



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ingeniería de Sistemas e Informática

Unidad de Posgrado

**Método de búsqueda eficiente para resolver el
problema de identificación de huella dactilar aplicando
Machine Learning**

TESIS

Para optar el Grado Académico de Magíster en Ingeniería de
Sistemas e Informática con mención en Ingeniería de Software

AUTOR

Maria Elena RUIZ RIVERA

ASESOR

Dr. Edgar Cruz RUIZ LIZAMA

Lima, Perú

2024



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Ruiz, M. (2024). *Método de búsqueda eficiente para resolver el problema de identificación de huella dactilar aplicando Machine Learning*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Maria Elena Ruiz Rivera
Tipo de documento de identidad	DNI
Número de documento de identidad	08960021
URL de ORCID	https://orcid.org/0000-0003-3300-7068
Datos de asesor	
Nombres y apellidos	Edgar Cruz Ruiz Lizama
Tipo de documento de identidad	DNI
Número de documento de identidad	08004558
URL de ORCID	https://orcid.org/0000-0001-9403-1358
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Carlos Edmundo Navarro Depaz
Tipo de documento	DNI
Número de documento de identidad	08482690
Miembro del jurado 1	
Nombres y apellidos	Ciro Rodriguez Rodriguez
Tipo de documento	DNI
Número de documento de identidad	06020241
Miembro del jurado 2	
Nombres y apellidos	Marcos Hernán Rivas Peña
Tipo de documento	DNI
Número de documento de identidad	09241816
Datos de investigación	

Línea de investigación	C.0.3.22. Ingeniería de Software
Grupo de investigación	Ingeniería de Software Basado en Algoritmos para Buscar Información en Grandes Base de Datos - ISBABINF
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	<p>Universidad Nacional Mayor de San Marcos</p> <p>Edificio: Facultad de Ingeniería de Sistemas e Informática</p> <p>País: Perú</p> <p>Departamento: Lima</p> <p>Provincia: Lima</p> <p>Distrito: Lima</p> <p>Centro poblado: Ciudad</p> <p>Universitaria Urbanización: Pando</p> <p>Manzana y lote:</p> <p>Calle: German Amezaga 375</p> <p>Latitud: -12.05819215</p> <p>Longitud: -77.0189181894387</p>
Año o rango de años en que se realizó la investigación	Obligatorio. Ejemplo: Marzo 2019 - marzo 2024
URL de disciplinas OCDE	<p>Sistemas de automatización, Sistemas de control</p> <p>https://purl.org/pe-epo/ocde/ford#2.02.03</p> <p>Ingeniería, Tecnología</p> <p>https://purl.org/pe-epo/ocde/ford#2.00.00</p> <p>Ciencias de la computación</p> <p>https://purl.org/pe-epo/ocde/ford#1.02.01</p>



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO DE MAGÍSTER
EN INGENIERÍA DE SISTEMAS E INFORMÁTICA CON MENCIÓN EN
INGENIERÍA DE SOFTWARE**

A los cinco (05) días del mes de marzo de 2024, siendo las 11:00 am., se reunieron en el Auditorio, Profesor: Alfredo Celso Alva Bravo, el Jurado de Tesis conformado por los siguientes docentes:

Dr. Carlos Edmundo Navarro Depaz (Presidente)
Mg. Marcos Hernán Rivas Peña (Miembro)
Dr. Ciro Rodriguez Rodriguez (Miembro)
Dr. Edgar Cruz Ruiz Lizama (Miembro Asesor)

Se inició la Sustentación invitando a la candidata a Magíster **MARIA ELENA RUIZ RIVERA**, para que realice la exposición oral de la tesis para optar el Grado de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software, siendo la Tesis intitulada:

**“MÉTODO DE BÚSQUEDA EFICIENTE PARA RESOLVER EL PROBLEMA DE
IDENTIFICACIÓN DE HUELLA DACTILAR APLICANDO MACHINE LEARNING”**

Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por la graduanda; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

..... DIECISEIS (16) BUENO

Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga a la Bachiller **MARIA ELENA RUIZ RIVERA** el Grado de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software.

Siendo las 11:57 horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.

Dr. Carlos Edmundo Navarro Depaz
(Presidente)

Mg. Marcos Hernán Rivas Peña
(Miembro)

Dr. Ciro Rodriguez Rodriguez
(Miembro)

Dr. Edgar Cruz Ruiz Lizama
(Miembro Asesor)



CERTIFICADO DE SIMILITUD

Yo Edgar Cruz RUIZ LIZAMA en mi condición de asesor acreditado con Dictamen N°000238-2021-UPG-VDIP-FISI/UNMSM de la tesis/monografía/informe de investigación/trabajo académico, cuyo título es MÉTODO DE BÚSQUEDA EFICIENTE PARA RESOLVER EL PROBLEMA DE IDENTIFICACIÓN DE HUELLA DACTILAR APLICANDO MACHINE LEARNING, presentado por el bachiller/magíster/egresado/licenciado/estudiante MARIA ELENA RUIZ RIVERA para optar el grado/título/especialidad de Maestría en Ingeniería de Sistemas e Informática con Mención en Ingeniería de Software CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de 19% de similitud, nivel PERMITIDO para continuar con los trámites correspondientes y para su publicación en el repositorio institucional.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado/ título/ especialidad correspondiente.

