

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE SISTEMAS

UNIDAD DE TITULACIÓN

**PROCESAMIENTO DE IMÁGENES DE TWITTER O URLS
ASOCIADOS, CLASIFICADOS COMO SOSPECHOSOS DE ESTAR
RELACIONADOS CON TRATA DE PERSONAS.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN (DEL TÍTULO DEL
GRADO DE MAGISTER EN SOFTWARE**

SERGIO LEONARDO GRANIZO PILATAXI

granizo.sergio@epn.edu.ec

Director: DRA. MYRIAM BEATRIZ HERNÁNDEZ ÁLVAREZ

myriam.hernandez@epn.edu.ec

Codirector: DRA. LORENA ISABEL BARONA LOPEZ

lorena.barona@epn.edu.ec

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación PROCESAMIENTO DE IMÁGENES DE TWITTER O URLS ASOCIADOS, CLASIFICADOS COMO SOSPECHOSOS DE ESTAR ASOCIADOS A TRATA DE PERSONAS MENORES DE EDAD, del estudiante de la Maestría de Software, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.

DRA. MYRIAM BEATRIZ HERNÁNDEZ ÁLVAREZ
DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, Sergio Leonardo Granizo Pilataxi, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

SERGIO LEONARDO GRANIZO PILATAXI

DEDICATORIA

A Dios sobre todas las cosas,

A mi madre y a mi familia,

A todos aquellos amigos y compañeros que hicieron posible este logro

AGRADECIMIENTO

Gracias en primer lugar a Dios y a todas las personas, amigos, familiares y en especial a mis directores de tesis por guiarme en este importante camino para lograr otra meta importante en la vida.

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS	i
LISTA DE TABLAS	ii
LISTA DE ANEXOS	iii
RESUMEN	iv
<i>ABSTRACT</i>	v
1. INTRODUCCIÓN	1
1.1. OBJETIVO GENERAL	2
1.2. OBJETIVOS ESPECÍFICOS.....	2
1.3. MARCO TEÓRICO	2
1.4. TRABAJOS RELACIONADOS.....	6
2. METODOLOGÍA.....	8
2.1. SCRAPING O RASTREO DE SITIOS WEB SOSPECHOSOS.....	9
2.1.1 DESCARGA MASIVA DE IMÁGENES.....	9
2.1.2 PREPROCESAMIENTO DE IMÁGENES	10
2.2 DISEÑO DE EXTRACCIÓN DE CARACTERÍSTICAS	11
2.2.1 ELECCIÓN POR VOTO MAYORITARIO	15
2.3 DISEÑO DEL SISTEMA DE CLASIFICACIÓN	16
2.3.1 DATOS DE ENTRENAMIENTO.....	16
2.3.2 CONFIGURACION DEL MODELO	17
2.3.3 GENERACIÓN DEL MODELO DE CLASIFICACIÓN	17
2.3.4 DATOS DE PRUEBA	17
2.3.5 RESULTADO DE LA CLASIFICACIÓN	17
2.4 ENTRENAMIENTO DEL SISTEMA DE CLASIFICACIÓN.....	18
3. CLASIFICACIÓN.....	19
3.1 CLASIFICACION CON MAQUINAS DE SOPORTE VECTORIAL.....	19
3.1.1 INDICADORES DE RENDIMIENTO DEL MODELO SVM	20
3.2 CLASIFICACIÓN CON REDES CONVOLUCIONALES	21
3.2.1 PARTES DE UNA RED CONVOLUCIONAL.....	22
3.2.1.1 BASE CONVOLUCIONAL.....	23

3.2.1.2	CLASIFICADOR CNN	23
3.2.1.3	ARQUITECTURA CNN DE 16 CAPAS	23
3.2.2	TIPOS DE ENTRENAMIENTO DE CNN.....	24
3.2.2.1	ENTRENAMIENTO COMPLETO	24
3.2.2.2	ENTRENAMIENTO PARCIAL	24
3.2.2.3	ENTRENAMIENTO SIN BASE CONVOLUCIONAL	24
3.2.3	INDICADORES DE RENDIMIENTO DEL MODELO CNN	25
4.	APLICACIÓN DE PROCESAMIENTO Y CLASIFICACIÓN DE IMÁGENES	27
4.1	ENTORNO DE DESARROLLO.....	27
4.1.1	PYTHON.....	27
4.1.1.1	TKINTER	27
4.1.1.2	SCRAPY.....	28
4.1.1.3	ENTORNOS VIRTUALES	29
4.1.1.4	OPEN CV	29
4.1.2	TENSORFLOW.....	30
4.1.2.1	KERAS	30
4.1.3	SKLEARN	30
4.2	INTERFAZ GRÁFICA	31
4.2.1	ANALIZAR SITIO WEB	31
4.2.2	SELECCIONAR ARCHIVO	31
4.2.3	RECONOCIMIENTO FACIAL	32
4.2.4	CLASIFICADOR CNN.....	33
4.2.5	CLASIFICADOR SVM.....	33
5.	RESULTADOS EXPERIMENTALES	35
5.1.	RESULTADOS SVM.....	35
5.1.1	EXPERIMENTO 1: CLASIFICACIÓN IMÁGENES CON ROSTRO.....	35
5.1.2	EXPERIMENTO 2: CLASIFICACIÓN IMÁGENES CON TORSO	37
5.2.	RESULTADOS CNN.....	40
5.2.1	EXPERIMENTO 1: CLASIFICACIÓN DE IMÁGENES CON ROSTRO..	40
5.2.2	EXPERIMENTO 2: CLASIFICACIÓN IMÁGENES CON TORSO	41
5.3	VALIDACION DE LOS MODELOS DE CLASIFICACIÓN.....	44
5.4	COMPARACIÓN ENTRE MODELOS DE CLASIFICACIÓN.....	44
5.4.1	COMPARACIÓN IMÁGENES CON ROSTRO	45

5.4.2	COMPARACIÓN IMÁGENES CON TORSO	45
	CONCLUSIONES	47
	AGRADECIMIENTOS	48
	REFERENCIAS BIBLIOGRÁFICAS	49
	ANEXOS	52

LISTA DE FIGURAS

Figura 1 - Representación Mundial de Trata de personas en el mundo	3
Figura 2 - Mapa de rutas de la trata Internacional.....	6
Figura 3 - Visión General del Sistema.....	8
Figura 4 - Extracción de características geométricas.....	11
Figura 5 - Descomposición de una imagen en k-ésimas áreas.....	12
Figura 6 - Detección de bordes usando clasificador HAAR.....	13
Figura 7 - Extracción de características usando modelo HAAR.....	14
Figura 8 - Detección Parte superior del cuerpo.....	15
Figura 9 - Diseño del sistema de clasificación.....	16
Figura 10 - Esquema SVM	20
Figura 11 - Convoluciones secuenciales.....	22
Figura 12 - Arquitectura típica de CNN	23
Figura 13 – Proceso de clasificación VGGNet CNN.....	23
Figura 14 - Tipos de entrenamiento	25
Figura 15 - Estructura de una ventana grafica utilizando tkinter.....	28
Figura 16 - Ventana Principal de la aplicación	31
Figura 17 - Imágenes descargadas a partir del archivo plano.....	32
Figura 18 - Detección de Rostros Alice y Bob.....	32
Figura 19 - Ventana CNN.....	33
Figura 20 - Ventana SVM.....	33
Figura 21 - Resultados clasificación SVM por género.....	36
Figura 22 - Resultados clasificación SVM por grupos de edad	37
Figura 23 - Indicadores clasificación SVM Género solo torso	38
Figura 24 - Indicadores clasificación SVM grupo edad solo torso.....	39
Figura 25 - SVM Rostro y Torso.....	39
Figura 26 - Clasificación CNN por género para imágenes con rostro	40
Figura 27 - Grupos de Edad usando CNN en imágenes con rostro	41
Figura 28 - Clasificación CNN por género imágenes solo torso.....	42
Figura 29 - Clasificación CNN por grupos de edad solo torso.....	43
Figura 30 - CNN Torso y Rostro.....	44
Figura 31 - Comparación CNN y SVM para imágenes con rostro.....	45

Figura 32 - Comparación SVM y CNN para imágenes con torso 45

LISTA DE TABLAS

Tabla 1 - Diagnóstico Situacional de Trata de Blancas en Ecuador.....	4
Tabla 2 - Lista de filtros HAAR para detectar características geométricas.....	12
Tabla 3 - Elección por voto mayoritario	15
Tabla 4 - Proporción por voto mayoritario	16
Tabla 5 - Esquema Matriz de confusión	20
Tabla 6 – Matriz de confusión Resultados SVM por genero con rostro.....	36
Tabla 7 – Matriz de confusión grupos de edad con rostro.....	36
Tabla 8 – Matriz de confusión SVM por género solo torso.....	38
Tabla 9 - Resultados SVM por grupos de edad solo torso	38

LISTA DE ANEXOS

Anexo 1 - Artículo 91 del Código Orgánico Penal del Ecuador	53
Anexo 2 - Artículo 92 del Código Orgánico Penal del Ecuador	54
Anexo 3 - Muestra de imágenes clasificadas	55
Anexo 4 - Muestra de imágenes clasificadas (continuación).....	56

RESUMEN

El uso de internet es parte esencial en la vida actual. Sin embargo, también se ha convertido en una puerta para extorsión, secuestro, estafa, tráfico de personas, entre otros delitos. En este contexto, la presente tesis de Maestría tiene la intención de procesar y clasificar imágenes publicadas por usuarios de Twitter con el objetivo de determinar si existe o no posible tráfico de personas para lo cual se detecta rostros y torsos en dichas fotos. Las imágenes fueron obtenidas de una lista de mensajes de Twitter que ya fueron etiquetados como sospechosos del delito de trata de personas, dentro del trabajo realizado en el proyecto de investigación PIS-17 10. En el mencionado proyecto semilla se trabaja en la detección de escenarios relacionados con el delito de tráfico de personas usando procesamiento de lenguaje natural y se enfoca exclusivamente en el grupo de menores de edad para identificar a personas que están siendo abusadas y que son promocionadas por medios digitales. De tal manera que en la presente tesis se propone realizar el procesamiento de imágenes que permitan clasificar a personas menores de 14 años a través de sus fotografías publicadas en Twitter que sirven para atraer “clientes”. Cabe indicar que, en la mayoría de países, la ley define que toda persona menor de 14 años que está siendo utilizada con fines de explotación sexual directamente se considera como víctima de trata de personas.

Palabras clave: CNN, Clasificación de Imágenes. Extracción de Características, Machine Learning, Scraping, SVM.

ABSTRACT

The use of the internet is an essential part of today's life. However, it has also become a door for extortion, kidnapping, scam, human trafficking, among others. This Master's thesis is intended to classify faces and torso to detect minors, in images presented by Twitter users. These images were already labeled as suspected of the crime of human trafficking, within the work carried out in the research project PIS-17 10. The mentioned seed project is working on the detection of scenarios related to the crime of human trafficking using natural language processing; focusing exclusively on the group of minors to identify people who are being abused and promoted by digital means. This thesis performs image processing that allows people under 14 years to be classified through their photographs posted on Twitter that serve to attract "customers." It should be noted that, in most countries, the law defines that every person under 14 years old, who is being used for sexual exploitation, is directly considered a victim of human trafficking. In this context, the project arises. The use of the internet is an essential part of today's life.

Keywords: CNN, Image Classification. Feature Extraction, Automated Learning, Scraping, SVM.

1. INTRODUCCIÓN

Hoy en día más del 53% de la población mundial tiene acceso a internet, pero con su auge también se han generado diversos peligros en este medio masivo de comunicación. La información digital que hoy en día se entrega en los sitios web y redes sociales van más allá del simple texto, ya que una página web, puede ser utilizada con fines mal intencionados.

Estudios realizados por Internet Watch Foundation (IWF) demostraron que cada mes las imágenes y vídeos sexuales que publicaban niños y jóvenes en determinadas webs (68), ascienden a 12.224 fotos en 47 horas (260 por hora, de media). Se rastrearon las imágenes y se observaron que la mayoría (88%) eran después republicadas en lo que denominaron sitios web parásitos creados con el objetivo concreto de mostrar imágenes de gente joven.¹ Esto genera numerosos escenarios relacionados con el delito de tráfico de personas en donde se encuentran víctimas de: Mensajería de Acoso, Sexting, Grooming, Cyberbullying, Sextorsión, entre otros. El presente trabajo se enfoca exclusivamente en el grupo de personas menores de 14 años de edad ya que por su ingenuidad y desconocimiento terminan siendo abusados, incluso siendo promocionados por medios digitales.

En este trabajo se propone realizar el procesamiento de imágenes que permite el reconocimiento de personas menores de 14 años en fotografías publicadas en Twitter que sirven para atraer “clientes”. Cabe indicar que, en la mayoría de países, la ley define que toda persona menor de 14 años que está siendo utilizada con fines de explotación sexual directamente se considera como víctima de trata de personas. En este contexto, el proyecto plantea las siguientes líneas de trabajo:

- Identificar sitios web que posiblemente realicen actividades relacionadas con tráfico de personas porque promocionan fotografías de menores de 14 años para atraer “clientes” de servicios sexuales.
- Realizar la extracción de las imágenes en Twitter o de los enlaces que dirigen hacia las imágenes que las contienen, esto se realizará utilizando técnicas avanzadas de escaneo de páginas web o Scraping.

Con este fin, se procesarán y clasificarán las imágenes con herramientas apropiadas para identificar por medio de un algoritmo de reconocimiento de imágenes las características de

¹ Sitio oficial de Watch Foundation (IWF): <https://www.iwf.org.uk/>

rostro y de torso según género y grupo de edad (menores de 14 años) que están asociadas a sitios que promocionan servicios sexuales. En ese grupo de edad se tendrán casos relacionados con trata de personas.

1.1. OBJETIVO GENERAL

Realizar el procesamiento de imágenes publicadas por usuarios de Twitter que pueden estar asociados a sitios web previamente clasificados como sospechosos en cuanto a temas relacionados con trata de personas menores de edad para analizar rostros y detectar si son menores de 14 años. El procesamiento de la información facial involucrará la codificación de la forma y tamaño de características individuales como los ojos, nariz y boca, como también la codificación de las relaciones espaciales y los ángulos entre estas características.

1.2. OBJETIVOS ESPECÍFICOS

- Analizar técnicas de Scraping en Twitter y sitios URLs direccionados por los usuarios de esa red social.
- Realizar descarga masiva y preprocesamiento de las imágenes de que contiene los sitios web previamente clasificados como sospechosos usando los resultados del Proyecto Semilla PII-17 10.
- Procesar las imágenes aplicando filtros en cascada para detectar si tales imágenes contienen torsos y/o rostros mediante la extracción de sus características geométricas.
- Realizar la clasificación de las imágenes usando el modelo SVM, con una fase de entrenamiento(estudio) y una fase de aprendizaje (prueba).

1.3. MARCO TEÓRICO

Según la ONU (Organización de las Naciones Unidas) en su Declaración de la Reunión de Alto Nivel sobre el Estado de Derecho (párr. 24). “La trata de personas es un delito serio y una violación grave de los derechos humanos, que constituye una amenaza para la seguridad nacional y menoscaba el desarrollo sostenible y el estado de derecho...” [2]

Muchas de estas víctimas son adolescentes, que a base de engaños son apresados y retenidos en contra de su voluntad para realizar varios trabajos de índole sexual [1],

actividades denigrantes como prostitución o esclavismo, los cuales muchas veces no son denunciadas. Por el contrario, las víctimas son amenazadas, golpeadas y maltratadas, por lo cual sienten temor de acusar a los explotadores, ya que su vida se encuentra en peligro y a merced de sus acosadores.

Este delito va más allá de las fronteras transnacionales incluso continentales, ya que muchos niños y adolescentes son vendidos como mercancía, mayormente desde Centroamérica y Latinoamérica hacia destinos europeos y asiáticos, en gran medida por la pobre aplicación y presión de los gobiernos seccionales contra este delito tan atroz. Los países donde se ha detectado el mayor número de delitos de trata de personas son: Honduras, Venezuela y El Salvador. En la Figura 1, se muestra, un cuadro estadístico emitido en el último informe de la ONU para la prevención y protección de trata de personas.

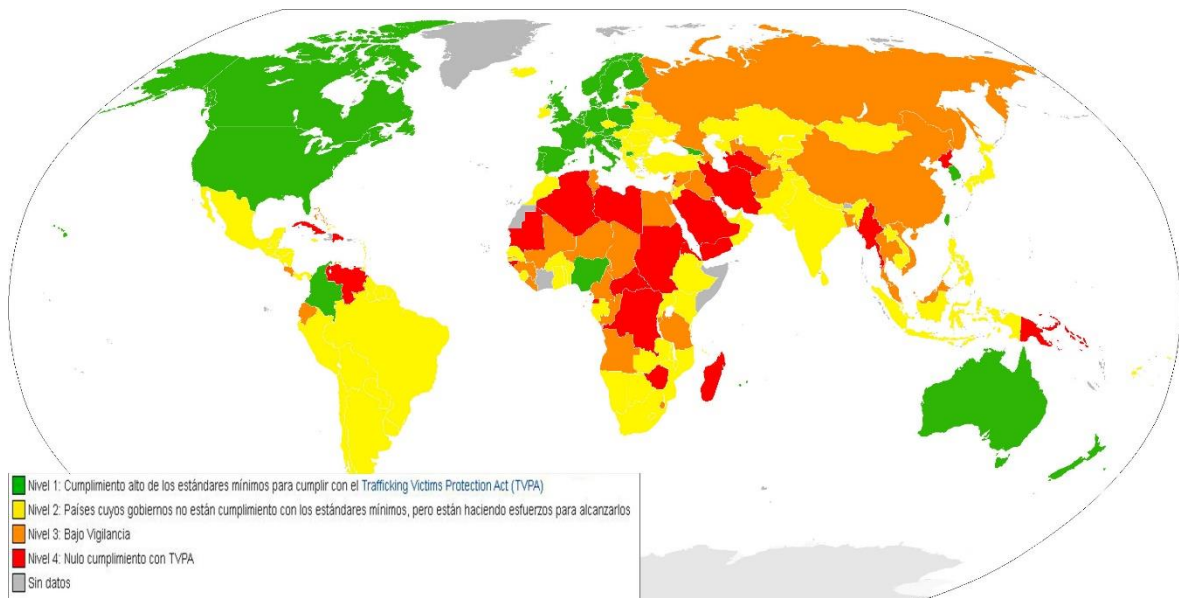


Figura 1 - Representación Mundial de Trata de personas en el mundo

Como se puede notar en la mayor parte de África y en Ecuador los estándares internacionales para evitar el tráfico de personas no se cumplen ni siquiera mínimamente. En Ecuador se presentan mayores índices de vulnerabilidad, pues según informes de la policía judicial ecuatoriana desde enero del 2017 se han registrado 332 víctimas de trata de personas, de los cuales 83 casos de detectaron en Quito y el resto en lugares rurales o suburbios en su mayoría ubicados en las provincias de Guayas.

