



Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y
Tecnología

Máster Universitario en Inteligencia artificial

Detección temprana de Bullying por medio de CCTV mediante multimodalidad

Trabajo fin de estudio presentado por:	Dylan Camilo Poveda Villamizar
Tipo de trabajo:	Desarrollo de software
Director/a:	Felipe Mirón Pozo
Fecha:	04/03/2026

RESUMEN

Este Trabajo Fin de Máster, desarrollado en el marco del Máster en Inteligencia Artificial de la Universidad Internacional de La Rioja (UNIR), comprende el diseño y desarrollo de una solución de software orientada a la detección temprana de conductas de acoso escolar mediante el uso de sistemas de videovigilancia. El propósito principal es proporcionar a los docentes una herramienta tecnológica que permita identificar de manera oportuna posibles focos de acoso escolar, facilitando la ejecución de intervenciones pedagógicas preventivas y oportunas dentro del entorno educativo. El desarrollo del sistema se llevó a cabo bajo la metodología ágil Scrum, lo que permitió una evolución iterativa del producto basada en la validación continua con expertos del dominio. A nivel técnico, se implementó una arquitectura híbrida que integra redes neuronales convolucionales (CNN) para el análisis espacial de imágenes y técnicas de análisis de patrones temporales para el procesamiento de secuencias de video, permitiendo así una evaluación más completa del comportamiento observado. Como resultado, se obtuvo un sistema que, bajo la supervisión y validación de expertos, cumple con las expectativas técnicas y funcionales definidas inicialmente. Esta solución contribuye al fortalecimiento de los procesos de prevención del acoso escolar, permitiendo abordar situaciones de hostigamiento de manera proactiva y favoreciendo la construcción de entornos educativos más seguros y adecuados para el desarrollo integral de los estudiantes.

Palabras clave: Acoso, Inteligencia artificial, Notificación, Multimodal, Vision Artificial

ABSTRACT

This Master's Final Project, belonging to the Master's Degree in Artificial Intelligence at the International University of La Rioja (UNIR), includes the design and development of a software solution aimed at the early detection of bullying behaviors through the use of video surveillance systems. The main purpose is to provide teachers with a technological tool that allows them to timely identify potential sources of school bullying, facilitating the implementation of preventive and timely pedagogical interventions within the educational environment. The system was developed under the Scrum agile methodology, which allowed for an iterative evolution of the product based on continuous validation with domain experts. At a technical level, a hybrid architecture was implemented that integrates convolutional neural networks (CNN) for spatial image analysis and temporal pattern analysis techniques for video sequence processing, thus enabling a more comprehensive evaluation of observed behavior. As a result, a system was obtained that, under the supervision and validation of experts, meets the technical and functional expectations initially defined. This solution contributes to strengthening school bullying prevention processes, enabling proactive handling of harassment situations and promoting the creation of safer and more suitable educational environments for the comprehensive development of students.

Keywords: Bullying, Artificial intelligence, Notification, Multimodal, Artificial Vision

Índice de contenidos

1. Introducción.....	9
1.1. Motivación	9
1.2. Planteamiento del trabajo.....	10
2. Contexto y estado del arte	11
2.1. Bullying.....	11
2.2. Prevención e identificación de bullying.....	12
2.3. Circuitos cerrados de televisión y legislación.....	13
2.4. Estado del arte.....	14
2.4.1. Desarrollo de un modelo de visión por computadora con modelos de redes recurrentes para identificación de situaciones de violencia.....	14
2.4.2. Detección de violencia en secuencias de imágenes utilizando redes neuronales.	15
2.4.3. Modelo de clasificación de imágenes violentas basado en un modelo de detección de objetos.....	15
2.4.4. Systematic mapping study on violence detection in video by means of trustworthy artificial intelligence	16
2.4.5. A spatio-temporal model for violence detection based on spatial and temporal attention modules and 2D CNNs	17
2.5. Conclusiones del estado del arte	18
3. Objetivos concretos y metodología de trabajo	20
3.1. Objetivo general.....	20
3.2. Objetivos específicos	20
3.3. Metodología	20
3.3.1. Introducción	20
3.3.2. Justificación	21

3.3.3.	Fases del Proyecto.....	22
3.3.4.	Herramientas y Técnicas.....	22
4.	Identificación de requisitos.....	25
4.1.	Contextualización del problema	25
4.1.1.	Naturaleza del problema	25
4.1.2.	Impacto y Relevancia.....	25
4.2.	Contexto de uso.....	26
4.2.1.	Escenarios	26
4.2.2.	Limitaciones y desafíos.....	26
4.3.	Metodología e identificación de requisitos	27
4.3.1.	Aporte de Expertos	27
4.3.2.	Análisis de Entrevistas.....	27
4.4.	Requisitos identificados	29
4.4.1.	Requisitos funcionales	29
4.4.2.	Requisitos no funcionales	29
4.5.	Conclusiones	30
5.	Desarrollo de la herramienta de software	31
5.1.	Desarrollo.....	31
5.1.1.	Diseño y Prototipado.....	31
5.1.2.	Desarrollo Iterativo	33
5.1.3.	Entrenamiento de modelo de detección de acoso	34
5.1.4.	Integración y Pruebas.....	39
5.1.5.	Lanzamiento.....	40
5.2.	Arquitectura	41
5.2.1.	Componentes principales	41

5.2.2.	Capa de usuario.....	43
5.3.	Funcionamiento.....	44
5.3.1.	Captura de video	44
5.3.2.	Procesamiento y análisis de video	45
5.3.3.	Mensajería.....	49
6.	Evaluación	53
6.1.	Evaluación Inicial	53
6.1.1.	Usabilidad	53
6.1.2.	Funcionalidad.....	54
6.1.3.	Rendimiento.....	55
6.1.4.	Contribución al campo.....	57
6.2.	Evaluación cualitativa.....	58
6.2.1.	Metodología de evaluación.....	58
6.2.2.	Preocupaciones éticas	58
6.2.3.	Conclusiones de expertos.....	59
7.	Conclusiones y Trabajo futuro	60
7.1.	Conclusiones.....	60
7.2.	Lineas de Trabajo Futuro.....	61
	Bibliografía	63
	Anexo A. Código	66

Índice de Figuras

Figura 1. Análisis estado del arte.....	19
Figura 2. Arquitectura de Prototipo.....	33
Figura 3. Desarrollo iterativo.....	34
Figura 4. Construcción Dataset propio.....	35
Figura 5. Evolución de Precisión.....	37
Figura 6. Análisis de overfitting.....	38
Figura 7. Proceso de desarrollo.....	40
Figura 8. Arquitectura de Software.....	43
Figura 9. Diagrama Integración de modelo.....	47
Figura 10. Notificación sistema de detección.....	51
Figura 11. Matriz de Confusión.....	56

Índice de Tablas

Tabla 1. Configuración de modelo.....	36
Tabla 2. Resultados Entrenamiento..	37
Tabla 3. Métricas en bullying.	55

1. INTRODUCCIÓN

1.1.MOTIVACIÓN

A lo largo de la historia la violencia siempre ha acompañado a la humanidad (Gil Villa, 2020), puede considerarse inútil la tarea de determinar el origen de esta con los eventos cotidianos donde se evidencia la violencia día a día (Masip, 2016). En cuanto a la violencia escolar, no solo responde a la agresión sino también a estructuras de poder y exclusión. Adicionalmente como lo plantea Gil villa, atacar estas conductas violentas debe hacerse desde la educación, sin embargo, la violencia también nace en entornos como son las instituciones educativas colegios, universidades, institutos, etc. (Gil Villa, 2020).

Por otra parte, para los docentes surge la necesidad de identificar y evitar que surjan actitudes de violencia tanto física como psicológicamente. La prevención o la intervención de bullying no deben limitarse a la reacción ante los casos visibles, sino ser parte del trabajo pedagógico y del clima escolar. Los docentes, al ser los primeros observadores de los estudiantes, desempeñan un papel clave en la detección temprana de comportamientos de violencia (Barri Batero, 2017).

Según los algunos estudios y análisis los hechos de violencia y bullying se presentan en distintos lugares de una entidad educativa como sitios de recreo, áreas de juego o canchas deportivas, pero también en sitios poco vigilados como pueden ser los pasillos, escaleras o donde se presente la ausencia de alguna autoridad (Boone, 2019).

Esta realidad motiva el desarrollo de herramientas y estrategias que fortalezcan la capacidad de los educadores para reconocer, analizar y prevenir el acoso, contribuyendo así a una educación más humana, inclusiva y libre de violencia.

1.2.PLANTEAMIENTO DEL TRABAJO

Lo que se propone en este trabajo es realizar un desarrollo de software a modo de prototipo que permita al profesorado de una institución educativa realizar una identificación rápida de focos de violencia o bullying dentro de la institución educativa. Esta sería una herramienta que permita a los profesores abordar más prontamente a su alumnado y así mediante pedagogía ir reduciendo el bullying dentro de las instituciones educativas.

Para el desarrollo de la herramienta se plantea construir un modelo de identificación de comportamientos mediante diferentes modelos de inteligencia artificial o multimodalidad que permitan interpretar el bullying desde diferentes contextos, como lo puede ser representar hostigamiento como pueden ser empujar, agarrar, golpear. Este modelo se plantea ser entrenado con un conjunto de imágenes y videos de violencia que pueda identificar estos comportamientos.

Posteriormente se plantea crear una aplicación que alerte al profesorado de comportamientos que puedan ser identificados como algún tipo de bullying. La aplicación tendría la capacidad de detectar en vivo alguna situación indebida y por medio de una notificación o alguna herramienta móvil alertar a los docentes.

2. CONTEXTO Y ESTADO DEL ARTE

2.1. BULLYING

El acoso, referido comúnmente como bullying, es un fenómeno que ha generado una gran preocupación y ha impulsado un rápido desarrollo en la investigación durante las últimas dos décadas. (Hershcovis, 2011) Se trata de un problema muy frecuente y alarmante en la sociedad, incluyendo el entorno escolar (Garaigordobil, 2018). Para abordarlo de manera efectiva, es fundamental llevar a cabo una identificación correcta de los casos y distinguir una situación de acoso de un simple conflicto entre pares, ya que la violencia escolar no siempre constituye abuso (Gippini, 2018). En términos generales, el bullying se define como una relación cotidiana en la que se ejerce un abuso sistemático de poder.

Otra característica distintiva es la intencionalidad de hacer daño. El agresor o agresores quieren hacer daño de forma consciente a la víctima. Aunque los perpetradores a menudo intentan justificar sus acciones mediante justificaciones vagas o en broma o simplemente que todos lo hacen, el bullying generalmente se asume como un comportamiento de alta intensidad con una clara intención de causar perjuicio. Estos actos negativos que se repiten pueden incluir la persecución física y/o psicológica, abuso constante, burlas ofensivas, ridículo, o la exclusión social (Hershcovis, 2011). Cuando el acoso ocurre, no se trata de un problema puntual, y sus secuelas pueden ser extremadamente graves, llegando a afectar el desarrollo psicológico del niño de por vida.

El acoso, referido comúnmente como bullying, es un fenómeno que ha generado una gran preocupación y ha impulsado un rápido desarrollo en la investigación durante las últimas dos décadas. Se trata de un problema muy frecuente y alarmante en la sociedad, incluyendo el entorno escolar (Garaigordobil, 2018). Para abordarlo de manera efectiva, es fundamental llevar a cabo una identificación correcta de los casos y distinguir una situación de acoso de un simple conflicto entre pares, ya que la violencia escolar no siempre constituye abuso. En términos generales, el bullying se define como una relación cotidiana en la que se ejerce un abuso sistemático de poder.

Dentro del bullying existen diferentes tipos de acoso:

- **Acoso físico:** El acoso físico se produce cuando el agresor realiza ataques físicos directos contra la víctima, los cuales pueden ir precedidos de ataques psicológicos en las etapas previas del acoso. El acoso físico forma parte de los tipos de maltrato o abuso entre iguales, definidos como una conducta de persecución que un estudiante ejerce contra otro, al que elige víctima de ataques repetidos (Galán, 2018).
- **Acoso gestual:** Este tipo de acoso ocurre cuando el agresor utiliza gestos o expresiones no verbales que buscan intimidar y causar miedo a la víctima (Galán, 2018).
- **Acoso social:** El acoso social se caracteriza porque el agresor o agresores aísla/n a la víctima del grupo. Esta exclusión social puede generar una disminución de la autoestima en la víctima, forjando una imagen negativa de sí misma y dificultándole la solicitud de ayuda (Galán, 2018).
- **Amenazas:** Cuando se ejerce acoso mediante amenazas, la víctima es obligada a realizar acciones que van en contra de su voluntad. Si la víctima no cumple con lo que se le exige, tendrá consecuencias negativas, lo que provoca que se sienta humillada y reduce las posibilidades de que hable con adultos, como padres o profesores, para poner fin a la situación (Galán, 2018).
- **Ciberacoso:** Cuando se ejerce acoso mediante amenazas, la víctima es obligada a realizar acciones que van en contra de su voluntad. Si la víctima no cumple con lo que se le exige, tendrá consecuencias negativas, lo que provoca que se sienta humillada y reduce las posibilidades de que hable con adultos, como padres o profesores, para poner fin a la situación (Galán, 2018).

2.2.PREVENCIÓN E IDENTIFICACIÓN DE BULLYING

La correcta identificación de una situación de acoso es fundamental, ya que es necesario distinguir cuándo se trata de una situación de bullying y cuándo es simplemente un conflicto entre iguales (Hershcovis, 2011). El acoso físico se produce cuando el agresor realiza ataques físicos directos contra la víctima, los cuales pueden ir precedidos de ataques psicológicos en etapas previas del acoso. Los casos más graves de este tipo de maltrato incluyen el abuso constante, los golpes y las palizas, o pueden llegar a tipificarse como actos de lesiones físicas. Para que una conducta sea considerada acoso escolar, debe poseer características específicas: debe ser una conducta sistemática y recurrente que se prolonga en el tiempo; debe existir un abuso o desequilibrio en el

uso del poder entre el agresor y la víctima; y debe haber una intencionalidad consciente de hacer daño por parte del agresor o agresores. El desequilibrio de poder es clave, ya que, si dos estudiantes de la misma fuerza se pelean, no se considera maltrato por acoso (Galán, 2018).

La prevención del bullying debe ser una labor constante y transversal durante todo el curso académico (Boone, 2019), con el objetivo de garantizar que el acoso no se va a producir. Dado que se ha comprobado que la violencia se aprende observando modelos agresivos, y que los niños sumergidos en ambientes familiares, escolares o comunitarios violentos tienden a reproducir acciones violentas, es crucial desarrollar acciones psicoeducativas y programas de intervención socioemocional que fomenten la convivencia pacífica. Específicamente, para prevenir las conductas físicas violentas, es importante que los estudiantes adquieran habilidades que les ayuden a enfrentarse a situaciones difíciles y de conflicto de forma pacífica y sin recurrir a la violencia. El refuerzo positivo de las actitudes adecuadas para la convivencia es fundamental. Cuando surgen conflictos, debe fomentarse la resolución mediante el diálogo y el método socioafectivo, ayudando a los niños a empatizar para que puedan ponerse en el lugar de la otra persona, lo que reduce la violencia como respuesta. Además, los recreos requieren especial atención para la prevención, ya que son lugares donde se produce el 76% de los maltratos en la escuela primaria; por lo tanto, es vital mejorar la supervisión en estos espacios y diversificar las actividades ofrecidas para evitar el aburrimiento y la agresión (Galán, 2018).

