

Jornadas de Automática

Integración ConvNeXt-YOLO mediante CVV para detectar caídas en robot social

Sánchez-Girón, C.^{a,*}, García-Gómez, M.^a, Duque-Domingo, J.^a, Gómez-García-Bermejo, J.^{a,b}, Zalama, E.^{a,b}

^aDISA-ITAP, Universidad de Valladolid, C/ Dr Mergelina s/n, 47011 Valladolid, España

^bCARTIF Centro Tecnológico 47151, Valladolid, España

To cite this article: Sánchez-Girón, C., García-Gómez, M., Duque-Domingo, J., Gómez-García-Bermejo, J., Zalama, E. 2024. ConvNeXt-YOLO integration using CVV to detect falls in social robots. *Jornadas de Automática*, 45. <https://doi.org/10.17979/ja-cea.2024.45.10788>

Resumen

Cada vez más personas mayores optan por vivir en su hogar, lo que ha generado una necesidad crítica de asegurar entornos seguros para esta población. El 50 % de las personas mayores de 80 años experimentan al menos una caída al año. Con este estudio se busca detectar caídas implementando un sistema de visión, proporcionando una respuesta rápida en caso de emergencia para que la persona caída tenga asistencia si sufre un accidente. Aquí se propone una metodología basada en el uso de modelos de aprendizaje profundo, específicamente utilizando la técnica de *Cross Validation Voting* (CVV), para mejorar la generalización y la precisión en la detección de caídas a partir de imágenes. El modelo planteado ha conseguido una exactitud de un 92.95 % y una pérdida del 0.1885 para el conjunto de test. El sistema de detección de caídas ha sido integrado en el robot social Temi, que se introducirá en el domicilio de los usuarios para monitorear continuamente su bienestar y proporcionar asistencia inmediata en caso de detectar una caída.

Palabras clave: Visión por computadora, Integración de datos, Aprendizaje por conjuntos, Robots móviles, Detección de caídas.

ConvNeXt-YOLO integration using CVV to detect falls in social robots

Abstract

More and more older adults are choosing to live at home, which has created a critical need to ensure safe environments for this population. 50 % of people over the age of 80 experience at least one fall per year. This study seeks to detect falls by implementing a vision system, providing a rapid response in case of emergency, so that the fallen person has assistance if he/she suffers an accident. Here we propose a methodology based on the use of deep learning models, specifically using the *Cross Validation Voting* (CVV) technique, to improve generalization and accuracy in detecting falls from images. The proposed model achieved an accuracy of 92.95 % and a loss of 0.1885 for the test set. The fall detection system has been integrated into the Temi social robot, which will be introduced in the users' home to continuously monitor their well-being and provide immediate assistance in case a fall is detected.

Keywords: Computer Vision, Data Integration, Ensemble Learning, Mobile robots, Fall detection.

1. Introducción

El número de personas mayores que eligen vivir en sus propios hogares, solas o con sus parejas, ha aumentado notablemente en los últimos años. Según el Instituto Nacional de Estadística (INE), en nuestro país, el número de personas mayores de 65 años que viven solas ha aumentado un 25.8 %

en la última década (INE, 2023). Esto refleja un cambio en los patrones de vivienda, donde cada vez más personas mayores optan por mantener su independencia en sus hogares. Es crucial asegurar que las personas mayores que viven solas estén en entornos seguros para evitar accidentes, especialmente caídas, que son la principal causa de lesiones en adultos ma-

*Autor para correspondencia: celia.sanchez-giron@uva.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

yores. Aproximadamente el 30 % de los mayores de 65 años y el 50 % de los mayores de 80 años sufren al menos una caída al año (Rodríguez-Molinero et al., 2015). Sin embargo, es importante destacar que las caídas no son inevitables y se pueden prevenir garantizando un entorno seguro para mejorar la calidad de vida de estas personas.

Para abordar este desafío, se han desarrollado nuevos métodos de detección de caídas, con el objetivo de reducir el riesgo de lesiones graves y mejorar la seguridad de los adultos mayores. Uno de los avances más prometedores es el uso de algoritmos de aprendizaje automático para identificar patrones de caídas a partir de imágenes.