Firma del Asesor _____

DNI: 08004558

Nombres y apellidos del asesor: Edgar Cruz RUIZ LIZAMA



**MÉTODO DE BÚSQUEDA EFICIENTE PARA RESOLVER EL
PROBLEMA DE IDENTIFICACIÓN DE HUELLA DACTILAR
APLICANDO MACHINE LEARNING**

Tesis presentada a la Universidad Nacional Mayor de San Marcos, (Lima - Perú), para obtener el Grado de Magíster en Ingeniería de Sistemas, con mención en Ingeniería de Software

Orientador: Dr. Ruiz Lizama, Edgar Cruz

Lima - Perú
Marzo - 2024

© María Elena Ruiz Rivera, 2024.

Todos los derechos reservados.

DEDICATORIA

Este trabajo se lo dedico a mi familia, a mi adorada madre, a mis 05 hermanos y en especial al Ingeniero Jesús Solano. A todos ellos que son el motor para concluir esta a investigación.

AGRADECIMIENTOS

A mi familia, que son el motor para concluir esta tesis, a ellos agradecerles también por la comprensión, al permitirme robarles no un poco sino mucho tiempo que debería compartir con ellos, y también por soportar mi temperamento en los momentos más estresantes frente a las dificultades que fueron apareciendo. A mis padres, en especial a mi adorada mamá Rosa, diciéndome ten cuidado, no llegues tarde, almuerza, llámame. Mis tres adoradas hermanas dándome tanto cariño como si fueran mi segunda mamá. Mis dos hermanos tan lindos protegiéndome siempre y ayudándome cuando más lo necesito.

Al Ing. Solano Cuyubamba, a quien considero mi maestro y guía, gracias por enseñarme a ser perseverante y nunca darme por derrotada ante los diferentes obstáculos que se presentan en la vida. A mi asesor el Dr. Edgar Ruiz Lizama por su orientación y dedicación y paciencia para que esta tesis de investigación cumpla con los objetivos trazados. Prácticamente sin su apoyo, hubiese sido imposible realizar este trabajo.

A Milagros y Liliana, gracias por liberarme de responsabilidad y recargarse ustedes del trabajo. A la Ing. Paola, por el apoyo en la parte administrativa de la UNMSM, y a todos los profesores de la facultad de Ingeniería de Sistemas e Informática y aquellas personas que directa o indirectamente me ayudaron para culminar este trabajo y que muchas veces constituyen un invaluable apoyo.

Y por encima de todo, dar gracias a Dios por darme fortaleza cada vez que aparecía el cansancio y por darme la sabiduría al iluminarme en los momentos que sentía no entender algunas cosas, y a mis dos angelitos Remigio mi sobrino y a Hugo Solís mi cuñado, por apoyarme desde el cielo y así concluir con mi investigación.

MÉTODO DE BÚSQUEDA EFICIENTE PARA RESOLVER EL PROBLEMA DE IDENTIFICACIÓN DE HUELLA DACTILAR APLICANDO MACHINE LEARNING

RESUMEN

El problema que tienen que ver con el reconocimiento de huellas es objeto de un extenso estudio en los últimos años dado su importancia para la identificación de personas. En situaciones de urgencia, es fundamental identificar de manera certera a la persona afectada, y esto se logra mediante la correcta identificación. En consecuencia, el propósito del presente trabajo es proponer una innovación en el proceso de búsqueda de una determinada huella dactilar. Esta innovación consiste en la creación de un método de búsqueda óptimo en una base de datos grande. Este método localiza una huella en menos tiempo puesto que categoriza las huellas en segmentos de acuerdo a las características que le son semejantes. Para lograr esto, se emplea Machine Learning, que clasifica las huellas dactilares. Posteriormente se aplica un algoritmo discreto para la búsqueda en secuencia en un determinado segmento, lo cual permite encontrar la huella en un tiempo significativamente menor, (Ruiz Rivera & Ruiz Lizama, 2021).

Palabras claves: Huella dactilar, sistema biométrico, Aprendizaje Automático, búsqueda de huella dactilar.

MÉTODO DE BÚSQUEDA EFICIENTE PARA RESOLVER EL PROBLEMA DE IDENTIFICACIÓN DE HUELLA DACTILAR APLICANDO MACHINE LEARNING

ABSTRACT

The problem related to fingerprint recognition is the subject of extensive study in recent years given its importance for the identification of people. In emergency situations, it is essential to accurately identify the affected person, and this is achieved through correct identification. Consequently, the purpose of this work is to propose an innovation in the search process for a specific fingerprint. This innovation consists of the creation of an optimal search method in a large database. This method locates a print in less time since it categorizes the prints into segments according to the characteristics that are similar to them. To achieve this, Machine Learning is used, which classifies fingerprints. Subsequently, a discrete algorithm is applied for the sequence search in a certain segment, which allows the fingerprint to be found in significantly less time, (Ruiz Rivera & Ruiz Lizama, 2021).