2.3.CIRCUITOS CERRADOS DE TELEVISIÓN Y LEGISLACIÓN

En Colombia, la instalación de cámaras de seguridad en las instituciones educativas oficiales está con el propósito de proteger la integridad de los estudiantes y disminuir los factores de inseguridad y violencia en los entornos escolares (Alcaldía Mayor de Bogotá D.C, 2011). La vigilancia se enfoca en las zonas comunes del y los entornos o alrededores de las instituciones. Sin embargo, el sistema en ningún momento está concebido para monitorear los salones de clase, ya que esto vulnera derechos fundamentales como el habeas data, la intimidad, la libertad de cátedra y la libre expresión (Corte Constitucional, 2025).

2.4.ESTADO DEL ARTE

2.4.1. Desarrollo de un modelo de visión por computadora con modelos de redes recurrentes para identificación de situaciones de violencia

El propósito general de este proyecto desarrollar una solución para la detección de situaciones de violencia por medio de videos tomados por cámaras de vigilancia, utilizando herramientas de visión por computadora. El proyecto buscaba automatizar la identificación de eventos violentos a partir del análisis de imágenes capturadas por cámaras de vigilancia. Para lograr este objetivo, se plantearon la obtención de una base de datos con videos de comportamientos violentos y no violentos, el desarrollo y la evaluación de un modelo de clasificación, y la implementación de un sistema de alarma (Bravo, 2023).

El proyecto se centró en la implementación de modelos de aprendizaje profundo, combinando Redes Neuronales Convolucionales (CNN) y Redes Neuronales Recurrentes (LSTM). La arquitectura específica utilizada integró MobileNetV2 con una red LSTM para permitir un análisis preciso y eficiente de las imágenes en tiempo real. MobileNetV2 se utilizó para la extracción de características espaciales, y la LSTM se encargó del análisis de la información temporal de los cuadros del video. Para el entrenamiento, se empleó una base de datos compuesta por cuatro conjuntos de datos diferentes: UCF Crimes, RWF 2000, Hockey Fight y Movies Fight (Bravo, 2023).

Al finalizar este trabajo se indica que se logró una precisión promedio del 81.6 %. El modelo implementado en la Raspberry Pi 4 demostró ser una opción viable, capaz de clasificar imágenes en un tiempo promedio de 5.180 segundos. Para futuros trabajos, se sugiere varias mejoras, incluyendo la optimización de recursos de hardware, la expansión de la plataforma de comunicación más allá de Telegram, el uso de múltiples cámaras y la implementación de multimodalidad, como la detección de objetos y la integración de la detección de sonido para mejorar la precisión de las alertas (Bravo, 2023).

2.4.2. Detección de violencia en secuencias de imágenes utilizando redes neuronales.

En este trabajo se realizó el diseño, implementación y evaluación de un sistema para la detección automática de situaciones violentas a partir de secuencias de video. Se incluyó el desarrollo de un sistema para la adquisición de información con una cámara RGBD ZED 2i y la evaluación de un modelo de deep learning para la detección de violencia, adaptándolo si era necesario (Baeza, 2024).

Las técnicas de identificación de imágenes utilizadas se basaron en la estimación de pose y el cálculo de la diferencia entre imágenes consecutivas. El modelo analiza dos flujos: uno para la información espacial utilizando redes neuronales convolucionales y otro para la información temporal usando una ConvLSTM. La base de datos utilizada para la validación fue RWF-2000. Para evaluar la robustez del sistema, se probó su funcionamiento reduciendo el número de puntos clave de los esqueletos sin reentrenar la red, presentando esqueletos de 18 puntos clave en lugar de los 25 utilizados en el entrenamiento original del modelo (Baeza, 2024).

Las conclusiones principales señalan que el modelo es robusto frente a alteraciones en los datos de entrada, ya que predice con una exactitud del 85,25%. Esta exactitud fue solo un 5% inferior a la obtenida con los 25 keypoints originales, indicando que el modelo funciona correctamente a pesar de los cambios. Como opciones de mejora, se plantea un reentrenamiento del modelo partiendo de sus pesos ya entrenados y utilizando los esqueletos de más puntos clave para mejorar la exactitud. Además, se propone la grabación de una nueva base de datos con violencia, usando la cámara ZED 2i y el sistema de adquisición desarrollado, para complementar las situaciones violentas del entrenamiento original y reforzar el aprendizaje del modelo con los nuevos esqueletos (Baeza, 2024).

2.4.3. Modelo de clasificación de imágenes violentas basado en un modelo de detección de objetos.

El propósito del modelo desarrollado fue que permitiera identificar la violencia mediante algoritmos de deep learning por medio de detección de objetos para facilitar la clasificación de imágenes violentas. El contexto específico de la clasificación de las imágenes fue el conflicto

armado en Colombia. Se recolectaron imágenes para construir el dataset, determinar los objetos violentos representativos, implementar y verificar modelos de clasificación, y comparar su eficiencia para determinar el modelo óptimo (Niño, 2022).

Las técnicas de identificación de imágenes se centraron en la detección de objetos. Se identificaron 15 objetos considerados violentos y representativos para la detección de violencia. Posteriormente, se etiquetaron manualmente estos objetos en cada imagen utilizando la herramienta LabelImg y el formato VOC pascal. Se implementaron y compararon tres modelos de detección de objetos de una etapa: YOLOv4, RetinaNet y SSD MobileNet. El entrenamiento se realizó con 4400 imágenes de uso libre (70% para entrenamiento y 30% para validación), con 9000 épocas de iteración (Niño, 2022).

Después del análisis de eficiencia se determinó que el modelo óptimo es YOLOv4, el cual alcanzó una precisión de aproximadamente el 90% después de las 9000 épocas de entrenamiento con un 92,35% en training y un 81,77% en validación. Además, YOLOv4 demostró un menor valor de pérdida (5.41) y un tiempo de convergencia más rápido (6 horas) en comparación con RetinaNet y SSD MobileNet. Como trabajo futuro, se propone fortalecer la recolección de imágenes diversificando las palabras clave para incrementar el dataset y así lograr un mejor entrenamiento. Adicionalmente, se sugiere contar con una máquina más robusta para la ejecución óptima de los modelos y mejorar los tiempos computacionales (Niño, 2022).

2.4.4. Systematic mapping study on violence detection in video by means of trustworthy artificial intelligence

El propósito de este estudio fue investigar si la predicción de violencia podía lograrse a partir de fotogramas individuales sin necesidad de análisis temporal utilizando solo redes neuronales convolucionales, evaluar la ventaja de combinar CNN con Bi-LSTM sobre CNN con LSTM, y examinar el impacto de los hiperparámetros y la comparación de VGG-19 vs. VGG-16 en la precisión (Negre, 2024).

La arquitectura VGG-19 preentrenada en el dataset ImageNet para la extracción de características espaciales. Se propusieron y probaron cuatro arquitecturas. VGG-19 solo el cual analiza fotograma por fotograma, VGG-19 con Lógica Manual definiendo un umbral mínimo de cuadros violentos

para clasificar el video, VGG-19 con capas LSTM VGG-19 con capas Bi-LSTM. Para el entrenamiento y prueba se utilizaron tres bases de datos: Hockey Fights, Violent Flow y RWF-2000. Para las arquitecturas con capas recurrentes, se modificaron los hiperparámetros para maximizar la precisión (Negre, 2024).

El trabajo concluye que el modelo VGG-19 con lógica manual superó en precisión a las combinaciones con LSTM y Bi-LSTM en dos de los tres datasets (RWF-2000 y Violent Flow), lo que sugiere que la detección de violencia podría no siempre requerir capas temporales complejas. Se lograron precisiones del 97% en Hockey Fights usando VGG-19 con Bi-LSTM. Se confirmó que las capas Bi-LSTM obtienen ligeramente mejores resultados que LSTM. Un hallazgo notable fue que, a pesar de tener más capas convolucionales, VGG-16 superó a VGG-19 en la detección de violencia en video, lo que sugiere que una mayor complejidad del modelo no siempre se traduce en mejor rendimiento. Adicionalmente para trabajos futuros se propone aplicar modelos de Inteligencia Artificial Confiable, específicamente modelos de explicabilidad como GradCAM, para comprender qué partes de la imagen son relevantes para la toma de decisiones, ya que los modelos basados en CNN y CNN+LSTM no proporcionan información sobre sus decisiones (Negre, 2024).

2.4.5. A spatio-temporal model for violence detection based on spatial and temporal attention modules and 2D CNNs

El propósito principal de este trabajo es desarrollar un sistema de detección de violencia automático y eficiente para sistemas de vigilancia, capaz de operar en tiempo real sin la alta complejidad computacional de las redes 3D. Para lograrlo, se propone un modelo espacio-temporal que utiliza una red neuronal convolucional CNN como núcleo, complementada con módulos diseñados para extraer características de video de manera efectiva. El sistema se divide en tres componentes clave: el módulo espacial de entropía para resaltar regiones de movimiento, el bloque de atención temporal de compresión para resumir la secuencia de video en tres canales procesables por la red neuronal. (Mahmoodi, 2024)

En cuanto a las técnicas de identificación, el módulo espacial de entropía utiliza un filtro de entropía aplicado a las diferencias entre tramas consecutivas para enfocar la atención en los píxeles

con mayor movimiento, lo cual es característico de las escenas violentas. Por otro lado, el bloque de atención temporal emplea agrupación promedio global y capas densas para asignar pesos de importancia a los fotogramas más informativos antes de su compresión. Para el entrenamiento, se utilizaron imágenes provenientes de cuatro conjuntos de datos de referencia de videos de hockey. Estas imágenes consisten en tramas de video redimensionadas y procesadas en formato de escala de grises para calcular las diferencias de movimiento y la entropía antes de ser alimentadas al modelo. (Mahmoodi, 2024)

Las conclusiones del estudio destacan que el modelo propuesto logra un rendimiento superior en términos de precisión en comparación con métodos tradicionales como redes 3D CNN actuales, alcanzando hasta un 99.7% en su dataset. El trabajo demuestra que es posible capturar información espacio-temporal compleja reduciendo significativamente el número de parámetros, lo que hace al sistema más práctico para aplicaciones reales. Como trabajo futuro, los autores planean proponer nuevas arquitecturas para entrenar modelos aún más confiables con menos parámetros y ampliar la investigación hacia una variedad de tareas de reconocimiento de acciones adicionales para diversos propósitos (Mahmoodi, 2024).

2.5. CONCLUSIONES DEL ESTADO DEL ARTE

El análisis de los estudios previos evidencia un avance significativo en el ámbito de la detección automatizada de violencia genérica en video, pero al mismo tiempo revela limitaciones fundamentales que justifican y demandan un enfoque específico para el acoso escolar. En primer lugar, los modelos que constituyen el actual estado del arte están predominantemente orientados hacia la identificación de agresiones físicas explícitas y de alta intensidad, como peleas abiertas, golpes contundentes o forcejeos evidentes. Esto resulta insuficiente para abordar la compleja realidad del bullying, donde las manifestaciones suelen ser más sutiles, contextuales y prolongadas en el tiempo como lo pueden ser empujones camuflados como accidentales, bloqueos de paso, aislamiento social deliberado, robos o daños sutiles de pertenencias. Esta brecha entre la violencia genérica y el acoso escolar implica la necesidad de trascender el modelo binario de clasificación de ser violento o no, hacia un sistema capaz de discernir y categorizar comportamientos específicos, como acoso físico indirecto, acoso social o intimidación psicológica, cada uno con sus propios patrones espaciotemporales.

Por otra parte, tras la revisión de los estudios señala una clara oportunidad la integración multimodal de señales complementarias. Investigaciones eficientes como la detección de objetos contextuales como maletas y dispositivos móviles ayudan a inferir agresiones, mientras que el uso del análisis postural para capturar dinámicas interpersonales también identifica estos actos violentos. Paralelamente, en otros trabajos se subraya la importancia crítica de la para dar transparencia a las decisiones del modelo. Por consiguiente, se hace imperativo desarrollar una aproximación multimodal que abarque y fusione diferentes componentes del bullying como son la acción corporal, el contexto material y la configuración grupal para lograr una comprensión holística y robusta de las situaciones de acoso.

Como aporte fundamental y síntesis de esta necesidad, este trabajo de grado propone el desarrollo de un sistema de detección que no se limite a la aplicación directa de arquitecturas existentes tal como se observa en la Figura 1, sino que diseñe una solución ad-hoc a partir de un dataset propio y contextualizado de bullying escolar. La novedad radicará en una arquitectura de fusión que combine, de manera interpretable, el análisis de gestos y posturas, la detección de objetos relevantes en el entorno escolar y el modelado de las dinámicas temporales de grupo, priorizando no solo la precisión, sino también la capacidad de generar alertas explicables y éticamente responsables para su implementación en entornos educativos reales.



Figura 1. Análisis estado del arte. Fuente: elaboración propia.

3. OBJETIVOS CONCRETOS Y METODOLOGÍA DE TRABAJO

3.1.OBJETIVO GENERAL

Desarrollar una herramienta que permita a los educadores de alguna institución educativa identificar de manera temprana hecho de hostigamiento o bullying entre estudiantes

3.2.OBJETIVOS ESPECÍFICOS

- O1 Analizar literatura científica y técnica más actualizada asociada a la lucha contra el bullying con apoyo de herramientas de inteligencia y visión artificiales.
- O2 Realizar la recolección y extracción de imágenes a partir de videos, aplicando técnicas de preprocesamiento, con el fin de generar un conjunto de datos adecuado para el entrenamiento.
- O3 Desarrollar un modelo de identificación de hostigamiento basada en técnicas de inteligencia artificial, utilizando algoritmos de visión artificial y Deep learning.
- O4 Integrar el modelo de clasificación e identificación de Bullying con una herramienta que permita informar a los docentes mediante alguna alerta a su dispositivo móvil sobre situaciones detectadas por el modelo.
- O5 Realizar análisis previo y evaluación del desarrollo de software con expertos en educación tales como docentes y personal de instituciones educativas.

3.3.METODOLOGÍA

3.3.1. Introducción

En el contexto actual de la información digital, la rápida difusión de contenidos a través de plataformas tecnológicas plantea importantes desafíos relacionados con la desinformación. Ante esta problemática, resulta fundamental contar con herramientas capaces de realizar procesos de monitorización, análisis y clasificación de contenidos de forma eficiente y en tiempo casi real. Por

eso se es necesario implementar una herramienta que ayude a los docentes de una institución educativa en el proceso de contar con información rápida y veraz. El desarrollo de este trabajo implica no solo la aplicación de técnicas avanzadas de Inteligencia Artificial, sino también la construcción de una solución tecnológica flexible, escalable y orientada al usuario. La naturaleza cambiante de los contenidos analizados y la necesidad de incorporar mejoras de manera continua hacen indispensable adoptar un enfoque metodológico que permita iterar, validar resultados y adaptarse a nuevos requerimientos conforme evoluciona el proyecto.

