuesta con el conjunto de entrenamiento. Por otro lado, la segunda consiste en evaluar la red neuronal por medio del conjunto de prueba, el cual es ajeno al conjunto de entrenamiento. Para medir el rendimiento de la red neuronal gruesa-fina se utilizaron métricas de exactitud.

Resultados del conjunto de datos UCI HAR: En la Figura 2a se muestra la evaluación de los conjuntos de entrenamiento y prueba. Se alcanzó el 100 % de recuperación y el parámetro de pérdida alcanzó el valor 0 % al evaluar el conjunto de prueba. El tiempo de entrenamiento de la red neuronal con el conjunto de entrenamiento fue de 4.83 minutos.

Resultados del conjunto de datos HAPT: HAPT resulta un conjunto de datos difícil, ya que no cuenta con un equilibrio en el número de instancias por clase (como se ve en la Tabla 8), sin embargo, se alcanzó una exactitud del 95.27 % con una pérdida de 4.73 % al evaluar el conjunto de prueba. En la Figura 2b se muestra la evaluación de los conjuntos de entrenamiento y prueba. El tiempo de entrenamiento de la red neuronal con el conjunto de entrenamiento fue de 5.30 minutos.

Resultados del conjunto de datos WISDM v1.1: Se alcanzó una recuperación de 98.12 % y valor de pérdida de 1.88 % al evaluar el conjunto de prueba.

En la Figura 2c se muestra la evaluación de los conjuntos de entrenamiento y prueba. El tiempo de entrenamiento de la red neuronal con el conjunto de entrenamiento fue de 7.06 minutos.

Resultados del conjunto de datos WISDM v2.0: WISDM v2.0 es una versión más grande que la v1.1, esto aumenta la dificultad. Los resultados obtenidos con el conjunto de datos WISDM v2.0 alcanzaron un valor de recuperación de 94.75 % y un valor de pérdida de 5.21 % al evaluar el conjunto de prueba. En la Figura 2d se muestra la evaluación de los conjuntos de entrenamiento y prueba. El tiempo de entrenamiento de la red neuronal con el conjunto de entrenamiento fue de 14.85 minutos.

Resultados del conjunto de datos MNIST: MNIST es una base de datos muy popular para la evaluación de modelos. Los resultados obtenidos alcanzaron un valor de exactitud de 99.47 % y un porcentaje de error de 0.53 % al evaluar el conjunto de prueba. La Figura 2e muestra la evaluación del conjunto de entrenamiento y del conjunto de prueba. El tiempo de entrenamiento de la red neuronal con el conjunto de entrenamiento y con 30 etapas fue de 30.75 minutos.

La Tabla 13 muestra una comparación de los resultados obtenidos con otros resultados en el estado del arte.

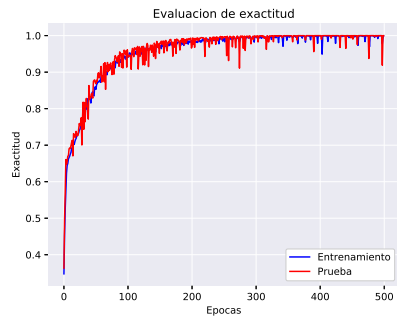
Tabla 13. Comparación de exactitud de diferentes técnicas del estado del arte

Base Datos	Propuesta	Avilés [5]	Zhang [33]	Ravi [24]	Abebe [1]	Taufeeq [28]	Li Wan [31]
UCI HAR	100	100	98.4	-	-	-	-
HAPT	95.27	-	93.1	-	-	100	-
WISD v1.1	98.12	100	97.0	98.6	-	-	-
WISD v2.0	94.75	-	-	92.7	97.9	-	-
MNIST	99.47	-	-	-	-	-	99.79