El uso de modelos neuronales para la detección de caídas ha crecido gracias a los avances de los procesadores gráficos, que han optimizado el entrenamiento de la inteligencia artificial. En el campo de la visión artificial, se han explorado diversas áreas, como el reconocimiento de imágenes, la detección de objetos y el seguimiento de movimiento (Anwary et al., 2022), gracias a la implementación de algoritmos cada vez más sofisticados. De hecho, es frecuente encontrar estudios que emplean modelos de detección de objetos en imágenes (Rao et al., 2008), permitiendo la identificación precisa de las coordenadas del objeto de interés en la escena, como YOLO o Fast R-CNN. Todo esto permite la detección temprana y precisa de caídas, lo que podría mejorar significativamente la seguridad de las personas mayores que viven solas.

A menudo encontramos modelos neuronales que alcanzan una elevada exactitud durante el entrenamiento a la hora de clasificar un conjunto de datos específico (Quinayás Burgos et al., 2020), pero luego muestran dificultades para generalizar de manera adecuada cuando se enfrentan a datos de test que no han visto previamente. Esta falta de generalización puede deberse a varios factores, como el sobreajuste a los datos de entrenamiento, la presencia de ruido en los datos, o la falta de diversidad en el conjunto de entrenamiento. Esto puede conducir a modelos que no son capaces de realizar predicciones precisas en situaciones del mundo real, lo que puede suponer un gran límite a la hora de aplicarlos a la vida real.

Para compensar esta falta de generalización, en la actualidad existen diferentes métodos que tienen como objetivo prevenir este sobreajuste que ocurre cuando el modelo memoriza los datos de entrenamiento. Una de estas técnicas es el *Cross Validation Voting* o CVV (Domingo et al., 2022) que permite mejorar la capacidad de generalización en modelos clasificadores. Aquí se evalúa el rendimiento del modelo con diferentes conjuntos de datos al crear diferentes *slots* o divisiones del conjunto de validación, por lo que se obtendrá una estimación más confiable del rendimiento general del modelo.

En la literatura se pueden encontrar varias investigaciones sobre la detección de caídas. Una de las mayores dificultades de estas investigaciones es obtener la capacidad de diferenciar entre una persona dormida en una cama o sentada en una silla y una persona caída en el suelo. En un trabajo previo, hemos planteado una metodología para afrontar esta dificultad (Sánchez-Girón et al., 2024). Aunque los resultados fueron adecuados, se han buscado mejoras para optimizar aún más su rendimiento.

En esta nueva propuesta, se ha trabajado con una base de datos más extensa y un nuevo planteamiento de detección de elementos en la imagen. Igualmente, para mejorar la generalización del modelo, se ha implementado la alternativa de *Cross Validation Voting* con el fin de acercarse más a lograr este objetivo.

El sistema de detección de caídas forma parte de un plan representado en la Figura 1, en el que un robot social, en particular el modelo Temi 3 (Temi, 2024), capturará imágenes de los individuos en sus domicilios y se enviarán al método de detección de caídas para evaluar su posición. En caso de detectar una persona caída, el robot preguntará al usuario si se encuentra bien. Si la respuesta es afirmativa entonces el robot comenzará la rutina diaria con los ejercicios y actividades programados para el usuario. Si el usuario responde que no se encuentra bien el robot realizará una llamada al contacto de emergencia del individuo que tendrá almacenado a nivel local.

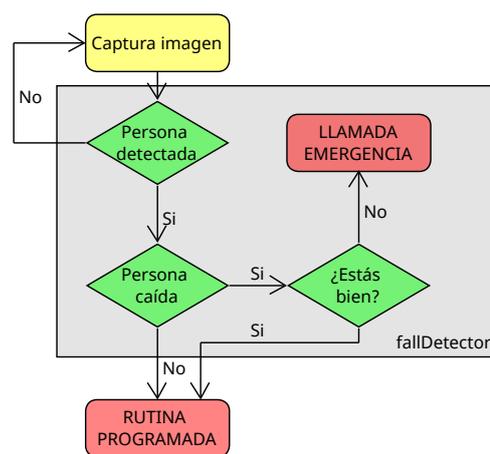


Figura 1: Plan de detección de caídas implementado en robot Temi

Este enfoque busca mejorar la precisión en la detección de caídas, especialmente en entornos donde las personas mayores residen solas y pueden necesitar asistencia inmediata. El artículo forma parte del proyecto EIAROB, que tiene como objetivo crear un ecosistema de Inteligencia Ambiental para apoyar el cuidado de larga duración en el hogar mediante robots sociales (EIAROB, 2022). Se propone integrar un sistema de detección de accidentes que combina sensores ambientales (Merino-Fidalgo et al., 2023) con la cámara que está integrada en un robot social, centrando este documento en la parte relacionada con la cámara del robot.