Según Tomás Guayasamín², director de Prevención de trata de Personas y Tráfico de Migrantes del Ministerio del Interior, la causa para este tipo de delitos se da por los escasos recursos económicos de las víctimas, quienes son engañadas con la promesa de un trabajo con ingresos económicos y beneficios de ley que le ayudarán a mejorar su situación financiera e incluso a mejorar la situación de sus familiares cercanos. El nivel de desconocimiento de estas personas a veces es tan alto que son los propios padres de las víctimas quienes permiten que sus hijos sean “vendidos” a cambio de promesas de cosas materiales y de un futuro mejor, el cual nunca llega. Otro dato importante que Guayasamín aportó es que el 89% de las víctimas ecuatorianas fueron mujeres y 17% hombres, aunque la mayoría de delincuentes se enfocan en niños y adolescentes, existen víctimas hasta de 35 años de edad.

Otra forma de que este delito llegue a consumarse es cuando el atacante se gana la confianza de la víctima o de su familia, esto se consigue a través de favores (prestando pequeñas cantidades de dinero, ayudando en la casa, ayudando en tareas escolares, etc.) para que nadie sospeche y así pasar inadvertido al momento de realizar el delito, que puede terminar con el secuestro de la víctima. Un caso que horrorizó a Ecuador, fue el de “La Niña Emilia” de apenas 9 años de edad, la cual presuntamente fue interceptada a la salida de la escuela por José Nero y dos cómplices más. Nero se convirtió en un supuesto amigo de la familia e impartía clases de baile a la niña, el móvil de este delito era traficar a la niña al extranjero, pero al poner resistencia, los delincuentes la asesinaron.

Según informes del Ministerio del Interior, las provincias en donde se han registrado mayor cantidad de casos de trata de personas en adolescentes u otro tipo de delito sexual son: Pichincha, Guayas y Manabí, tal como se muestra en la Tabla 1.

Tabla 1 - Diagnóstico Situacional de trata de Blancas en Ecuador

PROVINCIA	CANTON	2014	2015	2016	TOTAL
<i>Pichincha</i>	Quito	28	62	47	137
	Mejía		3		3
	Puerto Quito		2		2
	Los Bancos	1			1
	Rumiñahui			1	1
Total Pichincha		29	67	48	144

² Noticia Quito, 6 de marzo del 2018 con la información proporcionada por el Ministro Tomás Guayasamín en el auditorio de la Facultad Latinoamericana de Ciencias Sociales: (<https://www.ministeriodegobierno.gob.ec/ecuador-trabaja-en-tres-ejes-ante-el-delito-de-trata-de-personas-y- trafico-de-migrantes/>)

<i>Guayas</i>	Guayaquil	16	15	16	47
	Milagro	1	4	3	8
	Duran	1	3	1	5
	El Empalme	2	1		3
	Daule		1	1	2
	Yaguachi		1		1
	El Triunfo			1	1
Total Guayas		20	25	22	67
<i>Manabí</i>	Bolívar		1	1	2
	Chone		1		1
	El Carmen	1	1	2	4
	Flavio Alfaro			2	2
	Manta	1	6	5	12
	Portoviejo	5	7	1	13
	Puerto López		1		
	Tosagua		1		
	Pedernales	1			
	JipiJapa	1			
	Total Manabí	9	18	11	34
TOTAL PAIS	58	110	81	245	

Fuente: Ministerio del Interior (2011)

Las sanciones para el delito de tráfico de personas en Ecuador están tipificadas en los artículos 91 y 92 del Código Orgánico Penal del Ecuador (Anexo 1 y 2). Según datos de la DINAPEN (Dirección Nacional de Policía Especializada para Niños, Niñas y Adolescentes) facilitados a la FLACSO (Facultad Latinoamericana de Ciencias Sociales), en el proyecto: trata de personas, historias que invisibilizan verdades, muchas de las víctimas son transportadas fuera del país, en la Figura 2 se muestran los principales destinos de exportación de las víctimas.

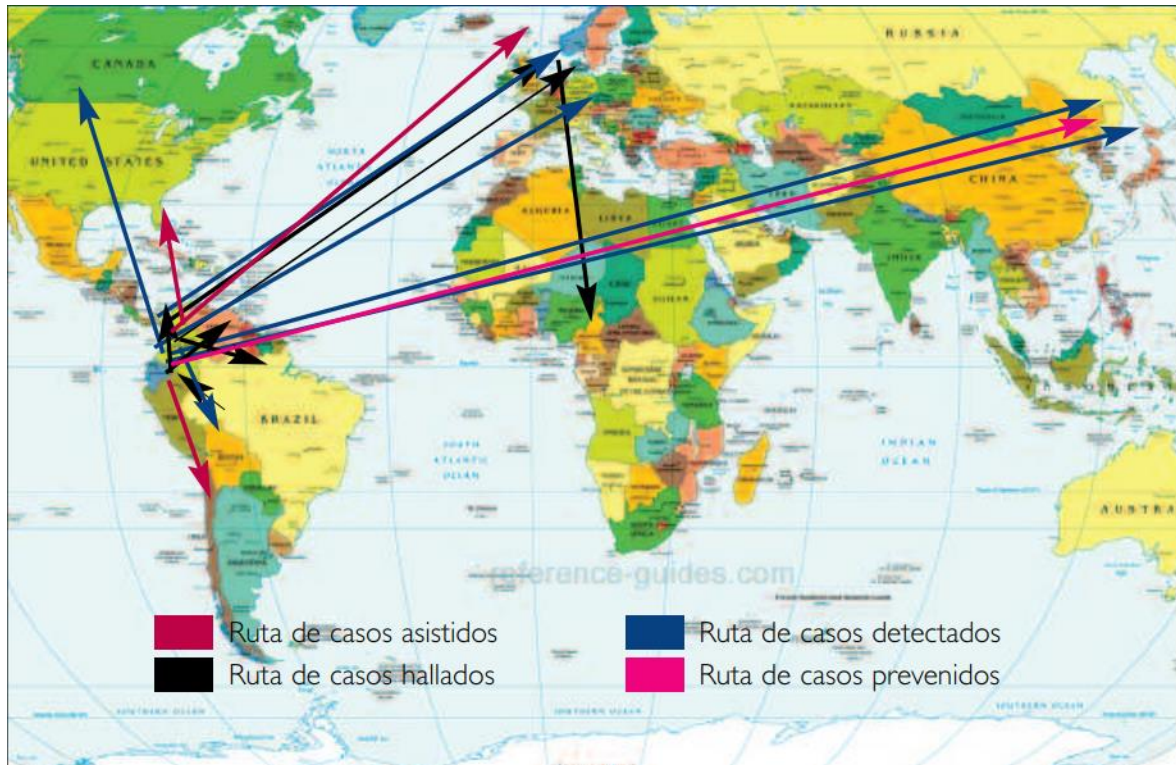


Figura 2 - Mapa de rutas de la trata Internacional

1.4. TRABAJOS RELACIONADOS

Existen trabajos de investigación relacionados con el engaño usando las redes sociales [4][10] y que hacen uso de la tecnología para cometer actividades ilícitas como tráfico de personas y explotación sexual[11], empleando ciertas técnicas novedosas [9] para engañar a los usuarios incautos. Gracias a estos estudios se han podido conocer algunos indicadores [13] del engaño online [14]. En el artículo "Detection of possible human trafficking in Twitter" [19] se realiza un estudio sobre los posibles mensajes que son comúnmente usados para captar la atención de "patrocinadores" que paguen por los servicios ilícitos que se ofertan, además se realiza un trabajo de clasificación sobre el dataset de los tuits relacionados.

Si bien existen trabajos de filtrado de tuits y clasificación de imágenes, en el mejor de nuestro conocimiento, existen trabajos como "A new algorithm for age recognition from facial images" presentado en [20] donde se muestra algoritmos para predecir grupos de edad con una exactitud aproximada de 86,64%, y "Deep convolutional neural networks and support vector machines for gender recognition" presentado en [21] donde se utilizan modelos de clasificación SVM y CNN para predecir el género de la persona, en el mejor de

los casos llegando a obtener una exactitud de 97,3%. En ambos trabajos únicamente se toma como referencia el rostro de la persona involucrada.

Trabajos como "Combating human trafficking with deep multimodal models"[22] han enfocado sus esfuerzos en análisis de tuits y anuncios publicados en la web, analizando el texto publicado, para encontrar patrones de tipo léxico como ofertas de altas remuneraciones económicas para mujeres jóvenes sin ningún tipo de experiencia. En esta propuesta se realiza la detección automática de anuncios sospechosos obteniendo como resultado que los mejores modelos de Machine Learning usados en ese proyecto son: Logístico Regression, Trafficking Deep Network y SVM con accuracy de 80,2%, 80,0% y 79,2% respectivamente, y valores F1 de 45,1%, 66,5% y 45,1% respectivamente.

Sin embargo, después de la revisión de la literatura actual se determina que no existen trabajos que consideren características de upper_body en las imágenes y su posterior clasificación y vinculación con el delito de tráfico de personas, siendo este el objetivo principal de este proyecto de tesis. Es importante mencionar que la expresión upper_body es utilizada para referirse al torso de la persona.

2. METODOLOGÍA

Para el presente proyecto se utilizan imágenes que están alojadas en sitios web los cuales ofrecen o insinúan servicios sexuales. En este trabajo se adaptará la metodología tradicional de desarrollo de machine learning para el caso específico de clasificadores de imágenes. Inicialmente como datos de entrada se utilizarán la lista negra de URLs de sitios web sospechosos obtenidos en el proyecto PIS-17 10, donde se analizó sitios web relacionados con usuarios sospechosos de estar involucrados en el crimen de trata de personas. El siguiente paso consiste en procesar las imágenes que se recopilarán realizando técnicas de Scraping sobre los websites relacionados y las imágenes que directamente se encuentren en los datos ya seleccionados en Twitter [7]. A continuación, en la figura 3, se presenta una vista general del proyecto, para su mejor comprensión.

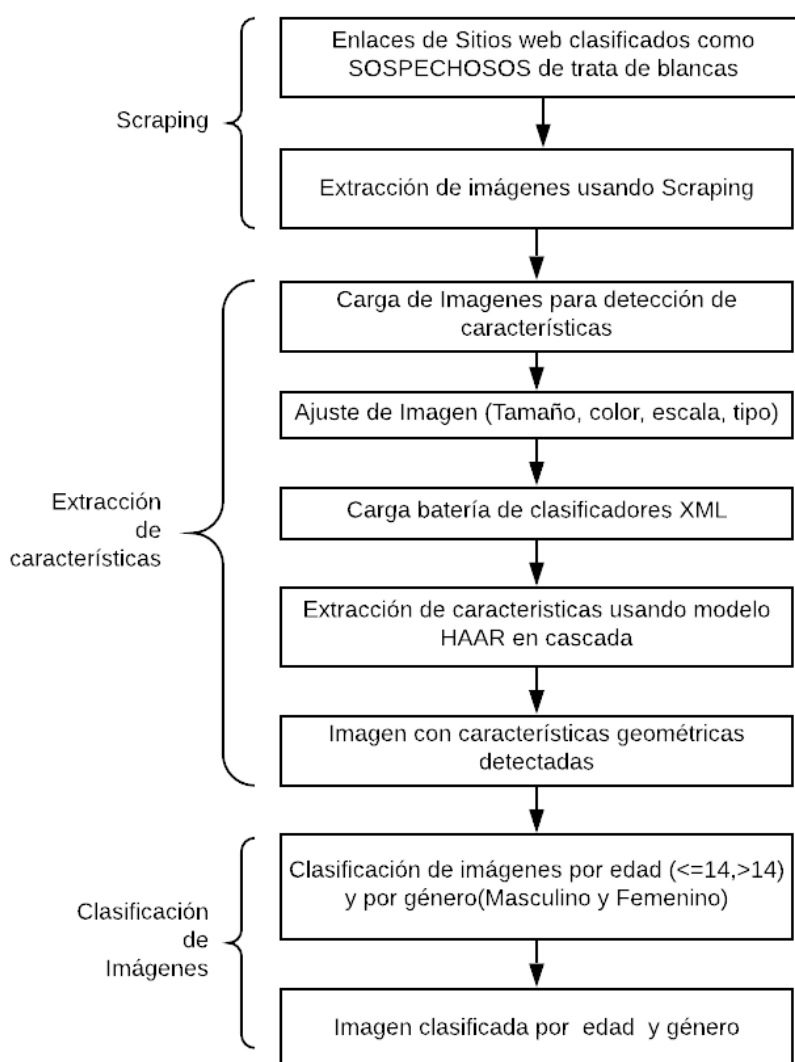


Figura 3 - Visión General del Sistema

Para el presente proyecto, se partió de una “lista negra” con todos los URLs de los sitios web que estuvieron enlazados en los tuits que previamente se catalogaron como sospechosos. Para generar la lista negra de sitios web sospechosos hubo que limpiar la data, a través de los siguientes procesos:

- Eliminar resultados duplicados
- Eliminar enlaces que dirijan a sitios web inexistentes
- Eliminar enlaces que dirijan a sitios web dados de baja

La lista negra depurada se utilizó para realizar el rastreo de los sitios web sospechosos.

2.1. SCRAPING O RASTREO DE SITIOS WEB SOSPECHOSOS

Existen sitios web que generan la concurrencia masiva de los usuarios a las redes sociales donde de manera deliberada se enlazan tuits que hacen alusión a comprar favores sexuales. Algunos de los sitios web que reiteradamente aparecen en los tuits sospechosos son: Infoscort, Punterking y Shinagawaesthe”. Se utilizó una lista negra de sitios web sospechosos para analizar su código fuente y de esta manera obtener los enlaces a imágenes que están mostrando en sus dominios, analizando el código html y css de dichos sitios web.

2.1.1 DESCARGA MASIVA DE IMÁGENES

Descargar las imágenes sospechosas manualmente demandará mucho tiempo (pues habría que hacerlo de una en una), por lo tanto, este proceso se realiza mediante un método automatizado que consiste en utilizar técnicas de Rastreo(Scraping) empleando algún framework configurado para este propósito. En este trabajo se utiliza el software Scrapy (Software gratuito desarrollado en Python), el cual para realizar el rastreo de un sitio web necesita como entrada el link o URL del sitio a analizar. Por ejemplo <https://www.shinagawa-esthe.jp/en> es un sitio web que ofrece servicios de scort y como resultado se devuelve un archivo plano, en donde están todos los enlaces a las imágenes del sitio web.

Después se descargará mediante un proceso automatizado las imágenes en base a la lista de links. Un inconveniente a tomar en cuenta es que sí los sitios web requieren suscripción

de pago no se puede descargar las imágenes internas pues la política de suscripción prohíbe acceder a las páginas subsiguientes de forma libre, razón por la cual Scraping está limitado a aquellas páginas que son de acceso libre, obviamente a menos que se disponga de una cuenta de pago.

Cabe mencionar que, aunque las técnicas de seguridad informática han evolucionado bastante, la utilización de certificados digitales [24] no necesariamente garantiza que las actividades de ciertos sitios web sean lícitas. Además, para esta clase de negocios ilegales suelen usarse frases aparentemente inofensivas como “caldo de pollo” o “club penguin” para intercambiar el material digital (fotos y videos)³.

2.1.2 PREPROCESAMIENTO DE IMÁGENES

Una vez descargado todo el set posible de imágenes se procede a realizar la limpieza y depuración de la data que consiste en lo siguiente:

- Descartar todas las imágenes que sean irrelevantes por ejemplo imágenes de iconos, imágenes de rótulos, de autos, de botones, etc.
- Descartar imágenes repetidas; para un mejor análisis es ideal no tener datos duplicados, entonces se procura que las imágenes no se repitan.
- Descartar imágenes que estén en escala de grises, pues para aplicar los respectivos modelos de clasificación se necesita que la composición de color de las imágenes sea en RGB (Rojo, verde, azul).
- Descartar imágenes con formatos desconocidos, para los modelos de clasificación el proyecto se enfoca en imágenes con formato *.jpg, si existen imágenes que no son *.jpg pero que aportan información, entonces previamente se realizara una conversión a *.jpg.
- Redimensionar imágenes para los modelos de clasificación se necesita que todas las imágenes tengan las mismas dimensiones, por lo tanto, se estandarizo el tamaño de todas las imágenes a 150x150 píxeles.