Por este motivo, se ha optado por complementar el enfoque general del trabajo con una metodología ágil basada en Scrum, la cual resulta especialmente adecuada para proyectos de desarrollo de software. Scrum permite estructurar el trabajo en ciclos iterativos e incrementales, facilitando la entrega progresiva de funcionalidades de una aplicación, así como la incorporación constante de retroalimentación. Este enfoque favorece la adaptación temprana a cambios, la priorización de requisitos y la mejora continua del producto, alineándose con los objetivos del proyecto y con la necesidad de evolución constante de la herramienta desarrollada.

3.3.2. Justificación

La elección de la metodología Scrum se basa no únicamente en los aspectos como su desarrollo incremental donde se puede empezar con un mínimo viable y realizar un escalamiento de la herramienta a partir de este punto, como también la flexibilidad y adaptabilidad al cambio, sino también a la experiencia que se cuenta de esta metodología proponiendo tiempos de sprint adecuados para el desarrollo de puntos de desarrollo de la herramienta.

Por este motivo, se ha optado por complementar el enfoque general del trabajo con una metodología ágil basada en Scrum, la cual resulta especialmente adecuada para proyectos de desarrollo de software. Scrum permite estructurar el trabajo en ciclos iterativos e incrementales, facilitando la entrega progresiva de funcionalidades dentro de la aplicación. Este enfoque favorece la adaptación temprana a cambios, la priorización de requisitos y la mejora continua del producto, alineándose con los objetivos del proyecto y con la necesidad de evolución constante de la herramienta desarrollada.

3.3.3. Fases del Proyecto

- Planificación y definición de requisitos:
 - a. Previo al desarrollo se realiza investigación de modelos relacionados y sugerencias dentro de literatura científica.
 - b. Se realizan entrevistas a expertos para la definición de requisitos y necesidades puntuales del caso.
- Diseño:
 - a. Arquitectura del sistema: Se realiza arquitectura base del proyecto con el cual se identifican los componentes clave del sistema y su interacción entre sí.
- Desarrollo:
 - a. Sprints de desarrollo: Se plantean sprints de dos semanas por cada avance iterativo, esto con el fin de llevar mejoras iterativas y continuas que garanticen el avance del proyecto.
 - b. Retrospectiva con expertos: Tras cada sprint se realiza una pequeña retroalimentación con expertos.
- Pruebas y validación con expertos:
 - a. Despliegue de la solución: Se realiza un despliegue de la solución para que pueda ser probado por expertos
 - b. Evaluación de expertos: Se realizará una evaluación con expertos con el fin de validar los resultados finales del sistema e identificar oportunidades de mejora futura. Este proceso se llevará a cabo mediante reuniones en las que se presentará el funcionamiento general de la herramienta. Durante estas sesiones, se comparará el comportamiento del sistema frente al listado de requisitos definidos, los cuales representan las necesidades planteadas por los expertos. Esta validación se apoyará en cuestionarios de satisfacción que permitirán medir el nivel de usabilidad, funcionalidad y cumplimiento de los objetivos de la herramienta.

3.3.4. Herramientas y Técnicas

- Lenguajes y Frameworks
 - a) Python:

Es un lenguaje de programación de alto nivel, interpretado y ampliamente utilizado en aplicaciones web, ciencia de datos y machine learning. Se destaca por ser eficiente, fácil de aprender gracias a su sintaxis similar al inglés, y por contar con una vasta biblioteca estándar de código reutilizable (Amazon Web Services, 2025).

b) OpenCV

Es una biblioteca de código abierto que integra cientos de algoritmos de visión artificial mediante una estructura modular. Facilita tareas que van desde el procesamiento básico de imágenes y video hasta la calibración de cámaras y reconocimiento de objetos (OpenCV, 2025).

c) PyYAML

Esta herramienta para Python funciona como un analizador y emisor de datos en formato YAML. Está diseñada para la serialización de datos de forma que sean legibles por humanos, siendo útil tanto para archivos de configuración complejos como para la persistencia de objetos (Simonov, 2025).

- Modelos Preconfigurados

a) YOLO

Se trata de un modelo avanzado de visión artificial especializado en la detección de objetos y segmentación de imágenes en tiempo real. Su arquitectura destaca por tratar la detección como un único problema de regresión, lo que permite un alto rendimiento en velocidad y precisión (Ultralytics, 2025).

b) MediaPipe

Es una herramienta que ofrece diversas soluciones tecnológicas relacionadas con la inteligencia artificial, destacando entre sus funciones la documentación específica para la estimación de poses. Al ser un repositorio público y colaborativo, permite a la comunidad de desarrolladores gestionar flujos de trabajo, realizar revisiones de código y participar activamente en el desarrollo de aplicaciones de IA (Google AI Edge, 2025).

- Hardware y Conectividad

a) Camara de Vigilancia:

Para el desarrollo e implementación de la solución se utilizará una cámara de seguridad TPLink la cual permite conexión directa con protocolo RSTP, la cual permite el procesamiento en vivo de la imagen.

b) Telegram

Telegram es una aplicación de mensajería enfocada en la velocidad y seguridad, esta aplicación también permite la integración de bots e interacciones con credenciales seguras para su uso. (Telegram, 2025)

- Herramientas de Desarrollo

a) Visual Studio Code

Visual Studio Code es un editor de código fuente gratuito, que permite escribir, depurar y probar código libremente, incluye librerías y extensiones que permiten el desarrollo en un entorno integrado y potente (Microsoft, 2025).

b) Git

Es un sistema de control de versiones que permite la manipulación de código plano. Al integrarse a plataformas como Github permite realizar la distribución de código con equipos de trabajo a la vez que permite la integración y despliegue continuos con otras herramientas en ambientes DevOps. (Github, 2025)

4. IDENTIFICACIÓN DE REQUISITOS

La identificación de requisitos es fundamental en el proceso del desarrollo de un software, ya que en estos se identifican las necesidades y más importantes las expectativas del usuario objetivo del desarrollo. Con ayuda de esta identificación se construye todas las etapas del desarrollo.

Para el caso puntual de la herramienta para la detección temprana de bullying por medio de CCTV, la identificación de los requisitos implica entender la problemática social del acoso escolar en diferentes escuelas y como un sistema de identificación de este pueda apoyar a la institución educativa a minimizar estos casos.

4.1.CONTEXTUALIZACIÓN DEL PROBLEMA

De acuerdo con el estado de arte realizado, el bullying es un problema muy presente en todos los países y principalmente el foco del trabajo se enfoca en las instituciones educativas, donde los estudiantes se encuentran en una etapa en donde se pueden seguir formando y a su vez con ayuda de profesionales como lo pueden ser psicólogos se pueden encontrar patrones que puedan deducir de donde es el origen del acoso a otros estudiantes.

4.1.1. Naturaleza del problema

El bullying puede presentarse en distintas categorías como se vio en el estado del arte, como lo pueden ser social, verbal, ciberacoso y físico entre otros. Sin embargo, la idea del proyecto contempla el uso de cámaras que se encuentren en una institución educativa para abordar este bullying y ser una respuesta temprana a estos comportamientos. Sin embargo, en estos casos el acoso que se puede observar a través de una cámara puede ser en muchos casos físico en el cual se enfocara este desarrollo.

4.1.2. Impacto y Relevancia

El acoso en las instituciones educativas no debe considerarse simplemente como un comportamiento natural entre los jóvenes, ya que también puede reflejar situaciones externas que afectan a los estudiantes, como problemas familiares o contextos ajenos al entorno escolar.

En las etapas tempranas de formación, el bullying puede generar graves consecuencias psicológicas en las víctimas y, al mismo tiempo, reforzar conductas de hostigamiento en los agresores. Estas acciones pueden repercutir de manera significativa en el desarrollo personal y en los resultados del proceso formativo de una persona.

4.2. CONTEXTO DE USO

Para el desarrollo de la herramienta de identificación adicional al entendimiento del problema es necesario contar con el contexto de los expertos, así mismo entender sus necesidades, expectativas y escenarios ideales en los cuales a la herramienta debería funcionar. Para ello es necesario identificar tres áreas fundamentales.

4.2.1. Escenarios

- **Tiempo de uso:** La herramienta debe estar activa en ciertos espacios de tiempo determinados teniendo en cuenta que en las instituciones educativas no siempre tendrá el flujo de estudiantes.
- **Lugares estratégicos:** Se requieren puntos estratégicos de identificación y ubicación de nuestros sensores en este caso las cámaras de seguridad que proporcionen una entrada efectiva al sistema.
- **Análisis de hallazgos:** La respuesta de la herramienta debe entregarse por algún medio digital que permita la captura y análisis del modelo de identificación de acoso sea alguna herramienta de mensajería o una aplicación dedicada.

4.2.2. Limitaciones y desafíos

- **Privacidad y ética:** De acuerdo con lo investigado en el estado del arte existen ciertas legislaciones que intervienen cuando se tratan de instituciones educativas tales como los sitios autorizados para que estas puedan ser instaladas.
- **Limitación de datasets:** Ya que se tienen pocos accesos a videos de niños se tienen pocos dataset clasificados donde se pueda identificar el acoso en estudiantes.

- **Diversidad de interacciones:** Al tratarse de una institución educativa se pueden observar múltiples interacciones entre estudiantes como el juego o situaciones que no corresponden a acoso escolar.

4.3.METODOLOGÍA E IDENTIFICACIÓN DE REQUISITOS

La identificación de los requisitos es un compendio de detalles técnicos, pero también características deseadas por los expertos para garantizar que la herramienta sea un sistema que apoye efectivamente el proceso pedagógico que se realice en las instituciones educativas y no exista sesgo a la hora de interpretar los resultados esperados de la herramienta.

4.3.1. Aporte de Expertos

La participación de los expertos en este trabajo fue fundamental ya que son una fuente invaluable de experiencia y conocimiento acerca del acoso en las instituciones educativas como en este caso lo fue los docentes y rector del Colegio Real de Cundinamarca, institución que se encuentra dentro del casco urbano de la ciudad de Bogotá capital de Colombia.

Para la identificación de requisitos se trabajaron diferentes varias sesiones en las cuales se realizaban las entrevistas pertinentes acerca de su experiencia con el acoso escolar, como se abordaba actualmente los casos de bullying y cuáles eran sus expectativas acerca de la herramienta a desarrollar.

4.3.2. Análisis de Entrevistas

El bullying o acoso escolar es una problemática frecuente en las instituciones educativas. Por esta razón, se realizó una entrevista al rector y los docentes de una institución educativa objeto de estudio ubicada en Bogotá D.C., Colombia. Durante la entrevista se realizaron varias preguntas que permitieron conocer desde un entorno real como se detecta el bullying y como lo afrontan.

Es importante resaltar que el entorno social y familiar en el que viven los estudiantes repercute directamente en su forma de relacionarse en diferentes situaciones con sus compañeros. Adicionalmente se evidencio que en colegios de menor tamaño es más sencillo identificar y controlar las situaciones de bullying, ya que se encuentran grupos reducidos de estudiantes lo que

facilita el seguimiento de los casos y tomar acciones oportunas antes de que se agraven dichas situaciones.

Se identifico que el bullying verbal es la forma de acoso más frecuente y ocasionalmente si no se detecta a tiempo se puede escalar progresivamente hacia agresiones físicas, si no se detectan a tiempo. Los principales mecanismos de detección son gracias a la observación directa del docente, tanto en el aula como en los recreos. El docente al estar en contacto permanente y cercano con los estudiantes son quienes pueden identificar y hacer seguimiento de la convivencia que se presenta entre ellos. Por esto se destaca que es importante que los docentes cuenten con formación y capacitación necesaria para saber reconocer y prevenir estas situaciones.

Por otro lado, el canal de reporte principal es el docente, seguido del estudiante y en ocasiones las familias. En ocasiones los estudiantes le cuentan al docente en confianza sus preocupaciones en situaciones de bullying, esto permite una intervención más rápida y efectiva y la atención a estos casos es de forma gradual, lo primero es siempre iniciar con el dialogo el cual puede o no escalar a sanciones o intervención psicológica, esto depende de la respuesta de los estudiantes. Sin embargo, se prioriza un enfoque pedagógico realizando actividades de integración, trabajo en equipo y dinámicas que fomenten la convivencia, si el problema de bullying no desiste se escala a psicología para trabajarlo en conjunto con las familias y, de no evidenciarse mejoras, se aplican medidas como sanciones disciplinarias, matrícula condicional y, en casos extremos, la no renovación del cupo escolar.

Un tema importante a tener en cuenta son los espacios donde más se presentan las situaciones de bullying y en ambas entrevistas se coincide que el recreo es el espacio donde hay mayor vulnerabilidad para la aparición de situaciones bullying. Esto puede deberse a que hay menos supervisión adulta y a su vez las interacciones son más espontaneas por lo que puede hacerse más visible las dinámicas de poder, exclusión, burlas o rechazo, los recreos son espacios donde hay libre expresión, sin embargo, aunque el aula de clase es un entorno más controlado por la presencia de un docente también se visibiliza señales en etapas tempranas de bullying solo que de formas más sutiles.

Por ultimo y no menos importante se resalta que el uso de la tecnología para la detección del bullying por medio de cámaras, puede servir como apoyo para la identificación de posibles

situaciones de bullying pero la intencionalidad, el contexto emocional y las causas del bullying deben hacerse con seguimiento y de la mano del docente, directivos y psicólogos, ya que las cámaras no interpretan la situación completa o no brindan la solución al problema, sería un trabajo conjunto entre tecnología y humanos. Así que se considera que un sistema de notificación mediante mensajería instantánea permitiría alertar a los docentes de manera oportuna y activar el conducto regular correspondiente.

4.4.REQUISITOS IDENTIFICADOS

De acuerdo con el estado de arte realizado y con base en las entrevistas con los expertos se identificaron los requisitos mínimos tanto técnicos como funcionales que el desarrollo de la herramienta de detección de acoso debería cumplir para llegar a considerarse usable y funcional.

4.4.1. Requisitos funcionales

Los requisitos funcionales describen las funcionalidades específicas que la herramienta debe entregar al usuario.

- **Captura en vivo:** La herramienta debe ser capaz de procesar el video captado por la cámara de vigilancia en tiempo muy cortos, esto con el fin de que el encargado tome acciones preventivas sobre alguna eventualidad.
- **Detección de personas:** El sistema debe ser capaz de identificar las personas que se encuentren implicadas en una situación de bullying.
- **Identificación de bullying:** El sistema debe realizar la interpretación visual del bullying que pueda ser captado en cámara dentro de los espacios disponibilizados para ello.
- **Notificación sencilla:** La herramienta debe generar un mensaje sencillo que disponga información precisa y útil a manera de alerta. Dicha notificación deberá hacerse por medio de alguna aplicación de mensajería instantánea.

4.4.2. Requisitos no funcionales

Por otra parte, los requisitos no funcionales se refieren a las características a nivel de calidad y eficiencia que la herramienta debe tener.

- **Rendimiento:** La herramienta debe ser capaz de procesar los videos capturados por la cámara y entregar la notificación en un corto periodo de tiempo.
- **Usabilidad:** La notificación entregada a los docentes debe contener información eficaz y veraz, tal que con esta información se pueda tomar acciones pedagógicas.
- **Funcionalidad:** La herramienta debe tener un índice de confiabilidad suficiente tal que no genere grandes cantidades de alertas que puedan ser falsos positivos.

4.5.CONCLUSIONES

Esta etapa fue fundamental para el desarrollo de la herramienta de software, en esta etapa nos brinda la información necesaria para la construcción de las tareas necesarias para la consecución del proyecto. Por otra, parte nos permite con ayuda de la metodología escogida que en este caso es scrum realizar un planteamiento de estos requisitos para ser desarrollados en sprints.

