



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación
y Desarrollo Tecnológico

Tesis de Maestría

Identificación de situaciones de interés centradas en
el humano

presentada por

Ing. Sagrario Garcia Garcia

como requisito para la obtención del grado de
Maestra en Ciencias de la Computación

Director de tesis
Dr. Raúl Pinto Elías

Cuernavaca, Morelos, México. Enero de 2024.



Cuernavaca, Mor., 05/diciembre/2023

OFICIO No. DCC/207/2023

Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de SAGRARIO GARCIA GARCIA con número de control M21CE057, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado "IDENTIFICACIÓN DE SITUACIONES DE INTERÉS CENTRADAS EN EL HUMANO" y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.


RAÚL PINTO ELÍAS
Director de tesis


NIMROD GONZÁLEZ FRANCO
Revisor 1
CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO


JONATHAN VILLANEVA TAVIRA
Revisor 2

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante

SEP TecNM
RECIBIDO
07 DIC 2023
LMZ
SUBDIRECCIÓN ACADÉMICA

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
07 DIC 2023
CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO
SERVICIOS ESCOLARES
RECIBIDO
EBN





Cuernavaca, Mor.,
No. De Oficio:
Asunto:

11/diciembre/2023
SAC/202/2023
Autorización de
impresión de tesis

**SAGRARIO GARCÍA GARCÍA
CANDIDATA AL GRADO DE MAESTRA EN CIENCIAS
DE LA COMPUTACIÓN
P R E S E N T E**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"IDENTIFICACIÓN DE SITUACIONES DE INTERÉS CENTRADAS EN EL HUMANO"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

*Excelencia en Educación Tecnológica®
"Conocimiento y tecnología al servicio de México"*



**CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO**

C. c. p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



DEDICATORIA

Gracias a Dios, porque todo lo que hago es por él y para él, por darme fortaleza durante este recorrido.

A mi madre María Sagrario, a mi padre Martín y mi hermana Hannia, por ser mi fuente de inspiración, creer en mí siempre y apoyarme incondicionalmente en cada uno mis sueños. Gracias a mi familia por todo el respaldo a la distancia.

A Brayan Alejandro, por ser mi mejor amigo y compañero de aventuras por mucho tiempo.

A mis mejores amigas Alondra, Alma y Alejandra por sus palabras de apoyo cuando lo necesitaba.

Este logro es tanto mío como suyo, los amo.

A la memoria de mi querida Zirahuén.

AGRADECIMIENTO

Agradezco al Consejo Nacional De Ciencia y Tecnología (CONACYT) por el apoyo económico brindado a lo largo de mis estudios de maestría por medio de una beca de posgrado. Al Tecnológico Nacional de México, campus Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), por otorgarme las facilidades para el desarrollo de mi crecimiento profesional.

Al Dr. Raúl Pinto Elías, director de tesis de ese trabajo, gracias por su apoyo, asesoramiento y enseñanzas durante la realización de esta investigación.

A mi comité revisor conformado por el Dr. Nimrod González Franco, Dr. Jonathan Villanueva Tavira. Por sus observaciones y sugerencias para la mejora de esta tesis.

Finalmente agradezco a mis compañeros de generación, por el tiempo compartido durante este trayecto.

RESUMEN

El reconocimiento de actividad humana (*Human Activity Recognition*, por sus siglas en inglés, HAR), brinda información sobre situaciones de interés específicas, de esta forma se obtiene el seguimiento de las actividades realizadas por una o más personas. El procedimiento general para que un sistema de supervisión comprenda la acción que se desarrolla, implica diversas tareas como la adquisición de datos, preprocesamiento, extracción y selección de características, clasificación y evaluación.

El desarrollo de este tema de estudio, puede contribuir a la mejora de la calidad de vida, para conocer el diagnóstico o seguimiento de actividades humana específicas. Gran parte de este crecimiento es el uso de dispositivos portátiles como lo son los celulares y relojes inteligentes.

En este trabajo se propone un sistema de reconocimiento automático de actividades humanas de interés, estas acciones son: estar de pie, caminar, sentarse, caer y “sin interés”. Por medio de un sistema de adquisición multimodal conformado por una cámara RGB y sensores inerciales, utilizando la combinación de modelos de aprendizaje automático para la clasificación de actividades. Como resultado se obtuvo un sistema automático de actividades, que es capaz de reconocer las actividades con un conjunto, combinación, o uso individual de sensores, mediante reglas definidas.

ABSTRACT

The recognition of human activity (Human Activity Recognition, for its acronym in English, HAR), provides information on specific situations of interest, in this way the monitoring of the activities carried out by one or more people is obtained. The general procedure for a monitoring system to understand the action that is being developed involves various tasks such as data acquisition, pre-processing, feature extraction and selection, classification and evaluation.

The development of this topic can improve the quality of life, from knowing the diagnosis or monitoring of specific human activities. Much of this growth is the use of portable devices such as cell phones and smart watches.

In this work an automatic recognition system of human activities of interest is proposed, these actions are: standing, walking, sitting, falling and "without interest". Through a multimodal acquisition system made up of an RGB camera and inertial sensors, using the combination of machine learning models for the classification of activities. As a result, an automatic activity system was obtained, which is capable of recognizing activities with a set, combination, or individual use of sensors, through defined rules.

ÍNDICE GENERAL

Resumen	IV
Abstract	V
Índice de figuras	IX
Índice de tablas.....	X
Acrónimos.....	XI
Capítulo 1 Introducción	1
1.1 Descripción del problema	2
1.1.1 Delimitación del problema específico	2
1.1.2 Complejidad del problema	3
1.2 Contexto teórico	3
1.2.1 Marco conceptual.....	3
1.2.1.1 Inteligencia artificial.....	3
1.2.1.2 Reconocimiento de patrones.....	4
1.2.1.3 Visión artificial	4
1.2.1.4 Reconocimiento de actividad humana.....	5
1.2.1.5 Actividades o acciones humanas	6
1.2.1.6 Tecnologías para el reconocimiento de actividad humana.....	6
1.2.1.6.1 Sensores de movimiento.....	7
1.2.1.7 Aprendizaje automático y profundo aplicado en HAR.....	8
1.2.1.7.1 Bosque aleatorios	8
1.2.1.7.2 YOLOV6	9
1.2.2 Antecedentes	10
1.2.3 Estado del arte.....	11
1.2.3.1 Un panorama del reconocimiento de actividad humana	11
1.2.3.2 Sensores en la identificación de actividades humanas.....	13
1.2.3.3 Sistemas HAR basados en dispositivos móviles	17
1.2.3.4 Exploración de enfoques clásicos y aprendizajes profundo en la identificación de actividades.....	18
1.2.4 Descripción de la solución	25
1.3 Discusión	26
Capítulo 2 Propuesta de solución	28
2.1 Planteamiento de la solución.....	28

2.1.1	Objetivos.....	28
2.1.1.1	Objetivo general.....	28
2.1.1.2	Objetivos específicos	29
2.1.2	Alcances y limitaciones	29
2.1.2.1	Alcances	29
2.1.2.2	Limitaciones.....	29
2.1.3	Justificación y beneficios.....	29
2.2	Modelo conceptual de la solución.....	30
2.3	Descripción de los elementos del modelo conceptual	31
2.4	Variables/ criterios de control.....	31
Capítulo 3 Análisis, diseño e implementación del sistema		33
3.1	Metodología de propuesta de solución	33
3.1.1	Adquisición de datos	36
3.1.1.1	Aplicación de adquisición de datos inerciales.....	36
3.1.1.2	Ejemplo de muestra recopiladas	38
3.1.2	Preprocesamiento de los datos	39
3.1.2.1	Preprocesamiento de las imágenes	39
3.1.2.2	Preprocesamiento de datos inerciales	41
3.1.3	Entrenamiento de algoritmos	42
3.1.4	Evaluación	44
3.1.5	Implementación del sistema.....	45
Capítulo 4 Validación, experimentación y análisis de resultados.....		48
4.1.	Validación y experimentación	48
4.1.1	Prueba 1	48
4.1.2	Prueba 2	50
4.1.3	Prueba 3	54
4.1.4	Prueba 4	57
4.1.5	Prueba 5	58
Capítulo 5 Conclusiones y trabajo futuro.....		61
5.1	Conclusiones generales	61
5.2	Objetivos alcanzados	62
5.3	Aportaciones	64
5.4	Trabajo Futuro.....	64
5.5	Productos generados	65
<u>Referencias</u>		<u>66</u>

Anexo A Pruebas de experimentación	72
Anexo B Constancias de participación	77

ÍNDICE DE FIGURAS

Figura 1.1 Etapas de sistema HAR (Minh Dang et al., 2020).....	5
Figura 1.2 Actividades cotidianas extraídas de un conjunto de datos público (Kim, Yun, et al., 2019).....	6
Figura 1.3 Categorías de tecnologías empleadas en HAR (Fereidoonian et al., 2020).....	7
Figura 1.4. Ejemplo de muestra de datos inerciales (Ferrari et al., 2021).	8
Figura 1.5. Ejemplo de diagrama de bosque aleatorio (<i>What Is Random Forest?</i> IBM, n.d.).....	9
Figura 1.6 Colocación de sensores en el cuerpo de estudio.....	14
Figura 1.7 Matriz de confusión (Bagate & Shah, 2019).....	22
Figura 1.8 Sistema propuesto en (Liu et al., 2021).....	23
Figura 1.9 Operación del algoritmo YOLO	24
Figura 1.10. Posición de los sensores empleados en investigación (Ranieri et al., 2021)	25
Figura 2.1 Modelo conceptual de la solución.....	30
Figura 3.1 Metodología empleada para el diseño de solución.....	35
Figura 3.2. Aplicación móvil Android y Wear Os, para recolección de datos	37
Figura 3.3. Interfaz para recepción de datos en PC	37
Figura 3.4. Ubicación de sensores de visión	38
Figura 3.5. Detección de elementos por medio de yolov5.....	39
Figura 3.6. Información generada por YOLOV5(localización de personas).....	40
Figura 3.7. Herramienta Labellmg	41
Figura 3.8. Líneas de código para entrenamiento de arquitectura YOLO.....	43
Figura 3.9. Matriz de confusión (Jukes, 2018).....	44
Figura 3.10. Interfaz gráfica desarrollada, pestaña fuera de tiempo real	46
Figura 3.11. Interfaz gráfica desarrollada, sistema en tiempo real.....	47
Figura 4.1 Métrica F1-confidence	51
Figura 4.2 Precision- confidence.....	51
Figura 4.3 Métrica Precision-Recall.....	52
Figura 4.4 Métrica Recall-Confidence	52
Figura 4.5 Matriz de confusión Yolov6 Tres clases	53
Figura 4.6 Matriz de confusión Random Forest.....	53
Figura 4.7 Ejemplo de actividades estar de pie, sentarse o caer, clasificadas erróneamente.....	54
Figura 4.8 Métrica Precision -Recall.....	55
Figura 4.9 Métrica Precision -Confidence.....	56
Figura 4.10 F1 -Confidence	56
Figura 4.11 Métrica Recall-Confidence	57
Figura 4.12 Ejemplo de operación de sistema en tiempo real	59
Figura B.1 Reconocimiento de presentación de póster en Escuela de inteligencia artificial	78
Figura B.2. Reconocimiento de presentación de poster 9na JCyTA.....	79
Figura B.3 Ponencia de artículo Human Activity Recognition en ICMEAE	80
Figura B.4 Reconocimiento de participación de ponencia.....	81

ÍNDICE DE TABLAS

Tabla 3.1 Características de datos propio.....	36
Tabla 4.1 Resultados de clasificación imágenes propias	49
Tabla 4.2. Resultados de clasificación de datos inerciales (métrica precisión), divididas en ventanas deslizantes de segundos.....	50
Tabla 4.3 Métricas evaluadas en modelo Random Forest	54
Tabla 4.4 Precisión obtenida entrenando los diferentes dispositivos	58
Tabla 5.1 Objetivos y comentarios	63
Tabla A.1. Resultados de la prueba no.1	72
Tabla A.2 Resultados de la prueba no.2	74
Tabla A.3 Resultados de la prueba no.3	75
Tabla A.4 Resultados de la prueba no.5	76

ACRÓNIMOS

Acónimo	Significado en inglés	Traducción
CNN	<i>Convolutional Neural Network</i>	Red neuronal convolucional
DBM	<i>Deep Boltzmann machine</i>	Máquina de Boltzmann profunda
DL	<i>Deep Learning</i>	Aprendizaje profundo
DT	<i>Decision Tree</i>	Árbol de decisión
FNN	<i>Feedforward neural network</i>	Red neuronal pre alimentada
HAR	<i>Human Activity Recognition</i>	Reconocimiento de actividad humana
HMM	<i>Hidden Markov Model</i>	Modelo oculto de Markov
IA		Inteligencia Artificial
IMU	<i>inertial measurement unit</i>	Unidad de medición inercial
KNN	<i>k-nearest neighbors</i>	K- Vecinos cercanos
LSTM	<i>Long short-term memory</i>	Red de memoria a corto plazo
ML	<i>Machine Learning</i>	Aprendizaje máquina
NB	<i>Naïve Bayes</i>	Bayesiano ingenuo
RF	<i>Random Forest</i>	Bosques Aleatorios
RGB	<i>Red Green Blue</i>	Rojo, Verde y Azul
RGB-D	<i>Red Green Blue and Depth</i>	Rojo Verde Azul y profundidad
RNN	<i>Recurrent neural network</i>	Red neuronal recurrente
SRU-GRU	<i>Simple Recurrent Unit- Gated recurrent units</i>	Unidades recurrentes simples y unidades recurrentes cerradas
SVM	<i>Support Vector Machine</i>	Máquinas de vectores de soporte
YOLO	<i>You Only Look Once</i>	Solo miras una vez

CAPÍTULO 1

INTRODUCCIÓN

El reconocimiento de la actividad humana es un tema donde se utiliza la inteligencia artificial (IA), se trabajan ramas como la visión por computadora, aprendizaje automático, estudio de señales y entre otras. El propósito es el análisis de actividades, mediante la adquisición de datos de sensores e interpretados a través de los algoritmos (Kim, Yun, et al., 2019), con el objetivo de comprender y clasificar situaciones de interés (actividades, acciones o movimientos), estas pueden ser “ambulatorias” y “no ambulatorias” (Vrigkas et al., 2015).

En la actualidad la supervisión automática tiene un papel relevante, un ejemplo de sus aplicaciones son en los campos de la salud (D’Angelo & Palmieri, 2021; Lu et al., 2019; Subasi et al., 2019), videovigilancia (Babiker et al., 2018), e interacción humano-máquina (Liu et al., 2021). El estudio de actividad humana contribuye al proceso de monitoreo y alerta de acciones realizadas por una persona en diversos espacios o lugares.

Sin embargo, la identificación de actividades humanas tiene desafíos importantes, como lo son la variabilidad de los datos obtenidos de las diferentes acciones ejecutadas por una o varias personas, la similitud entre las situaciones, el tipo de tecnología, la calidad y cantidad de muestras recopiladas, las metodologías empleadas, entre otros factores (Zhang et al., 2022) En este contexto, el presente trabajo de investigación, tiene como finalidad la detección automática de al menos tres situaciones de interés, efectuadas por una persona por medio de sensores y modelos computacionales.

1.1 Descripción del problema

El estudio de la actividad humana permite adquirir información del cómo se desarrolla o qué actividades se realizan en un entorno, con estos datos es posible hacer una evaluación, diagnóstico y/o simplemente el monitoreo continuo de una persona, con actividades o acciones de interés.

La supervisión de actividad humana en el presente es utilizada comúnmente en entornos domésticos, debido a que es una herramienta que repercute en la mejora de la calidad de vida de la persona, en el ámbito de cuidado de la salud y/o seguridad. Un ejemplo de estos sistemas, es el monitoreo de personas vulnerables, como aquellas que padecen alguna enfermedad, discapacidad, niños o adultos mayores, con la finalidad de aumentar el bienestar y autonomía (Alwakeel et al., 2016). Otro caso es la detección de actividades sospechosas en lugares específicos, con el fin de alertar o identificar acciones inseguras (Prati et al., 2019).

La identificación de actividad humana presenta algunos desafíos, como son la forma en que las personas ejecutan una acción; esto puede dificultar la clasificación. También factores como la edad, altura y peso. Por otro lado, los datos adquiridos por los sensores, cantidad de datos recopilados, la calidad de los mismos. Además, las muestras pueden ser afectadas por intrusión de los sensores vestibles o de visión. En la revisión del estado del arte, se identificaron estos retos que son los más frecuentes, esto implica que es un problema abierto, en el cual existen diversas alternativas de solución.

1.1.1 Delimitación del problema específico

El reconocimiento de actividad humana en la actualidad es un tema importante, dado que puede apoyar en las áreas previamente mencionadas. Existen diversas formas para el análisis de las acciones, esto sujeto a la aplicación, requisitos computacionales y tecnologías empleadas.

En esta investigación se pretende detectar una serie de actividades humanas con sensores que sean de fácil acceso, con el objetivo de clasificar las acciones que se llevan a cabo en un entorno determinado.

1.1.2 Complejidad del problema

La supervisión de actividades humanas presenta problemáticas como las siguientes. Sin embargo, no se limitan a las que se mencionan, a continuación:

- Variabilidad en la recolección de los datos de la situación desarrollada por una persona.
- Variabilidad de los datos al aumentar la muestra de datos, con un grupo de personas.
- Similitud entre la caracterización de las diferentes situaciones definidas, realizadas por una o varias personas.
- Presencia de ruido en la adquisición de datos provocando imprecisión en las muestras.
- Algoritmos de alto requerimiento computacional.

1.2 Contexto teórico

En la siguiente sección se presentan conceptos claves para el estudio del reconocimiento de actividad humana, además los antecedentes de este trabajo, así como la revisión del estado del arte que permitió el desarrollo de la investigación.

1.2.1 Marco conceptual

En esta subsección se exponen elementos relacionados con el tema de tesis, con el propósito de establecer definiciones teóricas que faciliten la comprensión de este trabajo de investigación. A continuación, se establecen conceptos como: inteligencia artificial, reconocimiento de patrones, visión artificial, el reconocimiento de actividad humana, actividad o acciones, tipos de tecnologías empleadas y algoritmos.

1.2.1.1 Inteligencia artificial

La inteligencia artificial (IA), se define como la ciencia e ingeniería de las máquinas que actúan de manera inteligente, la IA se ha convertido en una disciplina con un nivel de madurez alto, tiene un impacto en diversas áreas, debido a sistemas inteligentes capaces de apoyar en la toma de decisiones. Su objetivo la inferencia del modelo, a partir de un

conjunto de factores relacionados y su interacción a partir de los datos observados (Azuela & Cortés, 2021).

Algunas de las tareas donde se aplica la inteligencia artificial son:

- Reconocimiento de imágenes estáticas, clasificación y etiquetados.
- Detección y clasificación de objetos.
- Mejoras del desempeño de la estrategia algorítmica comercial.
- Procesamiento eficiente y escalable de datos de pacientes.
- Protección contra amenazas de seguridad cibernética (L. Ruhiainen, 2018).

1.2.1.2 Reconocimiento de patrones

Es una disciplina científica que tiene por objetivo el clasificar objetos en un número específico de categorías o clases. Los objetos pueden ser imágenes, sonidos, señales producto de mediciones que deben ser clasificadas. Está relacionada con procesos de ingeniería, matemáticos y de ciencias computacionales. Un sistema de reconocimiento de patrones se compone de sensores, mecanismo de preprocesamiento, mecanismo de extracción de características y algoritmo de clasificación (Dougherty, 2012).

1.2.1.3 Visión artificial

Puede ser definida como los procesos de obtención, caracterización e interpretación de información de imágenes tomadas desde un mundo tridimensional a partir de imágenes bidimensionales (Fu K.S et al., 1988). La visión artificial pretende percibir su entorno de manera que lo haría un ser humano o ser vivo, identificando y reconociendo objetos o ambiente, se emplea en el campo de la robótica, carros autónomos, campo de la medicina, videovigilancia entre otros. El proceso de la visión se puede dividir de la siguiente manera:

- Adquisición de la imagen: se consigue la imagen más adecuada, considerando cámara, tarjeta de adquisición de imagen, óptica, iluminación, entre otras.
- Preprocesamiento: reparación de imagen para eliminar ruidos.
- Segmentación: diferenciación de objetos del fondo y extracción de características.
- Reconocimiento: asignación de etiquetas basado en los clasificadores (Marcos et al., 2006).

1.2.1.4 Reconocimiento de actividad humana

En la literatura el reconocimiento de actividad humana, se hace mención por parte de los autores, que se refiere a la predicción automática de las actividades humanas que se desarrollan en la vida cotidiana, como caminar, correr, cocinar, realizar trabajos en la oficina, entre otras, por medio de datos obtenidos por sensores como secuencias de imágenes o datos inerciales. Con el fin de monitorear o regular actividades, para reducir el riesgo de alguna enfermedad o situación peligrosa (Deepak Jatkar Anil Surve, 2021). Este tipo de sistemas funcionan en etapas comenzando por la adquisición de datos, el procesamiento, extracción y selección de característica, entrenamiento y reconocimiento de actividad. En la Figura 1.1, se observan las etapas de un sistema HAR (Minh Dang et al., 2020).

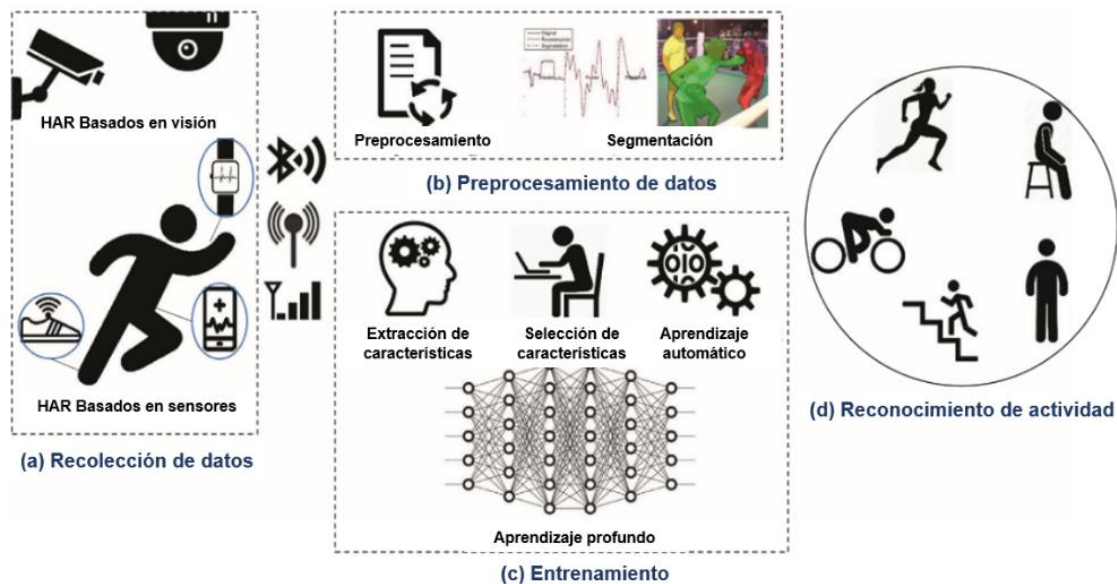


Figura 1.1 Etapas de sistema HAR (Minh Dang et al., 2020)

Algunos de los campos de investigación donde es aplicado HAR, es en seguridad y vigilancia, cuidado de la salud, conducción autónoma, interacción humana – máquina, casas inteligentes, el entretenimiento y entre otros. Con el objetivo de mejorar o apoyar en ciertas condiciones.