³ Hilo completo sobre intercambio de material digital ilegal en Twitter usando frases como “Caldo de Pollo” o “Club Penguin”:
<https://twitter.com/MyLifeAsThunder/status/1173267980811194368>

Las imágenes de los sitios web sospechosos forman el banco de imágenes para la data de pruebas. Para la data de entrenamiento se usaron las siguientes bases de datos abiertas:

- Flickr (<https://www.flickr.com>)
- UKBench (<https://archive.org/details/ukbench>)
- Deeplearning (<http://deeplearning.net/datasets/>)

2.2 DISEÑO DE EXTRACCIÓN DE CARACTERÍSTICAS

Aunque para un ser humano una imagen contenga mucha información detectable a simple vista, por ejemplo, una pintura famosa como la Gioconda, o la información en alguna radiografía o un plano, para el computador una imagen es únicamente una gigantesca matriz de puntos, donde cada punto recibe el nombre de pixel y contiene cierta información para la configuración RGB, es decir, para saber cuánta cantidad de rojo, verde y azul debe llevar cada punto de la imagen. Para extraer las características geométricas de una imagen, en este proyecto se utilizó el Modelo en cascada de HAAR [25]. De manera general el proceso de detección de características geométricas se muestra en la Figura 4.

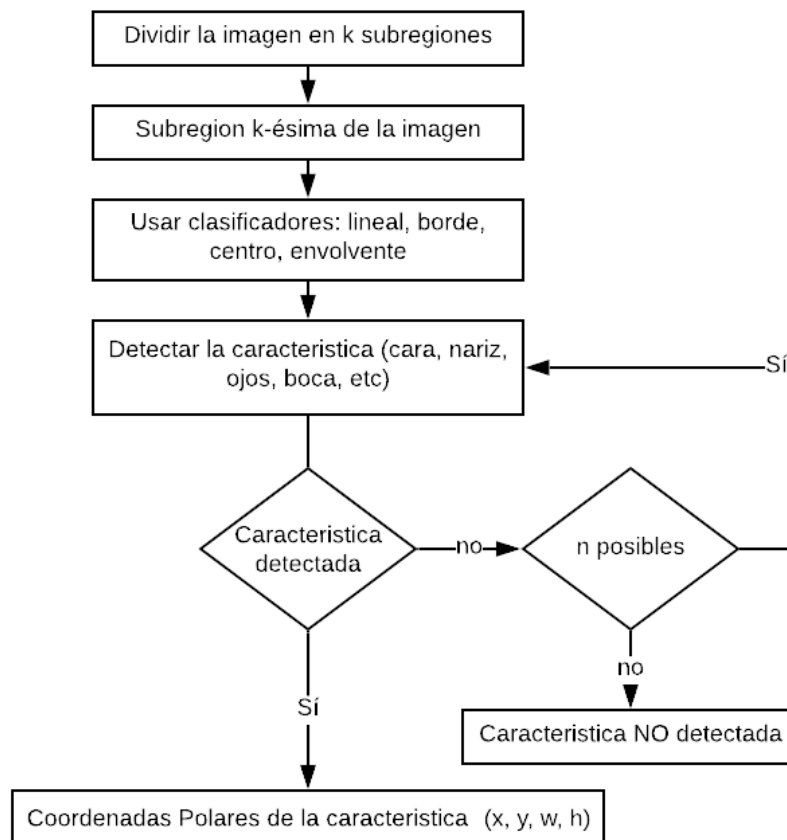


Figura 4 - Extracción de características geométricas

Para extraer la característica geométrica el primer paso consiste en cargar la imagen y luego separarla en matrices más pequeñas, generalmente de dimensiones 24x24px. (Figura 5).

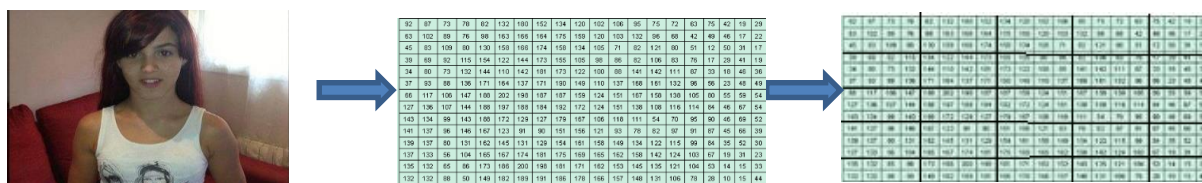






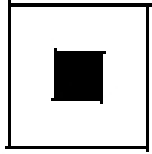



Figura 5 - Descomposición de una imagen en k-ésimas áreas

Como se muestra en la Figura 5, la imagen se divide en k regiones, luego a cada región k-ésima se aplica una batería de clasificadores HAAR en cascada, estos clasificadores son pequeños filtros matriciales de 3x3 que sirven para detectar características geométricas específicas como líneas, bordes y centros. En la Tabla 2, se describe brevemente los clasificadores de HAAR que son comúnmente utilizados.

Tabla 2 - Lista de filtros HAAR para detectar características geométricas

FILTRO	MATRIZ	DESCRIPCION									
	<table border="1"> <tr><td>-1</td><td>-1</td><td>5</td></tr> <tr><td>-1</td><td>-1</td><td>5</td></tr> <tr><td>-1</td><td>-1</td><td>5</td></tr> </table>	-1	-1	5	-1	-1	5	-1	-1	5	Detecta bordes verticales, la característica que se desea detectar está explícita en la tercera columna
-1	-1	5									
-1	-1	5									
-1	-1	5									
	<table border="1"> <tr><td>5</td><td>5</td><td>5</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> </table>	5	5	5	-1	-1	-1	-1	-1	-1	Detecta bordes horizontales, la característica que se desea detectar está explícita en la primera fila
5	5	5									
-1	-1	-1									
-1	-1	-1									
	<table border="1"> <tr><td>-1</td><td>5</td><td>-1</td></tr> <tr><td>-1</td><td>5</td><td>-1</td></tr> <tr><td>-1</td><td>5</td><td>-1</td></tr> </table>	-1	5	-1	-1	5	-1	-1	5	-1	Detecta líneas verticales, la característica que se desea detectar esta en la segunda columna
-1	5	-1									
-1	5	-1									
-1	5	-1									
	<table border="1"> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>5</td><td>5</td><td>5</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> </table>	-1	-1	-1	5	5	5	-1	-1	-1	Detecta líneas horizontales, la característica que se desea detectar esta explícita en la segunda fila
-1	-1	-1									
5	5	5									
-1	-1	-1									
	<table border="1"> <tr><td>-1</td><td>-1</td><td>5</td></tr> <tr><td>-1</td><td>5</td><td>-1</td></tr> <tr><td>5</td><td>-1</td><td>-1</td></tr> </table>	-1	-1	5	-1	5	-1	5	-1	-1	Permite detectar líneas inclinadas con pendiente positiva, la característica está explícita en la diagonal principal
-1	-1	5									
-1	5	-1									
5	-1	-1									

	<table border="1"> <tr><td>5</td><td>-1</td><td>-1</td></tr> <tr><td>-1</td><td>5</td><td>-1</td></tr> <tr><td>-1</td><td>-1</td><td>5</td></tr> </table>	5	-1	-1	-1	5	-1	-1	-1	5	Permite detectar líneas inclinadas con pendiente negativa, la característica está explícita en la diagonal secundaria
5	-1	-1									
-1	5	-1									
-1	-1	5									
	<table border="1"> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>-1</td><td>5</td><td>-1</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> </table>	-1	-1	-1	-1	5	-1	-1	-1	-1	Permite detectar centralizados, la característica está explícita en el centro de la matriz
-1	-1	-1									
-1	5	-1									
-1	-1	-1									
	<table border="1"> <tr><td>-1</td><td>5</td><td>-1</td></tr> <tr><td>5</td><td>5</td><td>5</td></tr> <tr><td>-1</td><td>5</td><td>-1</td></tr> </table>	-1	5	-1	5	5	5	-1	5	-1	Permite detectar puntos centralizados dentro de un contorno, la característica está explícita en el centro de la matriz
-1	5	-1									
5	5	5									
-1	5	-1									

Se utilizan matrices de 3x3 porque con ese tamaño es el que mejor funciona para detectar las características geométricas y se denominan en cascada, porque realizan un número n de intentos antes de descartar la característica buscada.

Para realizar la detección se realiza un producto escalar, se toma como centro cada píxel de la imagen y se va recorriendo progresivamente en toda la imagen. En caso de faltar filas y/o columnas se duplicará la última fila y/o columna adyacente de la imagen, cada convolución devolverá un valor único. A continuación un ejemplo en la Figura 6.

Región k-ésima	Clasificador HAAR	Resultado																																																																																																													
<table border="1"> <tr><td>7</td><td>7</td><td>6</td><td>5</td><td>5</td><td>6</td><td>7</td><td>7</td></tr> <tr><td>7</td><td>7</td><td>6</td><td>5</td><td>5</td><td>6</td><td>7</td><td>7</td></tr> <tr><td>6</td><td>6</td><td>4</td><td>3</td><td>3</td><td>4</td><td>6</td><td>6</td></tr> <tr><td>5</td><td>5</td><td>3</td><td>2</td><td>2</td><td>3</td><td>5</td><td>5</td></tr> <tr><td>5</td><td>5</td><td>3</td><td>2</td><td>2</td><td>3</td><td>5</td><td>5</td></tr> <tr><td>6</td><td>6</td><td>4</td><td>3</td><td>3</td><td>4</td><td>6</td><td>6</td></tr> <tr><td>7</td><td>7</td><td>6</td><td>5</td><td>5</td><td>6</td><td>7</td><td>7</td></tr> <tr><td>7</td><td>7</td><td>6</td><td>5</td><td>5</td><td>6</td><td>7</td><td>7</td></tr> </table>	7	7	6	5	5	6	7	7	7	7	6	5	5	6	7	7	6	6	4	3	3	4	6	6	5	5	3	2	2	3	5	5	5	5	3	2	2	3	5	5	6	6	4	3	3	4	6	6	7	7	6	5	5	6	7	7	7	7	6	5	5	6	7	7	<table border="1"> <tr><td>-1</td><td>-1</td><td>5</td></tr> <tr><td>-1</td><td>-1</td><td>5</td></tr> <tr><td>-1</td><td>-1</td><td>5</td></tr> </table>	-1	-1	5	-1	-1	5	-1	-1	5	<table border="1"> <tr><td>40</td><td>29</td><td>36</td><td>54</td><td>71</td><td>64</td></tr> <tr><td>29</td><td>19</td><td>27</td><td>45</td><td>67</td><td>59</td></tr> <tr><td>18</td><td>9</td><td>18</td><td>36</td><td>63</td><td>54</td></tr> <tr><td>18</td><td>9</td><td>18</td><td>36</td><td>63</td><td>54</td></tr> <tr><td>29</td><td>19</td><td>27</td><td>45</td><td>67</td><td>59</td></tr> <tr><td>40</td><td>29</td><td>36</td><td>54</td><td>71</td><td>64</td></tr> </table>	40	29	36	54	71	64	29	19	27	45	67	59	18	9	18	36	63	54	18	9	18	36	63	54	29	19	27	45	67	59	40	29	36	54	71	64
7	7	6	5	5	6	7	7																																																																																																								
7	7	6	5	5	6	7	7																																																																																																								
6	6	4	3	3	4	6	6																																																																																																								
5	5	3	2	2	3	5	5																																																																																																								
5	5	3	2	2	3	5	5																																																																																																								
6	6	4	3	3	4	6	6																																																																																																								
7	7	6	5	5	6	7	7																																																																																																								
7	7	6	5	5	6	7	7																																																																																																								
-1	-1	5																																																																																																													
-1	-1	5																																																																																																													
-1	-1	5																																																																																																													
40	29	36	54	71	64																																																																																																										
29	19	27	45	67	59																																																																																																										
18	9	18	36	63	54																																																																																																										
18	9	18	36	63	54																																																																																																										
29	19	27	45	67	59																																																																																																										
40	29	36	54	71	64																																																																																																										

Figura 6 - Detección de bordes usando clasificador HAAR

Al aplicar los filtros para saber si se ha logrado detectar líneas, bordes, centros y envolventes internamente en la matriz resultante se utiliza el algoritmo KNN (K-ésimos vecinos más cercanos) utilizando como métrica la distancia euclidiana entre los píxeles más próximos de cada subregión.

Al realizar sistemáticamente las operaciones escalares y luego aplicando KNN en la matriz resultante se define si existe o no la característica buscada, actualmente existen

clasificadores de HAAR configurados y optimizadores para focalizar la búsqueda de una determinada parte del cuerpo o cara. A continuación, se lista los archivos que contienen los clasificadores utilizados en este proyecto.

Archivos clasificadores XML para cuerpo:

- Cuerpo completo: haarcascade_fullbody.xml
- Parte inferior del cuerpo: haarcascade_lowerbody.xml
- Parte superior del cuerpo: haarcascade_upperbody.xml

Archivos clasificadores XML para cara:

- Cara de perfil: haarcascade_profileface.xml
- Cara de frente: haarcascade_frontalface.xml
- Par de Ojos: haarcascade_eyes.xml
- Ojo izquierdo: haarcascade_lefteye.xml
- Ojo derecho: haarcascade_righteye.xml

Luego se integran todos los resultados de las características encontradas como se muestra en la Figura 7.

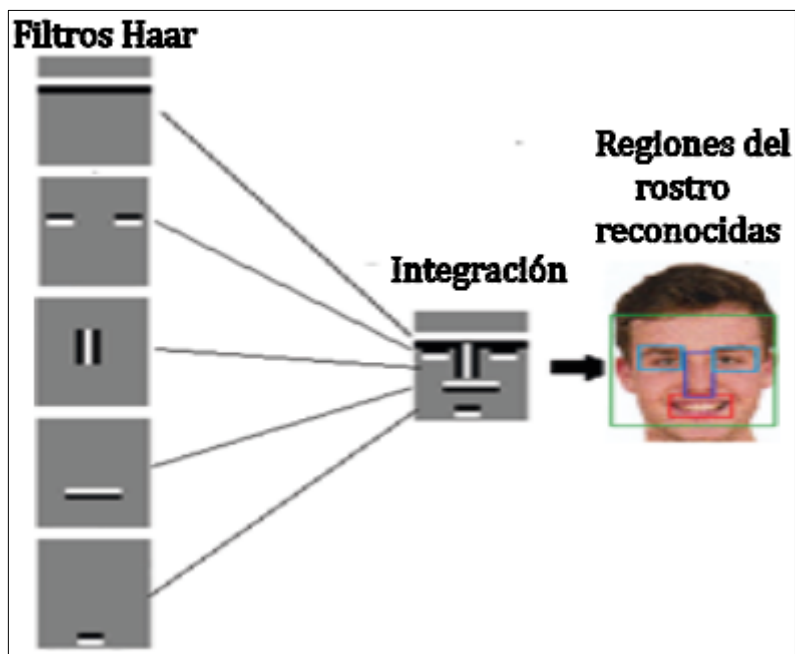


Figura 7 - Extracción de características usando modelo HAAR

Cuando se trata de buscar imágenes de personas menores de edad en los sitios web sospechosos existe el inconveniente que estas imágenes tienen el rostro pixelado, cubierto

o simplemente no se muestra en la foto, por esta razón también se realizará un método para la detección de torsos. De manera similar, se puede utilizar clasificadores para detectar el cuerpo o la parte superior del cuerpo de una persona en una imagen. En este caso, como se trata de detectar la parte superior del cuerpo el archivo clasificador XML es: haarcascade_upperbody.xml (Figura 8).

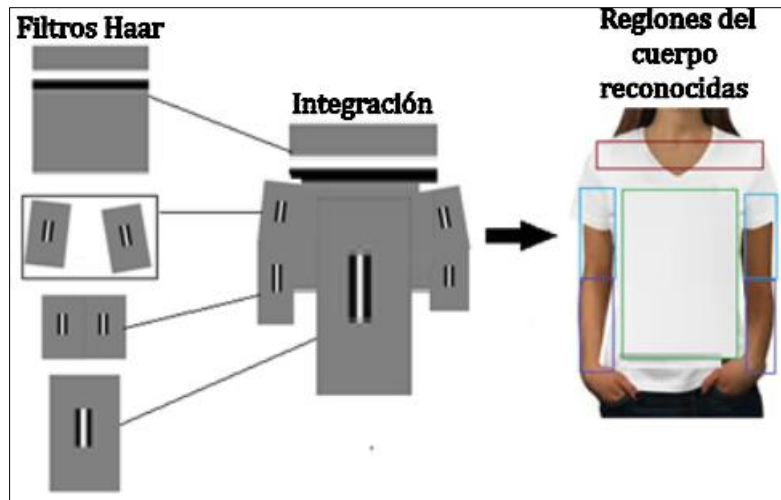


Figura 8 - Detección parte superior del cuerpo

En la Figura 8, se muestra la integración de distintos filtros HAAR individuales para la detección de la parte superior de la persona.

2.2.1 ELECCIÓN POR VOTO MAYORITARIO

En ciertas imágenes no es posible detectar todas las características geométricas individuales, debido a la mala calidad de la imagen, partes desenfocadas o posición del cuerpo. Para definir si en una imagen prevalece la detección de torso o del rostro se tomó en cuenta, la parte con el mayor número de características individuales detectadas, en caso de tener el mismo número de características detectadas, se seleccionará como parte determinante al torso como se detalla en la Tabla 3.

Tabla 3 - Elección por voto mayoritario

	Rostro	Torso	Parte predominante
Número de Características Geométricas individuales detectadas	2	3	Torso
	3	1	Rostro
	2	2	Torso
	0	0	Ninguno

Una vez extraídos torsos y rostros, del 100% de imágenes procesadas, se obtuvieron las respectivas proporciones. Se utilizó el voto mayoritario que se muestra en la Tabla 4 para etiquetar las imágenes.

Tabla 4 - Proporción por voto mayoritario

	% Rostros	% Torsos	% No detectado	Total
Dataset	64,19%	32,15%	3,66%	100%

En el presente proyecto de tesis las imágenes obtenidas de la detección de características (rostro y torso), se utilizan exclusivamente para conformar la data que servirá como entrada en el modelo clasificador SVM.