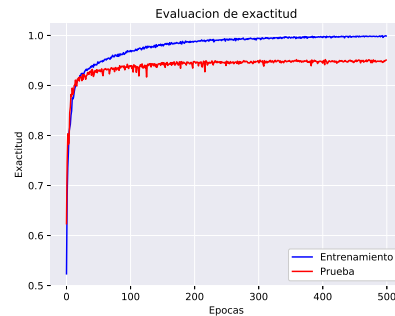
5. Conclusiones

En este artículo se presentó una nueva implementación de la propuesta de una red neuronal profunda previamente reportada en la literatura para aplicarse sólo para reconocimiento y clasificación de patrones HAR, tal que la estrategia se basa en la extracción paralela en tres etapas: fina, media y gruesa. Sin embargo, es necesario una mejora en la optimización de los parámetros, ya que como hemos presentado en la implementación para imágenes, como meta más ambiciosa y de mayor alcance para aplicaciones en visión por computadora, es necesario poder hacerle modificaciones que mejoren en rapidez y capacidad de extracción de características propias para imágenes como entidades numéricas más complejas. La implementación aquí presentada está ubicada en el nivel 25 del *Ranking*

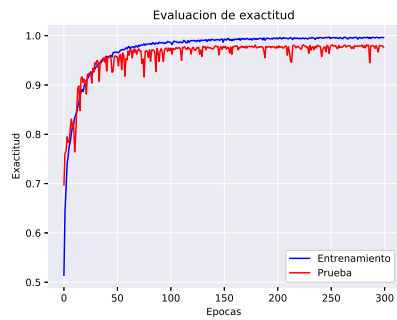
Implementación de una red neuronal profunda en tres etapas paralelas para el reconocimiento...



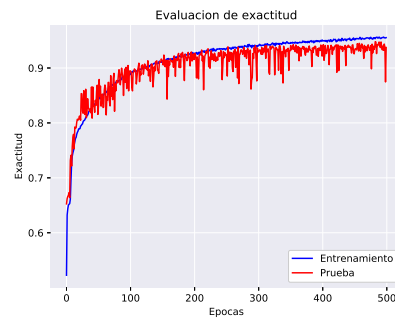
(a) UCI HAR: Exactitud de 100 %



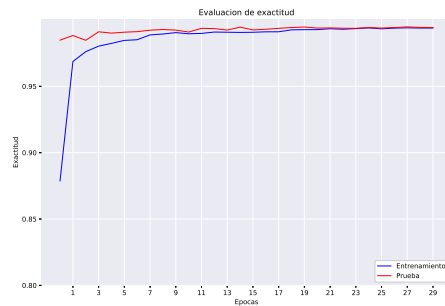
(b) HAPT: Exactitud de 95.27 %



(c) WISDM v1.1: Exactitud de 98.12 %



(d) WISDM v2.0: Exactitud de 94.75 %



(e) MNIST: Exactitud de 99.47 %

Fig. 2. Evaluación de la red neuronal propuesta con las bases de datos seleccionadas.

*classification datasets results*⁶, esta página se tomó como referencia para la

⁶ Ranking classification datasets results: https://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html

comparación de resultados de diferentes técnicas y enfoques del conjunto de datos MNIST, ya que para aplicaciones prácticas es factible su uso, pero consideramos que es posible su mejora si realizamos, como trabajo futuro, modificaciones a la capa de extracciones paralelas, buscando robustecer a los patrones orientados a problemas de imágenes más complejas.

Si bien la base de datos MNIST es una buena aproximación para validar posibles aplicaciones reales, existen problemas de imágenes más complejas como lo son el rostro humano y los escenarios naturales, donde en un futuro trabajo podremos mostrar los avances en tales campos.

Referencias

1. Abebe, G., Cavallaro, A.: Inertial-vision: Cross-domain knowledge transfer for wearable sensors. pp. 1392–1400 (10 2017)
2. Anguita, D., Ghio, A.: A public domain dataset for human activity recognition using smartphones. 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (April 2013)
3. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.: A public domain dataset for human activity recognition using smartphones. pp. 24–26. Computational Intelligence and Machine Learning (01 2013)
4. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.: A public domain dataset for human activity recognition using smartphones. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (2013)
5. Avilés-Cruz, C., Ramírez, A., Zúñiga López, A., Villegas Cortez, J.: Coarse-fine convolutional deep-learning strategy for human activity recognition. Sensors 2019 (03 2019)
6. Catal, C., Tufekci, S., Pirmitt, E., Kocabag, G.: On the use of ensemble of classifiers for accelerometer-based activity recognition. Applied Soft Computing 46 (01 2015)
7. Cho, H., Yoon, S.: Divide and conquer-based 1D CNN human activity recognition using test data sharpening. Sensors 1055 (06 2017)
8. Cireşan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (02 2012)
9. Cireşan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Deep, big, simple neural nets for handwritten digit recognition. Neural Computation 22(12), 3207–3220 (2010), pMID: 20858131
10. Dutt, V., Chaudhry, V., Khan, I.: Pattern recognition: an overview. American Journal of Intelligent Systems 2, 23–27 (01 2012)
11. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. The MIT Press (2016)
12. Graham, B.: Fractional max-pooling (12 2014)
13. HAN, X., YE, J., LUO, J., ZHOU, H.: The effect of axis-wise triaxial acceleration data fusion in cnn-based human activity recognition. IEICE Transactions on Information and Systems E103.D(4), 813–824 (2020)
14. Hancock, E., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.: Advances in pattern recognition methodology and applications. Pattern Recognition Letters 34, 359–360 (03 2013)