Dentro del análisis posterior a la captura de requisitos se pueden destacar algunos puntos específicos en los cuales debemos centrar el trabajo:

- **Complejidad del problema:** El acoso escolar no radica únicamente en un conflicto sencillo de personas en formación, este problema va mucho más allá de la identificación prematura del acoso, debe estar acompañado por una pedagogía acorde que promueva la sana convivencia dentro y fuera de las instituciones.
- **Experiencia y colaboración:** Para garantizar que cada uno de los requisitos técnicos y funcionales de la aplicación se cumplan, es necesario contar con la retroalimentación de los expertos, esto garantiza que las decisiones tomadas en el desarrollo tengan fundamento.
- **Información esencial:** La información entregada debe ser tan practica para que el encargado que pueda revisar esta notificación sea capaz de interpretar rápidamente una situación, esto es esencial ya que el propósito de este trabajo es identificar prematuramente hechos de violencia y acoso.

5. DESARROLLO DE LA HERRAMIENTA DE SOFTWARE

Ya que el acoso está presente en diferentes ámbitos de la vida cotidiana, puntualmente en las instituciones educativas como colegios o universidades donde los estudiantes aún están en proceso de formación, surge la necesidad de abordar, identificar y tener una respuesta rápida ante este tipo de situaciones.

La herramienta desarrollada dentro del marco de este trabajo aborda la necesidad del profesorado de una institución educativa en identificar oportunamente situaciones que puedan representar hechos de bullying dentro de la institución, esta herramienta cuenta con un modelo de inteligencia artificial que permite a los profesores ser alertados por medio de mensajería instantánea de dichos eventos y sea convertida en una herramienta de apoyo para disminuir y abordar el acoso de manera más oportuna.

En este capítulo se dará a conocer el desarrollo de la herramienta desde el planteamiento de la arquitectura, hasta su implementación y pruebas, dentro de los cuales vamos a ver los retos asumidos y superados que llevaron a la finalización del proyecto

5.1.DESARROLLO

5.1.1. Diseño y Prototipado

Con los requisitos y objetivos del proyecto definidos el siguiente paso fue el prototipado de la herramienta iniciando por una arquitectura inicial que nos permitiera inicial con el desarrollo del proyecto. Para esta fase se contempló los requisitos mínimos viables y las necesidades y alcances que podría llegar a tener una institución educativa como un colegio en este caso el Colegio Real de Cundinamarca.

El diseño de la herramienta se enfocó en el desarrollo de diferentes fases de desarrollo en los cuales en la última fase se lograrán integrar, la principal idea era que cada componente funcionara de manera independiente a las otras, con esto en mente se garantizaría la identificación de errores oportunamente y seguiríamos el patrón Scrum donde lograríamos realizar un mínimo viable entregable dentro de cada uno de los sprints realizados.

Bajo la arquitectura del prototipo el cual se encuentra en la Figura 2 se realizaron las primeras validaciones y planear los desarrollos a abordar y los puntos clave que se deberían atender primero para abordar ya que podrían implicar en desafíos como el análisis de imagen en tiempo real y su correcto procesamiento.

El desarrollo se llevó a cabo en diferentes fases de las cuales cada una de estas implicaba una necesidad y complejidad diferentes.

- **Captura de video:** El primer paso fue desarrollar una integración entre la captura del video y la capa backend garantizando entrega de información en tiempo real. Esta fase conlleva a cambios en cuanto a protocolos de comunicación efectivos y la captura de información efectiva para el modelo.
- **Construcción capa backend:** Una vez construida la capa de captura de imagen se tenía que construir la capa de backend la cual consistía en capturar y seccionar las imágenes para ser procesadas por el modelo de inteligencia artificial construido. Esta fase conlleva al procesamiento de la imagen en tiempo real.
- **Construcción modelo de identificación:** Para la construcción del modelo de identificación fue necesario analizar las técnicas necesarias para el procesamiento de las imágenes a procesar, así mismo fue necesario realizar el análisis de los videos que iban a ser procesados en el entrenamiento y pruebas de este. Con esto en mente se seleccionó un modelo que permitiera el manejo de secuencias para ser trabajado en tiempo real.
- **Envío de notificación:** Para esta fase fue necesario realizar el análisis de las diferentes herramientas de mensajería instantánea que permitieran la comunicación efectiva con nuestro backend y que fuera conocida popularmente, permitiera realizar una integración efectiva. Finalmente, se escogió una herramienta que cumplía con estos requisitos y a su vez el factor económico era beneficioso para el desarrollo del proyecto.

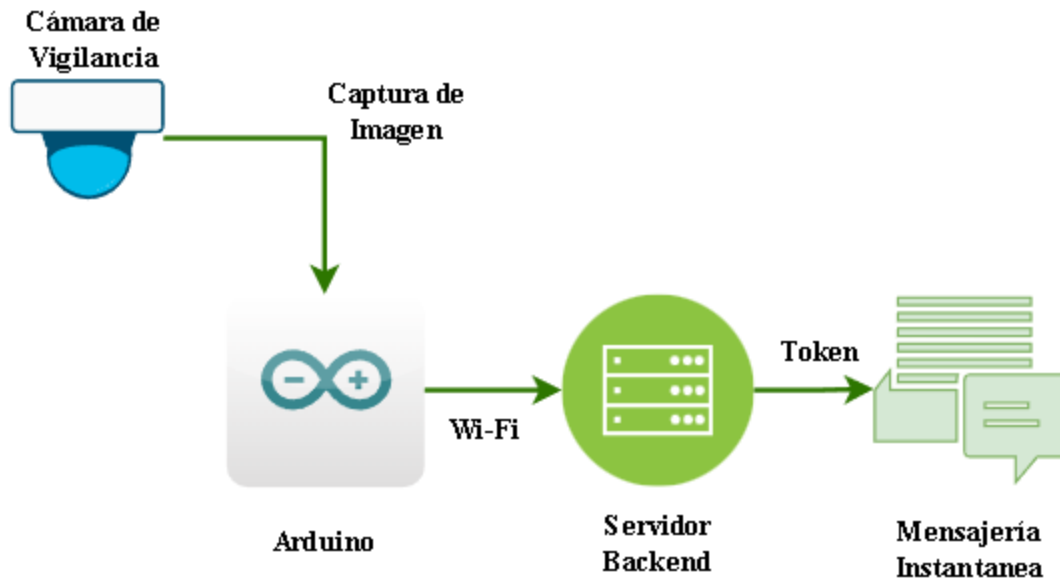


Figura 2. Arquitectura de Prototipo. Fuente: elaboración propia.

5.1.2. Desarrollo Iterativo

Ya que el proyecto está fundamentado por la metodología ágil scrum, el desarrollo iterativo fue crucial para plantear sprints realistas, garantizando que cada una de las etapas del desarrollo se llevara a cabo en los plazos propuestos. La metodología ágil nos permitió llevar a cabo la consecución del desarrollo del proyecto y posteriormente pruebas y puesta en marcha.

En Scrum la metodología nos permite dividir el proyecto en varios sprint lo cual en este caso permitía dividir el trabajo en periodos de dos semanas, lo cual permitía desarrollar un mínimo viable. Por otra parte, él también se debía contemplar los tiempos de ajuste de los desafíos encontrados por ellos se destinó una capacidad exclusiva para la atención de los hallazgos encontrados, esto se tomó inicialmente en un 20% de la capacidad destinada para este proyecto, posteriormente con los desarrollos más aterrizados a la realidad esta capacidad se redujo a un 10%.

Debido a que el desarrollo de la herramienta solo fue realizado por un desarrollador las sesiones de retrospectiva se tomaron para realizar feedbacks con expertos que en este caso actuaban como stakeholders, los cuales con sus comentarios ayudaban a orientar el desarrollo de la herramienta

hacia lo deseado. En la Figura 3 se puede encontrar la síntesis del desarrollo iterativo que se trabajará bajo el marco de trabajo Scrum.

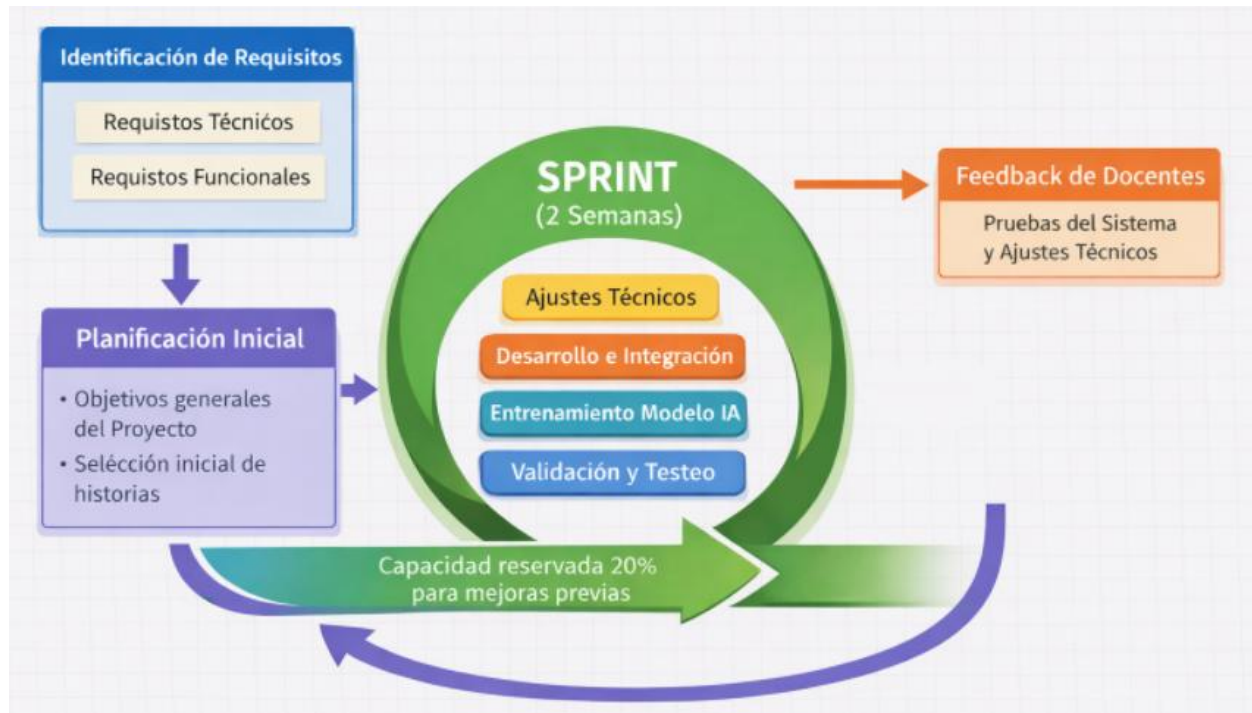


Figura 3. Desarrollo iterativo Fuente: elaboración propia

5.1.3. Entrenamiento de modelo de detección de acoso

El entrenamiento del modelo de detección de bullying se fundamenta en una arquitectura híbrida que combina un modelo de YOLO para la detección de objetos y personas con MediaPipe para el análisis de poses corporales. La metodología adoptada sigue un enfoque de aprendizaje supervisado donde se utilizan datasets etiquetados que contienen escenas de violencia, posturas agresivas y situaciones de bullying. El proceso de preparación de datos incluye la recopilación de videos de cámaras de seguridad simuladas, grabaciones controladas en entornos educativos y datasets públicos de detección de violencia.

La construcción del dataset requirió un proceso de etiquetado que involucró la colaboración con los docentes de la institución educativa tal como muestra la Figura 4, donde se puede evidenciar diferentes escenarios desde distintas posiciones. Se recopilaron aproximadamente 15,000 secuencias de video de 30 segundos cada una, provenientes de diferentes fuentes: simulaciones controladas en centros educativos, datasets académicos públicos como Real Life Violence Situations Dataset (Elesawy, 2025) y Violence Detection Video Classification (Unique Data, 2025) adaptados al contexto escolar.



Figura 4. Construcción Dataset propio. Fuente: elaboración propia

La fase de preprocesamiento de datos constituye un componente crítico del pipeline de entrenamiento, donde se implementan técnicas de augmentación de datos para incrementar la robustez del modelo ante variaciones de iluminación, ángulos de cámara y condiciones ambientales. Se aplican transformaciones como rotación ($\pm 15^\circ$), escalado (0.8-1.2x), ajuste de brillo y contraste ($\pm 20\%$), así como técnicas de recorte aleatorio para simular diferentes posiciones de cámara.

El balanceado de clases representa un desafío particular en este dominio debido a la naturaleza desbalanceada de los datos, donde las situaciones de bullying son significativamente menos frecuentes que las interacciones normales, por lo que los otros dataset son fundamentales para orientar el entrenamiento del modelo y validar la premisa de que el bullying se puede identificar como violencia ordinaria.

El entrenamiento del modelo YOLO se ejecuta utilizando transfer learning a partir de pesos preentrenados, aplicando fine-tuning específico para la detección de personas en contextos escolares y objetos potencialmente peligrosos como armas blancas o contundentes. Se emplea una función de pérdida compuesta que combina la pérdida de clasificación, localización y confianza, optimizada mediante el algoritmo AdamW.

Con el objetivo de optimizar el rendimiento del modelo híbrido CNN-LSTM propuesto, se diseñó y ejecutó un protocolo experimental sistemático que contempló cuatro configuraciones distintas de hiperparámetros tal como se muestra en la Tabla 1. Cada experimento fue concebido para evaluar el impacto de modificaciones específicas en la arquitectura y el proceso de entrenamiento, permitiendo identificar la configuración óptima que maximice la capacidad de generalización del modelo mientras se minimiza el sobreajuste. Las configuraciones experimentales incluyeron: (1) una configuración baseline que establece los parámetros de referencia, (2) un modelo de mayor capacidad con incremento en las unidades ocultas y capas LSTM, (3) una estrategia de pesos balanceados para abordar el desbalance de clases y (4) una tasa de aprendizaje elevada para acelerar la convergencia. Cada experimento se entrenó durante 30 épocas utilizando el optimizador Adam.

Configuración	Hidden Size	Layers	Dropout	Batch Size	Learning Rate	Epochs	Weights
1	256	2	0.5	4	0.001	30	0.4-0.6
2	512	3	0.5	4	0.001	30	0.4-0.6
3	256	2	0.5	4	0.001	30	0.5-0.5
4	256	2	0.5	4	0.005	30	0.4-0.6

Tabla 1. Configuración de modelo. Fuente: elaboración propia.

Posterior al entrenamiento se registraron métricas de precisión tanto en el conjunto de entrenamiento como en el de validación. A continuación, se presentan los resultados consolidados de cada configuración experimental, destacando la precisión máxima alcanzada en validación, la época en la que se obtuvo el mejor desempeño, y la precisión final en el conjunto de entrenamiento tal como se puede evidenciar en la Tabla 2, lo cual permite realizar un análisis comparativo exhaustivo del comportamiento de cada variante del modelo, realizando un análisis de los resultados obtenidos se puede determinar que incrementar la complejidad arquitectónica o la tasa de aprendizaje no garantiza una mejora en el desempeño, y que una configuración moderada con

adecuada regularización y balance de clases ofrece mejores resultados. El modelo híbrido CNN-LSTM alcanza su mayor capacidad de generalización bajo parámetros controlados, evidenciando la importancia de un ajuste cuidadoso de hiperparámetros para maximizar el accuracy y minimizar riesgos de inestabilidad o sobreajuste.