1.2.1.5 Actividades o acciones humanas

Las actividades o acciones humanas, se componen de movimientos ejecutados por uno o varios individuos, dentro del estado del arte, se dividen en ambulatorias y no ambulatorias, la primera consiste en el movimiento dinámico de la persona, por otra parte, la segunda ocurre cuando la persona se mantiene en una posición que no genera movimiento o se encuentra estático. También existe la siguiente clasificación: gestos, atómicas, interacciones humanas con objetos, humano a humano, agrupaciones, conductas y eventos (Vrigkas et al., 2015). Todo depende de la aplicación o uso del sistema.

Las acciones que con mayor frecuencia se estudian en artículos recientes son actividades cotidianas como estar de pie, caminar, sentarse, subir escaleras, bajar escaleras, acostarse, hasta acciones que involucran la interacción de objetos para saber si se está trabajando en la computadora, escritura en papel, realización de preparación de comida específica (Kim, Jalal, et al., 2019). Pero también actividades que pueden ser riesgosas para la persona, como caídas, desmayos, convulsiones y dolores en el cuerpo, por mencionar algunas (Kim, Yun, et al., 2019; Zia et al., 2021). En la Figura 1.2, se observa un conjunto de imágenes con actividades cotidianas.



(A) Sentado en una silla

(B) Recogiendo

(C) Acostado en la cama

Figura 1.2 Actividades cotidianas extraídas de un conjunto de datos público (Kim, Yun, et al., 2019)

1.2.1.6 Tecnologías para el reconocimiento de actividad humana

Las tecnologías empleadas para el reconocimiento de actividad humana, se categorizan en tres tipos, como se muestra en la Figura 1.3, donde existen los sistemas basados en visión, sensores y multimodal. En el primer aspecto son todo tipo de cámaras (ejemplo, RGB y de profundidad), los sistemas basados en sensores pueden ser inerciales (acelerómetro,

giroscopios y magnetómetros), sensores de temperatura, sensores ambientales, sensores fisiológicos y entre otros. Por último, el sistema multimodal es aquel que se compone de una combinación basados en visión y sensores (Chen et al., 2021).

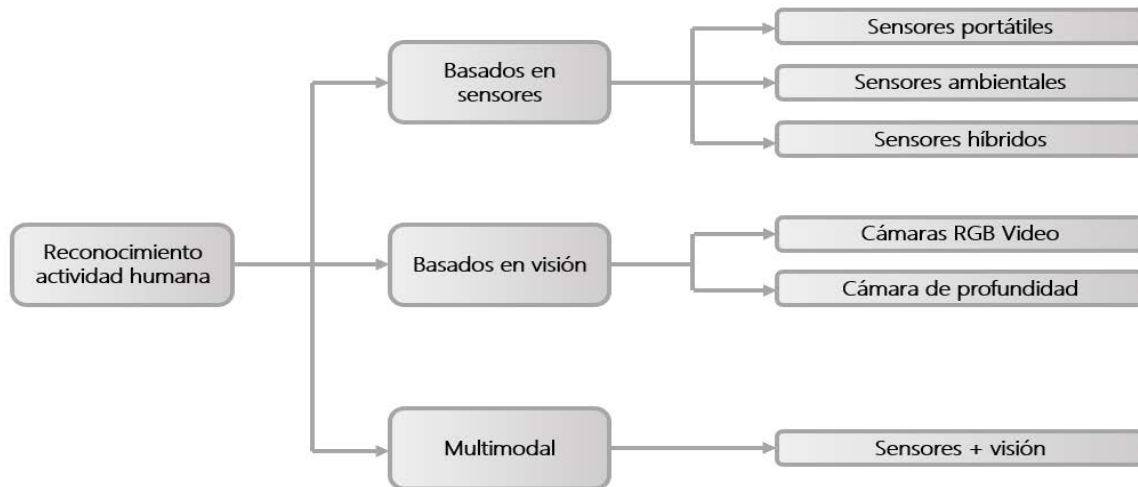


Figura 1.3 Categorías de tecnologías empleadas en HAR (Fereidoonian et al., 2020)

1.2.1.6.1 Sensores de movimiento

Dentro de la categoría de sensores que se emplean comúnmente en el área, en la adquisición y monitoreo de movimientos, son:

- **Acelerómetro:** mide el cambio de velocidad de un objeto, la unidad de medida es metros cuadrados sobre segundos cuadrados (m^2/s).
- **Giroscopio:** mide la velocidad angular triaxial, su unidad son grados sobre segundos ($grados/s$).
- **Magnetómetro:** mide el cambio de un campo magnético en un lugar determinado, se emplean los tres ejes ("x", "y" y "z")(Ferrari et al., 2021).

Estos se encuentran contenidos en diversos elementos como los sensores IMU, o dispositivos portátiles (teléfonos celulares, anillos portátiles y relojes inteligentes). Los cuales permiten que los datos sean recopilados fácilmente. En la Figura 1.4, se muestra un ejemplo de las muestras recopiladas por los sensores mencionados.

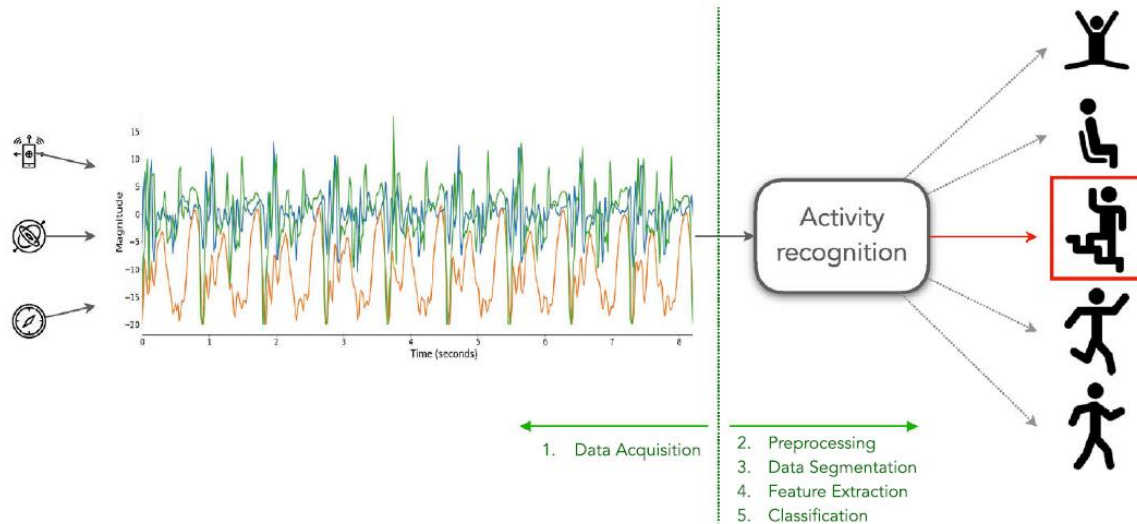


Figura 1.4. Ejemplo de muestra de datos inerciales (Ferrari et al., 2021).

1.2.1.7 Aprendizaje automático y profundo aplicado en HAR

Los algoritmos mayormente empleados en la supervisión de actividades humanas son los siguientes, en el caso de datos inerciales obtenidos por sensores de movimiento se encuentra *Random forest* (Bosque Aleatorio) y para trabajar la clasificación de imágenes las redes neuronales convolucionales el modelo más utilizado *YOLO* (*You Only Look Once*). A continuación, se describen cada uno de estos.

1.2.1.7.1 Bosque aleatorios

Es un algoritmo de aprendizaje automático de conjunto, que se usa para la clasificación y regresión, donde se utilizan un número mayor de árboles de decisión con el propósito de aumentar la exactitud de un modelo de clasificación. Cada árbol está entrenado a partir de un subconjunto de muestras de datos seleccionadas aleatoriamente, cada nodo toma un subconjunto de atributos al determinar la mejor separación. Se obtiene la salida final todos los modelos aportan una predicción individual. Una de las ventajas de este algoritmo es que tiene mejor generalización y menor afectación en el sobre ajuste (Pertuz, 2022). En la Figura 1.5, se observa un ejemplo de representación gráfica de bosque aleatorio (*What Is Random Forest? | IBM, n.d.*).

Los árboles no se podan después de la construcción, lo que les permite adaptarse parcialmente a su propia muestra de datos. Para diversificar aún más los clasificadores, en

cada rama del árbol, la decisión de en qué característica dividir se restringe a un subconjunto aleatorio de tamaño n , del conjunto completo de características (Jukes, 2018).

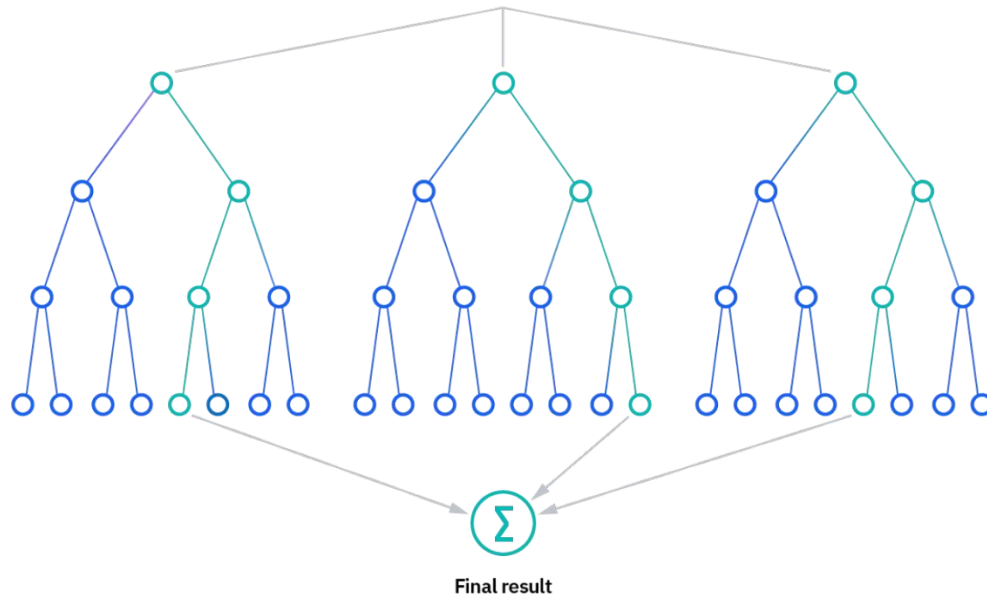


Figura 1.5. Ejemplo de diagrama de bosque aleatorio (*What Is Random Forest?* | IBM, n.d.).

1.2.1.7.2 YOLOV6

La arquitectura YOLO (*You Only Look Once*) es una red neuronal convolucional que puede realizar la tarea de detección y clasificación de imágenes, se destaca por la velocidad y precisión para localizar un elemento de interés en unas imágenes. En la actualidad este algoritmo se emplea en el campo de vehículos autónomos, sistemas de identificación y seguimiento de objetos, detección de matrículas, reconocimiento fácil, reconocimiento de actividad en vigilancia, deporte e interacción humano máquina (Terven & Cordova-Esparza, 2023).

La versión de YOLOV6 fue publicada en septiembre del 2022, realiza la clasificación de los objetos previamente ubicados, donde se determina en un cuadrado la región de interés y la probabilidad de que pertenezca a una clase. El modelo está compuesto de YOLOv5 (v6.0/6.1) consta de: - *Backbone: New CSP-Darknet53* - *Neck: SPPF, New CSP-PAN* - *Head: YOLOv3 Head*.

1.2.2 Antecedentes

En el Centro Nacional de Investigación y Desarrollo Tecnológico (TecNM/CENIDET), en el departamento de Ciencias Computacionales, existen trabajos de investigación previos relacionados al tema central esta tesis, reconocimiento de actividades humanas. A continuación, se describe brevemente cada uno de ellos.

En trabajo de tesis (Gloria Vanessa Urquiza Flores, 2022) “Detección de anomalías en el comportamiento de una persona”, el objetivo de este trabajo consistió en el análisis y compresión de patrones. Para conocer el comportamiento de una persona en rutinas cotidianas, con la finalidad de diferenciar una rutina normal de una anomalía. Todo esto por medio de un sistema automático que empleo un sistema multimodal.

Por otro lado, en el trabajo de tesis de (Alcantar Rafael, 2013), titulado “Caracterización Visual de Movimientos “Sospechosos” de Personas en Estacionamiento”, fue desarrollado un Sistema de visión inteligente, con la capacidad de identificar y caracterizar eventos sospechosos como lo son: merodear, agacharse y acostarse. Este sistema presenta la facultad de identificar a múltiples personas.

En “Caracterización Automática de una Muestra de Acciones en video” de (Dávila María Luisa, 2009), se identificaron y caracterización actividades relacionadas con la interacción de un entorno específico en este caso un estante, con las partes humanas como mano, cabeza, brazos y cuerpo. Por medio de autómatas y visualización gráfica.

Por último, en el trabajo de “Caracterización de Imágenes en Movimiento: Correr y Caminar” por (Saldaña Jorge Alfredo, 2007) se creó un sistema con técnicas de visión artificial que permite identificar acciones tales como correr y caminar de personas saludables. Por medio de un traje y una cámara se extrajeron características de los movimientos, para construir un modelo que identificará automáticamente las acciones previamente comentadas.

1.2.3 Estado del arte

En esta sección se presenta el análisis realizado del estado del arte, presentando los artículos más relevantes que contribuyeron a la solución de la problemática previamente expuesta. Los temas están relacionados con el reconocimiento de actividad humana automática.

1.2.3.1 Un panorama del reconocimiento de actividad humana

Trends in Humana activity recognition using smarthphone (Ferrari et al., 2021)

En el artículo se reporta el resumen de un conjunto de trabajos, sobre la importancia del uso de dispositivos inteligentes, que no requieren instalación técnica y no es intrusivo. También se presentan las etapas de un sistema de identificación de acciones humanas, los cuales describen pasos desde la adquisición, preprocesamiento, segmentación de datos, extracción de características y la clasificación, a continuación, se describen:

- **Adquisición:** Se emplean de cámaras, dispositivos portátiles y ambientales, el paso de los años ha permitido mejorar el costo y por lo tanto la capacidad de precisión. Un factor importante es la frecuencia de muestreo que son los datos registrados por segundo.
- **Preprocesamiento:** Se preparan los datos, lo que quiere decir que se hace un filtrado para aislar los datos con la finalidad de limpiarlos.
- **Segmentación de datos:** En esta etapa se dividen la muestra conjunta de las señales en segmentos pequeños, conocidos como ventanas, existen tres tipos destacables: ventana por actividad, por evento y deslizantes.
- **Extracción de características:** la obtención de características, para tener las más representativas y al mismo tiempo reducir los datos, que no representen a la actividad.
- **Clasificación:** El aprendizaje automático y profundo, han logrado un buen rendimiento dentro de la clasificación, el uso dependerá de la aplicación específica.

Machine Learning (ML) based Human Activity Recognition Model using Smart Sensors in IoT Environment (Navita & Mittal, 2022)

El trabajo de investigación presentado propone la monitorización de personas mayores con el fin de que estas vivan de una manera segura. Debido a que este sector de la población presenta muchas lesiones graves relacionadas por caídas, un artículo de Centros para el Control y Prevención de Enfermedades (CDC) menciona que este tipo de situaciones va en aumento en los últimos años. Se propone una arquitectura de aprendizaje automático integrando el internet de las cosas y sensores que no son intrusivos para el usuario. Los puntos importantes dentro de este desarrollo son: sensores para el seguimiento de la actividad, dispositivos de comunicación entre sensores y la nube, la nube, el análisis de datos por medio de aprendizaje automático y por último servicios médicos. Los algoritmos empleados fueron máquina de soporte vectorial (SVM), *Naive Bayes* (NB), KNN, *Random forest* (RF) y *Decision Tree* (DT). Donde se obtuvo como resultado que SVM obtuvo una precisión promediada de 0.9803%.

Detection of Smoking Events from Confounding Activities of Daily Living (Lu et al., 2019)

Se expone un campo de aplicación, que es la problemática del tabaquismo, esto causa problemas de la salud de leves a graves. El uso de sistemas de visión en algunas aplicaciones tiene desventajas debido a que es inexacto o invade privacidad del usuario. Se propone el uso de sensores de movimiento (acelerómetro y giroscopio), que se encuentra contenido en un Arduino con una tasa de muestreo de 20hz.

Los retos presentados son la diferenciación de actividades similares como lo son: fumar, beber, comer, rascarse la cabeza y morderse las uñas, todas ellas en común es la elevación de la mano a nivel de la cabeza. Para clasificar las actividades se analizaron los algoritmos de SVM y *Random Forest*. Las características que se extrajeron fueron la orientación de acelerómetro y giroscopio, característica de rotación, se comparó el rendimiento de cada sensor por separado y de la unión de estos. El clasificador que obtuvo mejores resultados fue *Random forest* con un F1-score de 93% la actividad de fumar.

Effective human activity recognition approach using machine learning (Abdul Lateef Haroon & Premachand, 2021)

El reconocimiento de actividad humana, tiene por objetivo reconocer, analizar y predecir situaciones humanas de interés. Por medio de sensores portátiles o dispositivos externos, cuando se utiliza una cámara de RGB o de profundidad la información que se requiere obtener de la secuencia de imágenes es características del movimiento.

Mientras más grande sea el conjunto de datos se tienen dos problemas que son uso adicional de recursos y el tiempo de procesamiento aumenta, lo cual lleva a que los requerimientos computacionales sean mayores

1.2.3.2 Sensores en la identificación de actividades humanas

Human activity recognition and fall detection (Deepak Jatkar Anil Surve, 2021)

El objetivo del estudio, fue la elaboración de un sistema de identificación de actividades que se relacionan directamente con la caída de personas, con el propósito de aplicarlo en un sector vulnerable como los adultos mayores. Las problemáticas presentadas son el monitoreo continuo por medio de sistemas de visión, la privacidad del individuo, la calidad de la imagen y los factores ambientales como la iluminación.

Se implementó un conjunto de cámaras, para adquirir información de la imagen y video, de las actividades caminar, trabajar en computadora, pararse, subir y bajar escaleras, saltar, correr, doblarse, aplaudir. Se procedió a la clasificación de los videos, por medio del algoritmo *K- Vecinos cercanos (KNN)*, el cual, toma los valores de ángulo, relación con el sujeto y otros como la entrada, se extrajeron características para detectar caídas al realizar las actividades previamente mencionadas. El trabajo realiza dos soluciones una es alimentar el sistema de videos y otro compara para predecir una caída.

Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning (Chung et al., 2019)

El análisis de este estudio indica la selección de sensores, para la adquisición de datos y el posicionamiento de estos, donde las principales ubicaciones son: muñecas, tobillos, muslo,

pecho, espalda y cadera. Además, el tipo de entrenamiento realizado para ejecutar la tarea de reconocimiento.

Para el entrenamiento y experimentación, se realizó un banco de pruebas, que se obtuvieron por medio de IMU MPU-9250, que fueron colocado en distintas ubicaciones del cuerpo con correas de velcro para sujetarlo y eliminar ruidos. Como se muestra en la Figura 1.6. Los datos recopilados se transmiten a un dispositivo móvil *Android*. Se almacenó la información de cinco personas, sin impedimentos corporales. El entorno fue controlado para tomar un conjunto de datos y otro en mundo real. Se fusionaron de los datos obtenidos de cada sensor para un mejor rendimiento. Para el reconocimiento se emplea la *red Long short-term memory (LSTM)* que puede trabajar con dependencias temporales, haciendo posible el reconocimiento de una actividad cuando se tiene una baja frecuencia.

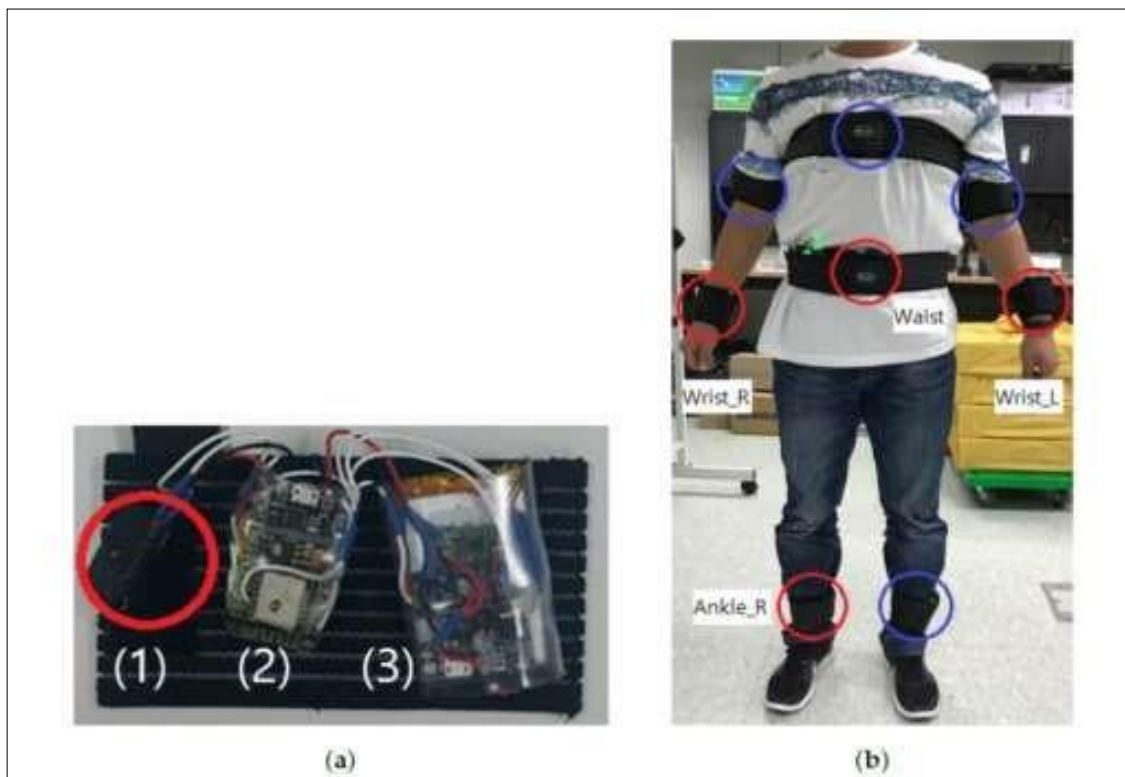


Figura 1.6 Colocación de sensores en el cuerpo de estudio

Fall Detection for the Elderly Based on 3-Axis Accelerometer and Depth Sensor Fusion with *Random forest Classifier* (Kim, Yun, et al., 2019)

Se presenta, un sistema de reconocimiento de actividades en tiempo real, que se compone de dos tipos de tecnologías, se utilizó un acelerómetro y RGB-D sensor, los dispositivos que los contenían es un reloj inteligente y un *kinect*®. Al implementar estos dos, la precisión de la identificación aumenta.

Se extraen características de movimiento del acelerómetro y pose del *kinect*®, ejemplo de ellas la velocidad, altura, ancho, profundidad, entre otras. Se aplica el algoritmo *random forest* para la clasificación, este ofrece ventajas como su rendimiento en un PC limitado y maneja un amplio número de variables. Con los datos recopilados de las actividades específicamente en las caídas hacia atrás, adelante y lateral, se simuló ese escenario. Se concluyó que el uso de estos dos tipos de sensores, mejora la identificación de actividades, en escenarios donde el sistema de visión no tiene campo de visualización, se complementa con el acelerómetro para conocer el estatus de la persona.