2. Metodología propuesta

En esta sección se va a explicar la metodología propuesta para el reconocimiento de caídas, así como el dataset utilizado y las diferentes tecnologías planteadas. El esquema del modelo se puede encontrar en la Figura 2. Como se observa en esta Figura, la propuesta presenta una primera parte enfocada a la detección de diferentes elementos en la imagen, una sección en la que se combina ConvNeXt con Fully Connected para realizar la predicción en cada modelo entrenado y una última parte donde se aplica Soft Voting como parte de la herramienta CVV para obtener la salida final.

2.1. Dataset

Para formar la base de datos con la que se ha entrenado el modelo de clasificación, se han unido imágenes de diferentes

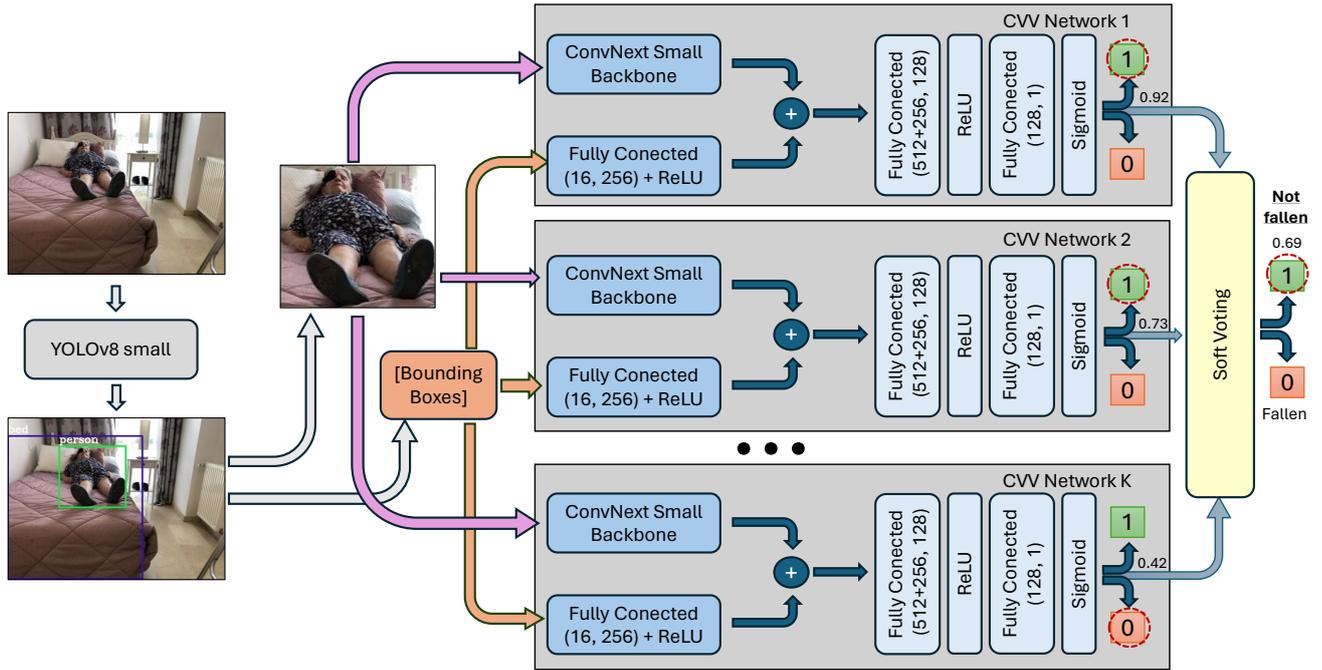


Figura 2: Esquema general del modelo

datasets públicos:

- **FPDS**: contiene imágenes con personas caídas y no caídas en diferentes ubicaciones interiores y exteriores. Estas imágenes han sido tomadas desde la cámara de un robot móvil. (Maldonado-Bascon et al., 2019)
- **setElderly**: conjunto de imágenes en las que aparecen personas mayores de 65 años caídas, tumbadas y en otras posiciones. (Maldonado-Bascon et al., 2019)
- **fallDataset**: imágenes en las que aparecen personas caídas y no caídas en diferentes perspectivas. (Kwolek and Kepski, 2014)
- **poseDetection**: contiene imágenes de personas tumbadas o dormidas mayoritariamente. (SAM, 2024) Se han incluido estas imágenes, ya que uno de los hitos que queremos conseguir es la distinción entre un persona tumbada y caída.
- **Test2 Dataset**: se han recopilado imágenes de personas mientras se están cayendo a partir de vídeos. (Workspace, 2024)