2.3 DISEÑO DEL SISTEMA DE CLASIFICACIÓN

Para la construcción del sistema de clasificación de las imágenes se tomará en cuenta el diseño del sistema de clasificación mostrado en la Figura 9.

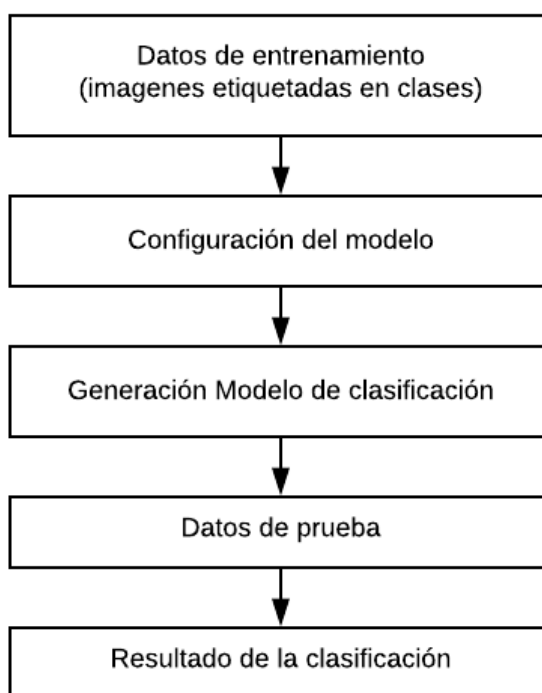


Figura 9 - Diseño del sistema de clasificación

2.3.1 DATOS DE ENTRENAMIENTO

El dataset a utilizar está conformado por 4096 imágenes de las cuales, 2048 imágenes son de personas menores de 14 años y 2048 imágenes corresponden a personas mayores de 14 años. Estas fotos usan composición RGB (Red, Green, Blue), en formato jpg.

2.3.2 CONFIGURACION DEL MODELO

Luego para iniciar el proceso de clasificación se configura el modelo de acuerdo a su tipo de configuración. En el caso de SVM, previamente se realizó una extracción de las características usando modelos en cascada de HAAR, para crear grupos de imágenes con rostro y torso, posteriormente se etiquetó los grupos de imágenes por género y grupos de edad. SVM es un modelo supervisado, eso significa que el usuario clasifica el dataset de entrenamiento en forma manual. El algoritmo se entrena con estos datos y crea un modelo que puede ser luego aplicado en un set de prueba. Como indicadores de rendimiento de SVM se utilizó, precisión, accuracy, recall y F-measures.

Por otro lado, CNN necesita aplicar capas convolucionales. Por lo tanto, se necesita 10 iteraciones y 16 capas para que el modelo vaya aprendiendo progresivamente. En este proyecto se utilizará un modelo de aprendizaje con enfoque no supervisado, por lo tanto, se requiere menos intervención del humano. Como indicadores principales se usó el error medio cuadrado (MSE) y la exactitud.

Tanto para el modelo SVM como para CNN, se utiliza la distribución 80% para datos de entrenamiento y 20% para datos de prueba. Además, la predicción se realizó sobre una variable objetivo tipo binario, es decir que al finalizar la clasificación solo se puede obtener uno de dos resultados posibles, ya que únicamente existen dos clases en cada experimento.

2.3.3 GENERACIÓN DEL MODELO DE CLASIFICACIÓN

A continuación, se genera un modelo que es capaz de predecir las imágenes en clases, gracias a las configuraciones de aprendizaje que se explicó previamente. Se puede generar varios modelos hasta obtener el modelo deseado que tenga mayor estabilidad y regularidad para clasificar.

2.3.4 DATOS DE PRUEBA

Para probar la eficiencia del modelo se requiere cargar datos de prueba, estos datos son las imágenes que previamente se ha descargado de Twitter o de los sitios web catalogados como sospechosos.

2.3.5 RESULTADO DE LA CLASIFICACIÓN

Como salida del sistema se obtiene los resultados de la clasificación que estarán representados en dos experimentos. El primer experimento consiste en la clasificación por

grupos de edad (Mayor o menor de 14 años). Mientras que en el segundo experimento la clasificación se realizará por género (Hombre o mujer).

2.4 ENTRENAMIENTO DEL SISTEMA DE CLASIFICACIÓN

Para la fase de entrenamiento o aprendizaje el modelo SVM recibió dos tipos diferentes de datos. En primer lugar, se entrenó el modelo con imágenes donde se detectó el rostro de la persona involucrada y luego se realizó dos los experimentos de clasificación con esta data.

Posteriormente, se entrenó el modelo con imágenes donde no se detectó el rostro de la persona, ya sea porque no exista o porque en la imagen está pixelada o cubierta de alguna manera; por esta razón en este segundo set de imágenes se toma como referencia el torso de la persona detectado por los filtros HAAR que se aplicaron previamente. Luego se realizó el experimento de clasificación requerido, por género y por grupo de edad.

El siguiente paso consiste en detallar el proceso de clasificación de imágenes, pero dado la amplitud del tema se realizó un capítulo completo para su desarrollo.

3. CLASIFICACIÓN

En este capítulo se tratará el tema de la clasificación de imágenes y el aprendizaje automatizado, para ello, se revisó el modelo de clasificación SVM como modelo principal. Adicionalmente se utilizó otro modelo basado en redes neuronales convolucionales o CNN, para corroborar y contrastar los resultados, con esto se tiene un mejor panorama al medir el nivel de exactitud al momento de realizar este tipo de clasificaciones. Una vez detectadas las características geométricas en las imágenes, se clasifican según los siguientes criterios:

Clasificación por género

- Hombres
- Mujeres

Clasificación por grupos de edad

- Menores de 14 años
- Mayores de 14 años

Para el presente trabajo se realizó dos experimentos. El primer experimento consiste en realizar estas clasificaciones tomando como referencia la detección del rostro y el segundo experimento consiste en realizar las clasificaciones usando como referencia la detección del torso.

Para generar el dataset se utilizó repositorios de imágenes que están libres en las siguientes bases de imágenes gratuitas.

- Flickr -> <https://www.flickr.com>
- UKBench -> <https://archive.org/details/ukbench>
- Deeplearning -> <http://deeplearning.net/datasets/>

Para complementar el dataset también se utilizaron imágenes descargadas directamente de Google Imágenes.

3.1 CLASIFICACION CON MAQUINAS DE SOPORTE VECTORIAL

El modelo de máquinas de soporte vectorial o SVM, es un modelo de clasificación muy eficiente basado en clasificadores lineales para variables objetivo de tipo booleana pues permite realizar una mejor discriminación entre los dos posibles resultados, minimizando el

sesgo o sobreajuste que se puede generar con otros modelos de clasificación de menor jerarquía.

En el presente proyecto, para el modelo SVM se usó el 80% como datos de entrenamiento (bases abiertas de imágenes) y el 20% como datos de prueba (imágenes descargadas de sitios sospechosos de trata de blancas). Además, como la variable objetivo es binaria, se usará un kernel no lineal. El modelo SVM (Figura 10), necesita la carga de imágenes etiquetadas en clases para la obtención de la imagen clasificada (Salida).

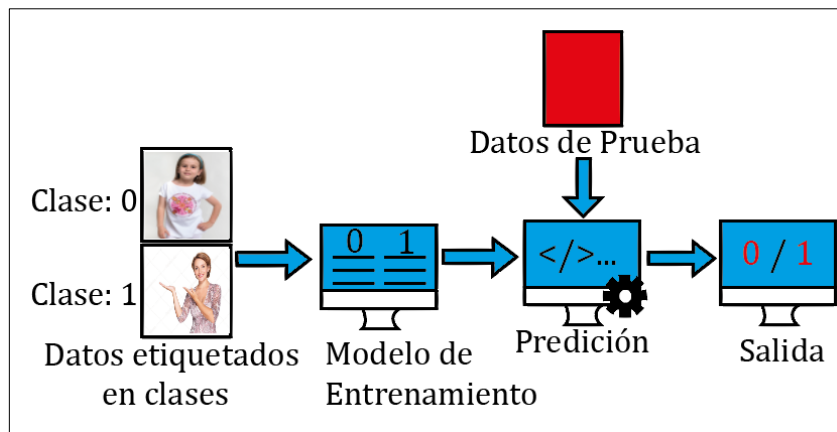


Figura 10 - Esquema SVM

3.1.1 INDICADORES DE RENDIMIENTO DEL MODELO SVM

Para medir el desempeño del modelo SVM se usó la matriz de confusión para contrastar los valores de clasificación reales contra los valores de clasificación predichos (Tabla 5).

Tabla 5 - Esquema Matriz de confusión

		Valores PREDICHOS	
		Clase 1	Clase 2
Valores REALES	Clase 1	Verdaderos Positivos	Verdaderos Negativos
	Clase 2	Falsos Positivos	Falsos Negativos

En la tabla 5 se muestra el esquema de una matriz de confusión. El éxito de una matriz de confusión es que el número de falsos positivos y verdaderos negativos tiendan a cero. Los indicadores usados en el proyecto son los siguientes: exactitud, precisión, recall y F_measures definidos en las ecuaciones (1), (2) y (3) y (4) respectivamente.

$$exactitudSVM = \frac{\text{verdaderos_positivos} + \text{verdaderos_negativos}}{\text{TOTAL}} \quad (1)$$

$$precisión = \frac{\text{verdaderos_positivos}}{\text{verdaderos_positivos} + \text{falsos_positivos}} \quad (2)$$

$$recall = \frac{\text{verdaderos_positivos}}{\text{verdaderos_positivos} + \text{falsos_negativos}} \quad (3)$$

$$F_measures = 2 * \frac{\text{precisión} * \text{recall}}{\text{precisión} + \text{recall}} \quad (4)$$

Se utilizó las características extraídas como entrada para un algoritmo SVM para la clasificación supervisada de grupo de edad y género. El clasificador SVM utiliza una función K de núcleo no lineal para construir la función de límite $f(x)$, definida en la ecuación (5), donde b es el valor de sesgo, y_i y α_i son los parámetros de optimización de Lagrange.

$$f(x) = \sum_{i=1}^m y_i \alpha_i K(x, x_i) + b \quad (5)$$

El modelo SVM como clasificador binario se utilizó en los dos experimentos tomando en cuenta los diferentes criterios de clasificación

3.2 CLASIFICACIÓN CON REDES CONVOLUCIONALES

Este modelo de clasificación comúnmente se ha aplicado para el análisis de imágenes digitales, que por su alto nivel de complejidad se denomina análisis profundo. CNN basa su funcionamiento asemejándose a los procesos biológicos de la naturaleza, en la forma en que los animales utilizan la corteza visual para observar, clasificar y diferenciar un objeto de otros.

CNN es un modelo de aprendizaje no supervisado que requiere un gran dataset de imágenes para clasificarlas en una u otra clase realizando varias iteraciones. La desventaja de CNN es que, por un lado, al ser un modelo de aprendizaje no supervisado, no se tiene control total sobre él. Por otro lado, este modelo requiere demasiado tiempo y recursos computacionales para realizar una buena predicción, ya que necesita muchas imágenes y varias iteraciones para aprender mejor. Para el proyecto actual, el tiempo aproximado requerido para cada iteración es 55 minutos, por lo tanto, realizar un experimento con 10 iteraciones, demandaría entre 9 y 10 horas, solamente para una clasificación binaria. A

pesar de ello se usó 10 iteraciones en cada clasificación de CNN para que el modelo trabaje de forma robusta. En la Figura 11 se muestra un ejemplo básico de CNN aplicando dos convoluciones secuenciales aplicada a una matriz de pixeles.

	Matriz de Entrada	Filtro	Matriz de Salida																																																													
Primera Convolución	<table border="1"> <tr><td>3</td><td>0</td><td>1</td><td>2</td><td>7</td><td>4</td></tr> <tr><td>1</td><td>5</td><td>8</td><td>9</td><td>3</td><td>1</td></tr> <tr><td>2</td><td>7</td><td>2</td><td>5</td><td>1</td><td>3</td></tr> <tr><td>0</td><td>1</td><td>3</td><td>1</td><td>7</td><td>8</td></tr> <tr><td>4</td><td>2</td><td>1</td><td>6</td><td>2</td><td>8</td></tr> <tr><td>2</td><td>4</td><td>5</td><td>2</td><td>3</td><td>9</td></tr> </table>	3	0	1	2	7	4	1	5	8	9	3	1	2	7	2	5	1	3	0	1	3	1	7	8	4	2	1	6	2	8	2	4	5	2	3	9	<table border="1"> <tr><td>1</td><td>0</td><td>-1</td></tr> <tr><td>1</td><td>0</td><td>-1</td></tr> <tr><td>1</td><td>0</td><td>-1</td></tr> </table>	1	0	-1	1	0	-1	1	0	-1	<table border="1"> <tr><td>-5</td><td>-4</td><td>0</td><td>8</td></tr> <tr><td>-10</td><td>-2</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>-2</td><td>-4</td><td>-7</td></tr> <tr><td>-3</td><td>-2</td><td>-3</td><td>-16</td></tr> </table>	-5	-4	0	8	-10	-2	2	3	0	-2	-4	-7	-3	-2	-3	-16
3	0	1	2	7	4																																																											
1	5	8	9	3	1																																																											
2	7	2	5	1	3																																																											
0	1	3	1	7	8																																																											
4	2	1	6	2	8																																																											
2	4	5	2	3	9																																																											
1	0	-1																																																														
1	0	-1																																																														
1	0	-1																																																														
-5	-4	0	8																																																													
-10	-2	2	3																																																													
0	-2	-4	-7																																																													
-3	-2	-3	-16																																																													
Segunda Convolución	<table border="1"> <tr><td>-5</td><td>-4</td><td>0</td><td>8</td></tr> <tr><td>-10</td><td>-2</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>-2</td><td>-4</td><td>-7</td></tr> <tr><td>-3</td><td>-2</td><td>-3</td><td>-16</td></tr> </table>	-5	-4	0	8	-10	-2	2	3	0	-2	-4	-7	-3	-2	-3	-16	<table border="1"> <tr><td>1</td><td>0</td><td>-1</td></tr> <tr><td>1</td><td>0</td><td>-1</td></tr> <tr><td>1</td><td>0</td><td>-1</td></tr> </table>	1	0	-1	1	0	-1	1	0	-1	<table border="1"> <tr><td>-13</td><td>-12</td></tr> <tr><td>-8</td><td>14</td></tr> </table>	-13	-12	-8	14																																
-5	-4	0	8																																																													
-10	-2	2	3																																																													
0	-2	-4	-7																																																													
-3	-2	-3	-16																																																													
1	0	-1																																																														
1	0	-1																																																														
1	0	-1																																																														
-13	-12																																																															
-8	14																																																															

Figura 11 - Convoluciones secuenciales

De la gráfica anterior se pueden deducir dos observaciones importantes:

1. Cada vez que se realiza una convolución a una imagen, la matriz de salida es más pequeña, por lo tanto, solamente se puede realizar un número limitado de convoluciones. Mientras la imagen sea más pequeña, menos convoluciones podrán realizarse, dicho en otras palabras, mientras la imagen sea más pequeña, será más difícil la detección de la característica que se está buscando.
2. Se puede notar que el filtro discrimina de mejor manera a los pixeles centrales, los pixeles de las esquinas no son procesados muy bien, esto puede provocar pérdida de la información. Si hay algún rostro en la esquina es probable que el modelo de reconocimiento facial no lo detecte correctamente, por lo tanto, el éxito de la clasificación dependerá mucho del dataset de imágenes que se ingrese al modelo CNN.

3.2.1 PARTES DE UNA RED CONVOLUCIONAL

Una CNN consta de dos partes: base convolucional y clasificador como se muestra en la Figura 12.

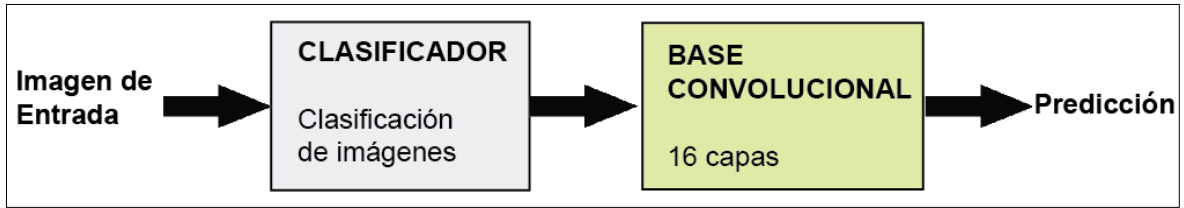


Figura 12 - Arquitectura típica de CNN

3.2.1.1 BASE CONVOLUCIONAL

Consta de 16 capas, en esta fase se extrae todas las características posibles de una imagen mediante un proceso automatizado que involucra un número finito de iteraciones. Lo recomendable es usar 10 iteraciones para no provocar sobreajuste, como salida se genera un modelo a partir de todas las características obtenidas en CNN.

3.2.1.2 CLASIFICADOR CNN

En esta fase se usa el modelo previamente generado para clasificar el dataset de imágenes en dos clases (hombres y mujeres o mayores y menores de 14 años) y se utilizará CNN con la arquitectura VGGNet. Actualmente VGGNet al momento de predecir o clasificar determinados elementos, objetos o características en una imagen es muy eficiente.

3.2.1.3 ARQUITECTURA CNN DE 16 CAPAS

La arquitectura CNN de 16 capas o VGGNET en 2014 obtuvo el primer lugar en el concurso de Visión y Reconocimiento artificial a nivel mundial llamado ILSVRC (ImageNet Large Scale Visual Recognition Competition), y ha sido utilizada en millones de proyectos de reconocimiento facial y Deep learning desde entonces. Esta arquitectura de 16 capas (VGGNet16), contiene algoritmos mejorados a su predecesor AlexNet, usado para reconocimiento de caracteres en los 90's. En la Figura 13 se muestra la arquitectura del modelo de aprendizaje CNN con sus 16 capas, separado en 6 etapas de procesamiento.