A Hybrid Deep Learning Model for Human Activity Recognition Using Multimodal Body Sensing Data (Gumaei et al., 2019)

Se utiliza un sistema multimodal, que consta de un conjunto de sensores, cómo se realiza la fusión de estos datos. Se presenta un enfoque para el reconocimiento de imágenes, con la adquisición multimodal con acelerómetro y giroscopio, empleando un modelo híbrido SRU-GRU (Unidades recurrentes simples y unidades recurrentes cerradas) contiene cuatro capas ocultas, capas de salida y entrada, la unión de SRU-GRU, la respuesta es rápida en tiempo real y se tiene una mejor clasificación. La muestra fue de diez personas las cuales portaban los sensores *SHIM-MER2*, la ubicación fue en muñeca derecha, tobillo izquierdo y pecho.

En la experimentación se elaboraron dos tipos de pruebas, usando técnica de retención (70% datos para entrenamiento y 30% para pruebas) este obtuvo 99.8% de precisión para el reconocimiento de imágenes y de validación cruzada se obtuvo 99.7%. Se comparó con un modelo SRU, resultando que el modelo propuesto es 99.80% más preciso.

Vision2Sensor: Knowledge Transfer Across Sensing Modalities for Human Activity Recognition (Radu & Henne, 2019)

Es otro ejemplo de la temática del uso de dos tipos de tecnologías para la adquisición de datos, la cual es oportuna para realizar una mejor precisión de los datos obtenidos, se muestra la manera en que los datos son transferidos para la identificación. El sistema *Visión2Sensor*, se divide en dos partes: sensores de teléfono inteligente (acelerómetro, giroscopio, magnetómetro, etc.) y visión, cuando empieza a operar el subsistema de sensores, inicia la comunicación con el de visión para sincronizar los datos.

Para el reconocimiento basado en visión, se genera la etiqueta para reconocer la región a estudiar, esto por medio de *MobilNet*, que está entrenada con *COCO*, se adoptó un clasificador CNN para el reconocimiento de actividad. Por otra parte, en el caso de los sensores se usó una red neuronal profunda. Para la validación de los modelos se utilizaron conjuntos de datos de conjuntos de datos como UCF101. El conjunto de estos dos subsistemas permite la identificación de actividades, aunque se encuentre fuera del campo de visión.

Los datos adquiridos por los autores fueron a través de una cámara fija y un teléfono móvil, se exploraron cinco actividades humanas simples: caminar, girar a la izquierda, de pie, giro a la derecha y sentado.

CZU-MHAD: A Multimodal dataset for Human Action Recognition Utilizing a Depth Camera and 10 Wearable Inertial Sensors (Chao et al., 2022)

En este artículo se presenta el uso de tipos de sensores multimodal, se emplean sensores como la cámara del *kinect* para obtener imágenes de profundidad y diez sensores portátiles, con el objetivo de recopilar información debido a que los conjuntos de datos que se encuentran disponibles carecen de información, de 22 acciones diferentes realizadas por cinco personas.

Como se mencionó anteriormente se emplearon estos sensores por sus características de bajo consumo de energía, bajo costo y sencillo de trabajar. Para la adquisición de datos los sensores portátiles están controlados por una *raspberry* y el *kinect* por una computadora y se sincronizan cuando se recopilan datos. Otro punto importante a

mencionar son las características del conjunto de datos, el que la persona realizara la acción varias veces, altura y peso, posición de sensores. La información se obtuvo de la recopilación de datos es posible utilizarla en investigaciones de reconocimiento de actividades humanas, estimación de pose, diagnóstico de enfermedades entre otras.

Efficient Human Activity Recognition Using a Single Wearable Sensor (Lu et al., 2020)

El uso de los sensores correctos previene que se obtengan mediciones incorrectas, que afecten el proceso de identificación de actividades humanas, esto puede ser grave si avanza a un diagnóstico, el cual puede resultar incorrecto y provocar algún efecto adverso en la salud. En este artículo las actividades que se detectan son fumar, beber, comer, rascarse la cabeza y frotarse la nariz. Se emplean dispositivos portátiles teléfono móvil y la placa *Arduino 101*. Los datos analizados consisten en un conjunto de datos adquirido públicamente y uno auto recopilado, a los cuales se les aplico un proceso de preprocesado para eliminar ruidos y de esta manera extraer la mayor información para las características.

En el proceso de selección de la mejor función se eliminaron todas aquellas características que no tenían aporte a la clasificación de las actividades. Se analizaron diversas arquitecturas de las cuales se obtuvo que el modelo GBDT, obtuvo mejores resultados. Los otros dos algoritmos que también funcionaron aceptables bajo ciertas condiciones fueron SVM y *Random Forest*.

1.2.3.3 Sistemas HAR basados en dispositivos móviles

Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances (Zhang et al., 2022)

Los avances actuales en el reconocimiento de actividad humana, permiten que se emplee en diversos campos, gran parte de este progreso es debido a los dispositivos portátiles. El desarrollo tecnológico ha permitido reducir el tamaño y peso de estos dispositivos. Los elementos más utilizados son: teléfonos móviles y relojes inteligentes, aunque también existen anillos inteligentes, zapatos inteligentes, brazaletes, entre otros. El trabajo se enfoca en un recopilado de diversas investigaciones donde se da a conocer un amplio panorama de los retos que enfrenta esta tarea, como la adquisición de los datos, desde la calidad y la cantidad, para una correcta descripción o caracterización de la actividad. Por otra parte, se

trabaja con algoritmos *Deep learning*, el enfoque de clasificación, que ha tenido un crecimiento los últimos años, el enfoque clásico ha sido superado por *Deep learning*, que mejora el rendimiento y desempeño del HAR, las redes neuronales pueden aprender de manera eficiente las características representativas de señales sin procesar.

Detection of Generalized Tonic Clonic Seizures and Falls in Unconstrained Environment Using Smartphone Accelerometer (Zia et al., 2021)

En esta investigación se menciona la elaboración de un sistema, que reconoce diferentes tipos de actividades clasificadas como estacionarias, ambulatoria ligera e intensa y anormal, en esta última son caídas y ataques epilépticos. Este estudio recopiló la información por medio de un teléfono móvil, colocado en la parte superior del brazo izquierdo de 23 personas. Después se efectúa el preprocesamiento con un análisis en dominio del tiempo y frecuencia, durante el desarrollo del sistema, se observó que la actividad de ataque epiléptico mostraba características muy significativas en los valores obtenidos. El artículo explica la relevancia, de la posición del sensor en el cuerpo y el tipo de algoritmo de clasificación que realizó de manera más eficiente el trabajo, en la detección de actividades, en este caso una caída.

1.2.3.4 Exploración de enfoques clásicos y aprendizajes profundo en la identificación de actividades

Human Activity Recognition by Wearable Sensors: Comparison of different classifiers for real-time applications (De Leonardis et al., 2018)

En este trabajo se realiza un análisis para la elección de un clasificador, para la identificación de ocho actividades, se elaboró la comparación de cinco clasificadores basados en aprendizaje automáticos utilizados comúnmente en el área, los cuales son: *K-Nearest Neighbor*, *Feedforward Neural Network (FNN)*, *Support Vector Machines (SVM)*, *Naive Bayes (NB)* y *Decision Tree (DT)*. Se examinó la precisión para la detección de actividades, las ventajas y desventajas de su implementación en un sistema en tiempo real.

El conjunto de datos adquirido es de 15 personas jóvenes y sanas, a través del sensor *MTx miniature magnetic and inertial measure unit (MIMU)* de *Xsens Technologies*. Las actividades fueron: sentarse, estar de pie, acostarse, caminar, bajar y subir escaleras. La duración de las muestras es de 20 segundos, con una repetición de 25 veces cada una.

Por medio de la adquisición de datos, extracción de características, construcción del conjunto de entrenamiento, se evaluó cada uno de los clasificadores, se desarrolló la comparación de los algoritmos, concluyendo que la mayoría de ellos, es capaz de reconocer las actividades en un 90% de exactitud. El mejor resultado en precisión fue KNN, pero centrado en ventajas y desventajas en un sistema en tiempo real, se concluyó que DT es el más adecuado para este tipo de aplicación, debido a sus requisitos computacionales.

Smartphone-Based Human Activities Recognition System using *Random forest Algorithm* (Radhika et al., 2022)

El uso de dispositivos móviles en el campo de reconocimiento de actividades se utiliza para el monitoreo de las condiciones de salud de las personas con el propósito de que su vida sea integral, dado a que el estudio de HAR tiene la capacidad de detectar condiciones fisiológicas. Se propone un sistema de reconocimiento de actividades por medio de un algoritmo de bosques aleatorios (RFA), trabaja de tal manera que contiene diversos árboles los cuales tienen la función de ser clasificadores, al final del proceso se vota para elegir el mejor resultado. El conjunto de datos empleado contiene datos de un teléfono inteligente, contiene el sensor de acelerómetro y sensor giroscopio. La toma de las muestras fue de 50 Hz, para seis actividades diferentes (acostarse, sentarse, pararse, caminar, bajar las escaleras y subir escaleras), de un grupo de 30 personas. De los datos recopilados se utilizó un 70% de estos para el entrenamiento y el 30% de pruebas. Se realizó la comparación del rendimiento del módulo propuesto contra KNN y SVM, obteniendo como resultado que la propuesta tiene 98.34% de precisión a comparación de los demás.

A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services (Park et al., 2016)

Se expone una técnica, para la identificación de actividades humanas, por medio de una cámara, un procesamiento para obtener la silueta de una persona, a través del entrenamiento con red neuronal. Sistema de reconocimiento con red neuronal recurrente (RNN) con la base de datos *microsoft research cambridge*. Se reconocen 12 actividades que fueron recolectadas en el base de datos antes mencionada de 30 sujetos, datos tomados por medio de una cámara de profundidad, se extrajeron características de las articulaciones

de 14 puntos (columna vertebral, cuello, antebrazo derecho, hombro derecho, el antebrazo izquierdo, hombro izquierdo, cadera derecha, muslo derecho e izquierdo, entre otros). De esta manera se obtuvo el ángulo de articulación. Se realizaron comparaciones entre los modelos de Markov oculto (HMM), *Deep Belief Network* (DBN) y RNN. Los resultados obtenidos demostraron que HMM tiene menor rendimiento en comparación con los demás, tomando como referencia la métrica de precisión, donde el modelo DBN mostró un 97.54% y RNN una precisión de 99.55%.

A robust human activity recognition system using smartphone sensors and deep learning (Hassan et al., 2018)

El estudio presenta la identificación de actividades, por medio de un teléfono móvil y la comparación de diversos algoritmos de clasificación. Los cuales, pueden ser tomados en cuenta para el tema de tesis. Se propone el reconocimiento de actividades, por medio de la adquisición de datos obtenidos de un teléfono inteligente que contiene sensores como acelerómetros triaxiales y giroscopios. Las actividades registradas son; pararse, sentarse, acostarse, caminar, caminar hacia arriba, caminar hacia abajo, pararse para sentarse, sentarse para pararse, sentarse para acostarse, acostarse para sentarse, pararse para acostarse y estar de pie. Se realizó la comparación de los algoritmos, para el entrenamiento y reconocimiento, se utilizaron red neuronal artificial (ANN), SVM y DBN. Resultando que este último, obtuvo la tasa de reconocimiento más alta con un 95.85%, por lo cual el porcentaje de error se reduce.

Human activity recognition by combining external features with accelerometer sensor data using deep learning network model (Varshney et al., 2022)

Se analizan tres tipos de modelo CNN, LSTM y ConvLST, para la identificación de actividades del conjunto de datos WISDM (*Wireless Sensor Data Mining*), datos dados de acelerómetro de teléfono inteligente. La frecuencia de muestreo va desde los 10hz hasta los 20hz. Las características utilizadas para la identificación de la posición y la aceleración, de esta manera se determina la clase.

Las características de las arquitecturas son las siguientes, en el caso de CNN, contiene cinco capas convolucionales, se dividen los datos en segmentos de series temporales del mismo tamaño; LSTM, de igual forma se dividen los datos, de estos el 80%

es para entrenamiento y el 20% para pruebas, este tipo de red, mantiene la información en la memoria; *Convolution Long-Short Term Memory* (ConvLSTM), la convolución maneja características locales y LSTM tiene la correlación temporal de los datos del sensor. Como resultado de la experimentación se tiene que el mejor, fue CNN-LSTM, se presentan los puntajes, obteniendo un 98% de precisión.

Bi-LSTM Network for Multimodal Continuous Human Activity Recognition and Fall Detection (Li et al., 2020)

El modelo de red Bi-LSTM, con la adquisición de los datos es multimodal, se utilizan tres sensores IMU colocados en muñeca, cintura y tobillo, además un sistema radar FMCW (del inglés *Frequency-Modulated Continuous Wave radar*), de 16 personas (quince hombres y una mujer), las actividades monitoreadas son caminar, sentarse, agacharse, beber agua, simular una caída frontal. La información obtenida se procesa por separado, posteriormente se extraen las características y se realiza la fusión de los datos obtenidos. Se enlazan los datos del radar con cada uno de los acelerómetros posicionados en partes del cuerpo. En la clasificación se utiliza la red BI-LSTM que se compone de una capa de entrada y múltiples capas Bli-LSTM y una capa *softmax*, los datos de entrada son del sensor ubicado en la muñeca y radar que obtuvieron mejor resultado en datos continuos.

Human activity recognition using RGB-D sensors (Bagate & Shah, 2019)

La propuesta presenta el reconocimiento de actividad humana por medio de cámara de profundidad RBG-D, este tipo de dispositivos son económicos. Se captura imágenes con colores y al mismo tiempo profundidad, además estos elementos extraen características en forma de esqueleto 3D. Las características espaciales que se extraen son las uniones para el esqueleto contemplando los ejes “x”, “y” y “z”. Se emplea un clasificador CNN, se aplica comúnmente para el análisis de imágenes, para detección de objetos, clasificación y entre otros. El conjunto de imágenes evaluadas es capturado con el sensor antes mencionado, registrando actividades de siete personas, la red neuronal convolucional se entrenó con paquetes *keras* y *tensorflow*. La salida del clasificador obtuvo como resultado la siguiente precisión, esto se ilustra en la matriz de confusión, que se encuentra en la Figura 1.7.

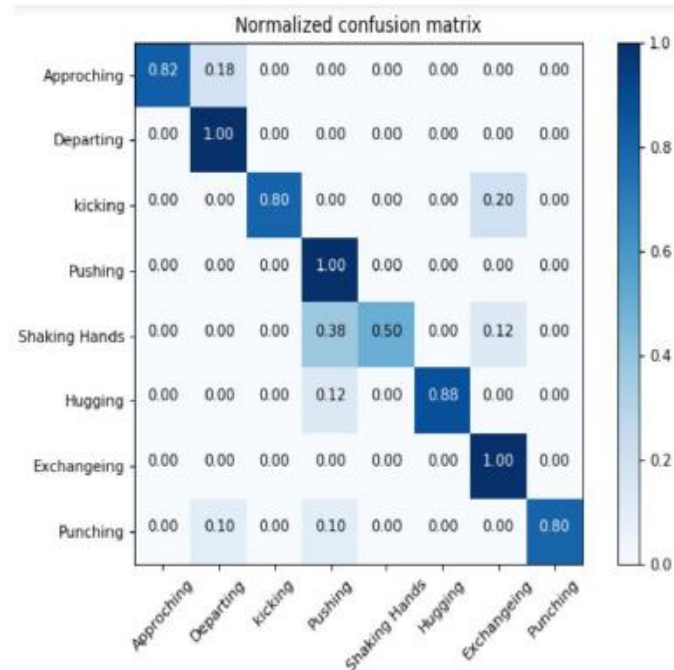


Figura 1.7 Matriz de confusión (Bagate & Shah, 2019)

Human Activity Recognition for Indoor Robotics: A Deep Learning Based Approach Using a Human Detection Stage (Luis et al., 2021)

En este trabajo por medio de datos de una cámara RGB, una arquitectura *Deep learning*, se analizaron tres tipos de estudios dos basados en imágenes y de esta manera extraer información de éstas, el tercero es un análisis desde la perspectiva de un esqueleto (unión de articulaciones y huesos). En los estudios de imágenes, en el primer caso se realizó un procesamiento de video para obtener un conjunto de fotogramas, se seleccionó una cantidad de 16 o 32 fotogramas, para después redimensionarlo, después de esto —pasa al proceso de extracción de características. El otro estudio es emplear la arquitectura YOLOv3 para recortar la sección de interés en la imagen. Para la identificación por medio del esqueleto que contiene las coordenadas 3D sobre las articulaciones. Después se evaluaron diferentes arquitecturas *mobileNetV1*, *mobileNetV2*, *ResNet18*, *ResNet50*, *ResNeXt*, *C3D*. Se hicieron dos categorías para el entrenamiento desde cero y pre entrenado, en el primer caso se obtuvo la exactitud más alta por parte de C3D y en modelo pre entrenado ResNet50 con 95% de exactitud.

Robot recognizing humans intention and interacting with humans based on a multi-task model combining ST-GCN-LSTM model and YOLO model (Liu et al., 2021)

Una de las aplicaciones actuales del reconocimiento de actividades humanas es la interacción humano robot, este trae beneficios ya que es capaz de monitorear a grupos vulnerables como personas mayores y/o con discapacidad. Las tres tareas más importantes de esta relación humana máquina, es reconocer la actividad, tomar decisiones a partir de la actividad y planificar la acción que realizará el robot.

Considerando anterior, se propone un modelo ST-GCN-LSTM (Redes Convolucionales de Gráficos Temporales Espaciales y Memoria a Largo Corto Plazo) para reconocer actividades y YOLOv3 para reconocer objetos que interactúan con las personas, identifica las coordenadas de los objetivos. Que combina datos de imágenes 2D y esqueleto 3D, en la Figura 1.8, se observan las características antes mencionadas.

El objetivo general del proyecto es apoyar a las personas y monitorear a personas cuando ocurran patrones extraños como una caída. Los datos recopilados son a través de un *Kinect v2*. Como resultados se obtuvo que la arquitectura YOLOv3 tuvo la capacidad de reconocer aceptablemente y la red de reconocimiento de actividad tiene una precisión de 86-89%, estos resultados son importantes para alimentar al robot para la relación entre humano y robots.

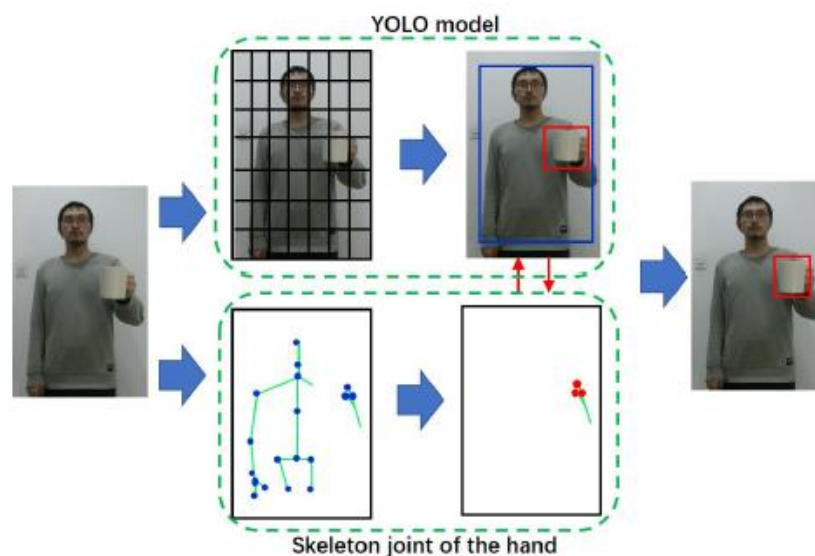


Figura 1.8 Sistema propuesto en (Liu et al., 2021)

Patient monitoring by abnormal human activity recognition based on CNN architecture (Gul et al., 2020)

El reconocimiento de patrones se emplea en problemas diarios, entretenimiento o aplicación médica, el enfoque de esta investigación consiste en visión artificial y el aprendizaje profundo para la identificación de actividades. El objetivo es clasificar actividades anormales tales como caída hacia atrás y enfrente, dolor de pecho, tos, desmayo, dolor de cabeza, tomar medicamentos y vómito. Para el desarrollo del proyecto se empleó CNN, específicamente YOLO, se requiere de una respuesta rápida en la inferencia, que se traduce en la monitorización en tiempo real de las actividades realizadas por las personas. Para la experimentación se creó un conjunto de datos de entrenamiento de las actividades mencionadas, las cuales fueron realizadas por ocho voluntarios, con repeticiones de tres veces por actividad, las secuencias de videos van de cuatro a cinco segundos.

La manera en que trabaja YOLO, es que las capas convolucionales extraen características de la imagen que recibe de entrada, las capas que están conectadas totalmente predicen las coordenadas donde se encuentra el objeto, como se observa en la Figura 1.9. Las métricas empleadas para evaluar el rendimiento son: *Precisión*, *recall*, *score F1* y exactitud, los resultados finales fueron que se reconoció aceptablemente el conjunto de actividades.

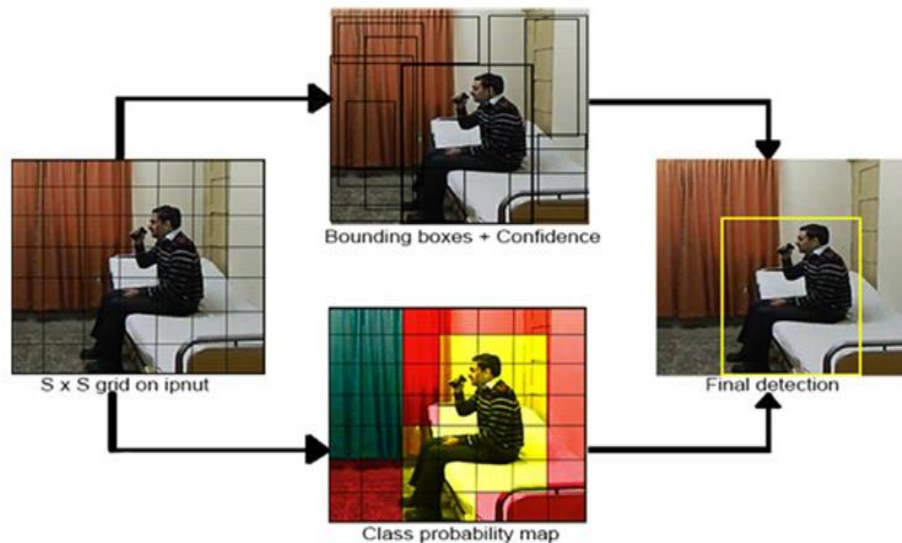


Figura 1.9 Operación del algoritmo YOLO

Activity recognition for ambient assisted living with videos, inertial units and ambient sensors (Ranieri et al., 2021)

La investigación aborda la problemática de reconocer actividades de la vida diaria concretamente en acciones domésticas, esto es parte de un trabajo futuro, para un proyecto de vida asistida por el ambiente. Por ejemplo, hogares inteligentes y robots de asistencia a una población específica para realizar un monitoreo de su vida diaria.