Configuración	Best Epoch	Best Accuracy	Final Accuracy
1	28	95.15	90.44
2	1	62.39	60.94
3	24	95.38	90.32
4	1	62.39	60.94

Tabla 2. Resultados Entrenamiento. Fuente: elaboración propia.

La selección del modelo 3 se fundamentó en un análisis del comportamiento de la precisión durante el proceso de entrenamiento a lo largo de las 30 épocas planificadas. Se monitoreó continuamente la evolución de la precisión tanto en el conjunto de entrenamiento como en el de validación para identificar el punto óptimo antes de que se manifestara overfitting,

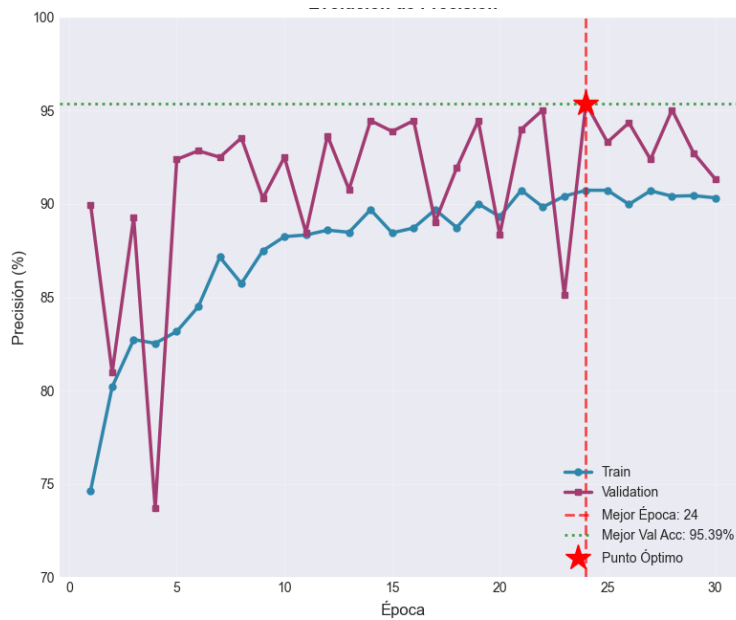


Figura 5. Evolución de Precisión. Fuente: elaboración propia

La Figura 5 muestra la evolución del entrenamiento a lo largo de las épocas del entrenamiento. La decisión de implementar early stopping permitió escoger el mejor modelo durante el entrenamiento

cuando la precisión de validación dejaba de mejorar durante épocas consecutivas, evitando así que el modelo memorizara los datos de entrenamiento en lugar de aprender patrones generalizables. Este enfoque garantizó que el modelo seleccionado mantuviera un balance adecuado entre capacidad de aprendizaje y generalización, resultando en un sistema capaz de detectar situaciones de bullying en escenarios reales no vistos durante el entrenamiento, sin sacrificar rendimiento por sobreajuste a los datos de entrenamiento.

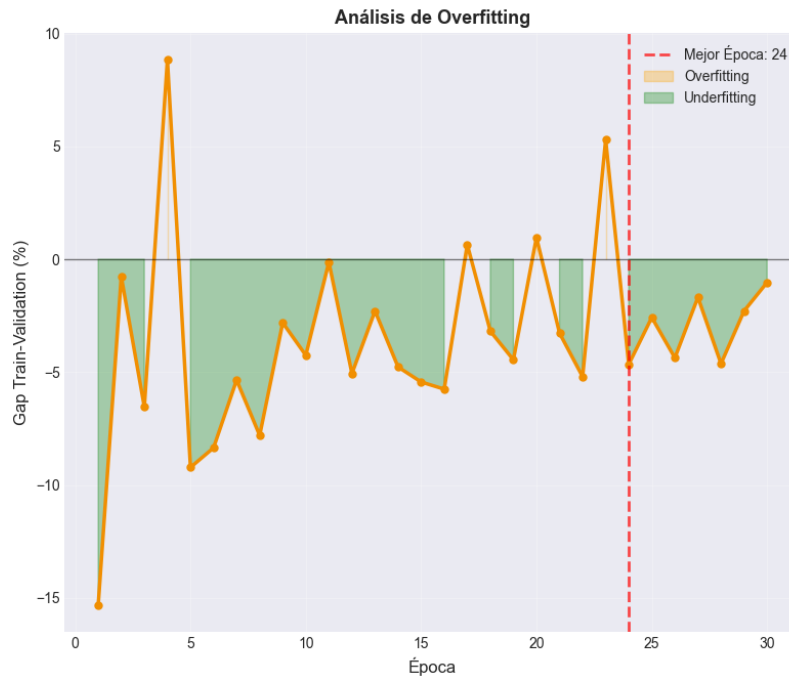


Figura 6. Análisis de overfitting. Fuente: elaboración propia.

La Figura 6 muestra el gap de entrenamiento vs validación a lo largo de las épocas, el cual representa la diferencia entre el desempeño del entrenamiento vs su validación, tal cual se puede observar en las primeras épocas se presenta una alta variabilidad, para posteriormente entre las épocas 5 a la 15 aproximadamente se presenta un registro negativo indicando que todavía el modelo no estaba sobre ajustado y se encuentra progresando, a partir de la época 20 se ven variaciones más marcadas donde se ve un sobreajuste leve. En la misma Figura 6, se puede observar que la mejor época es la época 24 donde el gap es modelado y alcanzo un buen equilibrio.

5.1.4. Integración y Pruebas

Una vez las cada una de las fases del proyecto fueron ejecutadas era necesario realizar la integración de cada uno de los componentes tal cual muestra la Figura 7. Realizada la integración de los componentes era necesario realizar pruebas de integración que garantizaría la sinergia de cada uno de los componentes desarrollados en las diferentes fases.

En la integración de cada uno de estos componentes se identificaron falencias y componentes de integración que no se habían profundizado mucho en cómo sería su interconexión, para ellos realizo un refinamiento de la arquitectura tal que estos componentes se logran comunicar entre sí.

Posterior al desarrollo y empalme de los componentes se realizan pruebas de integración y de funcionamiento para verificar cada uno de los funcionamientos en conjunto, incluyendo pruebas de captura de video, procesamiento de imagen, detección del modelo y notificación el usuario.

Realizado estas pruebas de integración se procede a realizar pruebas de uso en sitio de la aplicación realizando pruebas en un ambiente controlado realizando simulaciones de los eventos causantes del bullying en instituciones educativas, con esto en mente se realizan retroalimentación que alimenta y refina el desarrollo para un mejor funcionamiento.

Esta etapa es esencial en el proceso de refinamiento y optimización de la herramienta de software no solo a nivel tecnológico sino también a nivel funcional llegando a cumplir con todos los requisitos del usuario que en este caso son los docentes de una institución educativa.

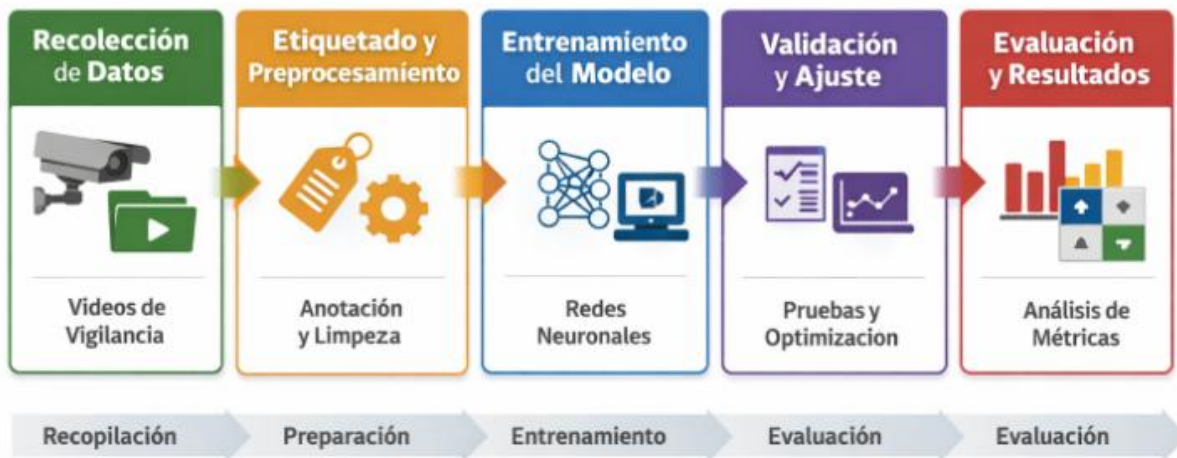


Figura 7. Proceso de desarrollo. Fuente: elaboración propia

5.1.5. Lanzamiento

Con las fases de integración y pruebas fueron finalizadas completamente, y tras superar las fases de validación y pruebas funcionales correspondientes, se procederá a su lanzamiento y despliegue en un entorno de producción controlado. Esta etapa contempla no solo la puesta en funcionamiento de la herramienta, sino también la definición de mecanismos de monitorización y seguimiento que permitan evaluar su desempeño y efectividad en un contexto educativo real.

La aplicación será configurada y desplegada garantizando la correcta comunicación entre la captura del video, el procesamiento de la imagen, el análisis de la secuencia de imágenes para posterior entregar la notificación a Telegram, a través de la cual se enviarán las notificaciones a los docentes. Se brindará una infraestructura que permita un funcionamiento en los horarios más requeridos, disponibilidad y capacidad de procesamiento para el análisis de las en tiempo casi real.

Durante el proceso de monitorización se realizará el seguimiento de métricas técnicas relevantes, tales como tiempos de respuesta del sistema, estabilidad de la comunicación entre la cámara – backend y backend - Telegram, consumo de recursos y registro de errores. De igual manera, se tendrán en cuenta métricas funcionales asociadas al uso de la herramienta, como la cantidad de alertas generadas, frecuencia de detecciones y patrones de interacción por parte de los docentes.

Con base en los resultados obtenidos de la monitorización y el feedback recopilado, se podrán planificar ajustes y mejoras progresivas en la aplicación, tanto a nivel del modelo refinado y entrenado como de la experiencia de uso.

5.2.ARQUITECTURA

La arquitectura de software se refiere al proceso de diseño y desarrollo a alto nivel de un sistema de software como también busca definir la estructura, vista y comportamiento del sistema. Teniendo en cuenta a cuáles son las soluciones necesarias para cumplir con los requisitos técnicos y comerciales de un sistema de software (UNIR, 2023), en este capítulo se describirá como fue construida la herramienta que permitirá solventar los requerimientos identificados en capítulos anteriores.

5.2.1. Componentes principales

Debido a que el desarrollo fue realizado por fases nos permite describir cada uno de los componentes principales de la aplicación los cuales se pueden ver representados en la Figura 8:

- 1. Captura de video:** En esta fase se trabajó la captura de información que iba a ser procesada por nuestro modelo de inteligencia artificial en este caso una cámara que permitiera protocolo RSTP, el cual permite controlar la transmisión de medios audiovisuales en tiempo real. El correcto funcionamiento de este componente garantiza el buen procesamiento de las fases posteriores, para ello se garantiza la correcta conexión y buen flujo de datos entre el backend y nuestro sensor que en este caso es una cámara Tplink C211.
- 2. Procesamiento de imagen:** Una vez realizada la captura del video es necesario realizar el procesamiento de la imagen para que pueda ser procesada por el modelo de detección. Para ello este procesamiento de imagen contempla el uso de la librería OpenCV para realizar la decodificación de frames en tiempo real, al mismo tiempo se contempla el uso de un buffer de imágenes para evitar el lag del video.

- 3. Modelo de inteligencia artificial:** En esta fase se trabaja en el corazón del desarrollo, el entrenamiento del modelo LSTM para detección de bullying requiere varios componentes esenciales que trabajan de forma secuencial. Primero, se necesita un dataset balanceado con videos etiquetados de comportamiento clasificado como acoso y otro clasificado como normal. Estos videos son procesados por para extraer las secuencias temporales, aplicando data augmentation (rotaciones de $\pm 5^\circ$ y ajustes de brillo) para aumentar la diversidad del dataset. La arquitectura del modelo combina un extractor de características ResNet18 preentrenado que procesa cada frame individualmente, seguido de una red LSTM bidireccional que analiza las dependencias temporales entre frames, y finalmente un clasificador fully-connected que produce la predicción binaria bullying/normal. Para el entramiento se hicieron uso de datasets públicos como los creados exclusivamente para este proyecto como lo fue el dataset de situaciones reales de violencia (Elesawy, 2025) y el dataset usado para clasificación de comportamientos violentos (Unique Data, 2025).
- 4. Servicio de notificación:** Una vez el modelo entrega el resultado binario del análisis de la imagen es necesario realizar una notificación a un grupo de Telegram en el cual se encontrarán los usuarios designados que reciben la notificación y alerta sobre un posible evento de acoso donde la cámara está instalada. Para ello se disponibilizó un Bot que recibe la información de la respuesta del modelo por medio de API keys. El mensaje se personaliza con la información que necesita el docente para realizar su respectiva inspección visual del hecho.
- 5. Base de datos:** Con el objetivo de obtener estadísticas se implementó una base de datos no relacional que capturaba el número de personas implicada y la fecha del evento incluyendo la hora, esto con el objetivo de obtener estadísticas reales de la detección de bullying de nuestro modelo y así mismo refinarlo con tal de suplir las necesidades de la institución.

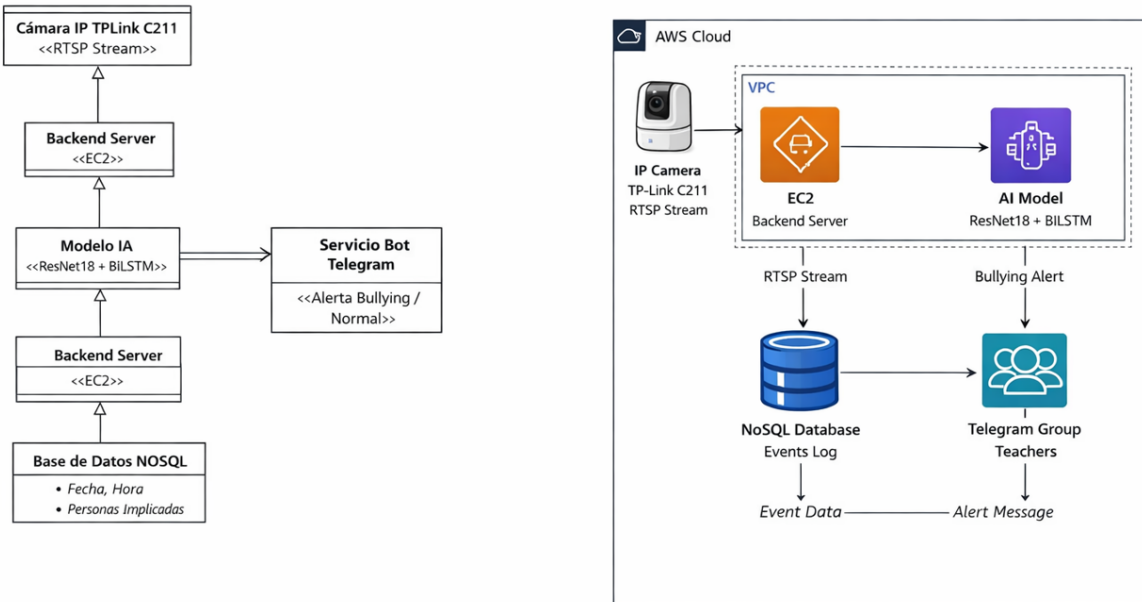


Figura 8. Arquitectura de Software. Fuente: elaboración propia

5.2.2. Capa de usuario

La Herramienta desarrollada principalmente y debido a las necesidades de los usuarios que en este caso requerían una interfaz de usuario que les permitiera tener la información era necesario el uso de una aplicación de mensajería instantánea, inicialmente se contempló el uso de WhatsApp ya que es la aplicación más popular dentro del contexto de Colombia lugar donde se encontraba el colegio objetivo. Sin embargo, se optó por el uso de Telegram por factores como el económico ya que permite la integración abierta y gratuita, tiene buen soporte a desarrolladores y contiene bibliotecas en varios lenguajes y a comparación de WhatsApp permite una integración rápida sin tener un workflow más burocrático como lo exige WhatsApp Bussiness.