A partir del conjunto de imágenes conseguido se ha entrenado el modelo de clasificación con las imágenes de las personas recortadas. En total se han empleado 1021 imágenes de personas caídas y 1811 imágenes de personas no caídas. Al crear un conjunto de datos a partir de diferentes bases e incorporar imágenes de personas en diferentes posiciones, se espera que el modelo adquiera una elevada capacidad de generalización y se comporte mejor en inferencia. Además, se han introducido varias imágenes de personas dormidas para

entrenar al modelo y que distinga si la persona está caída o simplemente tumbada sobre una cama o sofá.

2.2. Detección en la imagen y solapamiento

Una vez que se ha diseñado la base de datos, el paso siguiente consiste en procesar las imágenes utilizando un modelo especializado en la detección de objetos y personas. Para la metodología propuesta se empleó un enfoque que utiliza una red YOLO (*You Only Look Once*), en particular el modelo YOLOv8s (Jocher et al., 2023). YOLO es un modelo de detección de objetos en la imagen; para ello, nos devuelve el *Bounding Box* de estos objetos, que son 4 coordenadas que definen la ubicación del elemento en la imagen. La idea es emplear YOLO para obtener una imagen de la persona recortada a partir de las coordenadas del *Bounding Box* de la persona, así como obtener las coordenadas de otros objetos de interés. Se ha definido como elementos de interés sillas, camas y sofás, ya que se ha considerado que podrían influir en detectar si la persona está caída o en otra posición. A la hora de realizar esta detección se estableció un umbral de confianza para la identificación de las diferentes clases de objetos. Para ayudar a la red a distinguir si una persona está caída o tumbada sobre una cama o sofá, se integró una función para detectar el nivel de solapamiento entre la persona y el objeto en la imagen aplicando la técnica de *Intersection Over Union (IoU)*. Para ello se comprueba si el área del rectángulo que forma el *Bounding Box* de la persona segmentada interseca con el área del *Bounding Box* de cada objeto de interés detectado.

$$IoU = \frac{Area_{persona} \cap Area_{objeto}}{Area_{persona} \cup Area_{objeto}} \quad (1)$$

Si existe un solapamiento significativo, superior a un cierto umbral, se conservan las coordenadas de las personas en la

imagen y los objetos que están solapados con las mismas. Con estos asumimos una posible interacción de la persona con el objeto. Si el objeto de interés no es detectado o no está superpuesto con la persona, entonces indicamos que el *Bounding Box* de dicho objeto serán las coordenadas [0,0,0,0].

De esta forma, tras la detección obtendremos la imagen de la persona recortada respecto a la imagen original y cuatro conjuntos de cuatro coordenadas, al combinar el *Bounding Box* de la persona y los tres *Bounding Boxes* de los objetos presentes o no.

2.3. Cross Validation Voting

En los últimos años, gracias al desarrollo de los modelos de aprendizaje profundo, las redes neuronales y clasificadores de aprendizaje profundo se han aplicado de forma frecuente a problemas en los que hay que realizar una clasificación mediante visión computacional. A menudo, a pesar de que los resultados obtenidos son elevados, hay algunos modelos que no consiguen desarrollar una capacidad de generalización adecuada. Como consecuencia, a la hora de evaluar el modelo con datos desconocidos, los resultados pueden empeorar mucho con respecto al entrenamiento.