Una de las aportaciones es la creación de un conjunto de datos, con datos de cámaras (RGB y de profundidad), IMU (unidad de medida inercial) portátiles y sensores ambientales (temperaturas, luz y humedad. Al no existir ningún tipo de conjunto de datos, relacionado con el entorno doméstico, las actividades son: preparar una taza de té o sándwich y cereal, colocar la mesa, usar el teléfono, leer periódico, lavar platos y ordenar cocina. La posición de los sensores de visión fue en una ubicación fija, los sensores inerciales, se posicionaron en el brazo dominante y la cintura como se observan en la Figura 1.10. Otra de las contribuciones de este proyecto es el empleo del campo de *Deep Learning*, donde se empleó una arquitectura basada en movimiento, se combinaron módulos CNN y LSTM. Para el procesamiento las imágenes se cambian de tamaño, además se realiza aumento de datos, para los datos de sensores inerciales se sincronizan el muestro.



Figura 1.10. Posición de los sensores empleados en investigación (Ranieri et al., 2021)

1.2.4 Descripción de la solución

En la sección 1.1, se describe la problemática y la complejidad de esta. Por lo cual se plantea el desarrollo de un sistema con la capacidad de reconocer automáticamente las actividades desarrolladas por una persona, empleando un conjunto de sensores y modelos computacionales para la clasificación de actividades definidas en un entorno establecido.

1.3 Discusión

Para comprender de una manera precisa los conceptos claves para el estudio del reconocimiento de actividad humana, se realizó un análisis de definiciones e investigaciones recientes del tema. De tal manera, que se comprenda qué es el HAR, que como se menciona es la predicción de actividades desarrolladas por una o un grupo de personas, la finalidad dependerá de la aplicación del campo como es el área de la salud, deportes, industrial entre otros.

Además, se mencionan diversos trabajos del estado actual de esta área, los cuales dan respuesta a diversas interrogantes para el desarrollo de la propuesta de solución, se dan a conocer variables como: actividades estudiadas, sensores empleados, posición y número de sensores, así como técnicas y metodologías.

Uno de los aspectos relevantes que ha aportado el incremento de conocimiento es el uso de dispositivos móviles celulares, relojes inteligentes, anillos y entre otros elementos, debido a su avance tecnológico que permite el acceso a los sensores con los cuales se encuentra integrados en estos dispositivos, que reduce en primer lugar el costo de adquisición de sensores específicos, además en la actualidad son utilizados de una manera que no influyen en el desarrollo de actividades humanas. Los sensores más importantes son los de movimiento como acelerómetro, giroscopio y magnetómetro, pero algunos de estos cuentan con medición de ritmo cardiaco, presión arterial, oxígeno y entre otra información biométrica que podría ser empleada para la clasificación de actividades relacionadas directamente con la salud de las personas.

La adquisición de los datos inerciales, se logra a través de un enlace entre las aplicaciones desarrolladas para los dispositivos portátiles y un servidor donde son almacenadas. La frecuencia de muestreo que se sugiere o trabaja en la literatura es entre 10-50hz, condicionado a la capacidad computacional del elemento y tecnología de transferencia. Otro de los dispositivos que se emplea son los sensores de visión, cámaras RGB y RGB-D, que permiten adquirir datos sobre la interacción de las personas con el entorno, lo cual permite relacionar a los objetos con la posición o acción que se está desarrollando.

Los enfoques para la clasificación de las actividades humanas, van desde lo clásico como máquinas de vector soporte, KNN, *Random Forest* y entre otros, hasta arquitecturas que emplean redes neuronales. En el estado del arte, se expone que el uso de enfoques clásicos es frecuente para trabajar con datos de sensores, en el caso de las imágenes, se emplean redes neuronales preentrenadas, debido a que trabajan rápidamente y son ligeras.

Finalmente, se hace mención que el uso de diversos tipos de sistemas de adquisición (sensores y visión), en conjunto (multimodal), mejora el rendimiento del sistema, debido a que se obtiene mayor información de las partes del cuerpo que permiten una mejor caracterización.

CAPÍTULO 2

PROPUESTA DE SOLUCIÓN

En este capítulo, se detalla la propuesta de solución diseñada con el propósito de resolver la problemática descrita en el Capítulo 1. A continuación, se presentan los objetivos, alcances y limitaciones del trabajo. Además, la justificación y los beneficios asociados con la implementación del sistema. Por otra parte, se expone el modelado conceptual de la solución y una descripción detallada de sus elementos clave.

2.1 Planteamiento de la solución

El presente trabajo de tesis, propone el desarrollo de un sistema de identificación de situaciones de interés centradas en una persona, implementando la adquisición multimodal (celular, reloj inteligente y cámara RGB) y algoritmos inteligentes.

2.1.1 Objetivos

2.1.1.1 Objetivo general

Diseñar un sistema de identificación automático que utilice técnicas de visión artificial y reconocimiento de patrones para adquirir, preprocesar e interpretar información de al menos tres situaciones de interés realizadas por una persona.

2.1.1.2 Objetivos específicos

- Analizar y definir situaciones de interés a estudiar en esta investigación.
- Seleccionar un conjunto de sensores y/o cámara(s) para su posterior adquisición del movimiento del cuerpo humano respecto a las situaciones definidas.
- Modelar las situaciones de interés definidas.
- Definir los algoritmos para la identificación de las situaciones de interés.
- Integración del sistema.

2.1.2 Alcances y limitaciones

2.1.2.1 Alcances

- Identificar al menos tres situaciones de interés (se contemplaron las actividades, estar de pie, caminar, sentarse y caer).
- Definir las características de las situaciones de interés respecto a los sensores seleccionados.
- Adquirir los datos de al menos tres situaciones de interés realizadas por un mínimo de tres personas.
- Evaluar la utilidad y/o pertenencia de los conjuntos de datos reportados en el estado del arte.

2.1.2.2 Limitaciones

- Únicamente se realizará la identificación de las actividades seleccionadas.
- El sistema sólo funcionará para identificar las actividades una persona a la vez.
- El estudio no identificará patrones de comportamiento dentro de la rutina de una persona.
- No se garantiza que el sistema de reconocimiento de acciones funcione en tiempo real.

2.1.3 Justificación y beneficios

En la actualidad, el análisis de la actividad humana se ha aplicado en diversos campos con el objetivo de mejorar el bienestar de las personas. Esto puede ser el monitoreo, detección de situaciones de salud o de riesgo, hasta la autonomía de un grupo de individuos.

Las nuevas tecnologías tienen un papel muy importante en el desarrollo de este tema, debido a que en el presente se tiene la posibilidad de emplear sensores que se encuentran en dispositivos que se emplean día a día como los dispositivos móviles, los cuales permiten el desarrollo de las actividades de una manera natural.

Existen diversas técnicas y metodologías para la identificación de acciones o movimientos. Por este motivo se pretende desarrollar un sistema que emplee sensores de bajo costo y modelos computacionales, que pueda ser empleado en un entorno general.

2.2 Modelo conceptual de la solución

La solución propuesta es la creación de un sistema HAR compuesto de adquisición multimodal y modelos computacionales. Donde se trabajan con diferentes tipos de datos como lo son datos inerciales e imágenes, de esta manera se pretende aumentar la precisión de la clasificación. El diagrama del modelo conceptual en la Figura 2.1.

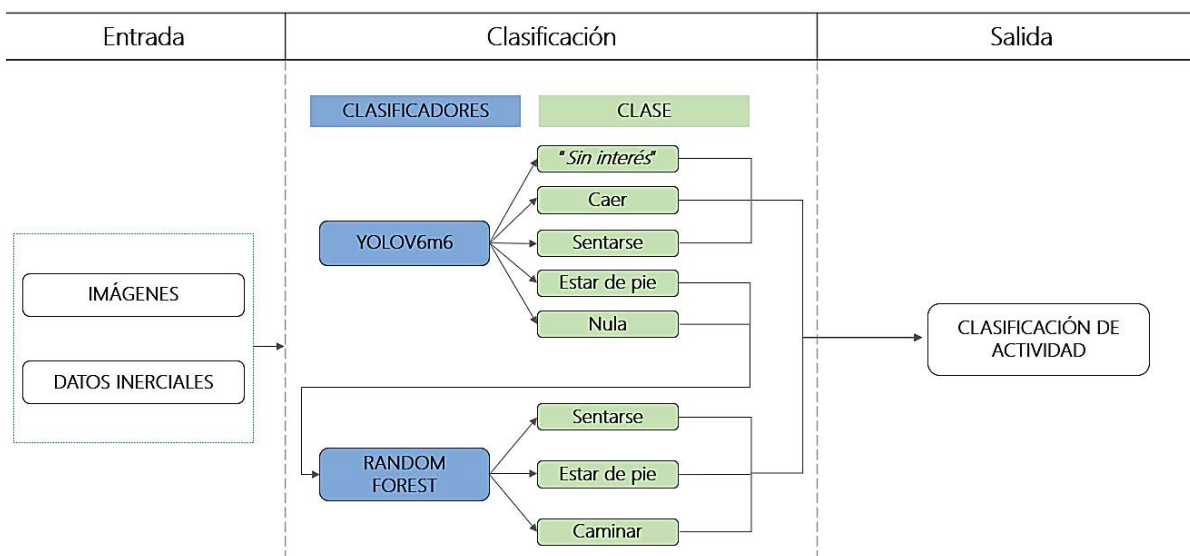


Figura 2.1 Modelo conceptual de la solución

2.3 Descripción de los elementos del modelo conceptual

Los elementos que conforman al modelo conceptual son los siguientes:

Entrada: Los datos que ingresan al sistema son adquiridos por sensores inerciales acelerómetro, giroscopio y magnetómetro, estos se encuentran en un celular móvil y en un reloj inteligente. Los cuales son portados por una persona en el bolsillo derecho de su pantalón y el reloj inteligente en la mano izquierda. Asimismo, se adquieren imágenes por medio de una cámara RGB, que puede ser colocada frontalmente o lateralmente.

Clasificación: Para la clasificación y detección de la persona en el caso de las imágenes se emplea del algoritmo YOLOV6, debido a que tienen un rendimiento adecuado al emplearlo en tiempo real y se encuentra reentrenado con la clase “*persona*”. Para la clasificación de los datos inerciales, se utiliza el algoritmo *random forest*, este trabaja con una gran cantidad de datos como los que ingresan al sistema que provienen de dos dispositivos.

Salida: Con un conjunto de reglas previas en la clasificación, se obtiene la salida de cuál es la actividad desarrollada. Esto es debido a que por medio de las imágenes y el algoritmo elegido no es capaz de determinar precisamente la actividad ambulatoria, en este punto entra la regla con mayor peso al clasificador *random forest* con los datos inerciales.

2.4 Variables/ criterios de control

Para un correcto funcionamiento del sistema propuesto se tienen que tomar en cuenta los siguientes puntos:

La posición de los sensores de adquisición: Como previamente es mencionado los dispositivos móviles deben ser colocados en partes específicas del cuerpo, las razones son que dentro de la literatura se expresa que son los puntos donde las lecturas de los datos presentan menos variaciones, en el caso del reloj inteligente es colocado conforme a los requisitos de su proveedor.

Las cámaras deberán ser colocadas de forma lateral o frontal, el dispositivo tiene que tener visibilidad de la persona a estudiar.

Las versiones de software de los dispositivos móviles: Además de la correcta colocación se necesitan que las versiones de software sean compatibles, debido a que las aplicaciones desarrolladas permiten una adquisición rápida y en conjunto. En el caso del celular *Android* 8 o superior y el reloj inteligente *Wear Os* 3.

Frecuencia de muestreo: La adquisición de los datos es en una frecuencia de 30 hz cada sensor.

Las actividades a estudiar: Por medio de un análisis realizada en una primera fase de la investigación se determinaron las siguientes actividades que son clasificadas en dos categorías, tomando en cuenta lo expresado en el estado del arte, que son ambulatorias: caminar y caerse; no ambulatorias sentarse y estar de pie.

Condiciones de algoritmos: Como se menciona anteriormente, cada uno de los modelos trabajan con un tipo de datos, los cuales generan dos salidas al final del proceso uno de los algoritmos toma un peso mayor, dependiendo de si es o no una actividad ambulatoria.

CAPÍTULO 3

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DEL SISTEMA

En este capítulo se detalla el análisis realizado para el diseño del sistema de reconocimiento automático de actividades propuesto, se describen las fases que se desarrollaron para la implementación del sistema final.

3.1 Metodología de propuesta de solución

Para el desarrollo de la solución y creación del sistema se llevaron a cabo las siguientes etapas presentadas en el diagrama de la Figura 3.1. Donde las fases o etapas consisten en:

Adquisición de datos: Recopilación de información para la creación de dataset propio, que cumpla con los objetivos planteados, conformado de dos tipos de datos: imágenes y datos inerciales.

Preprocesamiento: Preparación de datos recolectados, limpieza de datos, etiquetado de clases, análisis estadístico, a fin de obtener mejores resultados en el entrenamiento.

Entrenamiento: Selección de algoritmos para clasificación de las actividades de interés. Realizando diversas pruebas con el propósito de obtener un sistema que cumpla con los requerimientos propuestos.

Validación: Con las métricas correspondientes se evalúan los resultados de los entrenamientos para hacer comparaciones, analizando el rendimiento de cada caso de prueba.

Implementación: Creación de una interfaz gráfica que permite ejecutar el sistema completo para la detección de actividad de manera local y en tiempo real.

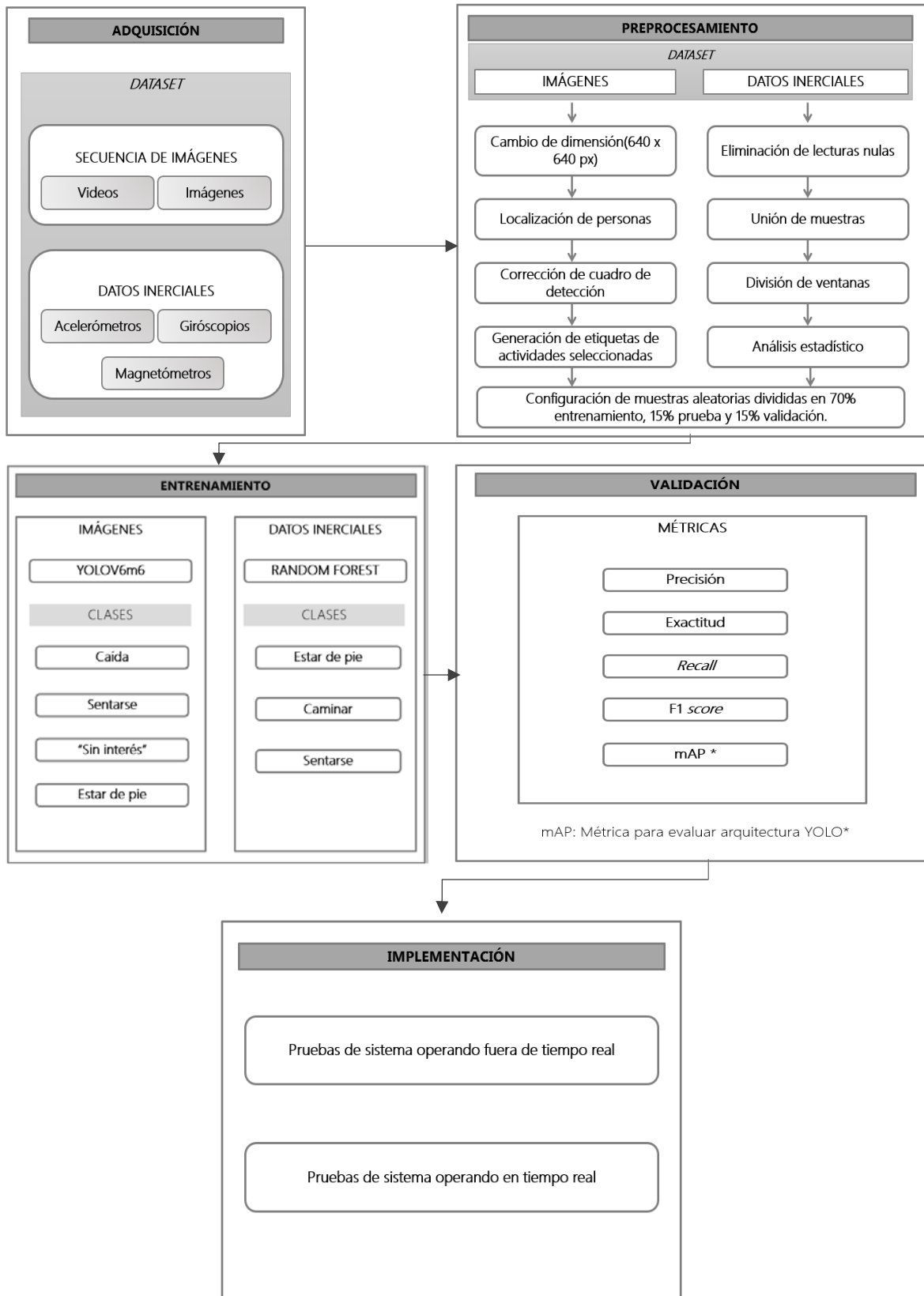


Figura 3.1 Metodología empleada para el diseño de solución

3.1.1 Adquisición de datos

Como previamente se mencionó, se trabajó con dos tipos diferentes de sensores: visión e inerciales. Con estos se realizó un conjunto de datos propio recolectando la información de ocho personas diferentes, esto con la finalidad de tener variedad de datos. En la Tabla 3.1, se presentan las características del conjunto de datos propio.

Las imágenes recopiladas en un primer muestreo se consiguieron un total de 39 videos, después para lograr una variedad de fondos y personas se agregaron imágenes extras. En el caso de los sensores inerciales se añadieron cinco mil datos adicionales de las actividades, con la peculiaridad de que las personas presentaban diversas alturas.

Tabl.1 Características de datos propio

Conjunto de datos propios	
Tipos de sensores	Visión y sensores inerciales (Acelerómetros, giroscopios y magnetómetros).
Dispositivos	Cámara RGB, celular y reloj inteligente.
Tipo de datos	Videos y datos numéricos (información inercial ejes "x", "y" y "z").
Posición de sensores	Cámara: vista lateral y frontal. Sensores inerciales: <ul style="list-style-type: none"> • Celular, bolsillo derecho. • Reloj, mano izquierda.
Actividades	Estar de pie, caminar, sentado, caída y "Sin interés".
Personas	Tres personas diferentes para videos Cinco personas diferentes para datos inerciales
Total, de archivos	39 videos con su respectivo muestreo inercial y 5 mil datos adicionales de datos inerciales.

3.1.1.1 Aplicación de adquisición de datos inerciales

La recolección de los datos, se ejecutó a través de una aplicación creada en *Python* en la computadora y otra realizada en *Android Studio*. La condición en que operan es que la computadora funciona como servidor para recolectar los datos en una frecuencia de 30 hz, los clientes son los dispositivos portátiles, celular y reloj inteligente. La conexión entre los

dispositivos es por medio del protocolo TCP (*Transfer Control Protocol*), se utiliza la IP de los elementos para la comunicación. En las Figuras 3.2 y 3.3, se muestra las aplicaciones desarrolladas en los dispositivos portátiles y en la computadora.



Figura 3.2. Aplicación móvil *Android* y *Wear Os*, para recolección de datos

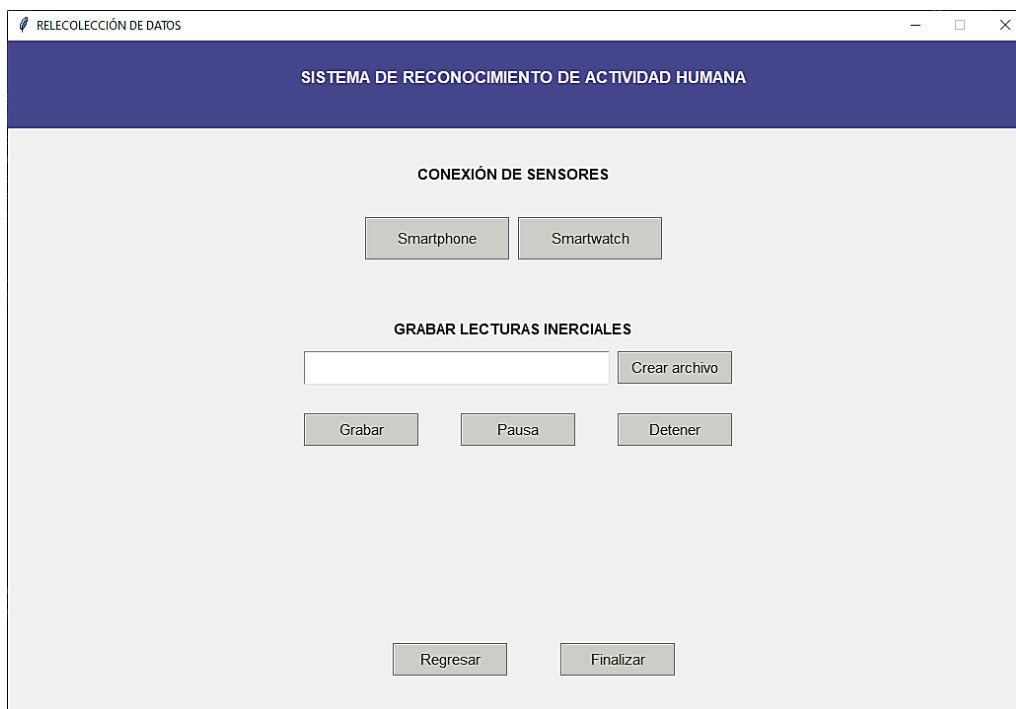


Figura 3.3. Interfaz para recepción de datos en PC

3.1.1.2 Ejemplo de muestra recopiladas

Enseguida se muestra un ejemplo del protocolo de muestras, en primer lugar, el usuario deberá portar los dispositivos portátiles, el reloj ubicado en la mano izquierda estas son las indicaciones del proveedor (*Galaxy Watch4 Bluetooth (44 Mm) | SM-R870NZKAMXO | Samsung México, n.d.*), el celular es colocado en el bolsillo derecho de la persona, tomando como referencia el estado del arte. Las cámaras se ubican en dos posiciones diferentes tratando de obtener la mayor información del cuerpo de la persona. Posteriormente la persona realiza la actividad un total de tres veces, recolectando de esta forma el video y 500 datos inerciales por actividad. En la Figura 3.4, se muestra la colocación de los sensores y dos vistas diferentes del sensor de visión.



Perspectivas diferentes con cámaras RGB

Figura 3.4. Ubicación de sensores de visión

3.1.2 Preprocesamiento de los datos

Para la etapa de entrenamiento, es recomendable efectuar un preprocesamiento previo con la información recopilada por los sensores, en cada caso es necesario aplicar criterios diferentes porque la información está compuesta por dos tipos de datos: imágenes y datos inerciales. Esto permite que el conjunto sea adecuado para el entrenamiento, esperando que los datos se encuentren en condiciones óptimas y homogéneas, para obtener un mejor resultado en la fase de entrenamiento.

3.1.2.1 Preprocesamiento de las imágenes

Para las imágenes, el primer paso fue dividir cada uno de los videos adquiridos en *frames*, se procedió a una limpieza de datos eliminando todos aquellos en los que no apareciera una persona en el cuadro de imagen y similitud de información, que un conjunto de imágenes contenga la misma posición sin cambios. Una vez realizado esto se obtiene una muestra representativa y funcional para el análisis de actividades de una persona.