15. Ignatov, A.: Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied soft computing* 62, 915–922 (9 2017)
16. Jalal, A., Kim, Y.H., Kim, Y.J., Kamal, S., Kim, D.: Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognition* 61 (08 2016)
17. Kwapisz, J., Weiss, G., Moore, S.: Activity recognition using cell phone accelerometers. *SIGKDD Explorations* 12, 74–82 (11 2010)
18. Lane, N., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.: A survey of mobile phone sensing. *IEEE Commun. Mag.* 48, 140 – 150 (10 2010)
19. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278 – 2324 (12 1998)
20. LeCun, Y., Cortes, C., Burges, C.: MNIST handwritten digit database. *ATT Labs* 2 (2010)
21. Liang, M., Hu, X.: Recurrent convolutional neural network for object recognition. pp. 3367–3375 (06 2015)
22. Martínez, F., González-Fraga, J., Cuevas-Tello, J.C., Rodríguez, M.: Activity inference for ambient intelligence through handling artifacts in a healthcare environment. *Sensors (Basel, Switzerland)* 12, 1072–99 (12 2012)
23. Quaid, M.A., Jalal, A.: Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm. *Multimedia Tools and Applications* pp. 1–23 (12 2019)
24. Ravì, D., Wong, C., Lo, B., Yang, G.Z.: A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE journal of biomedical and health informatics* PP (12 2016)
25. Reyes-Ortiz, J., Oneto, L., Ghio, A., Anguita, D., Parra, X.: Human activity recognition on smartphones with awareness of basic activities and postural transitions (01 2014)
26. San-Segundo, R., Lorenzo-Trueba, J., Martínez-González, B., Pardo, J.: Segmenting human activities based on hmms using smartphone inertial sensors. *Pervasive and Mobile Computing* 30 (01 2016)
27. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T., Kjærgaard, M., Dey, A., Sonne, T., Jensen, M.: Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. pp. 127–140 (11 2015)
28. Uddin, M.T., Billah, M.M., Hossain, M.F.: Random forests based recognition of human activities and postural transitions on smartphone. In: *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*. pp. 250–255 (2016)
29. Voicu, R.A., Dobre, C., Bajenaru, L., Ciobanu, R.I.: Human physical activity recognition using smartphone sensors. *Sensors* 19, 458 (01 2019)
30. Walse, K., Dharaskar, R., Thakare, V.M.: Performance evaluation of classifiers on wisdm dataset for human activity recognition (03 2016)
31. Wan, L., Zeiler, M., Zhang, S., Lecun, Y., Fergus, R.: Regularization of neural networks using dropconnect (01 2013)
32. Weiss, G., Lockhart, J.: The impact of personalization on smartphone-based activity recognition. *AAAI Workshop - Technical Report* (01 2012)
33. Zhang, Y., Zhang, Y., Zhang, Z., Bao, J., Song, Y.: Human activity recognition based on time series analysis using u-net (09 2018)
34. Zheng, Z., Du, J., Sun, L., Huo, M., Chen, Y.: Tasg: An augmented classification method for impersonal har. *Mobile Information Systems* 2018, 1–10 (12 2018)
35. Zhu, X., Qiu, H.: High accuracy human activity recognition based on sparse locality preserving projections. *PLOS ONE* 11(11), 1–18 (11 2016)