Una técnica muy empleada al trabajar con redes neuronales profundas es el *early stopping*, que detiene el aprendizaje del modelo cuando la exactitud del grupo de validación comienza a descender durante varias épocas de forma continuada. De esta forma, se detiene el entrenamiento asumiendo que el modelo ha alcanzado su comportamiento óptimo y se consigue una mejora de generalización por parte del modelo. Sin embargo, el *early stopping* puede provocar que el modelo se ajuste demasiado bien al conjunto de validación o *Validation Overfitting*, lo que puede afectar negativamente a la capacidad de generalización del modelo. Esto se expone en la investigación (Domingo et al., 2022): *"The early stopping approach followed to train neural networks makes the model favorable to the validation data, something known as validation overfitting. Although the evaluation of the model with respect to the test set is realistic, being totally independent unknown data, it may not be the best generalizing model."* Como consecuencia, puede que el modelo no sea el óptimo para predecir sobre datos completamente nuevos (conjunto de test) debido al ajuste excesivo a los datos de validación. Para enfrentarnos a este problema, una alternativa es el uso de *Cross Validation Voting* (CVV) (Domingo et al., 2022). Esta técnica se basa en la división del conjunto de entrenamiento en diferentes grupos o *slots*, como en el método de Cross Validation (Schaffer, 1993), de manera que cada grupo o *slot* tenga unos datos de validación diferentes a los demás, aunque siempre del mismo tamaño. A continuación se realizan varios entrenamientos, uno para cada *slot* diferente, obteniendo tantos modelos como entrenamientos se hayan hecho. Para el caso estudiado hemos elegido un modelo ConvNeXt para obtener las características de la imagen. Cada uno de los entrenamientos realizará *early stopping* respecto a su conjunto de validación propio. Por último, todos los modelos obtenidos se integran en un modelo de votación: *Hard Voting*, cada modelo genera un voto por clase, o *Soft Voting*, se emplean las probabilidades de cada modelo. En este caso se ha utilizado *Soft Voting*. Este método se puede aplicar utilizando clasificadores de diferente naturaleza, siempre que permitan la técnica de *early stopping* que es

crucial en este enfoque. Esta integración permite que el sistema se aproveche de lo mejor de cada clasificador, de forma que se compensen las deficiencias individuales de cada uno, logrando así un rendimiento global superior.

2.4. Modelo ConvNeXt

Una vez que el modelo haya detectado las personas en la imagen y los objetos que están solapados al área de las diferentes personas, los archivos de imagen y texto obtenidos se introducirán en cada uno de los modelos generados con CVV. Cada uno de los modelos presentan la misma estructura basada en una red ConvNeXt y una capa Fully Connected. Se ha elegido el modelo ConvNeXt, ya que varios artículos respaldan sus buenos resultados en cuanto a clasificación de imágenes (Liu et al., 2022).

Por un lado, se ha incorporado una capa Fully Connected a la que se pasa como entrada los cuatro conjuntos de coordenadas correspondientes a los *Bounding Boxes* obtenidos. Esta capa tendrá una entrada de tamaño 16, por estar constituida por el *Bounding Box* de la persona y los tres *Bounding Boxes* de los objetos (sofá, cama y silla) presentes o no. Por otro lado, se incorpora el modelo ConvNeXt de forma paralela para la clasificación de imágenes. Como entrada al modelo se pasará la imagen de la persona recortada obtenida a partir de la red YOLO. La salida de este modelo tendrá un tamaño de 512 como vector de características de la imagen. Una vez obtenidos los vectores anteriores, estos se concatenan y se ingresan a una capa densa. Esta capa produce un vector con dimensiones que coinciden con el número de clases, en nuestro caso, de tamaño uno, debido a que se trata de una salida binaria.

3. Resultados

El entrenamiento se realizó durante 50 épocas con una paciencia de 20 épocas para la pérdida del conjunto de validación. Para comprobar la eficacia del modelo a la hora de clasificar imágenes desconocidas, se realizó la evaluación y se calculó la matriz de confusión para el conjunto de test. El resultado se expone en la Figura 3. En esta figura se han representado los porcentajes de *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP) y *False Negatives* (FN).

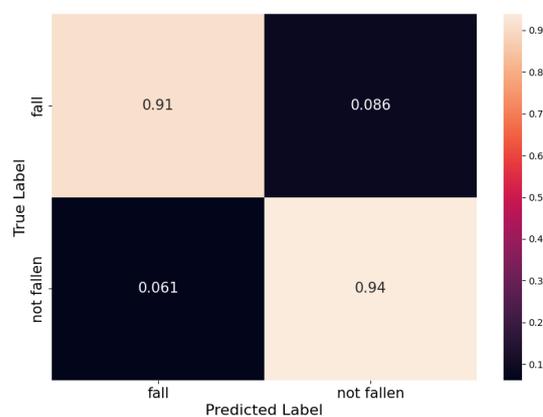


Figura 3: Matriz de confusión para el conjunto de test

A continuación, se calcularon diferentes métricas para comprender mejor el comportamiento del modelo, que se recogen en la Tabla 1.