Posteriormente, se hizo un cambio de dimensión a todas imágenes, ajustándolas a una resolución de 640 x 640 pixeles, según lo recomendado en la literatura (*GitHub - Ultralytics/Yolov5: YOLOv5 🚀 in PyTorch > ONNX > CoreML > TFLite*, n.d.). Después se llevó a cabo el etiquetado de las clases correspondientes (estar de pie, sentarse y caer), con la finalidad de optimizar la tarea de detección de los elementos de interés, se empleó la arquitectura YOLOv5m6, en la Figura 3.6, se observa la línea de código para la detección de elementos. Donde los elementos que lo componen son, el programa *detect.py*, la carpeta que contiene las imágenes, la clase, guardar archivo *.txt*, y modelo.

```
python detect.py --source "E:\nuevo_train\sentarse" --classes 0
--save-txt --max-det 1 --weights yolov5m6.pt --name sentarse
```

Figura 3.5. Detección de elementos por medio de yolov5

Como resultado se consiguieron los datos de localización de la persona, en formato YOLO para su entrenamiento. Estos datos se guardan en un archivo de texto (.txt), las componentes de este son: la etiqueta de la clase, la coordenada “x” y “y” punto centro de la caja predictora, por último, el ancho (w) y largo (h), de la caja predictiva en la Figura 3.6, se muestra un ejemplo del formato.

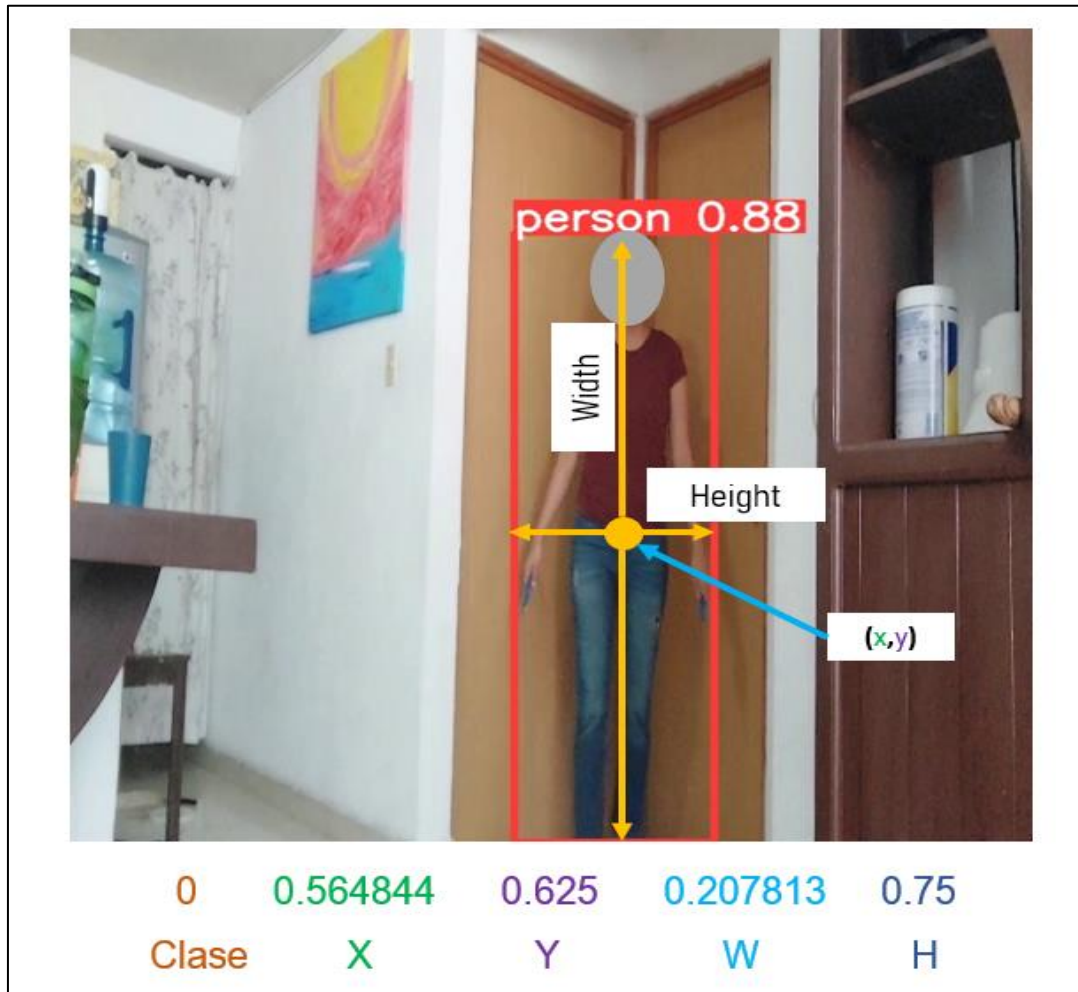


Figura 3.6. Información generada por YOLOV5(localización de personas)

Se realizó una revisión de las imágenes etiquetadas para verificar que la localización de las personas fuera correcta. En los casos donde se detectaron errores, se utilizó la herramienta *LabelImg* (*GitHub - Heartexlabs/LabelImg: LabelImg Is Now Part of the Label Studio Community. The Popular Image Annotation Tool Created by Tzutalin Is No Longer Actively Being Developed, but You Can Check out Label Studio, the Open Source Data Labeling Tool for Images, Text, Hypertext, Audio, Video and Time-Series Data.*, n.d.), la cual

trabaja con *Python 3.6*. Permite la edición de las cajas delimitadoras, la información adecuada se guarda nuevamente en el archivo de texto, previamente mencionado. En la Figura 3.7, se observa la corrección de un error de localización y la línea de ejecución de la herramienta. Finalmente, se re-etiquetaron cada una de las imágenes para asignar su clase correspondiente, para su posterior entrenamiento.

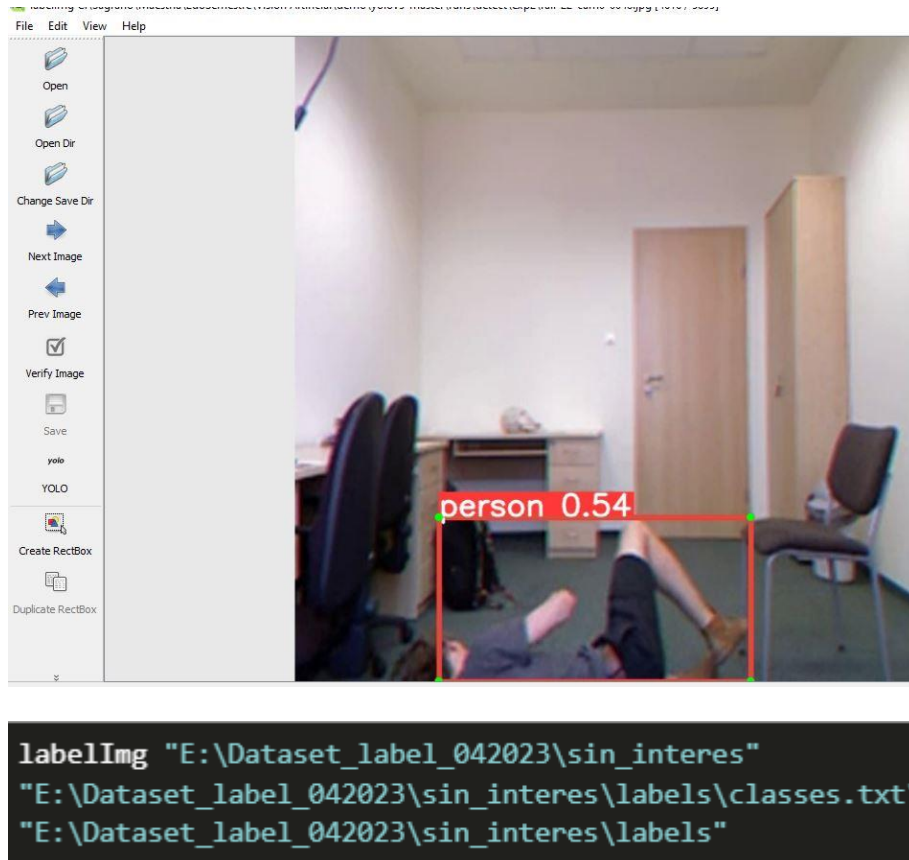


Figura 3.7. Herramienta LabelImg

3.1.2.2 Preprocesamiento de datos inerciales

En cuanto a los elementos inerciales, se eliminaron aquellas lecturas que no aportan características relevantes del fenómeno, como por ejemplo las primeras y últimas mediciones de la muestra. La información contenida en estos datos son los ejes “*x*,” “*y*” y “*z*”, de cada uno de los sensores acelerómetro, giroscopio y magnetómetro. Después de este proceso se realizó la unión de todas las muestras obtenidas.

En lo expuesto en el estado del arte, comúnmente los datos de señales se trabajan por medio de ventanas, esta técnica permite recopilar un rango de muestras para obtener información de un fragmento determinado. En este preprocesamiento se utilizan ventanas deslizantes los datos son divididos por tamaños fijos sin tener espacios entre ventanas.

La división del tamaño de las ventanas es de tres segundos, cinco segundos, siete segundos y diez segundos, tomando como referencia el estado del arte (Ferrari et al., 2021), en futuras menciones se utilizará la letra “s” para referirse a segundos. Por último, se realizó un análisis estadístico de los datos (Weiss, 2019), de esta forma se obtuvo un mayor número de características que permiten la correcta clasificación.

3.1.3 Entrenamiento de algoritmos

Condiciones de entrenamiento

El sistema de reconocimiento de actividad humana propuesto, consiste en un conjunto de modelos computacionales cada uno para cada tipo de datos, para las imágenes se usa la arquitectura YOLO (*You Only Look Once*) que consiste en un modelo de red neuronal convolucional, para los datos inerciales el algoritmo *Random Forest* brindó mejores resultados en la clasificación.

Para la experimentación de cada una de las pruebas, se desarrolló un programa en *Python* para seleccionar de manera aleatoria los elementos que conforman el entrenamiento, el cual se dividió en:

1. Entrenamiento 70%.
2. Test 15%.
3. Validación 15%.

Para el conjunto de imágenes, como anteriormente se mencionó se empleó la arquitectura YOLO, para la localización y clasificación de las actividades de interés. La entrada del modelo son las imágenes y los archivos de texto que contienen las coordenadas. Para el entrenamiento de esta arquitectura se siguen los siguientes pasos: creación de un ambiente *conda*, activación de ambiente, construir el archivo *yaml* esto

dependiendo de la versión de YOLO, por último, se construye la línea de código correspondiente, en la Figura 3.8, se observan las líneas de las versiones empleadas.

Los datos inerciales de igual forma se fragmentaron en tres conjuntos, recordando que los datos fueron adquiridos de dispositivos móviles (celular y reloj inteligente), estos fueron parte del entrenamiento del modelo *Random Forest*, en *Python* con la librería *SkLearn*.

Los parámetros utilizados para el entrenamiento del modelo inercial (*Random Forest*) fueron aquellos que se proporcionan por defecto, como lo son la cantidad de árboles que es igual a 100, criterio “*gin*” para la calidad de la división, la profundidad se expande hasta que las hojas sean puras y entre otros parámetros (*Sklearn.Ensemble.RandomForestClassifier* — *Scikit-Learn 1.3.0 Documentation*, n.d.) .

Yolov5m6

```
python train.py --img 640 --epochs 25 --data data/ssgv5.yaml --weights
weights/yolov5m6.pt --batch-size -1 --name 3clasesv5
```

Yolov6m6

```
python tools/train.py --data-path data/ssgv6o.yaml --conf-file
./configs/yolov6m6_finetune.py --epochs 25 --eval-interval 1 --name
3classbgY6 --batch-size 12
```

Yolov7

```
python train.py --weights yolov7.pt --cfg
cfg/training/dock_yolov7_sagrario.yaml --device 0 --batch-size 7 --data
data/dock_sgg.yaml --img 640 640 --name 3cv7_50 --epochs 50
```

Figura 3.8. Líneas de código para entrenamiento de arquitectura YOLO

3.1.4 Evaluación

Métricas

Para la evaluación del sistema creado, en la literatura se emplean métricas en general para medir el rendimiento de los algoritmos de aprendizaje automático, una de las principales herramientas que se emplean es la matriz de confusión, a partir de esta se pueden calcular métricas. Debido a que en esta se comparan los verdaderos positivos, falsos negativos, falsos positivos y verdaderos negativos. En la Figura 3.9, se encuentra un ejemplo de la matriz de confusión.

		Clase Asignada	
		Positivo	Negativo
Clase Actual	Positivo	TP	FN
	Negativo	FP	TN

Figura 3.9. Matriz de confusión (Jukes, 2018)

Los autores en el estado del arte, definen los siguientes parámetros:

- Verdaderos positivos (TP) número de objetos positivos clasificados como tales.
- Verdaderos negativos (TN) número de objetos negativos que se clasificaron como negativos.
- Falsos positivos (FP) número de objetos negativos que se clasificaron como positivos.
- Falsos negativos (FN) número de objetos positivos clasificados como negativos.

Las métricas que se utilizan son precisión, exactitud, *F1 score*, *recall* y *mAP*.

Precisión: Es la relación entre las predicciones clasificadas correctamente como positivas y el número total de predicciones correctas realizadas (se incluyen negativas y positivas). Para medir la calidad del modelo.

$$\text{Precisión} = \frac{TP}{TP + FP}$$

Exactitud: Consiste en la proporción entre las predicciones correctas y el número total de predicciones, para obtener el porcentaje de casos donde el modelo fue correcto.

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall: Número de elementos identificados correctamente como positivos del total de positivos verdaderos.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 score: Esta evaluación combina las medidas de precisión y sensibilidad en un solo valor, para comparar el rendimiento combinado de dichas métricas. El valor más alto en F1 es 1.0 esto indica que la precisión y recuperación son perfectas, si es más bajo la precisión disminuye.

$$F1 = 2 \cdot \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

mAP: Esta métrica se utiliza para la evaluación de modelos de detección de objetos, como en el caso del modelo YOLO utilizado en este trabajo. Esta se obtiene del promedio de la precisión de cada clase y después promediando varias.

$$mAP = \frac{\sum_{q=1}^Q \text{AveP}(q)}{Q}$$

3.1.5 Implementación del sistema

Para facilitar la operación del sistema de reconocimiento de actividad humana propuesto, se desarrolló una interfaz gráfica que permite al usuario interactuar con diversas opciones de dispositivos de adquisición y modelos, para identificar las actividades preestablecidas en este trabajo. La interfaz consta de la integración de los modelos inerciales y de visión, los cuales se detallarán en el capítulo 4, esta aplicación fue creada en *Python* con la librería

Tkinter, donde también se requirieron de diversos paquetes (para la comunicación de sensores, cargar modelos de aprendizaje entre otros.)

La manera en la que trabaja la interfaz es:

1. Fuera de tiempo real: donde se agregan los datos imágenes o datos inerciales, se inicia el sistema para reconocer la actividad, como se observa en la Figura 3.10.
2. En tiempo real: se conectan los dispositivos que se tengan disponibles, como cámara y/o dispositivos inerciales (celular y reloj inteligente), posteriormente el usuario elige los elementos que desee utilizar e inicia el sistema, en la Figura 3.11, se presenta la pantalla correspondiente.

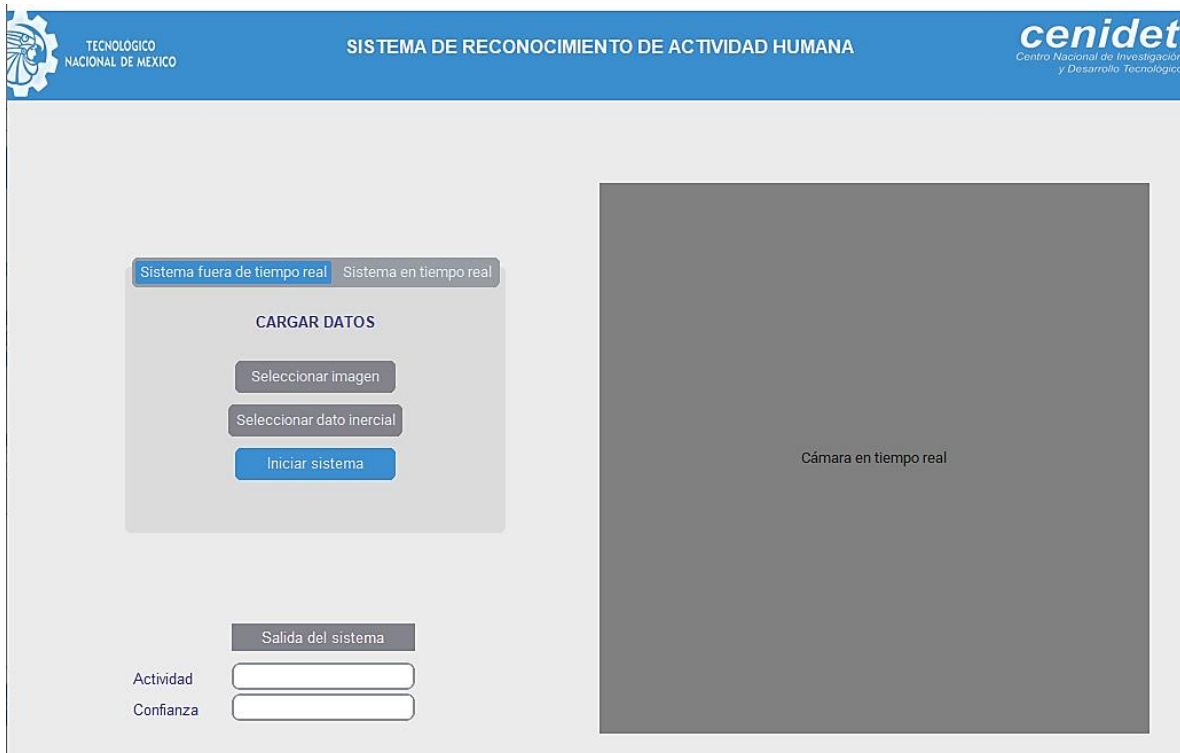


Figura 3.10. Interfaz gráfica desarrollada, pestaña fuera de tiempo real

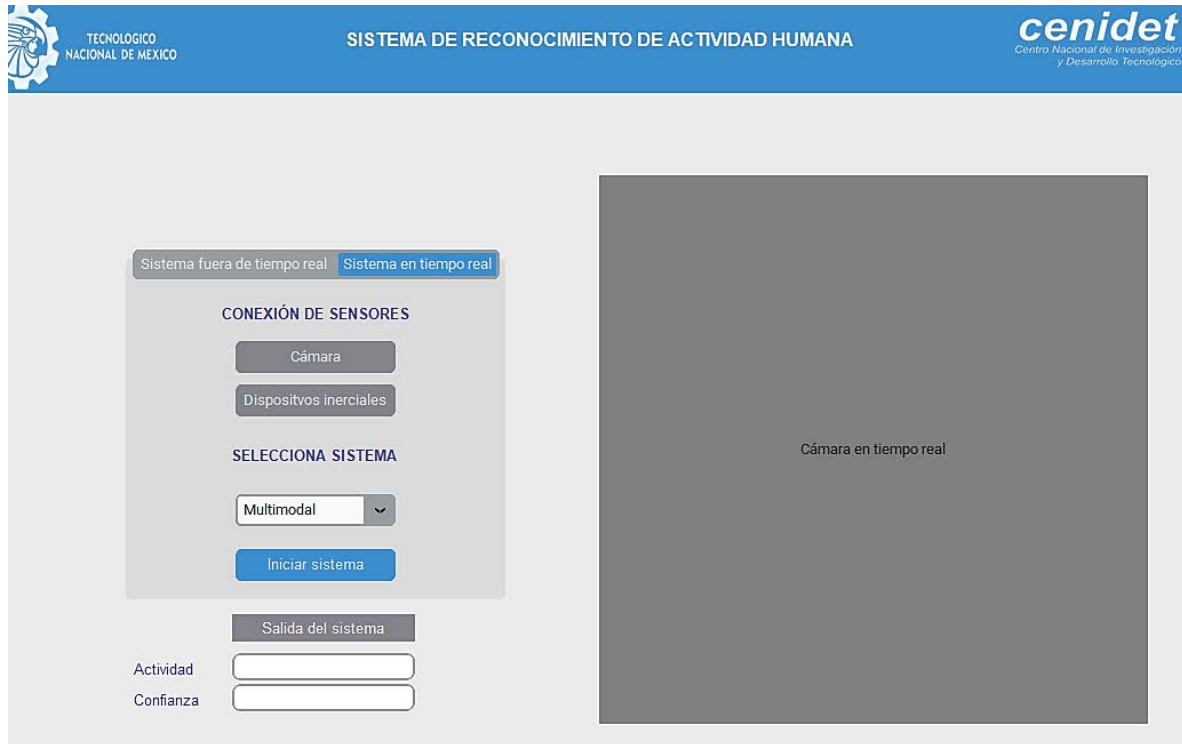


Figura 3.11. Interfaz gráfica desarrollada, sistema en tiempo real

CAPÍTULO 4

VALIDACIÓN, EXPERIMENTACIÓN Y ANÁLISIS DE RESULTADOS

En esta sección se describen las pruebas y resultados obtenidos como parte del proceso de experimentación para cumplir cada uno de los objetivos propuestos. A continuación, se detallan los pasos del análisis efectuado y la conclusión de los casos presentados.

4.1. Validación y experimentación

4.1.1 Prueba 1

Comparación de algoritmos para la clasificación de actividades de interés realizadas por una persona, con tres versiones diferentes de YOLO para imágenes y diversos clasificadores clásicos para los datos inerciales.

Objetivos: Analizar el rendimiento de los algoritmos que permitan una precisión alta.

Para esta prueba, se entrenaron las tres últimas versiones de YOLO (Yolov5m6, Yolov6m6 y Yolov7), con las configuraciones previamente mencionadas para seleccionar la de mejor resultado, así como la clasificación de los datos inerciales con algoritmos que se emplean con mayor frecuencia en el estado del arte, se utilizan diferentes tamaños de ventanas. En las Tablas 4.1 y 4.2, se muestran los resultados obtenidos.

Como resultado de esta prueba, se concluye que la arquitectura YOLO, presentó resultados altos para la detección y clasificación de actividades en sus tres versiones, de las cuales dos versiones (Yolov5 y yolov6), presentaron un *mAP* por arriba del 0.90. En este caso se seleccionó la arquitectura yolov6m6, que obtuvo un *mAP* 0.943, esta métrica permite conocer el rendimiento del detector de objetos.

Tabla 4.1 Resultados de clasificación imágenes propias

	CLASIFICADORES		
	yolov5m6	yolov6m6	yolov7m6
Métrica			
<i>Precision</i>	0.892	0.909	0.818
<i>Recall</i>	0.922	0.890	0.911
<i>F1 score</i>	0.900	0.900	0.810
<i>mAP0.5</i>	0.925	0.943	0.872

Por otra parte, en el caso de los datos inerciales, el algoritmo que mayor precisión obtuvo es *Random Forest*, con una ventana de 5 segundos, para las tres clases propuestas para este tipo de datos (estar de pie, caminar y sentarse). La principal observación, es capaz de diferenciar entre estar de pie y caminar de una manera acertada con 96.85 % de precisión total de las clases. En la Tabla A.1 en Anexo A, se encuentran una muestra de los resultados obtenidos para YOLOV6.

Tabla 4.2. Resultados de clasificación de datos inerciales (métrica precisión), divididas en ventanas deslizantes de segundos.