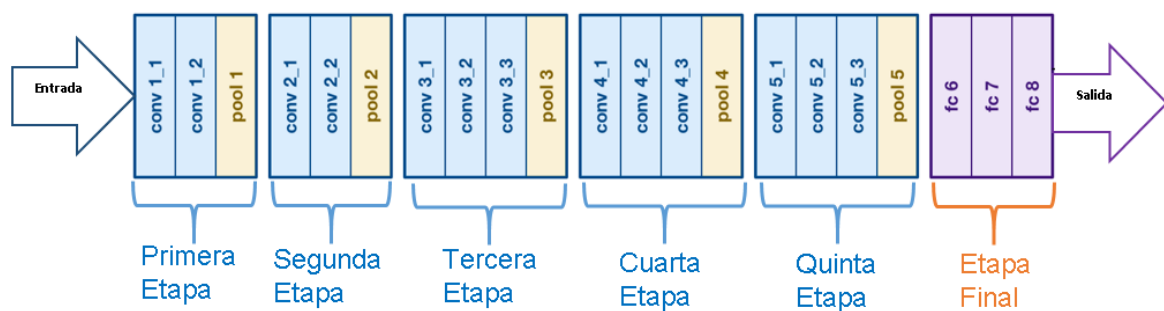


Figura 13 – Proceso de clasificación VGGNet CNN

En las capas de convolución el modelo va extrayendo las características geométricas, mientras que en las capas de agrupamiento(pool) se van concentrando las características para generar una característica global. El modelo va alternando dos capas de convolución y una de agrupamiento en las dos primeras etapas, luego realiza tres capas de convolución y una de agrupamiento en las tres siguientes etapas. En la última etapa utiliza tres capas de conexión total, para obtener como resultado la clase predicha, dependiendo de la variable objetivo como salida la imagen es clasificada por género o grupo de edad.

3.2.2 TIPOS DE ENTRENAMIENTO DE CNN

Independientemente de la arquitectura que se elija, se puede entrenar el modelo de acuerdo a las necesidades del experimento de tres formas diferentes, las cuales se detallan a continuación.

3.2.2.1 ENTRENAMIENTO COMPLETO

Consiste en entrenar todo el modelo, recomendado cuando básicamente se requiere que el modelo aprende desde cero, es muy útil, pero se requiere de una gran cantidad de datos y abundantes recursos computacionales, cada iteración del modelo con todo el conjunto de datos, se llama época, mientras más épocas se realice (contrastando con el tiempo de demora) el modelo mejorará su eficiencia y su precisión al momento de clasificar la imagen.

3.2.2.2 ENTRENAMIENTO PARCIAL

Se entrena algunas capas y otras se mantienen congeladas, es decir, no se tocan, este entrenamiento es recomendado si se dispone de un conjunto pequeño de datos, o parcialmente entrenado.

3.2.2.3 ENTRENAMIENTO SIN BASE CONVOLUCIONAL

Se mantiene la base convolucional en su estado natural, se usa cuando se tiene un grupo pequeño de datos y el modelo a usar está muy bien entrenado, por ejemplo, si se desea utilizar un modelo que ha sido ampliamente demostrado para clasificar perros y gatos; en este caso no es necesario modificar la base convolucional, sino simplemente usar la parte del clasificador para predecir los resultados. En la Figura 14, se muestran los tres tipos de entrenamiento que se puede realizar, dependiendo el problema.

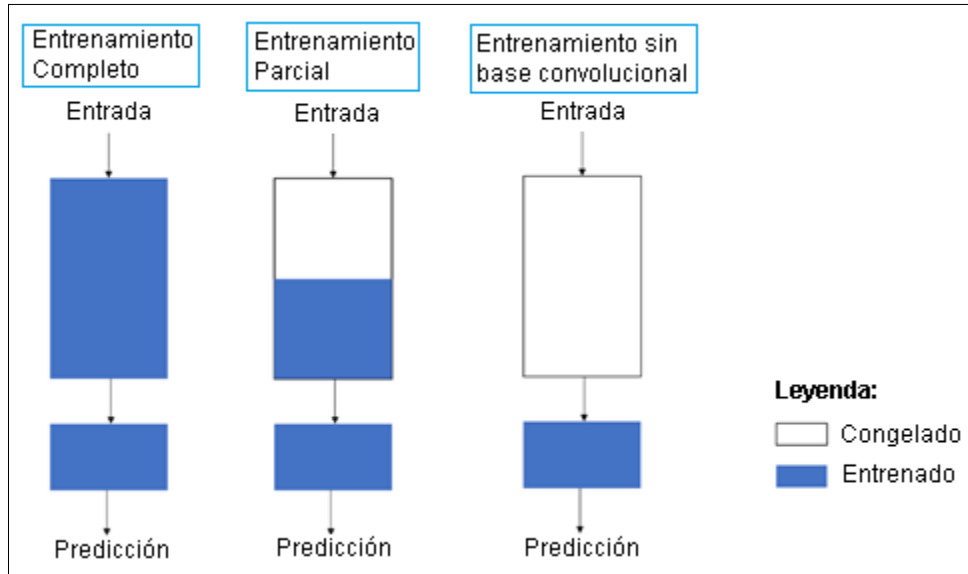


Figura 14 - Tipos de entrenamiento

En el proyecto actual se utilizó el entrenamiento completo, ya que se está creando el modelo desde cero.

3.2.3 INDICADORES DE RENDIMIENTO DEL MODELO CNN

Como indicadores de rendimiento del modelo CNN se utilizó Exactitud y EMC (Error Medio Cuadrado) para medir la clasificación de los elementos predichos como verdaderos y los que realmente son verdaderos, en este caso se dice verdaderos a aquellos elementos (imágenes) que fueron clasificados correctamente, ya sea por género o por grupo de edad.

La diferencia entre el valor real y el valor predicho ($\hat{Y}_i - Y_i$) se llama Error de la predicción EMC es el promedio de los errores en las predicciones, elevado al cuadrado; mientras que la exactitud es la razón entre todos los elementos bien clasificados y el total de la data, por lo tanto, para una óptima clasificación el valor de Exactitud debe tender a 1, mientras que el valor EMC debe tender a 0. Matemáticamente EMC y Exactitud se definen en (6) y (7) respectivamente.

$$EMC = \frac{1}{n} \sum_{i=1}^n (P_i - R_i)^2 \quad (6)$$

Donde:

P_i es el valor predicho, R_i es el valor real observado, i es la i -ésima imagen clasificada y n es el número de imágenes

$$exactitud_{CNN} = \frac{\text{verdaderos_positivos} + \text{verdaderos_negativos}}{\text{Total}} \quad (7)$$

Es importante recalcar que la extracción de características con modelos HAAR es exclusivamente para el modelo SVM, en el caso del modelo CNN se usó directamente el dataset con 4096 imágenes en formato jpg, casi todas las imágenes son frontales o ligeramente perfiladas, el resto son fotos de perfil, con sombreros, gorras, pareos, en fin, un poco de todo obviamente con la presencia de alguna persona en la imagen. Se realizará los dos experimentos. Aplicando los dos criterios de clasificación (por género y por grupo de edad).

4. APLICACIÓN DE PROCESAMIENTO Y CLASIFICACIÓN DE IMÁGENES

Una vez comprendido los fundamentos teóricos conceptuales del proyecto, en este capítulo se explicará el desarrollo de la aplicación, de acuerdo al alcance del proyecto, además se detallará las funcionalidades principales de la aplicación.

4.1 ENTORNO DE DESARROLLO

4.1.1 PYTHON

Python fue seleccionado por su versatilidad y aplicabilidad para el proyecto, en su versión 3.7.4⁴. Al momento de instalar se debe añadir el path o ruta de instalación de Python a las variables de entorno del Sistema Operativo, para que éste puede ejecutar sus comandos sin ningún inconveniente. También hay que recordar que Python es un lenguaje de programación de muy alto nivel y que es fuertemente tipado, por lo que se debe tener especial cuidado con el manejo de identaciones, tabulaciones y el correcto uso de mayúsculas y minúsculas.

Para crear la aplicación, se utilizó una arquitectura de tres capas, las cuales son el modelo, los controladores y para el entorno visual se hará uso de interfaces gráficas para tener una aplicación amigable. Además del Core Python, es necesario la integración de varias bibliotecas que permitirán agregar funcionalidad y mejorar la programación de la aplicación, a continuación, se describen las librerías que se utilizaron.

4.1.1.1 TKINTER

Es una librería que proporciona los elementos visuales necesarios (frames, labels, buttons, widgets, etc.) para crear todo el ambiente gráfico necesario y que la GUI (Interfaz gráfica de usuario), sea muy intuitiva y fácil de utilizar. La librería tkinter se ha ido actualizando periódicamente para mostrar una mejor calidad visual en el front-end de las aplicaciones. Tkinter es un “puente” entre la lógica de Python y el usuario. Para entenderlo de mejor manera en la figura 15 se muestra la estructura de una ventana en Python con tkinter⁵.

⁴ Página oficial de Python, para descargar el software: <https://www.python.org/downloads/release/python-374/>

⁵ Documentación Tkinter: <https://wiki.python.org/moin/TkInter>

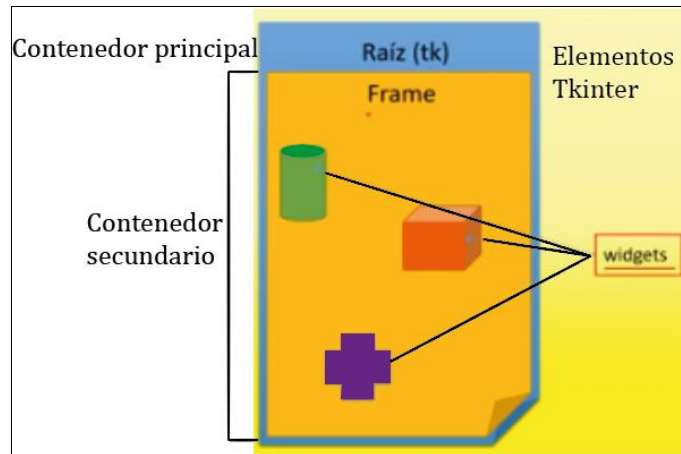


Figura 15 - Estructura de una ventana grafica utilizando tkinter

En la Figura 15 se muestra los elementos que ofrece Tkinter, los componentes que se utilizaron en la construcción de la interfaz del proyecto son:

- Raíz: Contenedor principal
- Frame: Contenedor secundario
- Widget: Botón, etiquetas, cajas de texto, etc.

4.1.1.2 SCRAPY

Esta biblioteca permite rastrear y extraer información de las páginas web, mediante el análisis de sus selectores html y css, en este caso primero se necesita extraer la lista de links o enlaces URL de las imágenes y luego con esta información extraer definitivamente las imágenes de los sitios web sospechosos para posteriormente procesarlos⁶. A continuación, se muestra una lista con algunos comandos de rastreo(Scraping) que se usaron en el proyecto:

- **bench**: Ejecute una prueba rápida de referencia
- **fetch**: Obtiene una URL usando el descargador Scrapy
- **genspider**: Genera un Nuevo spider a partir de plantillas predefinidas
- **runspider**: Ejecuta un spider auto-contenido sin necesidad de crear un proyecto
- **setting**: Obtiene los valores de configuración
- **shell**: Ejecuta la consola interactiva de Scraping
- **startproject**: Crea un Nuevo Proyecto
- **version**: Muestra la versión de Scrapy
- **view**: Abre el URL en el navegador

⁶ Documentación Scrapy: <https://docs.scrapy.org/en/latest/intro/tutorial.html>

Las funcionalidades de Scrapy se pueden utilizar desde una consola de comandos de Windows, por ejemplo, para averiguar la versión de Scrapy bastaría con escribir *scrapy version*, o para acceder a algún sitio web para analizarlo se puede usar *scrapy shell URL_que_desea_analizar*. Es importante digitar el URL directo, es decir, sin hacer uso de comillas simples o dobles. Algunos de los comandos con los que trabaja Scrapy son:

- **scrapy**: Muestra los objetos que contiene: scrapy.Request, scrapy.Selector, etc.
- **Crawler**: Muestra el objeto rastreador
- **Ítem**: Muestra la Lista de ítems
- **Request**: Solicita el recurso de internet, generalmente página web
- **Response**: Devuelve el recurso de internet, estado 200, para confirmar estado OK
- **Settings**: Muestra la configuración del objeto spider
- **Spider**: Muestra el spider predeterminado

Si desea profundizar el uso de técnicas Scrapy con el fin de potenciar su utilidad puede consultar [7].

4.1.1.3 ENTORNOS VIRTUALES

Un entorno virtual es un espacio de trabajo aislado que funciona de manera independiente con respecto a otras aplicaciones que pueden requerir los mismos recursos del Sistema Operativo como ciertos puertos frameworks y librerías propias de Microsoft en este caso. Además, un entorno virtual permite una mejor gestión de los proyectos de Python, ya que editarlos o eliminarlos no afectará de ninguna manera el resto de aplicativos. Para poder crear entornos virtuales, se usó la librería de Python *virtualenv*⁷.

Es importante recalcar que ocasionalmente las aplicaciones de alto nivel Python enfocadas en Scraping y Machine Learning presentan conflictos debido a que otras aplicaciones como XAMPP, Apache, Docker usan dependencias comunes del Sistema Operativo.

4.1.1.4 OPEN CV

Es una biblioteca de código libre estandarizada creada para trabajar conjuntamente con Python. Esta biblioteca es muy útil para realizar tareas de reconocimiento de regiones faciales tanto en imágenes como en videos multimedia, razón por la cual se la emplea mucho para temas de seguridad, control médico o afines.

⁷ Documentación virtualenv: <https://virtualenv.pypa.io/en/latest/>

La biblioteca OpenCV necesita ciertos archivos XML, que contiene los modelos de patrones prediseñados de reconocimiento para partes específicos del rostro como: nariz, frente, boca, ojos, etc.; estos archivos XML deberán ser descargados desde internet, extra oficialmente existe un proyecto en GitHub que cuenta con todos estos ficheros. OpenCV se utilizará única y exclusivamente para preprocesar las imágenes que entrarán al modelo SVM⁸.

4.1.2 TENSORFLOW

Es una librería de código abierto utilizado para aprendizaje automatizado, fue desarrollado por Google, pero se puede integrar muy fácilmente con los proyectos realizados en Python, C++, Java, Go y otros, gracias a sus múltiples APIs. Es muy útil para la creación y configuración de clasificadores de imágenes ya que el objetivo de Tensorflow es exclusivamente entrenar las redes neuronales. Tensorflow fue liberado para la comunidad en noviembre del 2015, esto aportó muchísimo a que los proyectos de aprendizaje automatizado por medio de redes neuronales mejoren su efectividad, en el proyecto actual TensorFlow se usará para procesar la data con el modelo CNN⁹.

4.1.2.1 KERAS

Keras es un API de redes neuronales de alto nivel escrito en Python exclusivamente y es usado en proyectos de aprendizaje automatizado como Machine Learning y Deep Learning, de manera rápida y sencilla. Además, Keras es muy intuitivo y se integra muy bien con Tensorflow para que los resultados de la investigación sean los más fiables posibles. Keras permite configurar los parámetros del modelo CNN de una forma mucho más organizada¹⁰.

4.1.3 SKLEARN

Es una biblioteca de código abierto que se puede utilizar con programas hechos en Python para realizar modelos SVM, para clasificar imágenes gracias a sus potentes APIs desarrolladas para machine learning. Si desea conocer más acerca de SKLearn puede revisar la documentación.¹¹

⁸ Para descargar la biblioteca OpenCV puede hacerlo siguiendo el enlace: <https://github.com/opencv/opencv>

⁹ Documentación Tensorflow: https://www.tensorflow.org/api_docs

¹⁰ Documentación Keras: <https://keras.io/>

¹¹ Documentación SKLearn en el siguiente enlace: <https://scikit-learn.org/stable/documentation.html>

4.2 INTERFAZ GRÁFICA

En la Figura 16 se muestra la interface gráfica de la aplicación que se desarrolló en el presente proyecto [23] y su funcionamiento, la cual permite procesar la información de una forma más intuitiva y amigable.



Figura 16 - Ventana Principal de la aplicación

La interface desarrollada tiene las siguientes funcionalidades:

4.2.1 ANALIZAR SITIO WEB

El usuario debe digitar el enlace URL de la página web que desea procesar, este campo obviamente es obligatorio, si el enlace está escrito correctamente y el sitio web está disponible, al presionar el botón “*Analizar Sitio Web*” la aplicación descargará en un archivo de texto plano (con extensión *.txt) todas las URLs de las imágenes alojadas en ese sitio web; si la url está mal escrita, si el sitio web no está disponible, o si el sitio web no contiene imágenes no se generará ningún archivo ya que la aplicación no pudo detectar enlaces existentes hacia las imágenes.

4.2.2 SELECCIONAR ARCHIVO

Una vez descargado el archivo plano de URLs, al presionar el botón “*Seleccionar Archivo*”, se despliega un explorador que permite cargar este archivo y a continuación se crea un

directorio local en el disco duro y se almacenan todas las imágenes posibles ligadas a la lista de enlaces (Figura 17).