Tabla 1: Métricas de la matriz de confusión

Resultados	
Accuracy	0.9295
Precision	0.9139
Recall	0.8967
F1 – Score	0.9052

Durante la evaluación del modelo utilizando *Cross Validation Voting*, se logra un índice de exactitud (*accuracy*) del 92.95 %, con una pérdida de 0.1885. Además, la puntuación F1 es notablemente alta, lo que sugiere un equilibrio adecuado entre la especificidad y la sensibilidad en el modelo. Estos resultados indican que el desempeño del modelo para la clasificación propuesta es bastante prometedor. Teniendo en cuenta el valor de exactitud obtenido para el conjunto de test, es de esperar que en inferencia presente una gran capacidad de generalización.

El resultado de esta metodología se observa en la Figura 4, en la que a partir de una imagen con una o varias personas, el modelo es capaz de distinguir el estado de cada una de ellas.

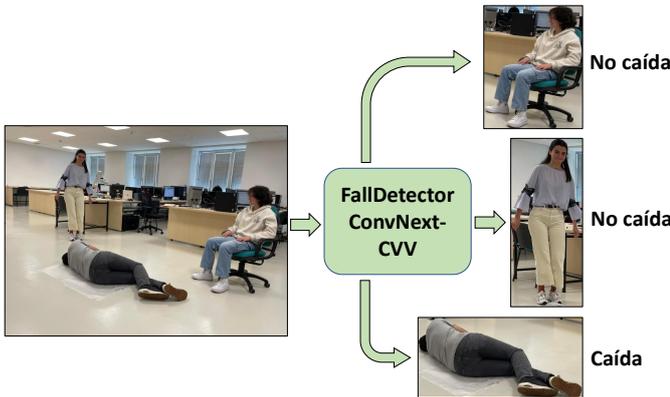


Figura 4: Ejemplo de clasificación con varias personas en una imagen

A la hora de realizar este modelo surgió la pregunta de que si el uso de *Bounding Box* en la metodología influía de forma significativa en la predicción final o, al contrario, el modelo solo tenía en cuenta las características de la imagen. Para probarlo se tomó una imagen de una persona sobre una cama que la red reconocía como no caída. Tras realizar la detección y localización de la persona y los objetos en la imagen se obtuvo como resultado la imagen de la persona recortada y un fichero de texto en el que se especificaban coordenadas del *Bounding Box* de la persona y las coordenadas de la cama en la imagen original. A continuación, se realizó de nuevo la predicción suprimiendo el *Bounding Box* del objeto de interés, de forma que las coordenadas de la cama fueran [0,0,0,0]. Cuando la red realizó la predicción original, devolvió una salida del 0.9920, donde 1 significa que la persona no está caída y 0 caída. Al eliminar el *Bounding Box* de la cama la predicción fue de 0.3529. Esto demuestra que el uso de *Bounding Box* en esta metodología tiene una repercusión significativa en la predicción.

Otro aspecto que se quería confirmar era la utilidad de la técnica de CVV en nuestra metodología. Por ello, se realizó la comparativa entre un modelo sin utilizar CVV, y el modelo utilizado en esta investigación con la técnica de CVV, entrenados con el mismo Dataset. En la Tabla 2 se muestran los resultados obtenidos con el conjunto de test en ambos modelos. Gracias al mecanismo de CVV, el efecto de los modelos que puedan fallar en la predicción se ven compensados, dando mejores resultados. Además, en la tabla se observa que el valor de F1-score es mayor, lo que indica que el modelo con CVV tiene un mejor equilibrio entre precisión y exhaustividad, logrando una clasificación más precisa y completa. Esto confirma la eficacia de la técnica de CVV para mejorar el rendimiento y la robustez de los modelos en tareas de clasificación complejas.

Tabla 2: Comparación modelos ConvNeXt con y sin CVV

	ConvNeXt	ConvNeXt – CVV
Accuracy	0.9124	0.9295
Loss	0.2117	0.1885
Recall	0.8767	0.8967
F1 – Score	0.8972	0.9052

En la Figura 5 se muestra un ejemplo en el que una persona apoyada sobre un sofá fue reconocida como caída por el modelo sin utilizar CVV. Al realizar la clasificación con el modelo que empleaba CVV y mostrar los resultados de cada modelo entrenado, se observó que había un modelo que estaba realizando una predicción errónea, sin embargo, el resto de modelos entrenados con diferentes *slot* clasificaron la imagen correctamente. Gracias a la técnica de *Soft Voting*, se realizó un promedio de las salidas de cada modelo y la predicción resultante fue no caída.