Clasificadores	Ventana deslizante			
	3s	5s	7s	10s
<i>Naive Bayes</i>	85.54	86.69	87.21	88.31
Perceptrón Multicapa	93.96	93.71	91.89	89.48
Máquina de vector soporte	56.16	53.13	53.12	57.36
<i>KNN (K- Vecinos cercanos)</i>	93.10	84.73	75.02	65.42
<i>Random Forrest</i>	89.78	96.85	87.07	84.26
<i>C 4.5</i>	93.37	91.89	89.07	89.39

4.1.2 Prueba 2

Probar sistema completo con la elección de modelo YOLOV6m6 y *Random Forest* con una ventana deslizante de 5s, realizado el reconocimiento de las situaciones de interés: estar de pie, caminar, sentarse y caer, la operación de este sistema no es en tiempo real.

Objetivo: Identificación de situaciones de interés

Se hicieron las pruebas del sistema completo, ingresando imágenes del conjunto de prueba. Las cuales son imágenes distintas al entrenamiento para comprobar la clasificación del sistema, los datos inerciales que se agregan con la ventana de 5s. En la Figuras 4.1-4.5 se observa las métricas de YOLOV6m6. Para los datos inerciales se aprecia la matriz de confusión en Figura 4.6, y en la Tabla 4.3 las métricas evaluadas.

Los resultados obtenidos, presentan que el sistema reconoce las actividades definidas de una manera correcta, empleando las imágenes y datos inerciales correspondientes a la secuencia de acciones especificada. En el caso de los datos inerciales con una precisión del 0.96 que es un complemento del modelo de imágenes que tiene un *mAP* 0.943. En el anexo A, específicamente Tabla A.2 se encuentran las grafico de la detección y clasificación del modelo YOLOV6.

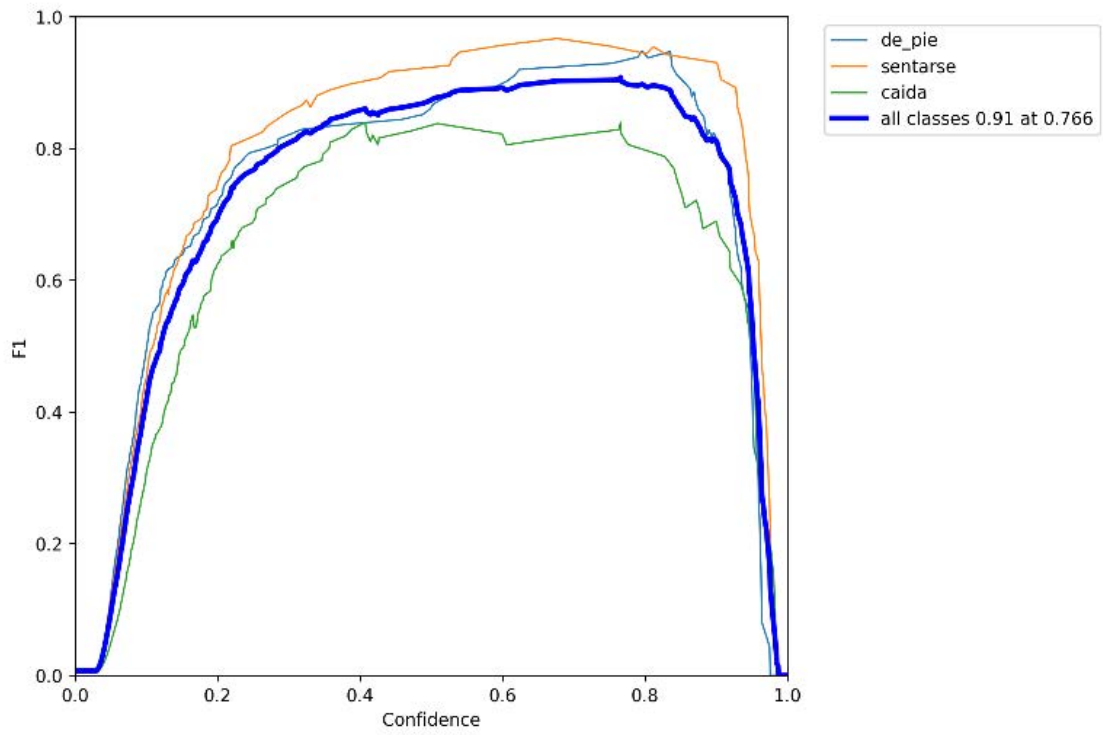


Figura 4.1 Métrica F1-confidence

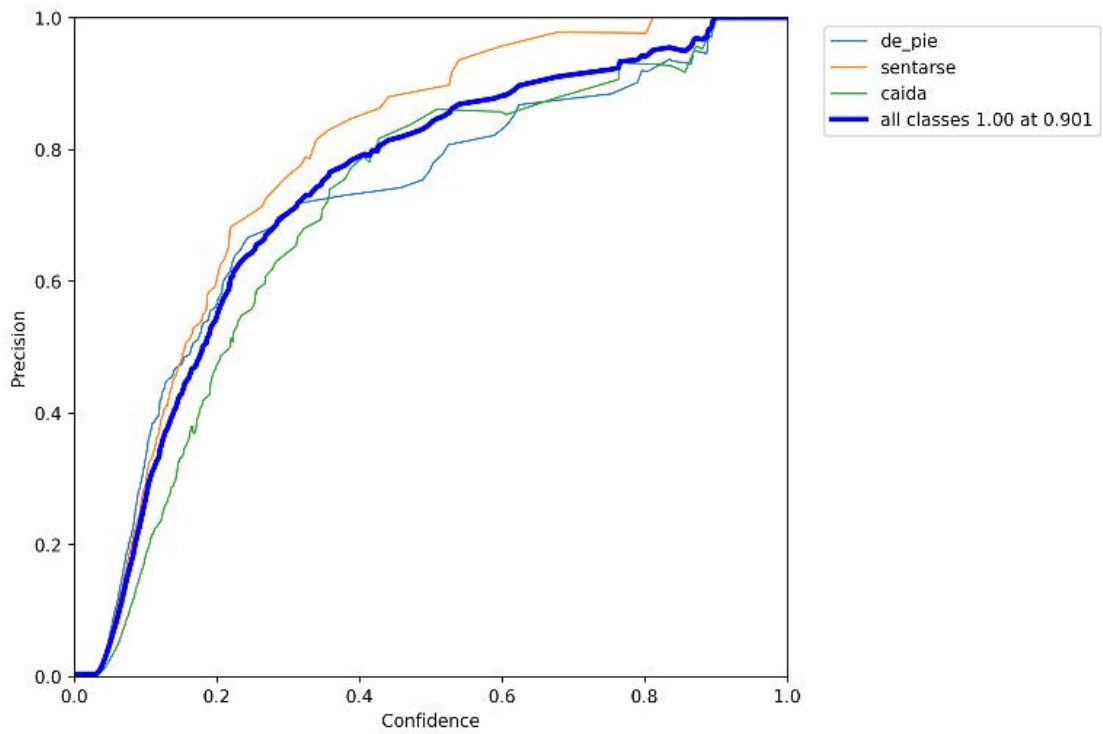


Figura 4.2 Precision- confidence

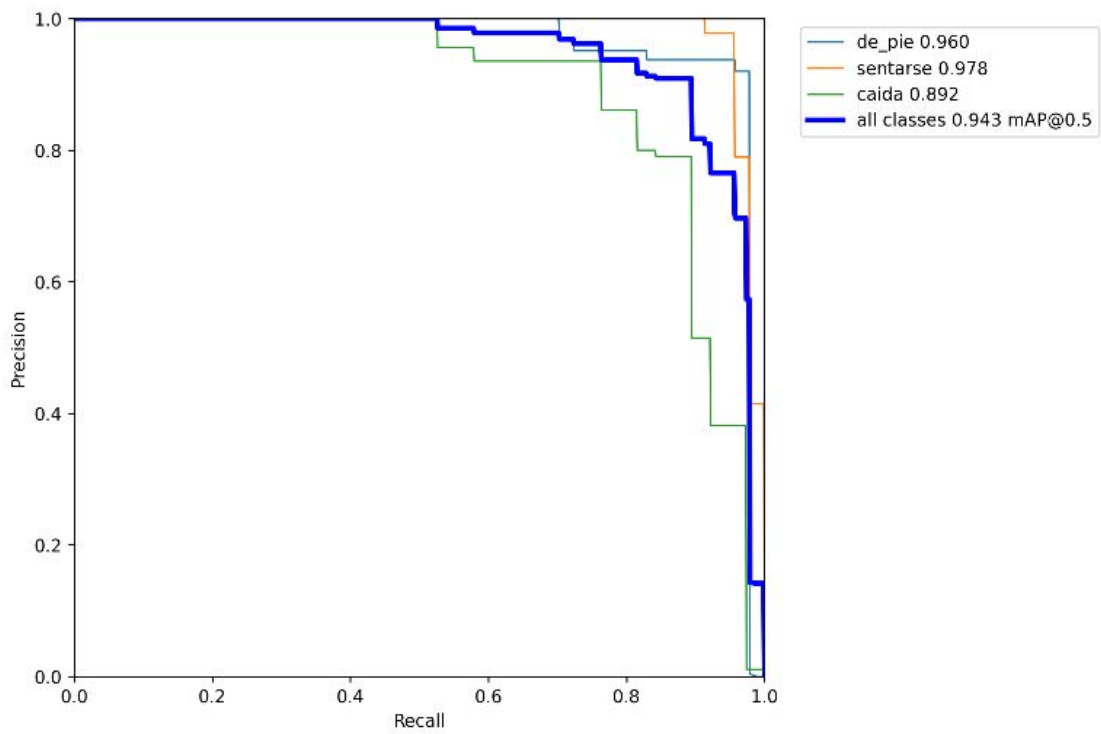


Figura 4.3 Métrica *Precision-Recall*

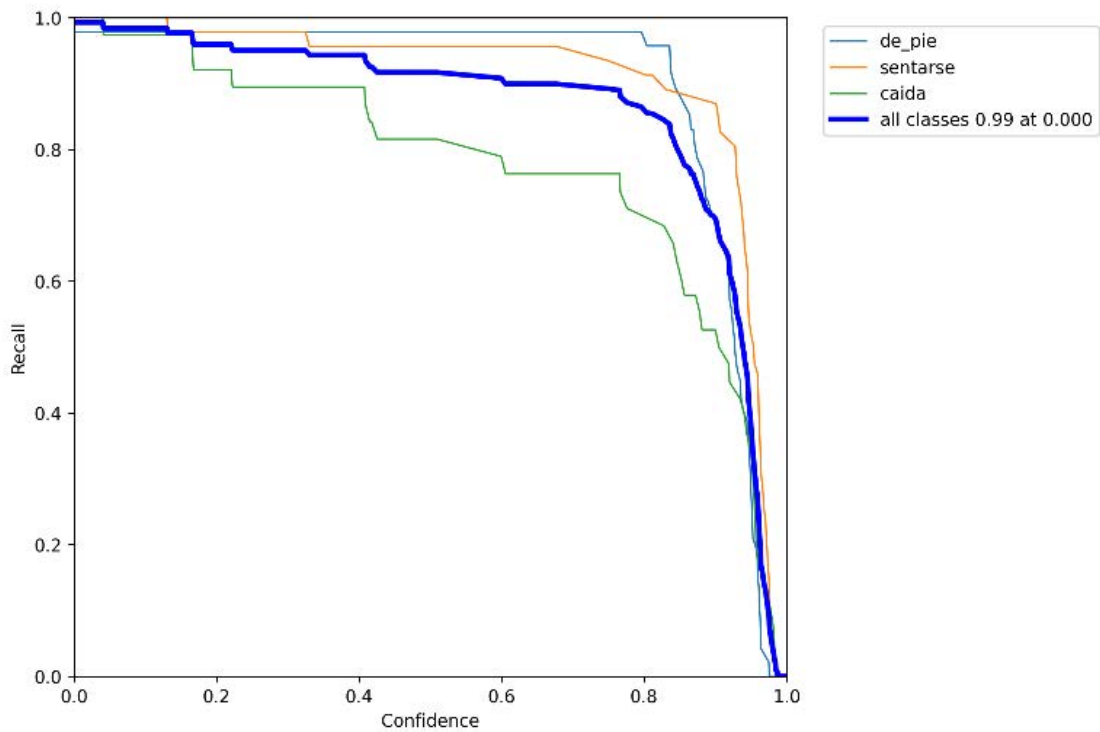


Figura 4.4 Métrica *Recall-Confidence*

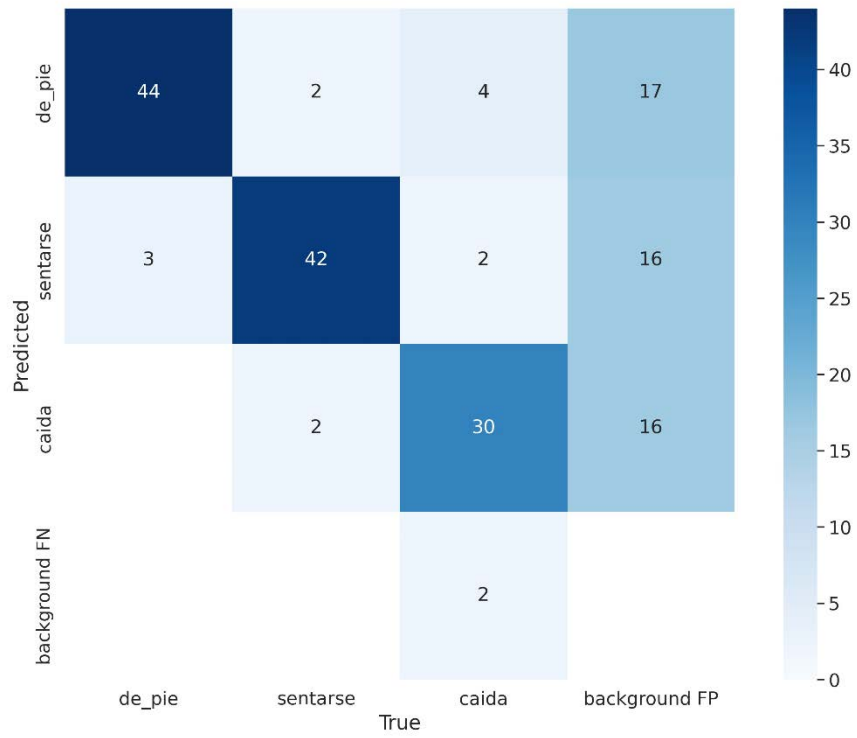


Figura 4.5 Matriz de confusión Yolov6 Tres clases

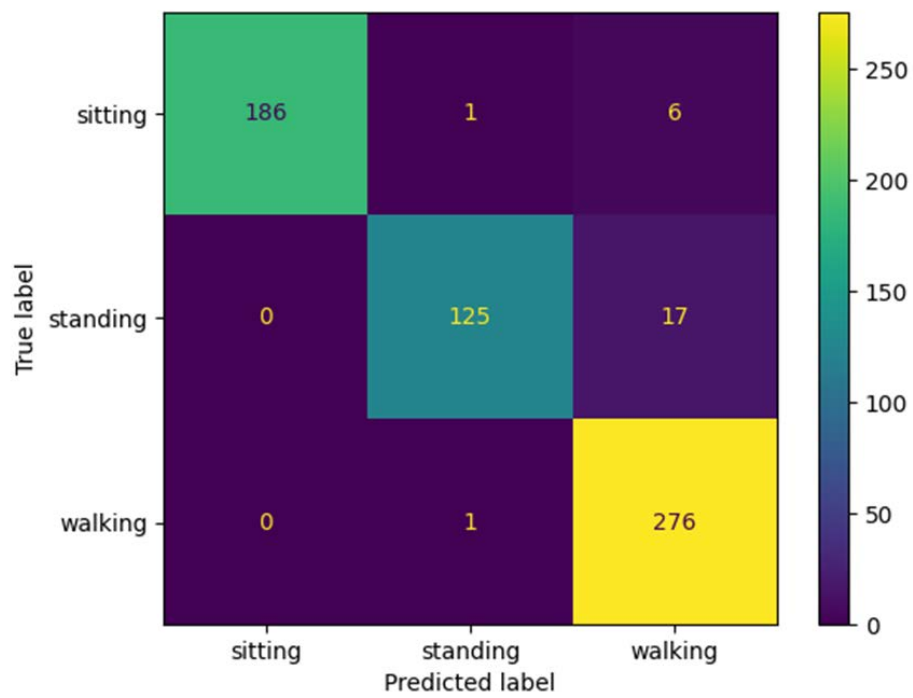


Figura 4.6 Matriz de confusión Random Forest

Tabla 4.3 Métricas evaluadas en modelo *Random Forest*

Clasificador	Ventanas deslizantes			
	<i>Precision</i>	<i>Accuracy</i>	<i>Recall</i>	<i>F1 Score</i>
Random Forest	0.9685	0.9647	0.9624	0.9622

4.1.3 Prueba 3

Adición de una clase para aumentar la precisión de la clasificación, cuando la actividad realizada no corresponde a las clases: estar de pie, caminar, sentarse y caer.

Objetivo: Evaluar sistema al agregar una clase adicional para mejorar la precisión.

En una primera evaluación se ingresaron imágenes donde las actividades no eran iguales a las previamente clasificadas, la respuesta del modelo de imágenes fue la siguiente:

- El sistema no realiza la clasificación, esto es correcto debido a que el algoritmo no fue entrenado para la nueva acción de entrada.
- El sistema clasificó la actividad, dentro de las categorías: estar de pie, sentarse o caer. Como se observa en la Figura 4.7.



Figura 4.7 Ejemplo de actividades estar de pie, sentarse o caer, clasificadas erróneamente

Para resolver esta problemática, se agregó una nueva clase, que lleva por nombre “sin interés”, los ejemplo que se añadieron fueron donde la persona desarrolla actividades como saltar, patear, posiciones de yoga, entre otros, que pueden ser confundidas con las categorías previamente entrenadas. Se obtuvieron los resultados, que se presentan en las Figuras 4.8-4.11.

Realizando las pruebas correspondientes, el sistema por medio de la arquitectura de imágenes es capaz de clasificar las actividades de interés con un mAP de 0.933. Donde se categorizan las poses o acciones que no se relacionan con las de prioridad, las cuales están contenidas en una clase especial. Finalmente en la Tabla A.3, se pueden observar el aumento de la clases.