Con el propósito de disponibilizar la información capturada en eventos de acoso en la institución educativa se creó un grupo en Telegram, en el cual se disponibilizaría toda la siguiente información a los docentes dentro del grupo.

- **Inicio de aplicación:** La aplicación con el objetivo de alertar a los docentes de que el sistema empezó a realizar su detección, envía un mensaje de inicio de detección.

- **Mensaje de detección:** Una vez se realiza la detección de un evento de acoso, se envía una notificación de alerta al canal del grupo, este mensaje indica la fecha, el nivel de confianza, es decir que tan probable es que sea bullying y la cantidad de personas implicadas.
- **Imagen del evento:** Para un entendimiento visual se brinda al usuario una imagen que captura el momento en el cual se realizó la detección, esta imagen viene acompañada de una guía de identificación de personas y la confirmación del mensaje de detección. No se realiza un ahondamiento en la información entregada al usuario con el fin de que la notificación entregada sea fácil de entender y el usuario pueda tomar acciones preventivas que eviten o escalen a un grado mayor el conflicto detectado.

5.3.FUNCIONAMIENTO

Para la descripción del funcionamiento de la herramienta se separó en los tres grandes componentes de la aplicación que son captura de información, procesamiento y análisis y notificaciones.

El aplicativo fue realizado por medio de lenguaje Python utilizando diferentes librerías para la construcción de los diferentes componentes dentro del aplicativo. En este capítulo se describirá cada uno de los artefactos desarrollados para el funcionamiento de cada uno de los elementos que componen esta herramienta de software.

5.3.1. Captura de video

Para la configuración y captura de información de la cámara, se realizó un artefacto encargado de la transmisión y procesamiento de la imagen llamado 'camera_capture'.py estos son los detalles del desarrollo:

1. **Importación de librerías:** Para este caso se realiza la importación de las diferentes librerías necesarias para la configuración y conexión de la cámara por medio de protocolo RTSP y así mismo el procesamiento de la imagen, en este caso se va a utilizar las librerías de OpenCV, Threading, Queue, Time.

2. **Configuración paramétrica:** Se realiza configuración paramétrica, con el fin de realizar la conexión de las credenciales, se realiza configuración del buffer para contener la cantidad de frames que iban a ser utilizados.
3. **Conexión RTSP:** La configuración mediante RSTP es muy sencilla ya que el sistema consiste en la exposición de un endpoint disponibilizado por un puerto en específico en el cual se indican las credenciales. Estas credenciales deben ser parametrizadas en la cámara, en este punto es importante recalcar que este tipo de conexión es posible ya que la cámara cuenta con tecnología para conexión por medio de Wifi.
4. **Obtención de frames:** Para la obtención de frames se realiza un tomado de video a 30FPS dato bajo el cual se realizaron pruebas de conectividad y lectura de imagen correcta, para ello se realizó a la lectura de una cola en la cual se capturaban los frames y nos servía de buffer para evitar lag a la aplicación y que a futuro el modelo lograra una mejor interpretación.
5. **Detención:** Se realiza un método de configuración para parar el proceso de transmisión con el fin de parar la obtención de frames.

5.3.2. Procesamiento y análisis de video

En esta sección se detallará el flujo de los frames obtenidos en la fase de captura del video para posteriormente ser analizada por un modelo de detección binaria de bullying.

1. **Recepción de imagen:** La recepción de imagen constituye el punto de entrada del sistema de procesamiento, donde cada frame capturado es recibido como una matriz RGB de dimensiones variables que debe ser estandarizada para el análisis posterior. El sistema implementa un buffer temporal circular mediante deque(maxlen=16) que mantiene una ventana deslizante de los últimos 16 frames consecutivos, redimensionando cada imagen a 1224x224 píxeles para cumplir con los requisitos de entrada del modelo ResNet18. Este buffer funciona como una cola FIFO donde cada nuevo frame desplaza al más antiguo,

asegurando que siempre se mantenga una secuencia temporal continua de exactamente 16 frames para el análisis de patrones de movimiento.

El preprocesamiento de cada frame incluye la conversión a tensor PyTorch y la aplicación de normalización ImageNet con medias [0.485, 0.456, 0.406] y desviaciones estándar [0.229, 0.224, 0.225], transformaciones críticas que permiten al modelo preentrenado ResNet18 interpretar correctamente las características visuales. La secuencia resultante se estructura como un tensor de dimensiones [1, 16, 3, 224, 224] representando `batch_size`, `secuencia_temporal`, `canales_color`, `altura` y `anchura` respectivamente. Este formato específico es fundamental para que la arquitectura híbrida CNN-LSTM pueda procesar tanto las características espaciales de cada frame individual como las dependencias temporales entre frames consecutivos, estableciendo la base para el análisis de comportamientos dinámicos como el bullying.

- 2. Procesamiento del modelo:** El procesamiento del modelo se ejecuta mediante una arquitectura híbrida de deep learning que combina extracción de características espaciales con análisis temporal secuencial, iniciando con el paso de cada uno de los 16 frames a través de ResNet18 que actúa como extractor de características convolucionales. La red ResNet18, con sus 18 capas profundas y conexiones residuales, procesa cada frame individualmente para generar vectores de características de 512 dimensiones que capturan patrones visuales complejos como posturas corporales, proximidad entre personas y movimientos agresivos, aprovechando el conocimiento previo adquirido durante su entrenamiento con millones de imágenes y videos. Estas características espaciales se concatenan temporalmente para formar una secuencia de 16 vectores de 512 dimensiones cada uno, creando una representación rica que preserva tanto la información visual de cada momento como la estructura temporal de la secuencia.

La secuencia de características extraídas por ResNet18 se alimenta posteriormente a una red LSTM bidireccional de 2 capas con 256 unidades ocultas, diseñada específicamente para capturar dependencias temporales a largo plazo y patrones de comportamiento que se desarrollan a lo largo del tiempo. La LSTM procesa la secuencia completa utilizando sus mecanismos de memoria selectiva para identificar patrones temporales característicos del

bullying como escaladas de agresión, movimientos coordinados entre múltiples personas o cambios súbitos en la dinámica grupal. El estado oculto final de la LSTM, que contiene la información temporal condensada de toda la secuencia, se pasa a través de capas densas con activación ReLU y dropout para regularización, culminando en una capa de salida con activación softmax que produce probabilidades para las dos clases: bullying (clase 0) y comportamiento normal (clase 1). La integración de estos modelos se puede visualizar gráficamente en la Figura 9, donde se puede evidenciar el flujo que realiza el modelo para su integración en la aplicación, para mayor profundidad en el detalle del entrenamiento del modelo construido se puede apreciar a detalle en el literal 5.1.3.

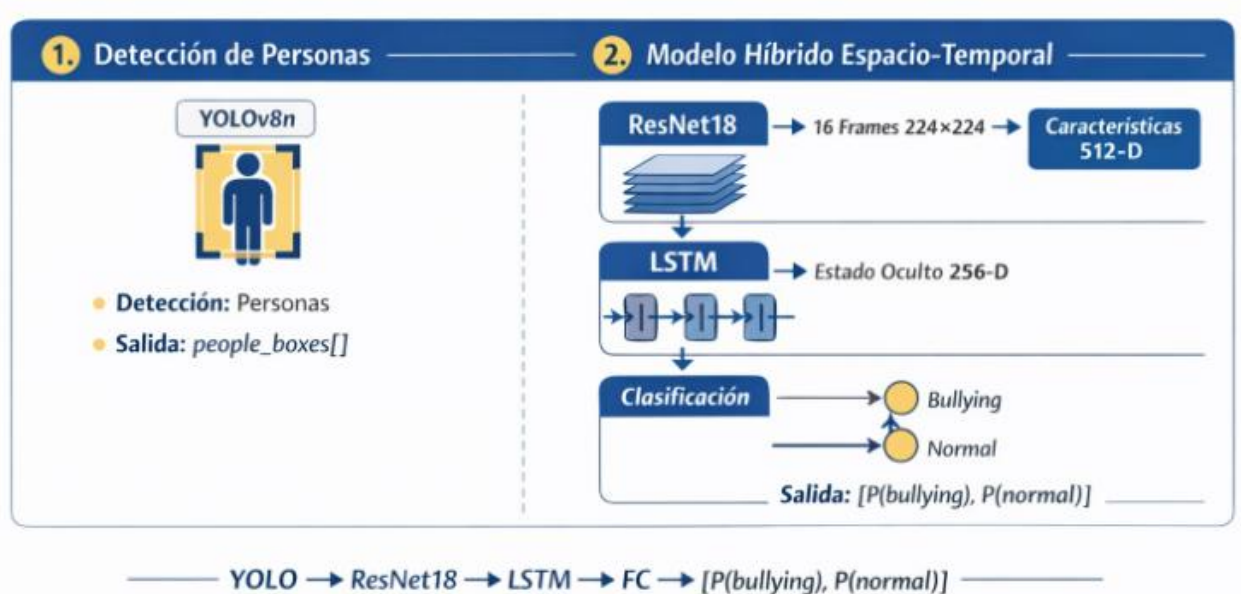


Figura 9. Diagrama Integración de modelo. Fuente: elaboración propia

- 3. Captura de Personas:** La captura de personas se realiza en paralelo al procesamiento temporal mediante el modelo YOLOv8s, una red neuronal convolucional de detección de objetos en tiempo real que analiza cada frame individual para identificar y localizar todas las personas presentes en la escena. YOLO procesa el frame completo en una sola pasada forward, generando simultáneamente las coordenadas de bounding boxes, las probabilidades de clase y los niveles de confianza para cada detección, filtrando únicamente aquellas detecciones correspondientes a la clase "persona" que superen el umbral de confianza configurado de 0.5. El sistema extrae las coordenadas de cada

bounding box detectado junto con su nivel de confianza, almacenando esta información en la estructura `people_boxes` que posteriormente se utiliza tanto para la visualización como para el conteo de personas involucradas en el incidente.

Esta información de detección de personas cumple múltiples funciones críticas en el sistema: primero, proporciona el conteo exacto de individuos presentes en la escena que se incluye en los reportes de incidentes; segundo, genera las coordenadas precisas para dibujar rectángulos de delimitación alrededor de cada persona detectada en la visualización en tiempo real; y tercero, aporta contexto cuantitativo que puede influir en la interpretación de la gravedad del incidente detectado. El sistema mantiene esta información sincronizada con los resultados del análisis temporal, permitiendo correlacionar el número de personas detectadas con la probabilidad de bullying calculada por el modelo LSTM, enriqueciendo así la información contextual disponible para los operadores del sistema y mejorando la precisión de los reportes automatizados

- 4. Detección de umbral:** La detección de umbral implementa un sistema de triple filtrado diseñado para minimizar falsos positivos mediante la aplicación secuencial de múltiples criterios de validación antes de confirmar un incidente de bullying. El primer nivel de filtrado evalúa si la probabilidad de bullying generada por el modelo LSTM supera el umbral de confianza individual configurado, estableciendo un criterio estricto que requiere alta certeza del modelo antes de considerar una detección como positiva. El segundo nivel implementa suavizado temporal mediante un buffer circular de las últimas 5 predicciones, calculando el promedio móvil de las probabilidades para reducir el impacto de fluctuaciones momentáneas y asegurar consistencia en las predicciones a lo largo del tiempo, mitigando así el ruido inherente en las predicciones de modelos de deep learning aplicados a video en tiempo real.

El tercer y más crítico nivel de filtrado requiere detecciones repetitivas y sostenidas a través de las últimas 30 evaluaciones, donde el sistema debe registrar al menos 30 detecciones positivas consecutivas antes de confirmar definitivamente un incidente de bullying. Este mecanismo de detección repetitiva está diseñado para capturar únicamente comportamientos agresivos sostenidos que representen verdaderos episodios de bullying, descartando interacciones breves, juegos normales o movimientos que podrían ser

malinterpretados por el modelo como agresivos. La configuración de estos parámetros establece un balance entre sensibilidad y especificidad, priorizando la reducción de falsos positivos sobre la detección de todos los posibles incidentes.

- 5. Respuesta final de procesamiento:** La respuesta final de procesamiento consolida todos los análisis realizados en una estructura de datos comprehensiva que encapsula tanto los resultados de la detección como la metadata contextual necesaria para la toma de decisiones y el registro de incidentes. El sistema genera un diccionario de resultados que incluye el estado binario de detección de acoso, el nivel de confianza promediado, el conteo de personas involucradas, las coordenadas de los bounding boxes, y métricas de seguimiento como la ratio de detecciones positivas que proporciona transparencia sobre el proceso de decisión del algoritmo. Esta estructura estandarizada permite que los componentes del sistema accedan de manera consistente a toda la información relevante del análisis.

Cuando se confirma un incidente de bullying, el sistema activa inmediatamente la cadena de respuesta, el registro del incidente en la base de datos con fecha-hora y metadata completa, y la actualización de la visualización en tiempo real con indicadores visuales de alerta. El sistema también implementa un mecanismo de cooldown de 30 segundos entre alertas consecutivas para evitar spam de notificaciones, mientras mantiene el logging continuo de todos los incidentes detectados para análisis posterior y generación de reportes estadísticos.

5.3.3. Mensajería

- 1. Activación del sistema de alertas:** La fase de mensajería se activa automáticamente cuando el sistema de detección confirma un incidente de bullying mediante el algoritmo de detecciones repetitivas, iniciando inmediatamente el proceso de notificación a través del módulo AlertSystem que gestiona la comunicación asíncrona con la API de Telegram Bot. El sistema verifica primero el estado del cooldown de 30 segundos implementado para prevenir spam de notificaciones, consultando la fecha de la última alerta enviada y comparándolo con el tiempo actual para determinar si es apropiado proceder con el envío. Si el cooldown está activo, el sistema registra el intento, pero no envía la notificación, mostrando en consola el tiempo restante hasta que se pueda enviar la próxima alerta,

manteniendo así un balance entre la respuesta inmediata a incidentes críticos y la prevención de saturación de notificaciones que podría desensibilizar a los receptores.