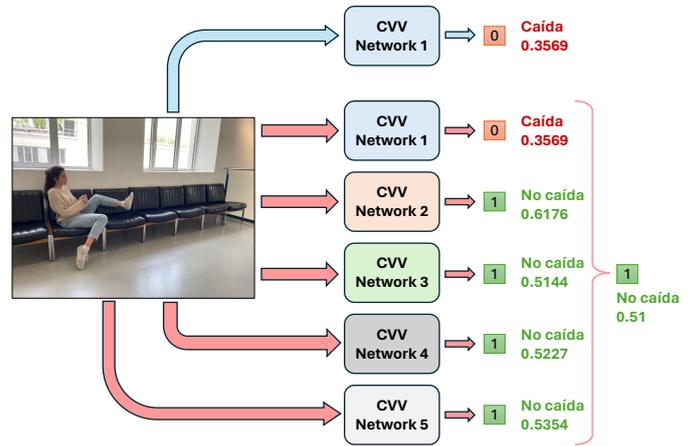


Figura 5: Predicción del modelo ConvNeXt con y sin CVV

Para comprobar el comportamiento del modelo con imágenes capturadas por el robot Temi, se realizaron pruebas con varias imágenes de personas caídas y no caídas tomadas por la cámara del robot. Esta misma verificación se realizó en el artículo (Sánchez-Girón et al., 2024), en el que se trabaja con un modelo Resnet, con el mismo conjunto de imágenes, y se obtuvieron resultados muy diferentes. En total se realizaron 118 pruebas con diferentes imágenes de personas caídas y no caídas. La exactitud para el modelo basado en Resnet y seg-

mentación fue de 87.39 % mientras que la exactitud del modelo ConvNeXt-CVV alcanzó un 96 %. Los resultados obtenidos superan a los de otros modelos como (Antonello et al., 2017) que emplea el dataset FPDS o el artículo (Martínez-Villaseñor et al., 2019) que emplea varias imágenes que hemos usado en el entrenamiento. Estos buenos resultados, que superan igualmente a los obtenidos con el modelo que utilizaba Resnet y segmentación, son debidos a la utilización del modelo ConvNeXt junto a la técnica *Cross Validation Voting*, que mejora la capacidad de generalización de los modelos neuronales al combinar los resultados de múltiples modelos entrenados con diferentes subconjuntos de datos. De esta forma, se mitiga el riesgo de sobreajuste y se consigue una evaluación más robusta y fiable del rendimiento del modelo.

4. Conclusiones

En el presente trabajo se propone una metodología completa para la detección de caídas utilizando la cámara integrada en un robot social. La detección se basa en el uso de modelos de aprendizaje profundo, específicamente utilizando la técnica de *Cross Validation Voting* (CVV), para mejorar la generalización y la precisión en la detección de caídas a partir de imágenes.

La metodología propuesta, que combina la detección de objetos con YOLO y el procesamiento mediante redes ConvNeXt, ha mostrado resultados muy satisfactorios en términos de precisión y capacidad de generalización. La integración de CVV ha sido clave para mitigar los problemas de sobreajuste y mejorar la robustez del modelo frente a datos desconocidos, lo que es crucial para su aplicación en entornos reales. Los resultados obtenidos durante la evaluación, con una exactitud del 92.95 %, indican que la metodología es efectiva para la clasificación de caídas, superando modelos anteriores.

Una ventaja de este proyecto es que en lugar de utilizar cámaras fijas, las fotos se capturan empleando un robot móvil. Como posible línea de investigación futura, se planea reubicar el robot en diferentes posiciones para que capture imágenes desde diferentes perspectivas. Esto podrá ser muy útil en caso de que la detección de personas o de caídas sea ambiguo y se podrá realizar una predicción más precisa.

En conclusión, la metodología propuesta no solo ofrece una solución técnica avanzada para la detección de caídas, sino que también contribuye significativamente a mejorar la seguridad y el bienestar de las personas mayores que viven solas. La implementación de este sistema en el proyecto EIA-ROB, que busca desarrollar un ecosistema de Inteligencia Ambiental con robots sociales, representa un avance importante hacia el cuidado y la autonomía de las personas mayores en sus hogares.