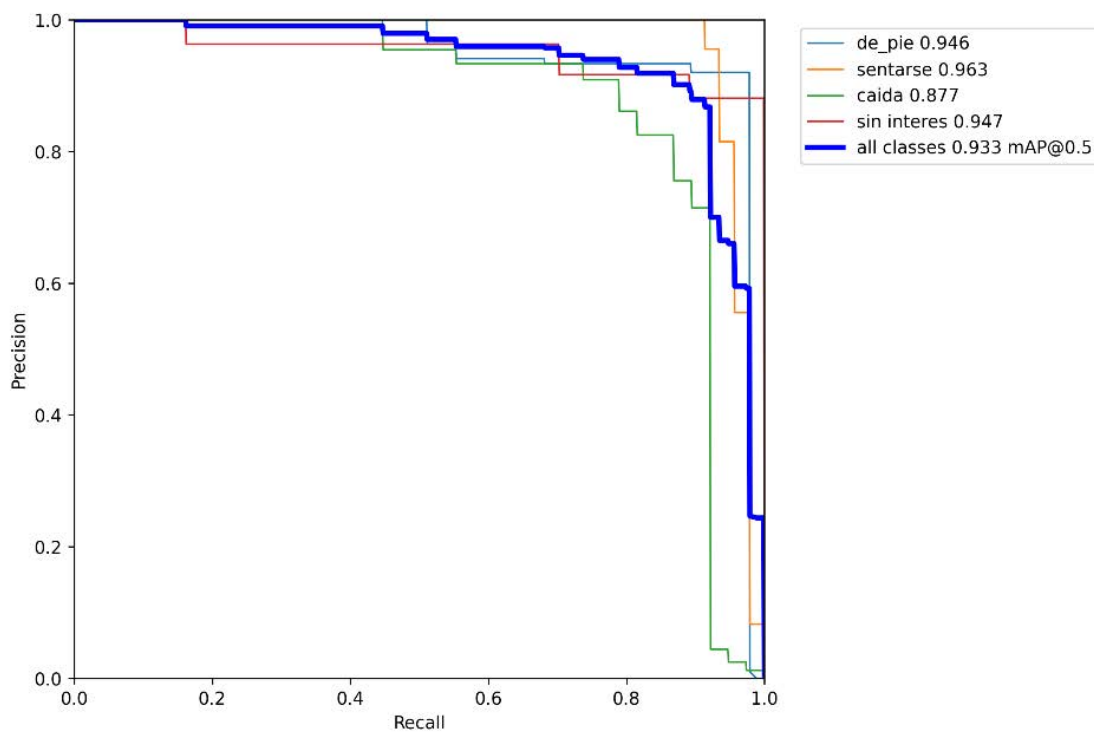


Figura 4.8 Métrica *Precision -Recall*

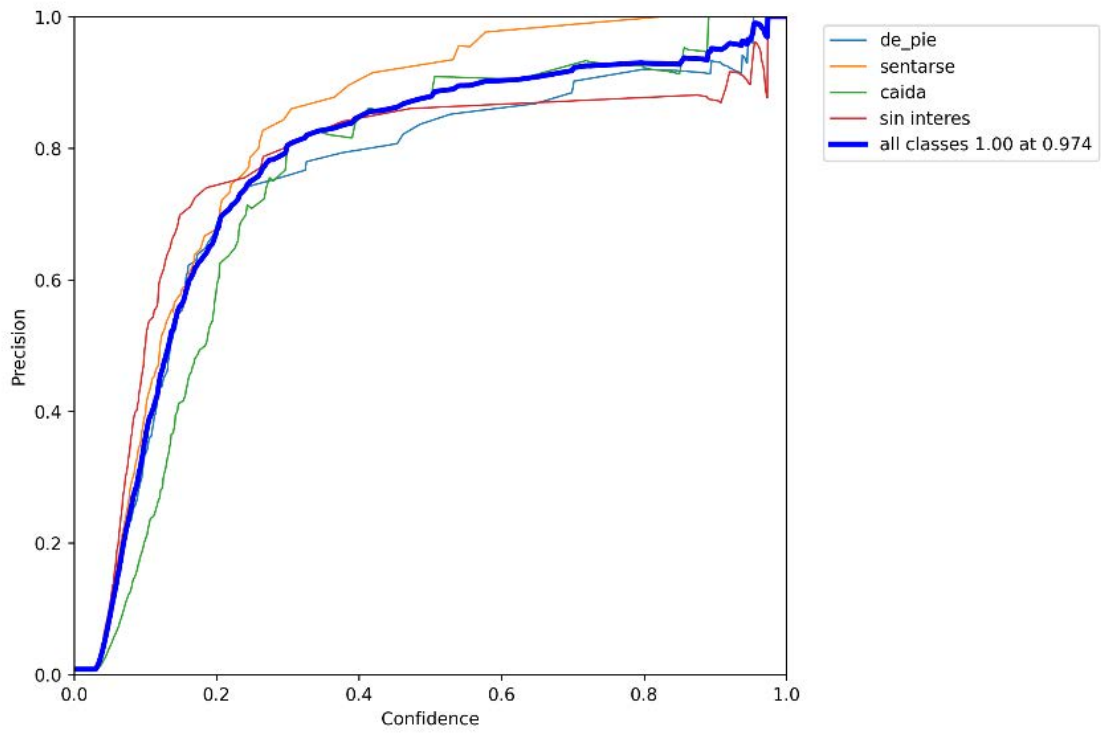


Figura 4.9 Métrica *Precision* -*Confidence*

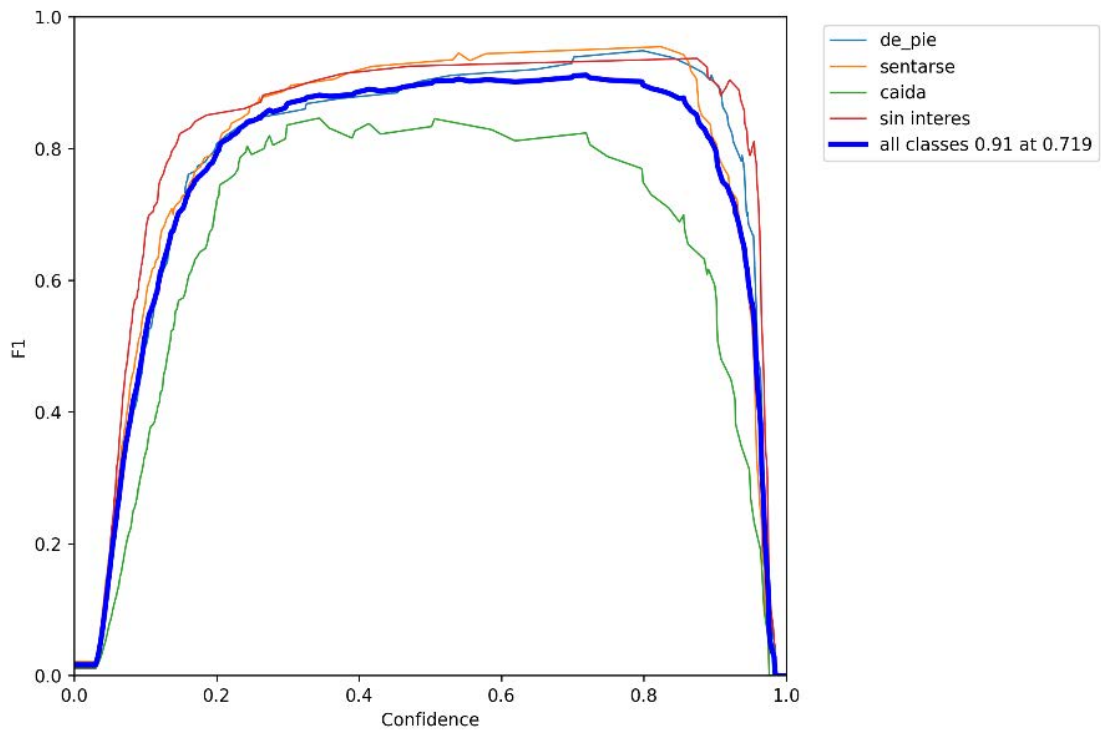


Figura 4.10 *F1* -*Confidence*

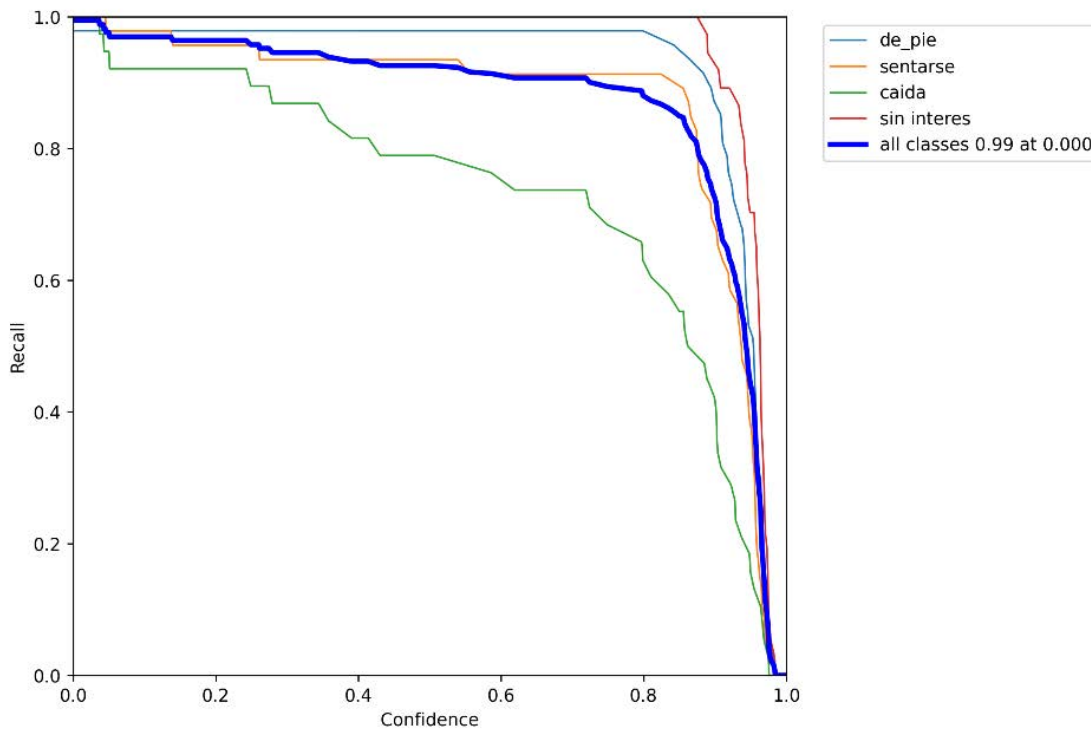


Figura 4.11 Métrica Recall-Confidence

4.1.4 Prueba 4

Probar el funcionamiento del sistema empleando diferentes modelos inerciales, dependiendo del dispositivo disponible.

Objetivo: evaluar el reconocimiento de actividades de interés, con uno o con la combinación de sensores.

Como se menciona en la sección de configuración de entrenamiento, se empleó el algoritmo *random forest* para la clasificación de los datos inerciales. En este experimento se realizó la separación de las muestras previamente etiquetadas como *device* para conocer el dispositivo al que pertenece, de esta manera se facilitó el entrenamiento de un modelo que combina celular y reloj inteligente, o que sólo se usara información individual.

Por separado el celular tiene una precisión del 98% y el reloj inteligente de 96%. En la Tabla 4.4, se muestran los resultados por actividad. Teniendo en cuenta las consideraciones previas que son que el dispositivo móvil celular, se coloque en el bolsillo

derecho del pantalón y el reloj atendiendo las instrucciones del proveedor que es la mano izquierda.

Tabla 4.4 Precisión obtenida entrenando los diferentes dispositivos

Actividad	Precisión
Estar de pie	99 %
Caminar	94 %
Sentarse	99 %

4.1.5 Prueba 5

Probar el funcionamiento del sistema en tiempo real

Objetivo: Reconocer las actividades de interés seleccionadas.

Como se mencionó anteriormente en alcances y limitaciones, el presente trabajo no está comprometido a que su funcionamiento sea en tiempo real, pero se trabajó en una sección en la interfaz, donde se puede realizar la conexión de dispositivos portátiles como lo son el celular y el reloj inteligente, por medio de este se adquiere información que es almacenada en ventana de 5s, para poder generar la clasificación, en el caso de las imágenes el modelo cargado es el generado a través de YOLOv6m6.

A continuación, en la Figura 4.12, se muestra una captura del resultado del sistema, teniendo en cuenta las condiciones que se han mencionado en el desarrollo del reporte. Donde la persona que realiza la actividad, está frente a una cámara, en este caso porta los dos dispositivos celulares y reloj inteligente. En la Tabla A.4, se adjuntan capturas de pantalla de la operación en tiempo real.

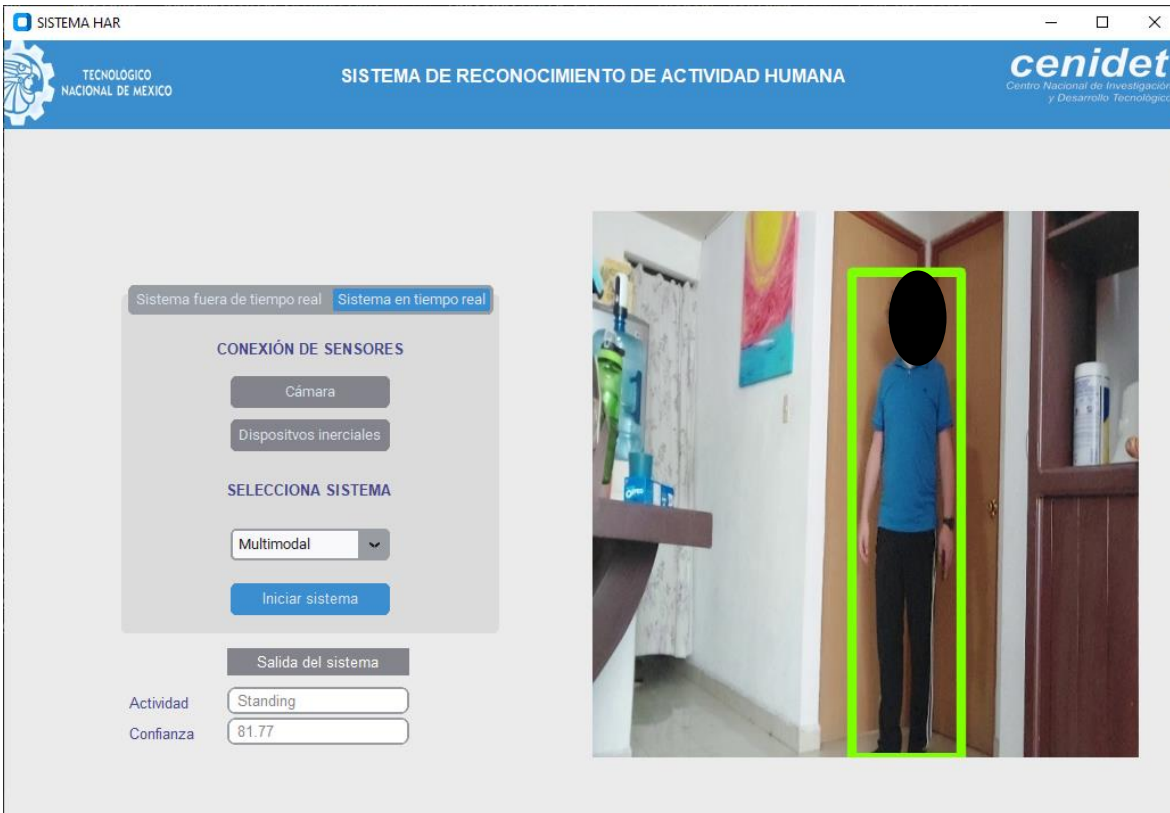


Figura 4.12 Ejemplo de operación de sistema en tiempo real

4.2 Análisis de resultados

Con el objetivo de una mejora en el sistema propuesto, se efectuó una serie de pruebas, se comenzó desde la más simple hasta la de mayor dificultad. En la primera fase se compararon los algoritmos más utilizados dentro del estado del arte, tanto para las imágenes y datos inerciales que previamente tenían un preprocesamiento. Se tomó la decisión de seleccionar aquellos que brindaran un mayor resultado en las métricas, posteriormente se fue ajustando finamente cada algoritmo seleccionado, como la experimentación individual por dispositivo, el incremento de una clase extra. Con el propósito de superar una de las limitaciones que es el funcionamiento del sistema en tiempo real, ejecutando la adquisición de datos tanto de dispositivos portátiles como de la cámara RGB.

Uno de los retos más complejos es la caracterización de la actividad humana, debido a que, por diferentes circunstancias o simple naturalidad, las personas tienden a hacer las

actividades de maneras diferentes, esto implica tomar un número alto de ejemplos de cómo se ejecuta una actividad. Otra de las complejidades, es extraer las características más relevantes dentro de los datos inerciales, los cuales aportan información importante para poder diferenciar las actividades.

CAPÍTULO 5

CONCLUSIONES Y TRABAJO FUTURO

5.1 Conclusiones generales

En esta investigación, se desarrolló un sistema de reconocimiento de actividad humana que cumpliera con el objetivo de reconocer al menos tres actividades de interés, estas se categorizan en dos tipos ambulatorias (caminar, caer y “sin interés”) y no ambulatorias (sentarse y estar de pie). Por medio de dos tipos de sensores (visión e inerciales) y empleando algoritmos de aprendizaje automático.

Analizando la literatura, con artículos recientes se detectaron primero cuáles eran las actividades con mayor frecuencia de estudio, los sensores más utilizados, así como su ubicación en el cuerpo, lo cual permite una mejor caracterización de las actividades, además de analizar el procesamiento de los datos y algoritmos con mejores resultados. La elaboración de este proyecto, se llevó en diferentes fases, donde se incluye la investigación de antecedentes, la adquisición de datos, preprocesamiento, la experimentación del sistema y la evaluación.

Los algoritmos empleados en este sistema, fueron la arquitectura Yolov6m6 y el algoritmo *random forest*. Los cuales se trabajaron en entornos virtuales *conda* con sus respectivas librerías, se llevaron a cabo seis pruebas, con la finalidad de aumentar la

precisión del sistema. En una primera fase el modelo de imágenes, había sido entrenado con las tres actividades planteadas en la propuesta (estar de pie, sentarse y caer), pero en algunas ocasiones este modelo entrenado generaba falsos positivos, donde detectaba acciones que no correspondían a ninguna clase, como lo son acciones intermedias o simplemente diferente al resto. Por esta razón se integró una clase llamada sin interés, de tal manera que todas aquellas actividades que no eran las de interés, tuvieran una clasificación con el fin de no afectar la clasificación.

Se desarrolló la integración del modelo *Random Forest*, con la finalidad tener una mayor precisión en las actividades estar de pie y caminar, como se mencionó el entrenamiento de la red convolucional, no es capaz de diferenciar entre estas dos situaciones. Los sensores empleados se encuentran dentro de dispositivos de fácil alcance actualmente, como lo son celulares y relojes inteligentes, que permiten cumplir las tareas cotidianas sin restricciones y adquirir datos reales.

Finalmente, con el objetivo de que la operación se realizara de manera intuitiva, se creó una interfaz, donde el usuario puede ejecutar el sistema no funciona en tiempo real, esto implica, ingresar las imágenes y/o datos inerciales para identificar la actividad, pero también se puede hacer uso de manera en tiempo real donde se puede realizar la selección de dispositivos disponibles para el reconocimiento de la actividad.

5.2 Objetivos alcanzados

Los objetivos que fueron definidos en la propuesta de solución fueron alcanzados cada uno de ellos, en diferentes fases de la investigación, desde el análisis de la problemática, identificación de variables de control y el desarrollo final del modelo, en la Tabla 5.1 se describen comentarios correspondientes a cada objetivo.

Por otra parte, cabe mencionar que una de las limitaciones fue superada, que es el uso del sistema en tiempo real, esto se logró gracias a la elección de los modelos ligeros, tanto en visión empleando cámara RGB, para los datos inerciales se emplea una conexión de transmisión de datos (TCP) que garantiza la llegada de los datos de una manera rápida y que sean procesados por el modelo correspondientes.

Tabla 5.1 Objetivos y comentarios

Objetivos	Comentarios
Analizar y definir situaciones de interés a estudiar en esta investigación	Se analizaron diferentes artículos dentro de la literatura, donde se observaron las actividades más estudiadas, a partir de esto se seleccionaron aquellas que se categorizaban en ambulatorias y no ambulatorias.
Seleccionar un conjunto de sensores y/o cámara(s) para su posterior adquisición del movimiento del cuerpo humana respecto a las situaciones definidas	Para una correcta caracterización se eligieron dos tipos de sensores, inerciales y de visión, estos permiten desarrollar las actividades de manera normal.
Modelar las situaciones de interés definidas.	Se procedió al etiquetado y análisis de cada actividad, extrayendo aquellas características importantes.
Definir los algoritmos para la identificación de las situaciones de interés	Por medio de un conjunto de pruebas se seleccionaron aquellos algoritmos con mejor rendimiento para la clasificación de actividades.
Integración del sistema.	Diseño de una interfaz, que permite al usuario un manejo accesible a la operación del sistema.

5.3 Aportaciones

Durante el desarrollo de la investigación, se cumplió con el objetivo de reconocer al menos tres actividades de interés, esto por medio de un sistema de reconocimiento automático, el cual generó:

- Un sistema de reconocimiento de actividad automático que identifica cinco situaciones de interés (estar de pie, caminar, sentarse, caer y “sin interés”). Compuesto de dos modelos uno de imágenes y otro de datos inerciales.
- Un conjunto de datos propios, que contiene imágenes y videos inerciales de ocho personas diferentes, desarrollando las actividades seleccionadas.
- Sistema de adquisición de datos inerciales para dos tipos de dispositivos (celular y reloj inteligente), en *Android studio* y *Python*.
 - Requerimientos mínimos:
 - Celular: *Android 8 “Oreo”*
 - Reloj inteligente: *Wear Os 3*
- Códigos fuente para preprocesamiento de datos, entrenamientos de los modelos e integración del sistema en una interfaz gráfica desarrollados en *Python*.

5.4 Trabajo Futuro

El desarrollo de esta investigación, permitió identificar situaciones interés con la integración de dos modelos computacionales. El modelo Yolov6m6, que es capaz de identificar actividades que no impliquen movimiento continuo, esto debido a su entramiento que es a través de imágenes, por lo cual se agregó, un sistema de datos inerciales, que se caracterizan por su precisión en actividades que implican movimiento. La integración de estos dos sistemas permite aumentar la precisión del reconocimiento de actividades. Sin embargo, se presentan diversas oportunidades para seguir mejorando el sistema de tal manera que sea más robusto. A continuación, se mencionan algunas:

- Detección de actividades ejecutadas por múltiples personas dentro del entorno de estudio, en el caso de la visión automática. Para realizar monitoreos de lugares específicos (ejemplo, escuelas, hospitales, estacionamientos y entre otros).

- Agregar nuevas actividades para su clasificación, incorporando situaciones relacionadas con el ambiente de estudio que se determine. Adquisición de ejemplos y datos, tanto en datos inerciales como en imágenes.
- Identificación de rostros, para saber qué sujeto efectúa una actividad, de esta manera se podría llevar a cabo un monitoreo por individuo, con algoritmo YOLO.
- Realizar un sistema de alerta a dispositivos móviles al usuario.

5.5 Productos generados

Durante el desarrollo del tema de tesis se generaron productos académicos, que se mencionan a continuación:

- a) Presentación de póster “Identificación de situaciones de interés centradas en el humano”, en la Escuela de Inteligencia Artificial y Robótica, del 16 al 20 en agosto 2022. Universidad Tecnológica Emiliano Zapata (UTEZ).
- b) Publicación y presentación de póster “Identificación de situaciones de interés centradas en el humano”, en la 9na Jornada de Ciencia y Tecnología Aplicada, celebrada del 16 al 18 de noviembre del 2022.
- c) Publicación y presentación del artículo titulado “*Human Activity Recognition implementing Yolo model*”, en ICMEAE 2022.
- d) Ponencia con conferencia “Identificación de situaciones de interés centradas en el humano”, en el simposio Internacional de Ingeniería en Sistema Computacionales – Sistemas Transversales. El 23 de marzo del 2023, Instituto Tecnológico De Cuautla.

Las constancias de cada uno de las actividades se encuentran el anexo B.

REFERENCIAS

- Abdul Lateef Haroon, P. S., & Premachand, D. R. (2021). Effective human activity recognition approach using machine learning. *Journal of Robotics and Control (JRC)*, 2(5). <https://doi.org/10.18196/jrc.25113>
- Alcantar Rafael. (2013). *Caracterización Visual de Movimientos “Sospechosos” de Personas en Estacionamiento*. Cenidet.
- Alwakeel, S. S., Alhalabi, B., Aggoune, H., & Alwakeel, M. (2016). A machine learning based WSN system for autism activity recognition. *Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015*. <https://doi.org/10.1109/ICMLA.2015.46>
- Azuela, J. H. S., & Cortés, F. R. (2021). *Inteligencia artificial aplicada a Robótica y Automatización*. Marcombo.
- Babiker, M., Khalifa, O. O., Htike, K. K., Hassan, A., & Zaharadeen, M. (2018). Automated daily human activity recognition for video surveillance using neural network. *2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017, 2017-November*. <https://doi.org/10.1109/ICSIMA.2017.8312024>
- Bagate, A., & Shah, M. (2019). Human activity recognition using RGB-D sensors. *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*. <https://doi.org/10.1109/ICCS45141.2019.9065460>
- Chao, X., Hou, Z., & Mo, Y. (2022). CZU-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and 10 Wearable Inertial Sensors. *IEEE Sensors Journal*, 22(7), 7034–7042. <https://doi.org/10.1109/JSEN.2022.3150225>
- Chung, S., Lim, J., Noh, K. J., Kim, G., & Jeong, H. (2019). Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors (Switzerland)*, 19(7). <https://doi.org/10.3390/s19071716>
- D’Angelo, G., & Palmieri, F. (2021). Enhancing COVID-19 tracking apps with human activity recognition using a deep convolutional neural network and HAR-images. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-021-05913-y>

Dávila María Luisa. (2009). *Caracterización Automática de una Muestra de Acciones en video*. Cenidet.

De Leonardis, G., Rosati, S., Balestra, G., Agostini, V., Panero, E., Gastaldi, L., & Knaflitz, M. (2018). Human Activity Recognition by Wearable Sensors : Comparison of different classifiers for real-time applications. *MeMeA 2018 - 2018 IEEE International Symposium on Medical Measurements and Applications, Proceedings*. <https://doi.org/10.1109/MeMeA.2018.8438750>

Deepak Jatkar Anil Surve, D. R. (2021). HUMAN ACTIVITY RECOGNITION AND FALL DETECTION. In *International Journal of Engineering Applied Sciences and Technology* (Vol. 6). <http://www.ijeast.com>

Dougherty, G. (2012). *Pattern recognition and classification: an introduction*. Springer Science & Business Media.

Ferrari, A., Micucci, D., Mobilio, M., & Napolitano, P. (2021). Trends in human activity recognition using smartphones. In *Journal of Reliable Intelligent Environments* (Vol. 7, Issue 3). <https://doi.org/10.1007/s40860-021-00147-0>

Fu K.S, Gonzalez R.C., & Lee C.S.G. (1988). *Robótica: Control, detección,visión e inteligencia*. McGraw-Hill.

Galaxy Watch4 Bluetooth (44 mm) | SM-R870NZKAMXO | Samsung México. (n.d.). Retrieved June 6, 2022, from <https://www.samsung.com/mx/watches/galaxy-watch/galaxy-watch4-black-bluetooth-sm-r870nzkamxo/>

GitHub - heartexlabs/labellmg: Labellmg is now part of the Label Studio community. The popular image annotation tool created by Tzutalin is no longer actively being developed, but you can check out Label Studio, the open source data labeling tool for images, text, hypertext, audio, video and time-series data. (n.d.). Retrieved October 25, 2022, from <https://github.com/heartexlabs/labellmg>

GitHub - ultralytics/yolov5: YOLOv5 🚀 in PyTorch > ONNX > CoreML > TFLite. (n.d.). Retrieved June 6, 2022, from <https://github.com/ultralytics/yolov5>

Gloria Vanessa Urquiza Flores. (2022). *Detección de anomalías en el comportamiento de una persona*. CENIDET TECNM.