Una vez confirmado que el cooldown ha expirado, el sistema actualiza la fecha de la última alerta y procede a construir el mensaje de notificación utilizando la información consolidada en el diccionario de resultados del modelo. El proceso incluye la validación de la configuración de Telegram almacenada en la configuración global del aplicativo 'config.yaml', inicializando el objeto Bot de la librería python-telegram-bot que establece la conexión HTTPS con los servidores de Telegram mediante el token de autenticación único proporcionado por @BotFather. El sistema implementa manejo robusto de errores para capturar posibles fallos de conectividad, tokens inválidos o problemas de permisos, registrando cualquier error en la consola con información detallada para facilitar el diagnóstico y mantenimiento del sistema.

- 2. Construcción y envío de mensaje:** La construcción del mensaje de alerta sigue un formato estructurado y estandarizado que maximiza la claridad y utilidad de la información para los receptores, comenzando con la generación de la fecha o en formato ISO que identifica exactamente cuándo ocurrió el incidente detectado. El mensaje incluye un emoji de advertencia seguido del encabezado "BULLYING CONFIRMADO" que inmediatamente comunica tanto la naturaleza del incidente como el hecho de que fue validado mediante detecciones sostenidas, diferenciándolo de alertas basadas en detecciones aisladas. La estructura del mensaje incorpora métricas clave como el nivel de confianza del modelo expresado como porcentaje y el número exacto de personas detectadas en la escena, información crucial para evaluar la magnitud y gravedad del incidente.

El envío del mensaje se ejecuta de manera asíncrona utilizando el chat ID configurado, que corresponde a un grupo de trabajo o también puede ser a un canal de difusión según las necesidades organizacionales de la institución. El sistema implementa logging detallado del proceso de envío, registrando en logs cada paso del proceso desde la construcción del mensaje hasta la confirmación de entrega exitosa, incluyendo información de debug como el chat ID de destino y la fecha de envío. En caso de errores durante la transmisión, el sistema captura la excepción completa con stack trace, permitiendo identificar rápidamente problemas de conectividad, configuración incorrecta o limitaciones de la API de Telegram,

mientras continúa operando normalmente para no interrumpir el monitoreo continuo del sistema.

- 3. Transmisión de Evidencia:** Simultáneamente al envío del mensaje de texto, el sistema procesa y transmite evidencia fotográfica del incidente capturando el frame exacto en el momento de la detección confirmada, guardándolo temporalmente en la carpeta temp/ con un nombre único basado en su fecha para evitar conflictos de archivos. La imagen se codifica en formato JPEG con compresión optimizada para reducir el tamaño del archivo sin comprometer la calidad visual necesaria para la identificación de personas y análisis posterior del incidente. El sistema utiliza 'bot.send_photo()' para transmitir la evidencia visual junto con una descripción contextual que incluye la ubicación de la cámara y detalles relevantes del incidente, proporcionando a los receptores tanto la alerta inmediata como la documentación visual necesaria para evaluar la situación y tomar decisiones informadas sobre la respuesta apropiada tal como muestra la Figura 10.

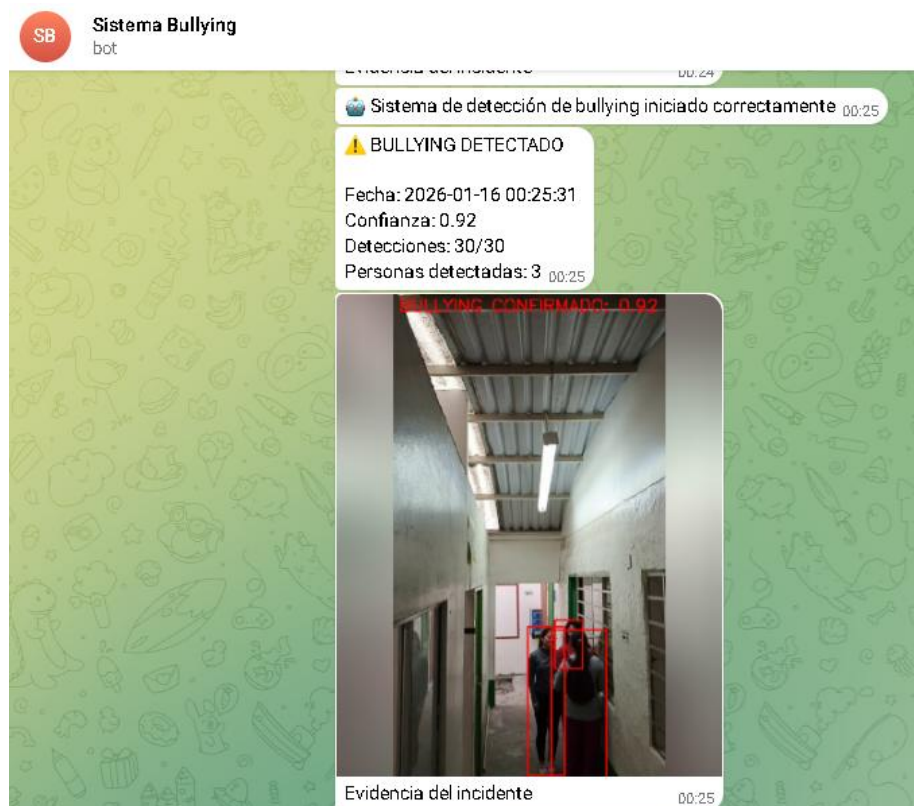


Figura 10. Notificación sistema de detección Fuente: elaboración propia

Paralelamente a la notificación vía Telegram, el sistema ejecuta el registro permanente del incidente en el archivo de log local incidents.log y opcionalmente en la base de datos configurada, almacenando información completa que incluye la fecha, nivel de confianza, número de personas involucradas, ratio de detecciones. Este registro dual asegura tanto la notificación inmediata para respuesta en tiempo real como la documentación permanente para análisis estadístico, generación de reportes, auditorías de seguridad y cumplimiento de protocolos institucionales. El sistema limpia automáticamente los archivos temporales de evidencia después de la transmisión exitosa para optimizar el uso del espacio de almacenamiento, mientras mantiene referencias en los logs para trazabilidad completa de todos los incidentes procesados por el sistema de detección de bullying.

6. EVALUACIÓN

En la etapa de la evaluación en el ciclo primordial del proyecto de desarrollo. En esta etapa se realiza la introspección para verificar si el desarrollo cumple con las necesidades de en este caso el profesorado.

6.1.EVALUACIÓN INICIAL

De acuerdo con los requisitos funcionales y técnicos registrados en este trabajo es necesario abarcarlos para realizar un análisis de su utilidad como herramienta para combatir el acoso en instituciones educativas. Para realizar este análisis es necesario desprender en varios ámbitos del funcionamiento como lo son la usabilidad que se refiere a que tan intuitiva y que tan fácil puede ser para los docentes entender los datos suministrados para realizar un abordaje temprano a este tipo de situaciones.

Por otra parte, también se debe evaluar la funcionalidad, la cual es el aspecto en el cual se centra el desarrollo de la herramienta y para la cual fue construida, en este caso la capacidad de detección temprana de acoso mediante cámaras de vigilancia. Por otra parte, es necesario también validar el rendimiento de la aplicación en un entorno de trabajo normal, en el cual va a tener mayor flujo de datos como también información con la cual el modelo no fue entrenado.

Por último, es necesario realizar una introspección de la herramienta desarrollada y la contribución realizada al campo de la educación y el abordaje al acoso apoyado con herramientas tecnológicas en este caso con modelos de inteligencia artificial.

6.1.1. Usabilidad

Teniendo en cuenta que este desarrollo es una versión inicial y puede llevarse a tener muchas mejoras tanto a nivel técnico como funcional y partiendo del conocimiento previo donde tenemos usuarios que se enfrentan por primera vez a una tecnología que hace de ayuda al análisis previo a una situación de bullying. Se lleva a cabo un análisis de su usabilidad teniendo en cuenta los feedbacks de los mismos usuarios.

Para este caso particular, se realiza una demostración de funcionamiento y uso del aplicativo al rector del Colegio Real de Cundinamarca en la ciudad de Bogotá. En esta demostración se le realiza una breve introducción al desarrollo, donde se le muestra el flujo de datos del aplicativo, desde la captura del video, hasta la notificación entregada al dispositivo móvil por medio de la aplicación de Telegram.

Para este caso, el feedback del usuario nos permitió tomar la decisión de solo mostrar información relevante al docente, lo cual permitía abordar prontamente una situación de posible acoso y llegar a validarlo en caso de que la detección fuera correcta. Esto permite no saturar de información al usuario y desgastarlo en la interpretación de los datos mostrados en un aplicativo.

6.1.2. Funcionalidad

La evaluación de funcionalidad del aplicativo se fundamentó en un análisis exhaustivo de métricas de rendimiento que demostró una precisión suficiente para garantizar respuestas efectivas por parte de los receptores de notificaciones en entornos educativos reales. El análisis integral del sistema confirmó las fortalezas identificadas previamente en el estado del arte, particularmente la efectividad de la arquitectura híbrida que combina redes neuronales convolucionales con modelos LSTM para abordar la complejidad inherente en la interpretación de comportamientos infantiles, donde actividades lúdicas normales pueden presentar características visuales similares a episodios de agresión física o acoso, requiriendo así un análisis temporal sofisticado que capture las sutilezas contextuales que diferencian el juego del comportamiento agresivo real.

Por otro lado, el sistema de notificaciones desarrollado proporciona información clara y estructurada que facilita la toma de decisiones rápidas por parte del personal docente, incluyendo métricas de confianza, evidencia fotográfica y datos contextuales necesarios para evaluar la gravedad del incidente detectado. La integración con plataformas de mensajería como Telegram representa una ventaja operacional significativa, ya que distribuye la responsabilidad de monitoreo entre múltiples miembros del equipo educativo en lugar de depender de una sola persona, permitiendo respuestas más ágiles mediante la disponibilidad de varios supervisores que pueden realizar inspección visual inmediata y coordinar intervenciones apropiadas según la naturaleza de cada situación detectada por el sistema de inteligencia artificial.

6.1.3. Rendimiento

La evaluación del rendimiento en la captura de video demostró que el sistema mantiene una operación estable y eficiente durante el procesamiento continuo de comunicación RTSP desde cámaras Tapo C211, logrando una tasa de procesamiento consistente de aproximadamente 30 frames por segundo sin pérdida significativa de calidad visual. El módulo de captura implementado mediante hilos asíncrono y buffer circular de 10 frames garantiza que el sistema pueda manejar fluctuaciones en la latencia de red y variaciones en el ancho de banda sin interrumpir el flujo de análisis, mientras que el mecanismo de comunicación de RTSP a HTTP asegura continuidad operacional ante problemas de conectividad. Las pruebas de estrés realizadas durante períodos extendidos de monitoreo confirmaron que el sistema mantiene un uso optimizado de memoria mediante la gestión automática del buffer FIFO, descartando frames antiguos para acumulación y garantizando que siempre se procesen las imágenes más recientes para detección en tiempo real.

Para realizar el análisis del rendimiento del modelo se contemplo una sección de los videos para prueba con el objetivo de estimar únicamente los videos que son bullying exclusivamente omitiendo los videos que eran netamente de violencia. Esto nos permite realizar un acercamiento al objetivo de este trabajo de fin de máster.

Accuracy	86.29
Recall	75.24%
F1-score	84.8%
Specificity	97.91%

Tabla 3. Métricas en bullying. Fuente: elaboración propia

El análisis de métricas del modelo BullyingLSTM de la Tabla 3, reveló un rendimiento robusto con una accuracy del 86.29%, specificity del 97.91%, recall del 75.24%, y F1-score del 84.8%, indicando que el sistema prioriza efectivamente la minimización de falsos positivos sobre la detección exhaustiva de todos los posibles incidentes. La especificidad del 97% confirma que el modelo no clasifica erróneamente comportamientos normales como bullying, aspecto crítico para mantener la confianza del personal educativo en las alertas generadas, mientras que el recall del 75.24%, sugiere que el sistema detecta aproximadamente siete de cada diez episodios reales de bullying, proporcionando un balance apropiado entre sensibilidad y especificidad para aplicaciones en entornos escolares, para mayor entendimiento se puede observar la matriz de

confusión de la Figura 11, donde se detalla el número de casos que para nuestra evaluación se utilizaron y se indica gráficamente la proporción de casos falsos negativos como positivos.

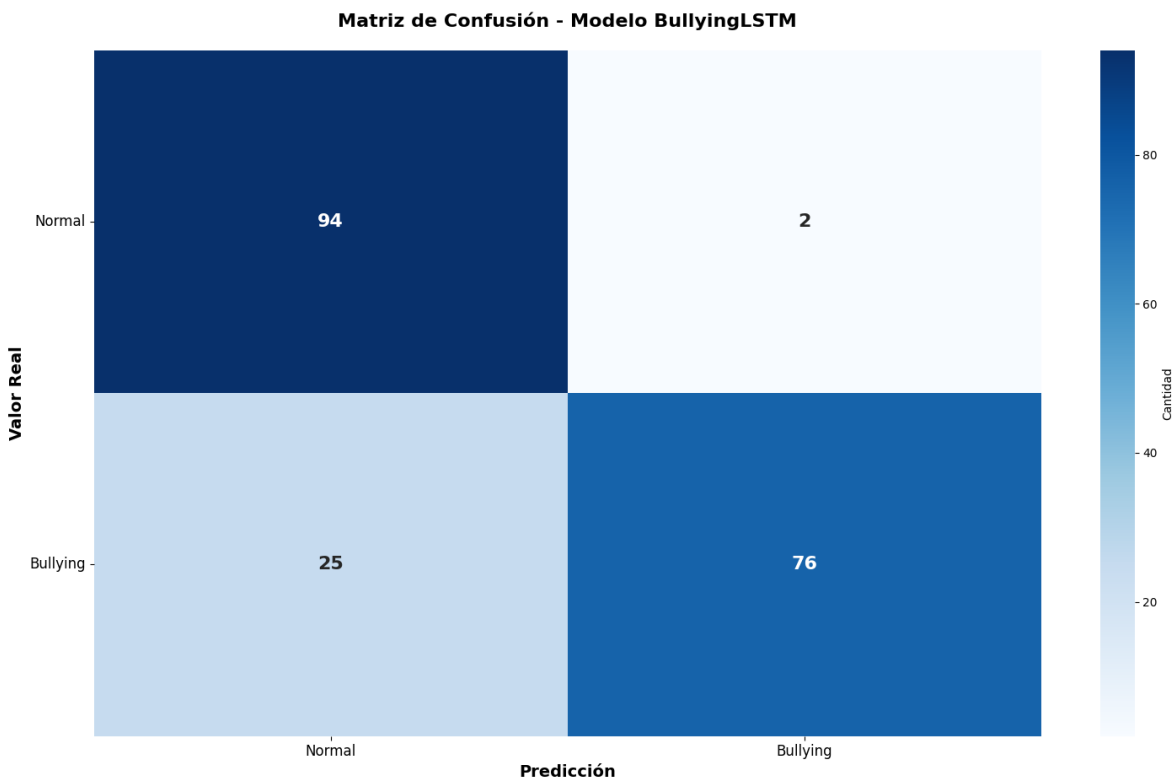


Figura 11. Matriz de Confusión. Fuente: elaboración propia

De acuerdo con la figura 10, se logra mostrar un desempeño sólido en la identificación de contenido clasificado como comportamiento normal, mostrando una baja tasa de clasificaciones erróneas en esta categoría. Sin embargo, se observa una mayor dificultad en la detección exhaustiva de casos de bullying, ya que una proporción significativa de estos fue clasificada incorrectamente como contenido normal. Este comportamiento sugiere una ligera inclinación del modelo hacia la clase mayoritaria o menos sensible, priorizando la reducción de falsas alarmas sobre la detección total de conductas de acoso. En consecuencia, el modelo demuestra un funcionamiento adecuado en términos generales, sin embargo, explica el ajuste realizado orientado

a mejorar la sensibilidad frente a la clase bullying, con el fin de minimizar la omisión de casos reales.