Agradecimientos

La investigación que se presenta en este trabajo ha recibido financiación del proyecto ROSOGAR PID2021-123020 OB-I00 financiado por MCI-

N/AEI/10.13039/501100011033/FEDER, UE, y del proyecto EIA-ROB Financiado por Consejería de Familia of the Junta de Castilla y León - Next Generation EU.

Referencias

- Antonello, M., Carraro, M., Pierobon, M., Menegatti, E., 2017. Fast and robust detection of fallen people from a mobile robot. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp. 4159–4166.
- Anwary, A. R., Rahman, M. A., Muzahid, A. J. M., Ul Ashraf, A. W., Patwary, M., Hussain, A., 2022. Deep learning enabled fall detection exploiting gait analysis. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). pp. 4683–4686. DOI: 10.1109/EMBC48229.2022.9871964
- Domingo, J. D., Aparicio, R. M., Rodrigo, L. M. G., 2022. Cross validation voting for improving cnn classification in grocery products. IEEE Access 10, 20913–20925.
- EIA-ROB, 2022. Ecosistema de inteligencia ambiental para el apoyo a los cuidados de larga duración en el hogar mediante uso de robots sociales. <https://www.itap.uva.es/en/eiarob/>.
- INE, 2023.
- Jocher, G., Chaurasia, A., Qiu, J., Jan. 2023. Ultralytics YOLO. URL: <https://github.com/ultralytics/ultralytics>
- Kwolek, B., Kepski, M., 2014. Human fall detection on embedded platform using depth maps and wireless accelerometer. Computer methods and programs in biomedicine 117 (3), 489–501.
- Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986.
- Maldonado-Bascon, S., Iglesias-Iglesias, C., Martín-Martín, P., Lafuente-Arroyo, S., 2019. Fallen people detection capabilities using assistive robot. Electronics 8 (9), 915.
- Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., Peñafort-Asturiano, C., 2019. Up-fall detection dataset: A multimodal approach. Sensors 19 (9), 1988.
- Merino-Fidalgo, S., Zalama, E., Gómez-García-Bermejo, J. and Duque-Domingo, J., Gómez, R., Viñas, P., García, D., Urueña, H., 7 2023. Sistema de monitorización no intrusiva para vivienda de personas mayores. Jornadas Nacionales de Robótica y Bioingeniería 2023: Libro de actas, 115–121. No publicado. DOI: 10.20868/UPM.BOOK.74896
- Quinayás Burgos, C. A., Quintero Benavidez, D. F., Ruíz Omen, E., Narváez Semanate, J. L., 2020. Sistema de detección de caídas en personas utilizando vídeo vigilancia. Ingeniare. Revista chilena de ingeniería 28 (4), 684–693.
- Rao, S., Pramod, N., Paturu, C. K., 2008. People detection in image and video data. In: Proceedings of the 1st ACM workshop on vision networks for behavior analysis. pp. 85–92.
- Rodríguez-Molinero, A., Narvaiza, L., Gálvez-Barrón, C., de la Cruz, J. J., Ruíz, J., Gonzalo, N., Valldosera, E., Yuste, A., 2015. Caídas en la población anciana española: incidencia, consecuencias y factores de riesgo. Revista Española de Geriatria y Gerontología 50 (6), 274–280.
- SAM, apr 2024. pose detection dataset. <https://universe.roboflow.com/sam-vcqdz/pose-detection-gwapj>, visited on 2024-05-16. URL: <https://universe.roboflow.com/sam-vcqdz/pose-detection-gwapj>
- Schaffer, C., 1993. Selecting a classification method by cross-validation. Machine learning 13, 135–143.
- Sánchez-Girón, C., García-Gómez, M., Duque-Domingo, J. and Gómez-García-Bermejo, J., Zalama, E., 5 2024. Detección de caídas con un robot social aplicando visión artificial. Jornadas Nacionales de Robótica y Bioingeniería 2024: Libro de actas. No publicado.
- Temi, R., 2024. URL: <https://www.robotemi.com/product/temi/>
- Workspace, jan 2024. Test2 dataset. <https://universe.roboflow.com/workspace-3pvv7/test2-fqv2c>, visited on 2024-05-16. URL: <https://universe.roboflow.com/workspace-3pvv7/test2-fqv2c>