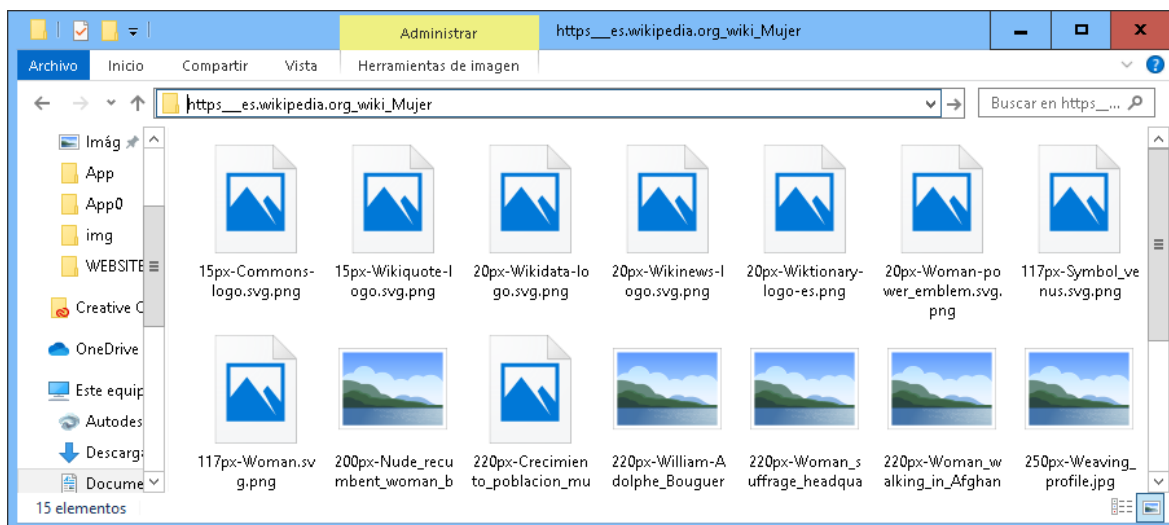


Figura 17 - Imágenes descargadas a partir del archivo plano

4.2.3 RECONOCIMIENTO FACIAL

Esta es la funcionalidad más potente de todas, al presionar el botón “Reconocimiento Facial” el usuario puede cargar cualquier imagen en formato .jpg (Figura 17), luego el algoritmo empieza a procesar la imagen, en caso de no detectar el rostro parecerá un mensaje indicando que no se ha detectado ningún rostro, caso contrario la aplicación devolverá la imagen detectando el rostro y mostrando el género de la persona, además se indicará si el rostro pertenece o no a una persona menor de 14 años de edad. En la Figura 18, se puede observar un reconocimiento facial de dos muchachos (llamados ficticiamente Alice y Bob). La niña Alice es detectada como **M,<14** una mujer menor de 14 años y el chico Bob es detectado como **H,14** un hombre menor de 14 años.

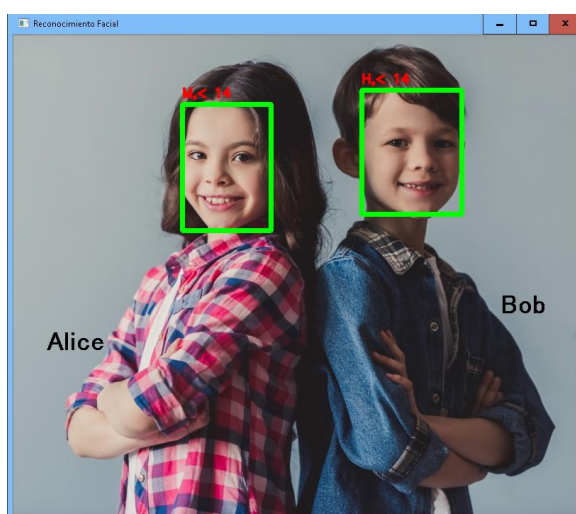


Figura 18 - Detección de Rostros Alice y Bob

4.2.4 CLASIFICADOR CNN

Al presionar el botón “Clasificador CNN” se despliega la siguiente ventana (Figura 19).

REDES NEURONALES CONVOLUCIONALES

Datos de Entrenamiento

Datos de Validación

Número de Iteraciones

----- Si ya generó el modelo, proceda a cargar los datos de prueba -----

Datos de Prueba

Figura 19 - Ventana CNN

En la figura 19 se muestra la ventana donde se carga los datos de entrenamiento, de prueba y se indica el número de iteraciones que se va a realizar, luego se procede a generar el modelo CNN; una vez generado el modelo, se cargan los datos de prueba que se van a clasificar, al presionar el botón *EJECUTAR MODELO CNN*, devuelve un archivo plano csv con los resultados obtenidos de la clasificación.

4.2.5 CLASIFICADOR SVM

Al presionar el botón “Clasificador SVM” se despliega la siguiente ventana (Figura 20).

MAQUINAS DE SOPORTE VECTORIAL

Datos de Entrenamiento

Datos de Prueba

Kernel Utilizado

Figura 20 - Ventana SVM

En la Figura 20 se muestra la ventana donde se carga los datos de entrenamiento y los datos de prueba, luego al presionar el botón *EJECUTAR MODELO SVM*, se selecciona el kernel, para el proyecto, el kernel con el que se obtuvo mejores resultados es el kernel *lineal*, finalmente se aplica el modelo no supervisado de máquinas de soporte vectorial (estudiado en los capítulos anteriores), como resultado se obtendrá la matriz de confusión, para medir el rendimiento del modelo.

5. RESULTADOS EXPERIMENTALES

En este capítulo, se detalla los resultados obtenidos en los dos experimentos (Rostro y Torso) para la clasificación de las imágenes usando ambos modelos. Los diferentes experimentos para el modelo SVM se analizará con los resultados de la matriz de confusión para que sean lo más objetivos posible; mientras que para el caso de CNN los resultados de rendimiento se analizaran con los indicadores EMC (Error Medio Cuadrado) y exactitud.

Las clasificaciones realizadas utilizando SVM y CNN están estructuradas en el siguiente orden:

- Dataset de imágenes detectando rostro
 - Clasificación por género
 - Clasificación por grupos de edad
- Dataset de imágenes detectando torso
 - Clasificación por género
 - Clasificación por grupos de edad

Primero se muestra los resultados de CVM y luego los de CNN

5.1. RESULTADOS SVM

Se necesita que la data se encuentre bien clasificada y etiquetada por clases para proceder a aplicar el modelo SVM. Para medir la calidad de la clasificación, se utilizó la matriz de confusión como base para obtener la exactitud del modelo. La desventaja que presenta SVM es que, mientras más grande sea la data y existan más clases resultantes, tomará más tiempo realizar la clasificación manual para entrenar al modelo por primera vez.

5.1.1 EXPERIMENTO 1: CLASIFICACIÓN IMÁGENES CON ROSTRO

En esta etapa de clasificación se utilizaron grupos de imágenes donde el modelo HAAR detectó el rostro de las personas en las imágenes.

5.1.1.1 CLASIFICACIÓN POR GÉNERO

En la Tabla 6 se muestra la matriz de confusión y luego se obtuvieron los indicadores pertinentes para la clasificación por género en imágenes donde se muestra el rostro de la persona.

Tabla 6 – Matriz de confusión Resultados SVM por genero con rostro

		VALOR PREDICHO	
		hombres	mujeres
VALOR REAL	hombres	160	20
	mujeres	40	100

Como se puede observar en la Figura 21, el valor de exactitud del modelo es 81,25% mientras que el valor EMC es 3,52%.

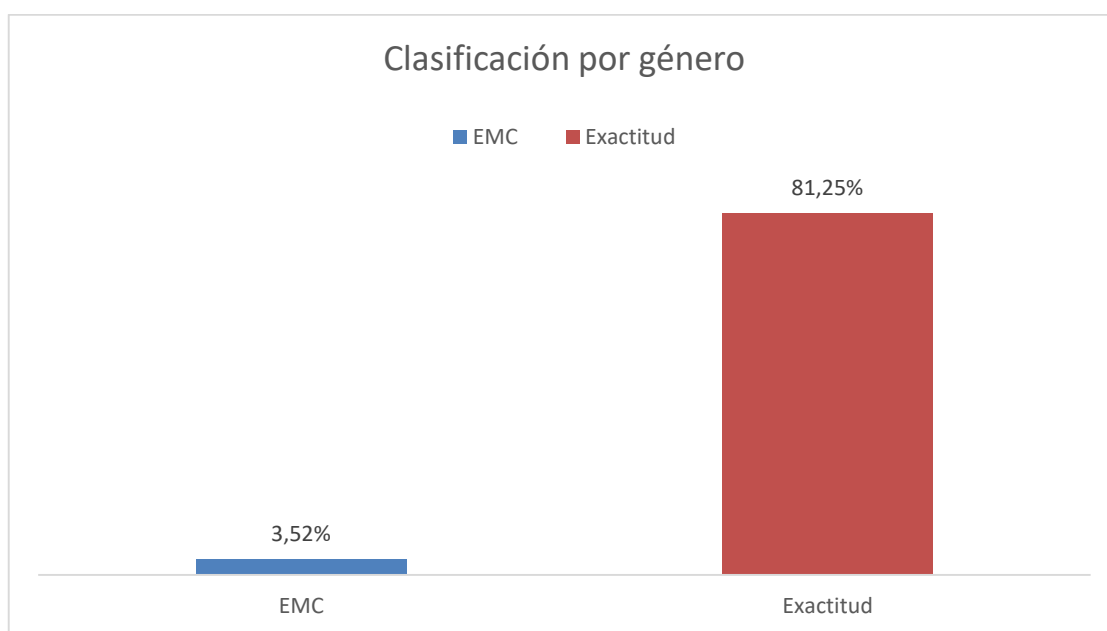


Figura 21 - Resultados clasificación SVM por género

Por los resultados mostrados en la Figura 21 se comprueba que en este caso específico de clasificación SVM clasifica correctamente al menos 8 de cada 10 individuos.

5.1.1.2 CLASIFICACIÓN POR GRUPOS DE EDAD

El siguiente experimento consiste en realizar la clasificación por grupos de edad. La matriz de confusión y los resultados se muestran en la tabla 7.

Tabla 7 – Matriz de confusión grupos de edad con rostro

		VALOR PREDICHO	
		Mayores14	Menores14
VALOR REAL	Mayores14	170	20
	Menores14	40	80

Como se puede apreciar en la Figura 22, el valor de exactitud del modelo es 80,65% mientras que el valor EMC es 3,75% para la clasificación por grupo de edad. SVM clasifica correctamente a 8 de cada 10 individuos.

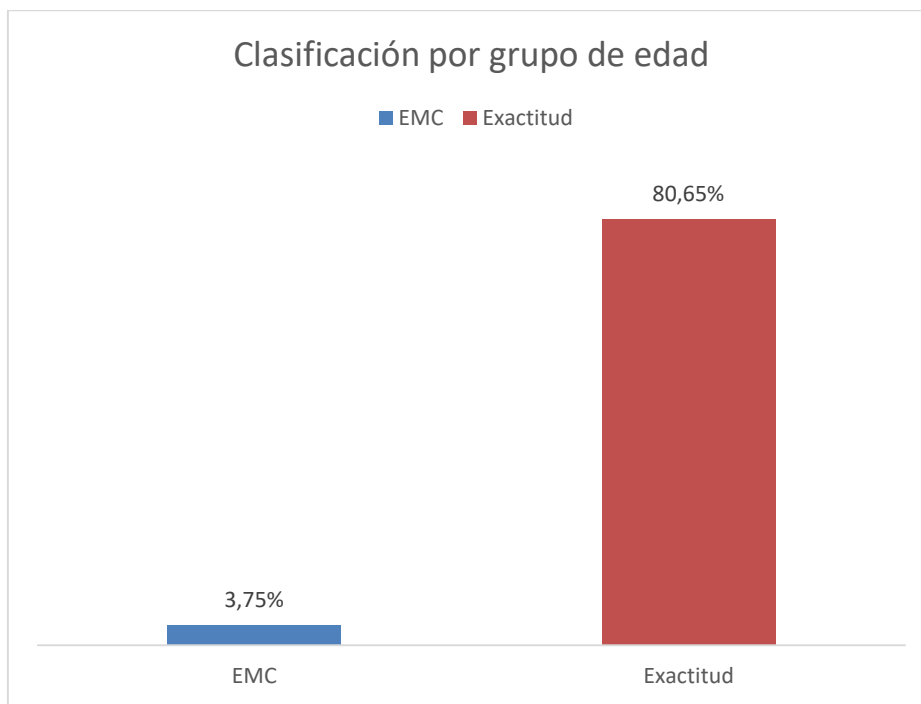


Figura 22 - Resultados clasificación SVM por grupos de edad

A manera de conclusión se puede decir que los resultados obtenidos al clasificar por género y por edad son bastante aceptables 81,25% y 80,65% respectivamente, lo cual indica que SVM es un modelo de clasificación aceptable y recomendable para clasificar imágenes con rostros, pues de cada 10 individuos, al menos 8 son clasificados correctamente.

5.1.2 EXPERIMENTO 2: CLASIFICACIÓN IMÁGENES CON TORSO

En la segunda etapa de la clasificación usando SVM, se trabajó con imágenes donde se detectó el torso de las personas. Este caso de estudio es de mucha importancia para el proyecto porque muchas imágenes son manipuladas digitalmente para que no se pueda reconocer el rostro de las personas.

5.1.2.1 CLASIFICACIÓN POR GÉNERO

El primer experimento consiste en clasificar por género. Se obtuvo la matriz de confusión de la Tabla 8, con los respectivos indicadores.

Tabla 8 – Matriz de confusión SVM por género solo torso

		VALOR PREDICHO	
		Mayores14	Menores14
VALOR REAL	mayores14	220	50
	menores14	50	180

Como se puede observar en la Figura 23 el valor de exactitud es 81,63% mientras que el valor EMC es 3,37% para la clasificación por género.

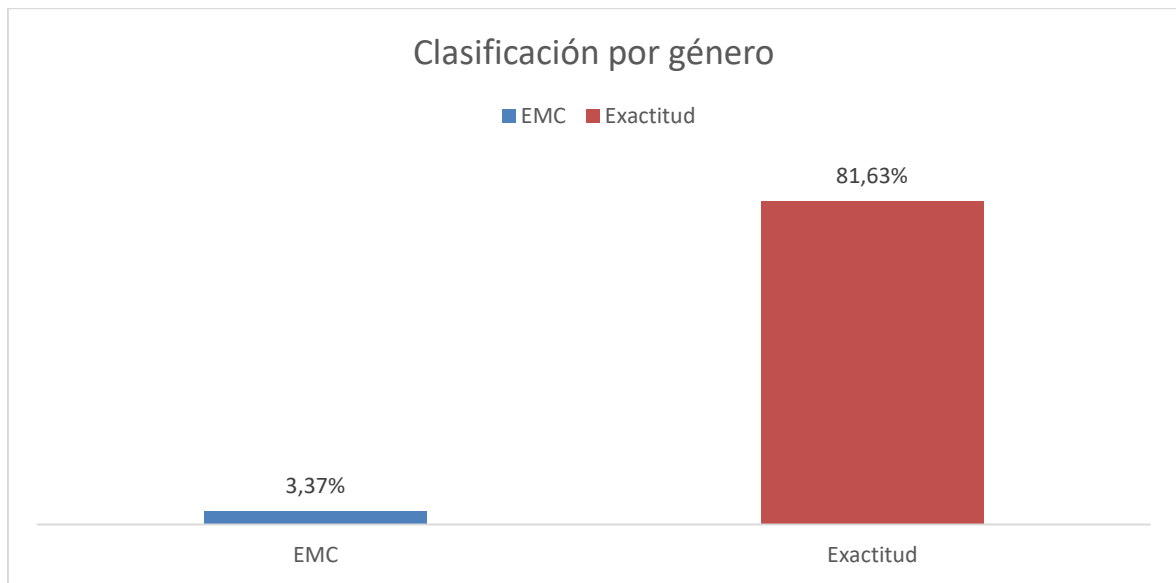


Figura 23 - Indicadores clasificación SVM Género solo torso

5.1.2.2 CLASIFICACIÓN POR GRUPOS DE EDAD

En el siguiente experimento se procedió a clasificar las imágenes por grupos de edad, generando la matriz de confusión que se muestra en la Tabla 9.

Tabla 9 - Resultados SVM por grupos de edad solo torso

		VALOR PREDICHO	
		Mayores14	Menores14
VALOR REAL	Mayores14	110	30
	Menores14	20	120

Como se puede observar en la Figura 24, el valor de exactitud es de 82,14% mientras que el valor EMC equivale a 3,19% cuando se clasifica torsos por grupo de edad.

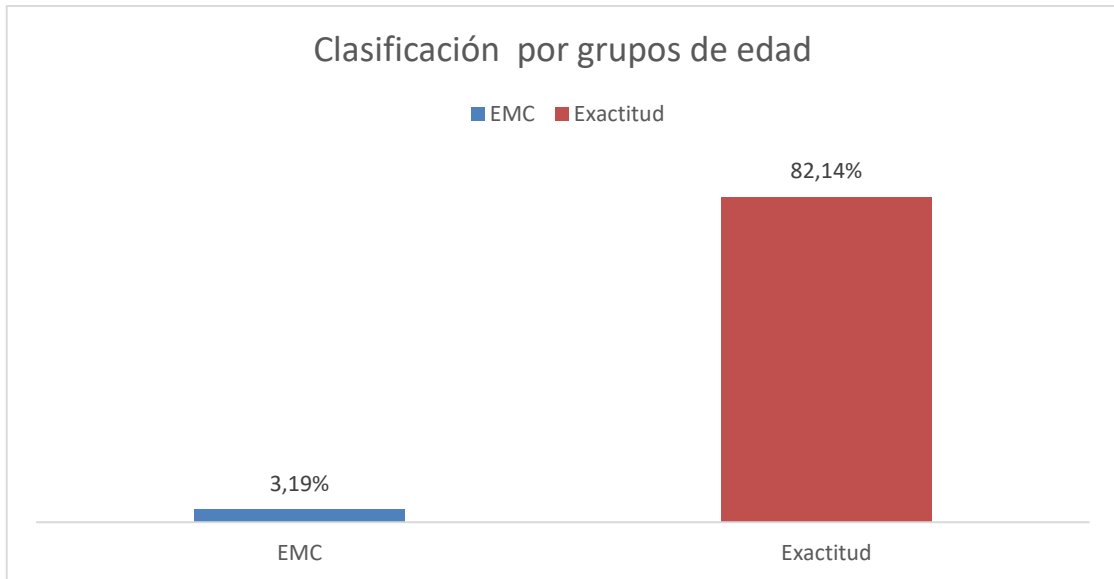


Figura 24 - Indicadores clasificación SVM grupo edad solo torso

De estos experimentos (Figuras 23 y 24) se puede concluir que los resultados obtenidos al clasificar por género y por edad tienen niveles exactitud de 81,63% y 82,14% respectivamente, lo cual indica que SVM es un buen modelo para clasificar imágenes con torsos, aunque no está del todo afinado este modelo predice correctamente al menos 8 individuos de cada 10.