- Gul, M. A., Yousaf, M. H., Nawaz, S., Rehman, Z. U., & Kim, H. (2020). Patient monitoring by abnormal human activity recognition based on CNN architecture. *Electronics (Switzerland)*, 9(12). <https://doi.org/10.3390/electronics9121993>
- Gumaei, A., Hassan, M. M., Alelaiwi, A., & Als Salman, H. (2019). A Hybrid Deep Learning Model for Human Activity Recognition Using Multimodal Body Sensing Data. *IEEE Access*, 7. <https://doi.org/10.1109/ACCESS.2019.2927134>
- Hassan, M. M., Uddin, M. Z., Mohamed, A., & Almogren, A. (2018). A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, 81. <https://doi.org/10.1016/j.future.2017.11.029>
- Jukes, E. (2018). Encyclopedia of Machine Learning and Data Mining (2nd edition). *Reference Reviews*, 32(7/8). <https://doi.org/10.1108/rr-05-2018-0084>
- Kim, K., Jalal, A., & Mahmood, M. (2019). Vision-Based Human Activity Recognition System Using Depth Silhouettes: A Smart Home System for Monitoring the Residents. *Journal of Electrical Engineering and Technology*, 14(6). <https://doi.org/10.1007/s42835-019-00278-8>
- Kim, K., Yun, G., Park, S. K., & Kim, D. H. (2019). Fall Detection for the Elderly Based on 3-Axis Accelerometer and Depth Sensor Fusion with Random Forest Classifier. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. <https://doi.org/10.1109/EMBC.2019.8856698>
- L. Ruhlainen. (2018). INTELIGENCIA ARTIFICIAL 101 COSAS QUE DEBES SABER HOY SOBRE INTELIGENCIA ARTIFICIAL. *Inteligencia Artificial*, 1.
- Li, H., Shrestha, A., Heidari, H., Le Kernec, J., & Fioranelli, F. (2020). Bi-LSTM Network for Multimodal Continuous Human Activity Recognition and Fall Detection. *IEEE Sensors Journal*, 20(3). <https://doi.org/10.1109/JSEN.2019.2946095>
- Liu, C., Li, X., Li, Q., Xue, Y., Liu, H., & Gao, Y. (2021). Robot recognizing humans intention and interacting with humans based on a multi-task model combining ST-GCN-LSTM model and YOLO model. *Neurocomputing*, 430. <https://doi.org/10.1016/j.neucom.2020.10.016>

- Lu, J., Wang, J., Zheng, X., Karmakar, C., & Rajasegarar, S. (2019). Detection of Smoking Events from Confounding Activities of Daily Living. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3290688.3290743>
- Lu, J., Zheng, X., Sheng, M., Jin, J., & Yu, S. (2020). Efficient Human Activity Recognition Using a Single Wearable Sensor. *IEEE Internet of Things Journal*, 7(11). <https://doi.org/10.1109/JIOT.2020.2995940>
- Luis, H., Garrote, L., & Nunes, U. J. (2021). Human Activity Recognition for Indoor Robotics: A Deep Learning Based Approach Using a Human Detection Stage. *2021 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2021*. <https://doi.org/10.1109/ICARSC52212.2021.9429774>
- Marcos, A. G., de Pisón Ascacibar, F. J. M., Elías, F. A., Limas, M. C., Meré, J. B. O., González, E. P. V., & others. (2006). Técnicas y algoritmos básicos de visión artificial. *Técnicas y Algoritmos Básicos de Visión Artificial*.
- Minh Dang, L., Min, K., Wang, H., Jalil Piran, M., Hee Lee, C., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 108. <https://doi.org/10.1016/j.patcog.2020.107561>
- Navita, & Mittal, P. (2022). *Machine Learning (ML) based Human Activity Recognition Model using Smart Sensors in IoT Environment*. <https://doi.org/10.1109/confluence52989.2022.9734152>
- Park, S. U., Park, J. H., Al-Masni, M. A., Al-Antari, M. A., Uddin, M. Z., & Kim, T. S. (2016). A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services. *Procedia Computer Science*, 100. <https://doi.org/10.1016/j.procs.2016.09.126>
- Pertuz, C. M. P. (2022). *Aprendizaje automático y profundo en python*. Ra-Ma Editorial.
- Prati, A., Shan, C., & Wang, K. I. K. (2019). Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. In *Journal of Ambient Intelligence and Smart Environments* (Vol. 11, Issue 1). <https://doi.org/10.3233/AIS-180510>

- Radhika, V., Prasad, Ch. R., & Chakradhar, A. (2022). *Smartphone-Based Human Activities Recognition System using Random Forest Algorithm*. <https://doi.org/10.1109/iconat53423.2022.9726006>
- Radu, V., & Henne, M. (2019). Vision2Sensor: Knowledge Transfer Across Sensing Modalities for Human Activity Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3). <https://doi.org/10.1145/3351242>
- Ranieri, C. M., Macleod, S., Dragone, M., Vargas, P. A., & Romero, R. A. F. (2021). Activity recognition for ambient assisted living with videos, inertial units and ambient sensors. *Sensors (Switzerland)*, 21(3). <https://doi.org/10.3390/s21030768>
- Saldaña Jorge Alfredo. (2007). *Caracterización de Imágenes en Movimiento: Correr y Caminar*. Cenidet.
- sklearn.ensemble.RandomForestClassifier* — *scikit-learn 1.3.0 documentation*. (n.d.). Retrieved September 3, 2023, from <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- Subasi, A., Khateeb, K., Brahimi, T., & Sarirete, A. (2019). Human activity recognition using machine learning methods in a smart healthcare environment. In *Innovation in Health Informatics: A Smart Healthcare Primer*. <https://doi.org/10.1016/B978-0-12-819043-2.00005-8>
- Terven, J., & Cordova-Esparza, D. (2023). A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *ArXiv Preprint ArXiv:2304.00501*.
- Varshney, N., Bakariya, B., Kushwaha, A. K. S., & Khare, M. (2022). Human activity recognition by combining external features with accelerometer sensor data using deep learning network model. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-021-11313-0>
- Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A review of human activity recognition methods. In *Frontiers Robotics AI* (Vol. 2, Issue NOV). <https://doi.org/10.3389/frobt.2015.00028>
- Weiss, G. M. (2019). WISDM Smartphone and Smartwatch Activity and Biometrics Dataset. *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set*, 7.

What is Random Forest? | IBM. (n.d.). Retrieved August 25, 2023, from <https://www.ibm.com/topics/random-forest>

Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors*, 22(4). <https://doi.org/10.3390/s22041476>

Zia, S., Khan, A. N., Zaidi, K. S., & Ali, S. E. (2021). Detection of Generalized Tonic Clonic Seizures and Falls in Unconstraint Environment Using Smartphone Accelerometer. *IEEE Access*, 9. <https://doi.org/10.1109/ACCESS.2021.3063765>

ANEXO A

Pruebas de experimentación

En el siguiente anexo se presentan los resultados obtenidos en la detección y clasificación, estos se exponen a través de imágenes generadas por el modelo entrenado YOLOV6. Estos resultados corresponden a las pruebas 1, 2, 3 y 5 en las tablas A.1-A.4.

Tabla A.1. Resultados de la prueba no.1







Prueba 1: Comparación de algoritmos para la clasificación de actividades de interés realizadas por una persona, con tres versiones diferentes de YOLO para imágenes.	
Clases	De pie, sentarse y caída
Resultados de detección y clasificación	
De pie	
	
Yolov5	Yolov6
Confianza 0.81	Confianza 0.84
	
Yolov7	Confianza 0.57
Sentarse	
	
Yolov5	Yolov6
Confianza 0.56	Confianza 0.75
*Doble etiquetado con estar de pie	
	
Yolov7	Confianza 0.34

Tabla A.1. Resultados de la prueba no.1 (Continuación)

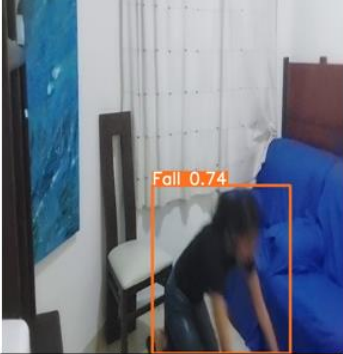
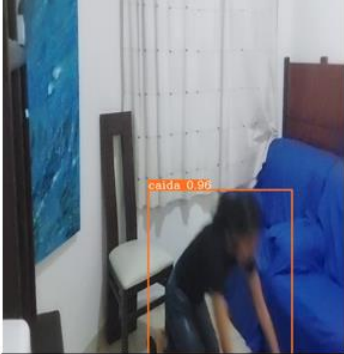

Prueba 1: Comparación de algoritmos para la clasificación de actividades de interés realizadas por una persona, con tres versiones diferentes de YOLO para imágenes.	
Clases	De pie, sentarse y caída
Resultados de detección y clasificación	
Caída	
	
Yolov5 Confianza 0.74	Yolov6 Confianza 0.96
	
Yolov7 Confianza 0.44	

Tabla A.2 Resultados de la prueba no.2







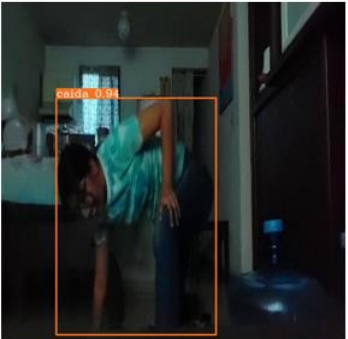
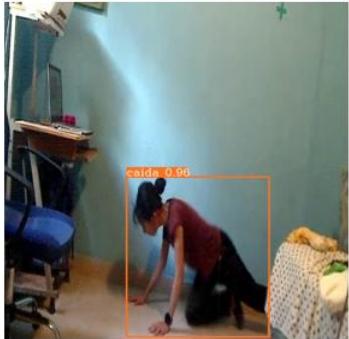

Prueba 2: Probar sistema completo con la elección de modelo YOLOV6m6, la operación de este sistema no es en tiempo real.	
Clases	De pie, sentarse y caída
Resultados de detección y clasificación	
De pie	
	
Yolov6	Confianza 0.80
Yolov6	Confianza 0.90
	
Yolov6	Confianza 0.84
Sentarse	
	
Yolov6	Confianza 0.96
Yolov6	Confianza 0.98
	
Yolov6	Confianza 0.97
Caída	
	
Yolov6	Confianza 0.94
Yolov6	Confianza 0.96
	
Yolov6	Confianza 0.92

Tabla A.3 Resultados de la prueba no.3





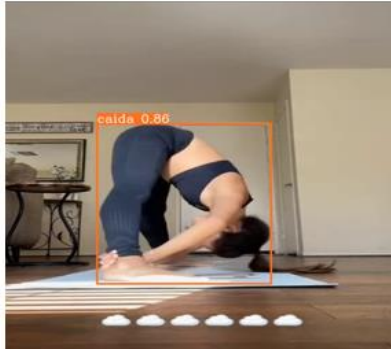
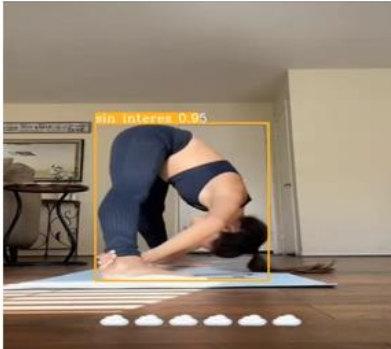
Prueba 3: Adición de una clase para aumentar la precisión de la clasificación	
Clases	De pie, sentarse, caída y "Sin interés"
Resultados de detección y clasificación	
<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Tres clases (de pie, sentarse y caída)</p> </div> <div style="text-align: center;"> <p>Cuatro clases (de pie, sentarse, caída y "sin interés")</p> </div> </div>	
 <p>de pie 0.95</p>	 <p>sin interés 0.96</p>
 <p>caída 0.76</p>	 <p>sin interés 0.96</p>
 <p>caída 0.86</p>	 <p>sin interés 0.95</p>
<p>De pie</p> <p>Confianza 0.95</p>	<p>Sin interés</p> <p>Confianza 0.96</p>
<p>Caída</p> <p>Confianza 0.76</p>	<p>Sin interés</p> <p>Confianza 0.96</p>
<p>Caída</p> <p>Confianza 0.86</p>	<p>Sin interés</p> <p>Confianza 0.95</p>

Tabla A.4 Resultados de la prueba no.5

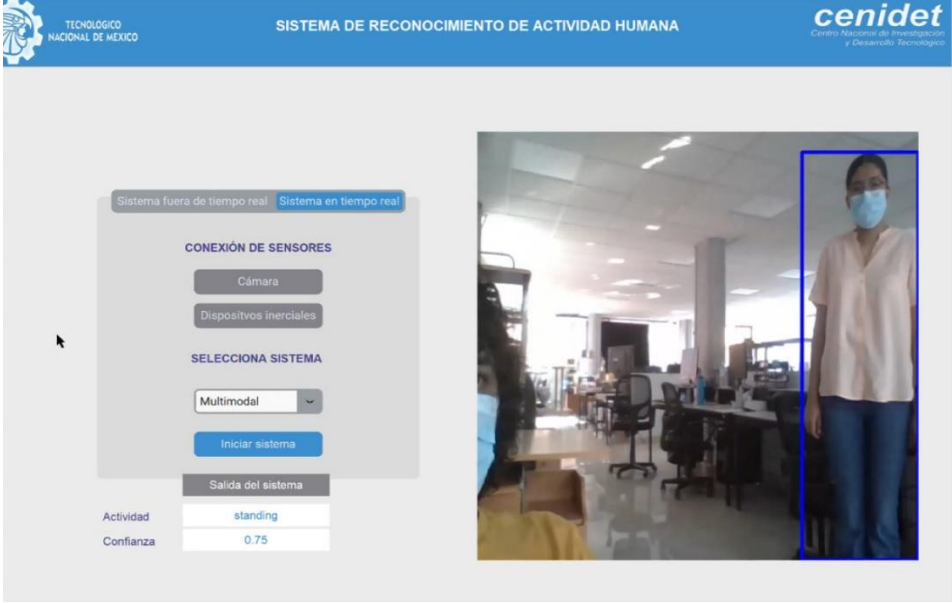
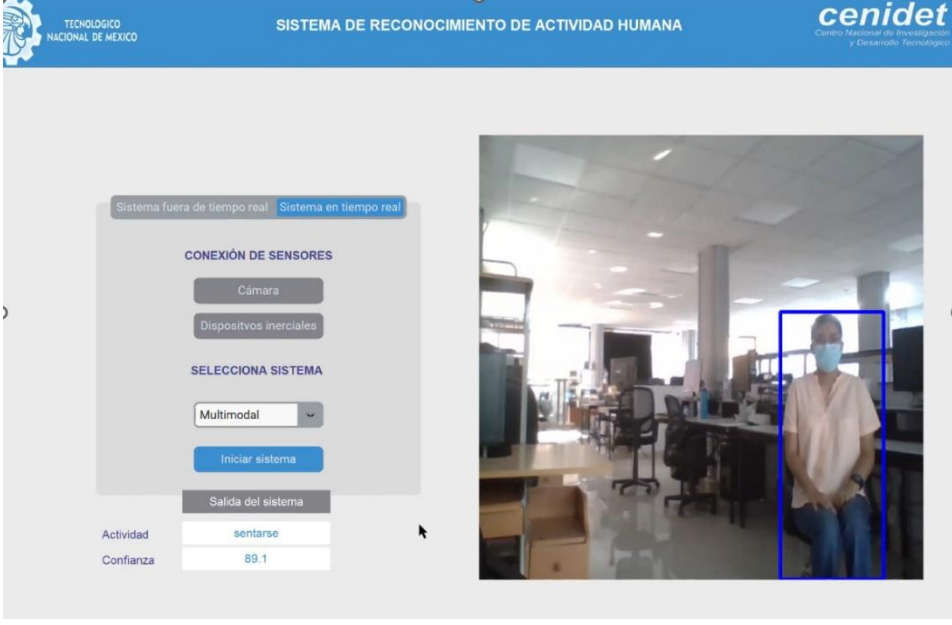
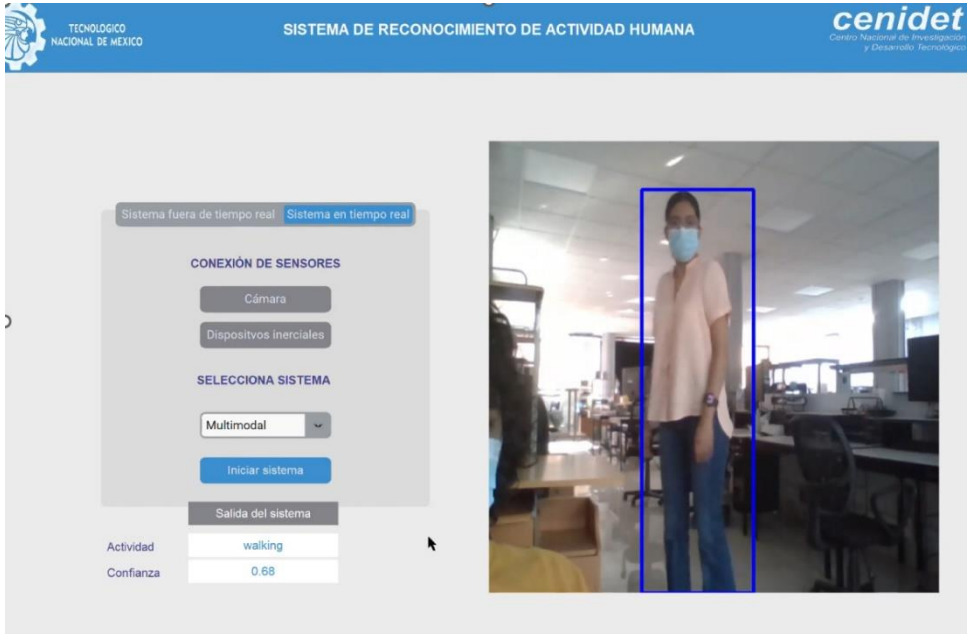
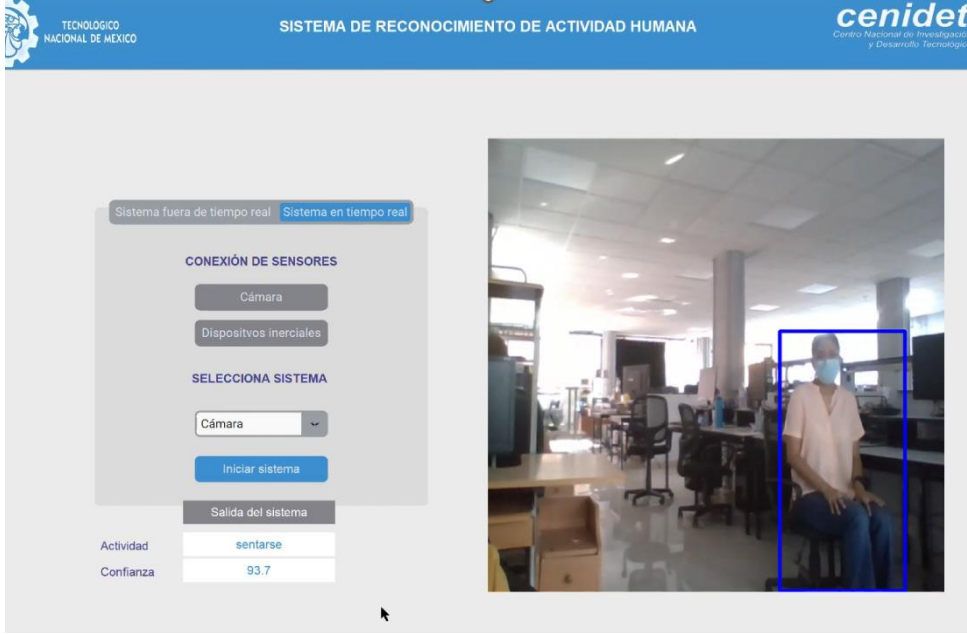
Prueba 5: Probar sistema completo con la elección de modelo YOLOV6m6 en tiempo real	
Clases	De pie, sentarse, caída y "Sin interés"
Resultados de operación de sistema en tiempo real	
	
	

Tabla A.5 Resultados de la prueba no.5 (Continuación)

Prueba 5: Probar sistema completo con la elección de modelo YOLOV6m6 en tiempo real	
Clases	De pie, sentarse, caída y "Sin interés"
Resultados de operación de sistema en tiempo real	
 <p>The screenshot shows the system interface for human activity recognition. On the left, there is a control panel with the following elements: <ul style="list-style-type: none"> Buttons for 'Sistema fuera de tiempo real' and 'Sistema en tiempo real'. 'CONEXIÓN DE SENSORES' section with 'Cámara' and 'Dispositivos inerciales' buttons. 'SELECCIONA SISTEMA' section with a dropdown menu set to 'Multimodal' and an 'Iniciar sistema' button. 'Salida del sistema' button. Activity display: 'Actividad' is 'walking' and 'Confianza' is '0.68'. </p> <p>On the right, a video feed shows a person in a light-colored shirt and blue pants walking in an office environment. A blue bounding box is drawn around the person.</p>	
 <p>The screenshot shows the system interface for human activity recognition. On the left, there is a control panel with the following elements: <ul style="list-style-type: none"> Buttons for 'Sistema fuera de tiempo real' and 'Sistema en tiempo real'. 'CONEXIÓN DE SENSORES' section with 'Cámara' and 'Dispositivos inerciales' buttons. 'SELECCIONA SISTEMA' section with a dropdown menu set to 'Cámara' and an 'Iniciar sistema' button. 'Salida del sistema' button. Activity display: 'Actividad' is 'sentarse' and 'Confianza' is '93.7'. </p> <p>On the right, a video feed shows the same person sitting in the office environment. A blue bounding box is drawn around the person.</p>	

ANEXO B

Constancias de participaciones



Figura B.1 Reconocimiento de presentación de póster en Escuela de inteligencia artificial



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

**EL TECNOLÓGICO NACIONAL DE MÉXICO
A TRAVÉS DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO**

OTORGA EL PRESENTE

RECONOCIMIENTO

A

SAGRARIO GARCIA GARCIA

CENIDET

POR LA PRESENTACIÓN DEL ARTICULO:
IDENTIFICACIÓN DE SITUACIONES DE INTERÉS CENTRADAS EN EL HUMANO
EN EL MARCO DE LA 9ª JORNADA DE CIENCIA Y TECNOLOGÍA APLICADA, CELEBRADA
DEL 16 AL 18 DE NOVIEMBRE DE 2022, EN EL TECN/CENIDET

CUERNAVACA, MORELOS, 16-18 DE NOVIEMBRE DE 2022



GS0104522

<http://constancias.cenidet.tecnm.mx>

DRA. YESICA IMELDA SAAVEDRA BENÍTEZ
**DIRECTORA DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO**

Sello Digital:

*dSzUpw1V81H8u9Qt1p5R39nWocFV8s4/diPR1aZtyc7k31B3Iw7r0iWnLOHvDcyIXV4132oAeFjuDlaPggFax
EH01u13Nzs6d2Bj1855Jhh9WwuUAde2F9wFrqa8IW2R24hmMKqGUa3cVkb75N0dLW1Xp8MOoiV58BR6pJDrJf
m05XhrV5sHOPpVXC0Ac/CE5XJPTG75i15UkH58CdwBPSi8TXRnj6VNrtrb8o88mb0UMwR/qA3atenYK768SMqZ1
fShrQf9ZU0wtpAaFdFUX920hyYcxuZ5ZHL0JYp1+fPwKsy8nNtTmipxM2bkVGAErB37RhBv1TMo3PvtSiqsw==*

cenidet
Centro Nacional de Investigación
y Desarrollo Tecnológico



Figura B.2. Reconocimiento de presentación de poster 9na JCyTA



Figura B.3 Ponencia de artículo Human Activity Recognition en ICMEAE



Figura B.4 Reconocimiento de participación de ponencia