Por otra parte, el sistema de notificaciones demostró tiempos de respuesta rápidos, con latencias promedio de menos de 2 segundos desde la confirmación de un incidente hasta la entrega del mensaje de alerta en Telegram, incluyendo tanto el texto estructurado con métricas del incidente como la evidencia fotográfica adjunta. Las pruebas de conectividad realizadas bajo diferentes condiciones de red confirmaron la robustez del sistema de alertas, manteniendo tasas de entrega exitosa superiores al 98, mientras que el mecanismo de cooldown de 30 segundos previene efectivamente la saturación de notificaciones sin comprometer la capacidad de respuesta ante incidentes múltiples o prolongados. La integración con grupos de Telegram facilita la distribución inmediata de alertas a equipos multidisciplinarios, permitiendo que múltiples supervisores reciban simultáneamente la información necesaria para coordinar intervenciones rápidas, con evidencia visual que elimina ambigüedades y acelera la toma de decisiones en situaciones que requieren atención inmediata del personal educativo.

6.1.4. Contribución al campo

El desarrollo de una herramienta de detección de acoso mediante modelos de inteligencia artificial profundiza en el campo de la identificación del comportamiento humano. En este ámbito ya se venían realizando investigaciones relacionadas con la violencia en general; sin embargo, estas no estaban enfocadas de manera específica en agresiones o comportamientos que ocurren dentro de una institución educativa. Este aspecto resulta especialmente relevante, ya que una de las premisas principales del trabajo consistía en validar la relación directa entre la violencia común y el acoso. Bajo esta premisa, se llevó a cabo el entrenamiento del modelo utilizando videos exclusivos de violencia y, adicionalmente, videos enfocados exclusivamente en situaciones de acoso dentro de instituciones educativas.

Este tipo de investigaciones abre el debate sobre la implementación de este tipo de prácticas tecnológicas y, al mismo tiempo, permite preparar las bases para desarrollos más profundos. A partir de estos avances, se podrían diseñar sistemas más complejos que integren capacidades tecnológicas con enfoques pedagógicos, generando sinergias que puedan ser aplicadas en

diferentes instituciones educativas, según las estrategias pedagógicas que cada una llegue a implementar.

6.2.EVALUACIÓN CUALITATIVA

La evaluación cualitativa de la herramienta se llevó a cabo mediante la participación de dos expertos pertenecientes al Colegio Real de Cundinamarca: el rector del colegio y un docente con experiencia directa a la convivencia escolar. Su participación fue continua durante el desarrollo del proyecto, lo que permitió incorporar retroalimentación progresiva y contextualizada a lo largo de los diferentes sprints.

6.2.1. Metodología de evaluación

Desde la fase inicial del proyecto, los expertos participaron en cada sprint se presentaron los avances alcanzados, permitiendo ajustar el desarrollo en función de las observaciones recibidas. Esta dinámica que fortaleció la solución tecnológica y las necesidades reales se realizó por medio de:

- Demostraciones funcionales
- Entrevistas
- Validación progresiva de requisitos
- Retroalimentación al final de cada sprint

6.2.2. Preocupaciones éticas

En conversaciones con los docentes se manifestó preocupaciones relevantes que enriquecieron el desarrollo del proyecto. La principal inquietud se relacionó con la interpretación del contexto, especialmente en escenarios donde los estudiantes pueden estar jugando o realizando actividades físicas sin intención de agresión. Se reconoció que el modelo identifica patrones visuales asociados a violencia física, pero no necesariamente puede inferir la intencionalidad emocional o social detrás del comportamiento.

Por otra parte, existen consideraciones importantes sobre privacidad y marco legal, dado que el sistema opera sobre imágenes de menores de edad. Los expertos señalaron que cualquier implementación real debe garantizar:

- Cumplimiento de la normal en el marco de la ley colombiana
- Uso de espacios autorizados por la ley.
- Transparencia frente a la comunidad

Estos puntos son fundamentales para un análisis de trabajo futuro que se enfoque en la investigación de este marco legal en distintas ciudades donde se pueda aplicar.

6.2.3. Conclusiones de expertos

En términos generales, durante las fases donde el experto manifestó su feedback realiza una notable aceptación del desarrollo por parte de los docentes, quienes manifestaron conformidad con el funcionamiento en el contexto educativo. Tanto el rector como el docente coincidieron en que la herramienta representa un apoyo innovador y viable para la identificación temprana de posibles situaciones de acoso en espacios de recreo, destacando especialmente la rapidez de notificación y la claridad visual de las alertas generadas. Si bien señalaron consideraciones éticas y contextuales que deben contemplarse eventualmente, no se identificaron objeciones técnicas significativas, y el sistema fue percibido como una solución adecuada para una institución con recursos tecnológicos limitados.



Figura 12. Mapa mental conclusiones expertos. Fuente: elaboración propia

7. CONCLUSIONES Y TRABAJO FUTURO

7.1. CONCLUSIONES

Este trabajo fin de máster se ha centrado en abordar la violencia en otro ámbito como lo es el escolar, donde se presentan con frecuencia hechos de bullying dentro y fuera de las instituciones educativas. La inclusión de la tecnología y nuevas herramientas como modelos de inteligencia artificial no reemplazan el trabajo pedagógico y psicológico que un hecho de acoso pueda conllevar o pueda representar conflictos más profundos tanto de la víctima como el acosador.

Para el desarrollo de la herramienta fue necesario la verificación de fuentes fiables y trabajos relacionados, para lo cual se encontraron trabajos asociados y modelos construidos bajo diferentes enfoques, como lo son las agresiones físicas explícitas y de alta intensidad, como peleas abiertas, golpes contundentes o forcejeos evidentes. Teniendo en cuenta que se realizó el trabajo bajo la premisa de que el acoso hace parte del conjunto de violencias, se realizó una captura de información fuente de datasets y videos exclusivos de violencia y comportamiento normal. Sin embargo, se dedujo que esta información no es suficiente para el refinamiento y categorización del bullying como objeto de identificación en nuestro modelo, por ello fue necesario realizar un levantamiento de diferentes comportamientos previamente identificados con expertos, en este caso los docentes y directores de una institución educativa. Para ellos se realizaron variedad de grabaciones con diferentes casos de acoso lo cual fue fundamental para la finalización y éxito del trabajo.

Posterior a la construcción del modelo de identificación de bullying mediante el uso combinado de algoritmos CNN y LSTM, se garantiza un análisis integral de las imágenes obtenidas a partir del procesamiento de video utilizando técnicas de visión artificial. Este enfoque permite no solo la detección de patrones visuales en cada fotograma, sino también el análisis del comportamiento a lo largo del tiempo, lo que resulta fundamental para identificar situaciones de acoso que se desarrollan de manera progresiva dentro de un entorno educativo. A partir de este procesamiento, se realiza la extracción de métricas relevantes que permiten evaluar el desempeño del modelo y validar que cumple con los requisitos mínimos necesarios para su implementación en escenarios reales.

En este contexto, el modelo se proyecta como una herramienta tecnológica orientada a la identificación temprana de posibles situaciones de acoso escolar. La detección oportuna permite generar alertas mediante una herramienta de mensajería instantánea, facilitando la notificación rápida al personal encargado de la supervisión o acompañamiento. Una vez recibida esta información, el responsable puede tomar acciones pedagógicas y preventivas de manera oportuna, promoviendo estrategias de intervención que fortalezcan la convivencia, el respeto y el desarrollo social adecuado de las personas en proceso de formación. De esta manera, la solución no solo se enfoca en la detección tecnológica del problema, sino también en apoyar procesos educativos orientados a la prevención y mejora del entorno escolar.

Finalmente, el trabajo colaborativo y la retroalimentación continua con los expertos resultaron fundamentales para garantizar que los requisitos técnicos y funcionales del sistema se cumplieran de manera adecuada. Durante esta etapa, se identificaron diversos desafíos propios del entorno educativo, ya que dentro de una institución es común encontrar situaciones que, aunque puedan parecer agresivas desde una perspectiva puramente visual, no corresponden necesariamente a escenarios reales de violencia o acoso, como sucede en juegos físicos, actividades recreativas o interacciones habituales entre estudiantes. Este contexto evidenció la necesidad de incorporar criterios de análisis más precisos que permitieran diferenciar comportamientos normales de posibles situaciones de riesgo.

Como resultado de este proceso, se desarrollaron componentes adicionales orientados a validar los eventos detectados mediante niveles de concurrencia asociados a la detección positiva de acoso. Este enfoque permitió disminuir la probabilidad de generar falsos positivos, los cuales podrían provocar ruido en la operación del sistema, pérdida de confianza en la herramienta y una percepción de ineficiencia del modelo. De esta manera, se fortaleció la confiabilidad de la solución, asegurando que las alertas generadas tengan un mayor nivel de precisión y utilidad para la toma de decisiones dentro del entorno educativo.

7.2. LINEAS DE TRABAJO FUTURO

Este trabajo fin de máster es fundamental y abre la posibilidad de profundizar más en el ámbito de la pedagogía enfocada al acoso escolar o incluso en ámbitos no necesariamente donde una persona

se esté formando. A continuación, podremos ver algunas líneas en las cuales este trabajo puede servir como punto de referencia.

1. **Integración Multimodal:** En este trabajo se logró observar que es posible integrar diferentes variantes de modelos para la interpretación de un evento de acoso o violencia. Se sugiere, que en la medida de posible se realicen integraciones con modelos que agreguen una capa de interpretación a lo que pueda considerarse un tipo de violencia, teniendo en cuenta que cada tipo de violencia es específica y requiere una serie de datos para que un modelo pueda ser entregado correctamente
2. **Supervisión del sistema:** El sistema desarrollado contempla el entrenamiento del modelo con videos de violencia que fueron encontrados en diferentes fuentes, adicionalmente a estos se construyó un dataset propio que sirvió como fuente de entrenamiento. Sin embargo, este sistema podría contar un método de respuesta al sistema que permita la mejora continua y mejor efectividad al momento de detectar el acoso.
3. **Ampliación de detección:** En este caso se propone la identificación de violencia y acoso en instituciones educativas. Sin embargo, este desarrollo abre la puerta a la investigación y desarrollo de nuevas herramientas para la detección de diferentes tipos de acoso, como pueden ser social, verbal o cibernético. Se deben contemplar diferentes estrategias para interpretar estas situaciones.
4. **Integración a terceros:** Para la notificación este trabajo contemplo una aplicación de mensajería instantánea como es Telegram, sin embargo y dependiendo la necesidad podría llegar a surgir la necesidad de integración con otro tipo de alertas y no necesariamente digitales.
5. **Nuevas tecnologías:** Podría evaluarse la incorporación de modelos multimodales que integren texto, audio e imagen para una detección más robusta del acoso. El avance constante en técnicas abre la posibilidad de mejorar el rendimiento con menores requerimientos de datos etiquetados. Finalmente, la adopción de infraestructuras en la nube y herramientas de despliegue con aprendizaje continuo permitiría mantener el modelo actualizado frente a nuevas dinámicas del lenguaje digital.

Bibliografía

- Google AI Edge. (12 de 12 de 2025). Obtenido de Github: <https://github.com/google-ai-edge/mediapipe>
- Alcaldía Mayor de Bogotá D.C. (2011). *Boletín No. 77*. Bogotá D.C.
- Amazon Web Services. (01 de 12 de 2025). *Amazon*. Obtenido de <https://aws.amazon.com/es/what-is/python/>
- Baeza, J. (2024). Detección de violencia en secuencias de imágenes utilizando redes neuronales. Universidad de Alcalá Escuela Politécnica Superior.
- Barri Batero, F. (2017). *Acoso escolar o bullying: Guía imprescindible de prevención e intervención*. Altaria Editorial.
- Boone, C. (. (2019). *Bullying in schools : Prevention, gender differences and perspectives*. Nova Science Publishers, Incorporated.
- Bravo, W. (2023). Desarrollo de un modelo de visión por computadora con modelos de redes recurrentes para identificación de situaciones de violencia. Pontificia Universidad Javeriana.
- Corte Constitucional. (2025). *Sentencia T-170/25*.
- Elesawy, M. (05 de 10 de 2025). *Kaggle*. Obtenido de <https://www.kaggle.com/datasets/mohamedmustafa/real-life-violence-situations-dataset>
- Galán, C. H. (2018). Modelos de prevención del bullying: ¿qué se puede hacer en educación infantil? *SciELO Analytics*.
- Garaigordobil, L. M. (2018). *Bullying y Cyberbullying : Estrategias de Evaluación, Prevención e Intervención*. UOC.
- Gil Villa, F. (2020). *El bullying que no cesa: las bases de la violencia escolar*. Ediciones Octaedro, S.L.

Gippini, R. (2018). Acoso escolar. *ELSEVIER*.

Github. (01 de 12 de 2025). *Documentación de GitHub*. Obtenido de Github:
<https://docs.github.com/>

Hershcovis, M. S. (2011). "Incivility, social undermining, bullying. . .oh my!": A call to reconcile constructs within workplace aggression research. *Journal of Organizational Behavior*.

Mahmoodi, J. (2024). A spatio-temporal model for violence detection based on spatial and temporal attention modules and 2D CNNs. *Pattern Analysis and Applications*, 27:46.

Masip, F. G. (2016). El giro viopolítico: Violance y desconstrucción. *Política y Cultura*, 33-53.

Microsoft. (03 de 11 de 2025). *Visual Studio Code*. Obtenido de Visual Studio:
<https://code.visualstudio.com/>

Negre, P. (2024). Systematic mapping study on violence detection in video by means of trustworthy artificial intelligence. *SSRN Electronic Journal*.

Niño, A. (2022). Modelo de clasificación de imágenes violentas basado en un modelo de detección de objetos. Universidad Distrital Francisco Jose de Caldas.

OpenCV. (20 de 11 de 2025). *OpenCV*. Obtenido de <https://docs.opencv.org>

Simonov, K. (05 de 11 de 2025). *PyYAML*. Obtenido de Pypi: <https://pypi.org/project/PyYAML/>

Telegram. (30 de 12 de 2025). *Telegram*. Obtenido de <https://telegram.org/faq/es#p-que-es-telegram-que-puedo-hacer-aqui>

Ultralytics. (1 de 12 de 2025). *Ultralytics YOLO Docs*. Obtenido de Ultralytics:
<https://docs.ultralytics.com>

Unique Data. (07 de 10 de 2025). *Kaggle*. Obtenido de <https://www.kaggle.com/datasets/trainingdatapro/aggressive-behavior-video-classification>

UNIR. (25 de 10 de 2023). *UNIR*. Obtenido de UNIR LA UNIVERSIDAD EN INTERNET:
<https://colombia.unir.net/actualidad-unir/arquitectura-software-y-sistemas/>

Anexo A. Código

El código utilizado para el desarrollo de este proyecto se encuentra alojado en el siguiente repositorio público.

<https://github.com/dylanpoveda/Early-Bullying-Detection.git>