Posteriormente se comparó los resultados de exactitud clasificando rostros y torsos con SVM tal como se muestra en la Figura 25.

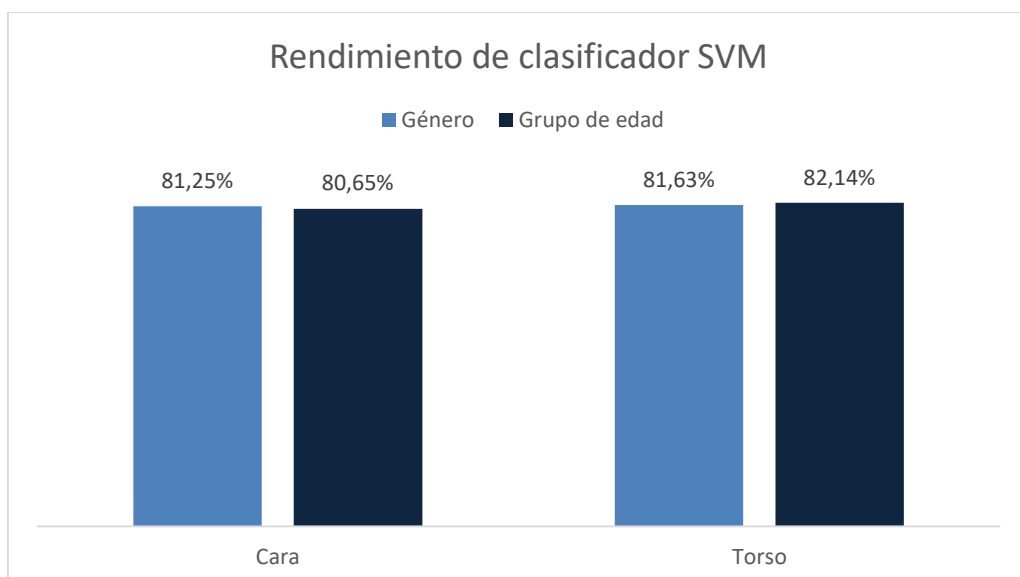


Figura 25 - SVM Rostro y Torso

Como se puede observar en la Figura 25, los valores de exactitud en torsos son 81,63% clasificando Género y 82,14% clasificando Grupos de edad, mientras que en imágenes con rostro el valor de exactitud es 81,25% clasificando Género y 80,65%, clasificando Grupos de edad, el rendimiento del modelo es bastante regular y estable para todos los casos de clasificación tratados en este proyecto.

5.2. RESULTADOS CNN

A continuación, se muestran los diferentes experimentos y los resultados obtenidos al aplicar el modelo CNN. Como indicadores principales se usaron el Error Medio Cuadrado y la exactitud para corroborar el rendimiento del modelo.

5.2.1 EXPERIMENTO 1: CLASIFICACIÓN DE IMÁGENES CON ROSTRO

En este experimento para la clasificación se utilizaron grupos de imágenes donde se usa como referencia el rostro de las personas.

5.2.1.1 CLASIFICACIÓN POR GENERO

Se realizaron 10 iteraciones y se muestra el EMC y el valor de precisión obtenido en cada iteración, como se muestra en la Figura 26.

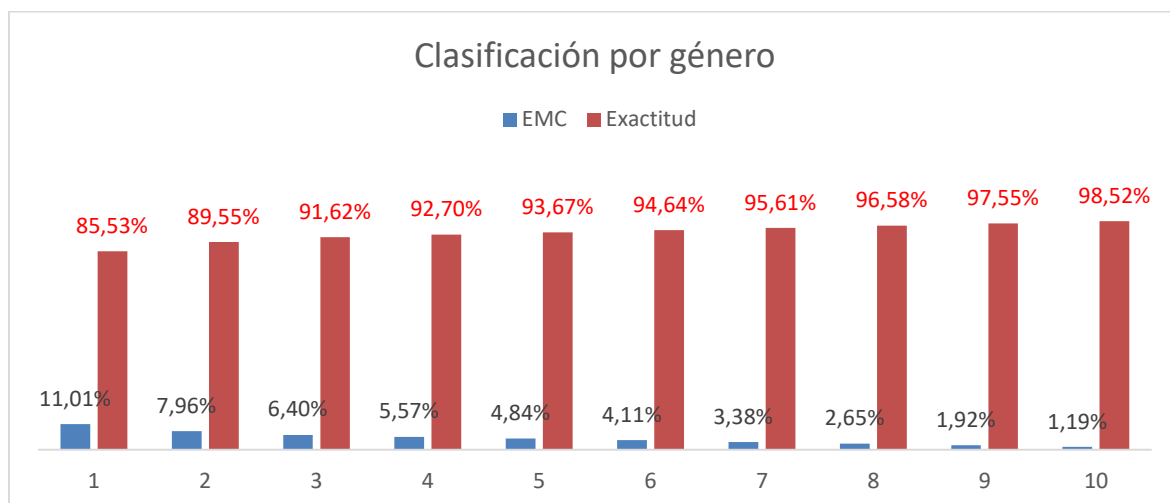


Figura 26 - Clasificación CNN por género para imágenes con rostro

Como se puede observar en la Figura 26, el valor de exactitud es 85,53% en la primera iteración, luego va aumentando progresivamente hasta llegar al valor de 98,52% en la última iteración. Para medir la precisión global de este modelo, se tomará como referencia el valor final de exactitud de las 10 iteraciones, equivalente a 98,52%, para este caso el valor de exactitud es excelente.

5.2.1.2 CLASIFICACIÓN POR GRUPOS DE EDAD

Ahora se usará un grupo de imágenes donde se busca predecir el grupo de edad de las personas como se muestra en la Figura 27.

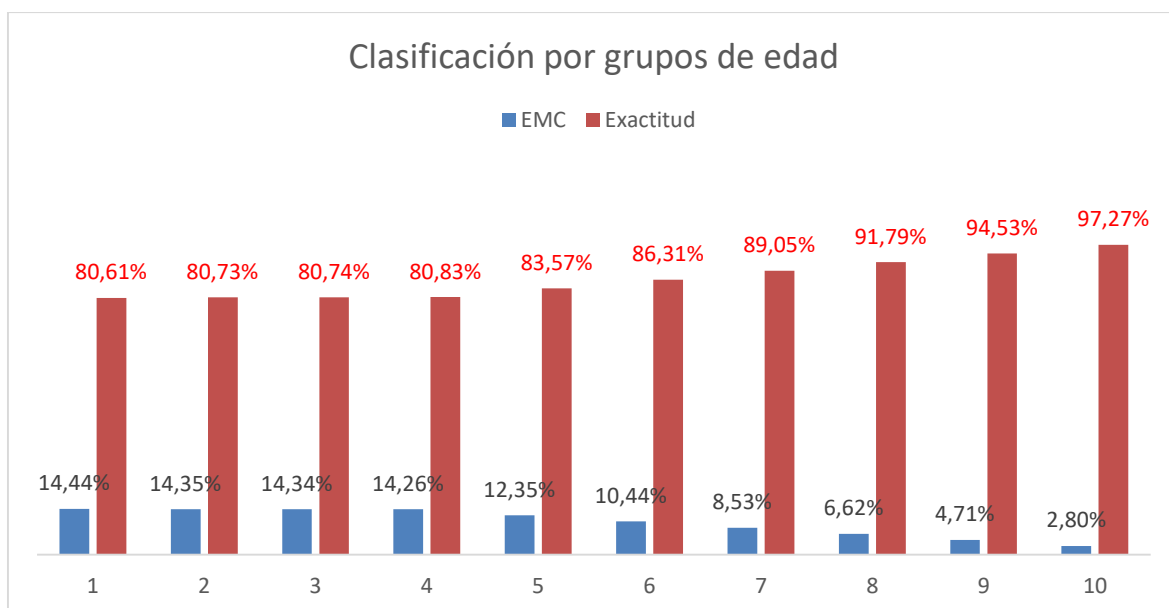


Figura 27 - Grupos de Edad usando CNN en imágenes con rostro

Como se muestra en la Figura 27 el valor de exactitud es 80,61% en la primera iteración, luego empieza a incrementar debido a que el modelo sigue aprendiendo con el transcurso de las iteraciones, de tal forma que al llegar a la décima iteración el valor de exactitud es de 97,27%.

5.2.2 EXPERIMENTO 2: CLASIFICACIÓN IMÁGENES CON TORSO

Para el siguiente experimento se realizó la clasificación utilizando imágenes donde se toma como referencia las características geométricas del torso de las personas. Este modelo permitirá clasificar imágenes donde no se muestre o este distorsionado la cara de la víctima, ya que cuando se trata de personas menores de edad sus rostros son pixelados digitalmente para ocultar su identidad.

5.2.2.1 CLASIFICACIÓN POR GÉNERO

Como se puede observar en la Figura 28, el valor de exactitud en la primera iteración es 50,66% luego aumenta periódicamente hasta llegar a 64,21% en la última iteración.

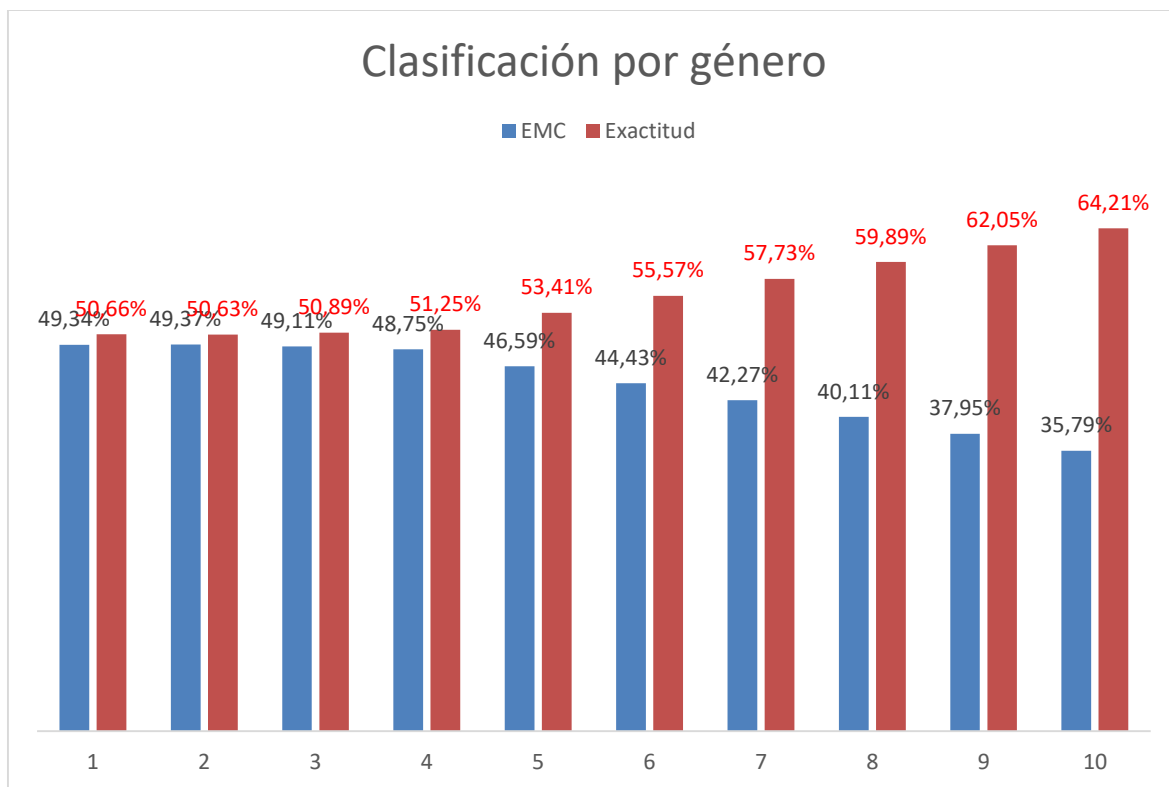


Figura 28 - Clasificación CNN por género imágenes solo torso

Se tomó el valor final de exactitud de las 10 iteraciones, equivalente a 64,21% que en términos generales indica que el modelo tiene muchas limitaciones al realizar esta clasificación.

5.2.2.2 CLASIFICACIÓN POR GRUPOS DE EDAD

Como se puede observar en la Figura 29, el valor de exactitud en la primera iteración es de 51,13% y se mantiene prácticamente constante durante las 10 iteraciones.

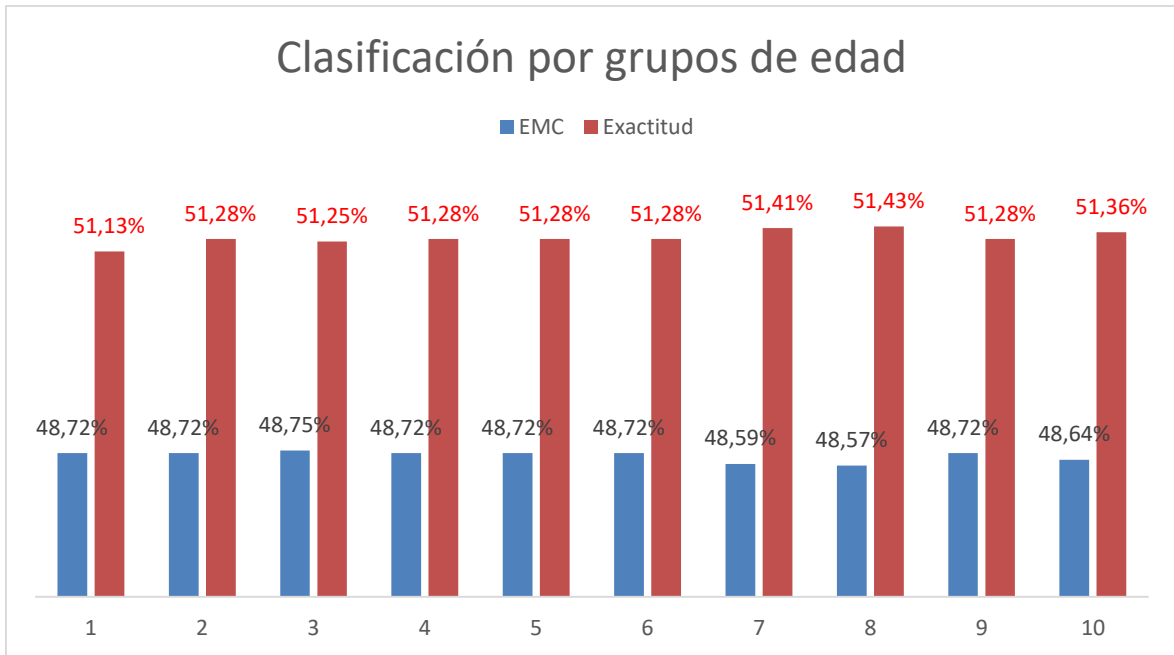


Figura 29 - Clasificación CNN por grupos de edad solo torso

Se puede observar que la exactitud de la clasificación está muy por debajo de lo esperado, el valor final de exactitud en la última iteración es de 51,13%, que prácticamente es como seleccionar las imágenes al azar lanzando una moneda.

Al observar los valores de exactitud clasificando por género y por grupos de edad 64,21% y 51,13% respectivamente donde únicamente se muestra el torso, se puede apreciar que CNN falla drásticamente, esto se debe a que CNN tiene muy buenos algoritmos para clasificar rostros, pero no para clasificar torsos.

Adicionalmente se contrastaron los resultados de exactitud clasificando rostros y torsos (experimento 1 y 2) obteniendo los siguientes resultados (Figura 30).

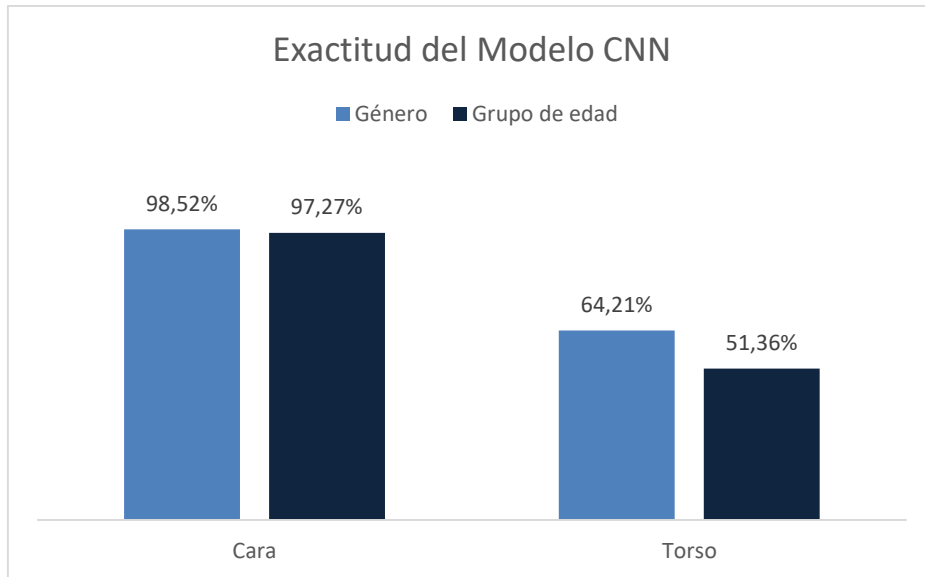


Figura 30 - CNN Torso y Rostro

Como se puede observar en la Figura 30, bajo cualquier criterio de clasificación (género o grupo de edad) el modelo CNN es mucho más eficiente, única y exclusivamente cuando existen rostros (características faciales) en la imagen, mientras que en grupos de imágenes que solo tengan torsos este modelo resulta ineficiente.

5.3 VALIDACION DE LOS MODELOS DE CLASIFICACIÓN

Para validar los modelos de clasificación se usó el criterio Ground truth que consiste en comparar los resultados obtenidos del modelo con los resultados proporcionados por observación y etiquetado directo. La validación del modelo se realizó exclusivamente para las imágenes con torso, porque el objetivo del proyecto está focalizado en clasificar este tipo de imágenes, se trabajó con 1024 imágenes etiquetadas por género y por grupo de edad.

5.4 COMPARACIÓN ENTRE MODELOS DE CLASIFICACIÓN

Al comparar la exactitud de los dos modelos de clasificación, CNN y SVM, se obtienen los resultados de rendimiento. Para ello, se realizó la comparación tomando un grupo de fotos donde se reconoce el rostro de la persona y otro grupo de imágenes donde solo se encuentra el torso de la persona.

5.4.1 COMPARACIÓN IMÁGENES CON ROSTRO

En la Figura 31 se puede notar que el modelo CNN tiene una exactitud ligeramente superior al momento de clasificar imágenes, 98,52% y 97,27% para género y grupos de edad, respectivamente. Mientras que SVM clasifica aceptablemente por genero con un 81,25% y por grupos de edad, con un valor 80,65% de exactitud.

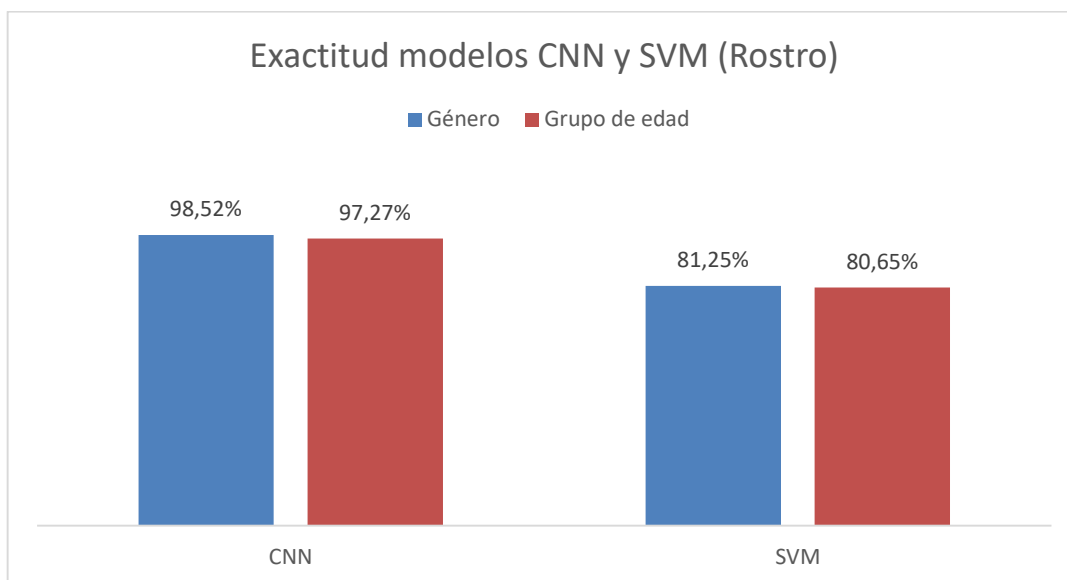


Figura 31 - Comparación CNN y SVM para imágenes con rostro

5.4.2 COMPARACIÓN IMÁGENES CON TORSO

En la Figura 32, se puede notar que SVM tienen un buen rendimiento al clasificar por género 81,63% y por grupos de edad 82,14%, mientras que CNN clasifica pobremente con un rendimiento de 51,36% para grupos de edad y 64,21% para género.

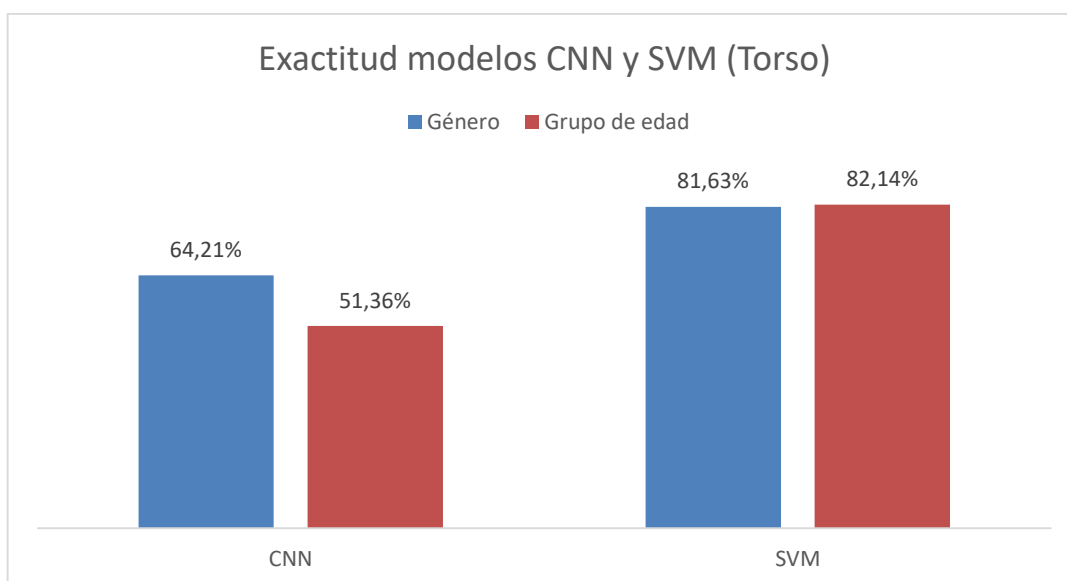


Figura 32 - Comparación SVM y CNN para imágenes con torso

Por los resultados obtenidos anteriormente, en especial en la Figura 31 y dado que el objetivo principal de éste proyecto es detectar el género y el grupo de edad en imágenes donde no se muestra el rostro de la persona, porque presumiblemente se trata de menores de edad que están sufriendo algún tipo de extorsión, queda demostrado que el modelo idóneo para utilizar en nuestro proyecto es SVM, por su exactitud y regularidad en todos los experimentos donde se trabajó con torsos.

Finalmente se debe recalcar que el proyecto realizado es pionero en el ámbito de clasificación de imágenes sin rostro, se ha revisado investigaciones y avances realizados en artículos científicos y no existen trabajos similares enfocados en el procesamiento de imágenes asociando el torso de la persona involucrada. Además, los resultados del proyecto pueden ser aplicados a investigaciones sobre trata de personas, desaparición o secuestro. Esta información puede ser de mucha utilidad para la policía o contrapartes oficiales.

TRABAJOS A FUTURO

Dentro de los trabajos futuros derivados del presente proyecto se plantean las siguientes ideas:

- El estudio de las características que tengan en cuenta las diferencias étnicas en las proporciones del cuerpo y rasgos faciales.
- Ampliar la funcionalidad del proyecto para poder extraer características geométricas del cuerpo (no faciales) de otros formatos de imágenes, incluso en videos en todas sus variantes.
- Detección de problemas médicos analizando características extraídas de imágenes de torso, espalda, piernas, etc.

CONCLUSIONES

- La tecnología de los algoritmos de reconocimiento facial y de aprendizaje automatizado han ido mejorando con el pasar de los años, en el último evento ILSVRC 2014 se logró una precisión estimada de 90% +- 5%. En condiciones óptimas esta medida puede superar incluso a la capacidad de reconocimiento visual que utiliza el ser humano.
- Al realizar las experimentaciones se pudo observar que los algoritmos se encuentran más afinados para clasificar género y grupo de edad, cuando detectan el rostro humano.
- Al realizar varias pruebas e iteraciones se comprobó que los algoritmos de clasificación son un poco más efectivos al momento de predecir el género (hombres y mujeres), antes que el grupo de edad.
- Combinando distintos métodos de reconocimiento es posible obtener mejores resultados que trabajar por separado. En el proyecto para la extracción de características se complementó el clasificador SVM con el modelo HAAR.
- Al analizar imágenes hay muchos factores a tomar en cuenta como: tamaño, color, opacidad, resolución, tipos de formatos de imágenes, canales de color, etc.; la calidad de la data incide directamente con la calidad de resultados obtenidos. Por tal motivo, en la presente tesis se realizó un pre procesamiento de las imágenes.
- La detección de torsos por medio de redes convolucionales es un tema poco investigado, razón por la cual hay muchos proyectos de investigación que tratan constantemente de mejorar su aplicabilidad para que sea de utilidad a una mayor cantidad de usuarios; mientras que SVM es un modelo clásico ampliamente aceptado.

AGRADECIMIENTOS

Se agradece públicamente por su auspicio a:

La Escuela Politécnica Nacional del Ecuador

La Facultad de Sistemas

Las Dras. Myriam Hernández y Lorena Barona por su auspicio en el proyecto PIS-17-10: Minería de Datos, Extracción de características y obtención de modelo con reconocimiento de patrones y aprendizaje automatizado para detección de escenarios relacionados con el delito de tráfico de personas.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Hughes, D. (2008). Wilberforce can be free again: Protecting trafficking victims. Recuperado de <https://www.nationalreview.com/2008/03/wilberforce-can-free-again-donna-m-hughes/>
- [2] Naciones Unidas (2014), Informe mundial sobre la trata de personas. Recuperado de https://www.unodc.org/documents/data-and-analysis/glotip/GLOTIP14_ExSum_spanish.pdf
- [3] Geng, X., Zhou, Z., H., & Smith-Miles, K. (2007). Automatic age estimation based on facial agings patterns. IEEE Transactions on pattern analysis and machine intellisence, 29(12), 2234-2240.
- [4] Hewahi, N., Olwan, A., Tubeel, N., El-Asar, S., & Abu-Sultan, Z. (2010). Age estimation based on neural networks using face features. Journal of Emerging Trends in Computing and Information Sciences, 1(2), 61-67.
- [5] Van de Wolfshaar, J., Karaaba, M. F., & Wiering, M. A. (2015, December). Deep convolutional neural network and support vector machines for gender recognition. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 188-195). IEEE.
- [6] Peter Harrington. Machine learning in action. Manning Publications Co., 2012 (pp. 384)
- [7] Ravi Janani, (2018), Extracting Structured Data from the Web Using Scrapy _ Pluralsight Recuperado de <https://app.pluralsight.com>.
- [8] Dehshibi, M.M., & Bastanfard, A. (2010). A new algorithm for age recognition from facial images. Signal Processing, 90(8), 2431 – 2444.
- [9] Diederik Kingma, Jimmy Ba, (2015), Adam: A method for stochastic optimization Recuperado de <https://arxiv.org/pdf/1412.6980.pdf>

- [10] Godoy Daniel, (2018) Understanding binary cross-entropy / log loss: a visual explanation. Recuperado de <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>
- [11] Brendan Klare, Scott Klum, Joshua Klontz, Emma Taborsky, Tayfun Akguly, Anil Jain, (2014) Suspect Identification Based on Descriptive Facial Attributes
- [12] Lacey Best-Rowden, Shiwani Bisht, Joshua C Klontz, and Anil KJain. Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. In IEEE International Joint Conference on Biometrics, pages 1–8. IEEE, 2014.
- [13] Brendan F Klare, Austin Blanton, and Benjamin Klein. Efficient face retrieval using synecdoches. In IEEE International Joint Conference on Biometrics, pages 1–7. IEEE, 2014.
- [14] Spyros Gidaris and Nikos Komodakis. Object detection via a multi-region and semantic segmentation-aware cnn model. In Proceedings of the IEEE international conference on computer vision, pages 1134–1142, 2015.
- [15] Priyadarshini Adyasha Pattanaik, Zelong Wang, and Patrick Horain. Deep cnn frameworks comparison for malaria diagnosis. arXiv preprint arXiv:1909.02829, 2019
- [16] Hyojoo Son, Changwan Kim, and Hyunchul Choi. High-quality as-is 3d thermal modeling in mep systems using a deep convolutional network. Advanced Engineering Informatics, 42:100999, 2019
- [17] Hao Yin, Zuhong Ou, Shengquan Huang, and Anbo Meng. A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition. Energy, 189:116316, 2019
- [18] Tyldum, G., “Limitations in research on human trafficking,” International Migration, 48(5), 2010, pp. 1-13.
- [19] Myriam Hernández-Álvarez. Detection of possible human trafficking in twitter. In 2019 International Conference on Information Systems and Software Technologies (ICI2ST), pages 187–191. IEEE, 2019.

[20] Dehshibi, M.M., & Bastanfard, A. (2010). A new algorithm for age recognition from facial images. *Signal Processing*, 90(8), 2431 – 2444

[21] Jos van de Wolfshaar, Mahir F Karaaba, and Marco A Wiering. Deep convolutional neural networks and support vector machines for gender recognition. In *2015 IEEE Symposium Series on Computational Intelligence*, pages 188–195. IEEE, 2015.

[22] Edmund Tong, Amir Zadeh, Cara Jones, and Louis-Philippe Morency. Combating human trafficking with deep multimodal models. *arXiv preprint arXiv:1705.02735*, 2017

[23] Granizo L. Sergio (2019) Procesamiento de Tuits usando modelos de clasificación de Machine Learning Recuperado de: <https://github.com/sergiogranizo2014/ProcesamientoImágenes>

[24] Mario E Cueva Hurtado and Diego Javier Alvarado Sarango. Análisis de certificados ssl/tls gratuitos y su implementación como mecanismo de seguridad en servidores de aplicación. *Enfoque UTE*, 8(1):273–286,2017

[25] Marta Lucía Guevara, Julian David Echeverry, and William Ardila Urueña. Detección de rostros en imágenes digitales usando clasificadores en cascada. *Scientia et technica*, 1(38), 2008

ANEXOS

Anexo 1 - Artículo 91 del Código Orgánico Penal del Ecuador

Artículo 91.- Trata de Personas

La captación, transportación, traslado, entrega, acogida o recepción para sí o para un tercero, de una o más personas, ya sea dentro del país o desde o hacia otros países con fines de explotación, constituye delito de trata de personas.

Constituye explotación, toda actividad de la que resulte un provecho material o económico, una ventaja inmaterial o cualquier otro beneficio, para sí o para un tercero, mediante el sometimiento de una persona o la imposición de condiciones de vida o de trabajo, obtenidos de:

1. La extracción o comercialización ilegal de órganos, tejidos, fluidos o material genético de personas vivas, incluido el turismo para la donación o trasplante de órganos.
2. La explotación sexual de personas incluida la prostitución forzada, el turismo sexual y la pornografía infantil.
3. La explotación laboral, incluido el trabajo forzoso, la servidumbre por deudas y el trabajo infantil.
4. Promesa de matrimonio o unión de hecho servil, incluida la unión de hecho precoz, arreglada, como indemnización o transacción, temporal o para fines de procreación.
5. La adopción ilegal de niñas, niños y adolescentes.
6. La mendicidad.
7. Reclutamiento forzoso para conflictos armados o para el cometimiento de actos penados por la ley.
8. Cualquier otra modalidad de explotación.

Anexo 2 - Artículo 92 del Código Orgánico Penal del Ecuador

Artículo 92.- Sanción para el delito de trata de personas

La trata de personas será sancionada:







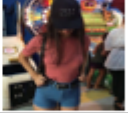



1. Con pena privativa de libertad de trece a dieciséis años.
2. Con pena privativa de libertad de dieciséis a diecinueve años, si la infracción recae en personas de uno de los grupos de atención prioritaria o en situación de doble vulnerabilidad o si entre la víctima y el agresor ha existido relación afectiva, consensual de pareja, conyugal, convivencia, de familia o de dependencia económica o exista vínculo de autoridad civil, militar, educativa, religiosa o laboral.
3. Con pena privativa de libertad de diecinueve a veintidós años, si con ocasión de la trata de personas, la víctima ha sufrido enfermedades o daños psicológicos o físicos graves o de carácter irreversible.
4. Con pena privativa de libertad de veintidós a veintiséis años, si por motivo de la trata de personas se produce la muerte de la víctima.

La trata se persigue y sanciona con independencia de otros delitos que se hayan cometido en su ejecución o como su consecuencia.

Anexo 3 - Muestra de imágenes clasificadas

Image Number	Gender	Group Age	Image
1	Man	Under 14 years old	
2	Woman	Over 14 years old	
3	Man	Over 14 years old	
4	Woman	Under 14 years old	
5	Woman	Over 14 years old	
6	Woman	Under 14 years old	
7	Woman	Under 14 years old	
8	Woman	Over 14 years old	
9	Man	Under 14 years old	
10	Man	Over 14 years old	

Anexo 4 - Muestra de imágenes clasificadas (continuación)

11	Man	Over 14 years old	
12	Woman	Under 14 years old	
13	Woman	Under 14 years old	
14	Woman	Under 14 years old	
15	Woman	Under 14 years old	
16	Woman	Under 14 years old	
17	Woman	Under 14 years old	
18	Man	Under 14 years old	
19	Man	Under 14 years old	
20	Woman	Under 14 years old	

Acerca del Autor



Sergio Leonardo Granizo nació en Quito, Ecuador, en 1988. Ingeniero informático graduado de la Universidad Central del Ecuador y Master en Software por la Escuela Politécnica Nacional. Trabajó como Desarrollador de software y Analista de software en Kruger Corporation y LLACSAA.SA, participación en varios proyectos de Machine Learning, Minería de datos, procesamiento de imágenes y programación de Python, además de consultoría e informática en varias empresas ecuatorianas. (<https://orcid.org/0000-0001-7220-7308>)