Key word: Fingerprint, biometric, Machine Learning, fingerprint search.

ÍNDICE

INDICE DE FIGURAS	10
INDICE DE TABLAS	11
CAPTITULO I	12
1. INTRODUCCION	12
1.1 Antecedentes del problema	12
1.2. Definición del Problema	13
1.3 Objetivos	14
1.3.1 Objetivo principal	14
1.3.2 Objetivo secundarios	14
1.4 Metodología	14
1.5 Justificación	15
1.6 Alcances	15
1.7 Hipótesis	15
1.8 Discusión	16
1.9 Propuesta a la solución	16
1.10 Organización de la tesis	17
CAPÍTULO II	19
2. Marco Teórico	19
2.1 Biométrica	19
2.2 Sistema Biométrica	20
2.2.1 Identificador Biométrico	21
2.2.2. Características de un indicador biométrico	21
2.3 Huella dactilar	22
2.3.1 Característica de la huella dactilar	23
2.3.2 Componentes básicos de los sistemas basados en huella dactilar	23
2.3.3 Verificación biométrica de huella dactilar	24
2.3.4 Identificación Biométrica de la huella dactilar	25
2.4 Procesos de sistema biométrico	26
2.5 Método de Clustering	27
2.6 Algoritmo k-MEANS	29
2.7. Machine Learning	29
2.8 Método de Búsqueda Secuencial	31

2.8.1 Interfaz gráfica de búsqueda secuencial.....	32
2.8.2 Funcionamiento interno de Sistema de búsqueda secuencial	33
2.8.3. Búsqueda en tablas sin índices	33
CAPITULO III	36
3. Estado del arte	36
3.1 Clasificación Huellas Dactilares Usando Curvas de Flujo de Campo de Orientación	36
3.2 Una técnica eficaz de clasificación y búsqueda de huellas dactilares	38
3.3 Un sistema de emparejamiento en tiempo real para grandes bases de datos de huellas dactilares	41
3.4 Categorización de las Imágenes de Huellas Dactilares Usando la técnica de Redes Neuronales	42
CAPITULO IV	44
4. Desarrollo de la propuesta.....	44
4.1 Metodo Propuesto.....	44
4.1.1 Fase de entrenamiento del modelo	46
4.2 Implementación en Phytton	47
4.3 Base de Datos	49
4.3.1 Configure su base de datos.....	49
4.3.2 Verificando	50
4.4 Modelamiento	50
4.4.1 Usuarios a quién va dirigido.....	50
4.4.2 Modelo de Datos.....	52
4.4.3 Caso de Uso.....	52
4.4.4 Diagrama de Secuencia.....	53
4.5 Funciones y módulos	54
CAPITULO V	57
5. Resultados de la Investigación.....	57
5.1 Resultados Obtenidos.....	57
5.2 Interfaz	60
CONCLUSIONES Y RECOMENDACIONES	62
REFERENCIAS	63

ÍNDICE DE FIGURAS

Figura 1. Estructura del Trabajo de Investigación (Elaboración propia)	<u>18</u>
Figura 2. Clasificación de las huellas digitales. Base de datos especial del Instituto Nacional de Estándares y Tecnología (NIST4) (Dass & Jain, 2004)	<u>23</u>
Figura 3. Proceso de registro de las imágenes de huella dactilar	<u>24</u>
Figura 4. Esquema del proceso de verificación de huella dactilar	<u>25</u>
Figura 5. Esquema del proceso de identificación de huella dactilar	<u>26</u>
Figura 6. La arquitectura de los procesos de un sistema biométrico (Elaboración propia)	<u>27</u>
Figura 7. Clustering	<u>28</u>
Figura 8: Algoritmo K-Means	<u>30</u>
Figura 9: Agrupación de datos bajo criterio del algoritmo K-Means	<u>30</u>
Figura 10. Modelo de Búsqueda de Huellas Dactilares [Elaboración propia]	<u>31</u>
Figura 11. Interfaz gráfica de usuario	<u>32</u>
Figura 12 Código para conectarse a la base de datos	<u>33</u>
Figura.13 Código para la tabla SQLite (Query planning)	<u>34</u>
Figura 14. Diseño lógico de la tabla “FP_BSTRING”	<u>34</u>
Figura 15. Full Table Scan	<u>35</u>
Figura 16. Clasificación de las Huellas Digitales	<u>37</u>
Figura 17. Query based search over clustered fingerprint	<u>40</u>
Figura 18. Búsqueda de la Huella Dactilar, según su Clasificación en Segmentos utilizando Machine Learning	<u>45</u>
Figura 19. Fase de Entrenamiento del Modelo, utilizando algoritmo K-Means	<u>46</u>
Figura 20. Representación matricial de una imagen de huella digital	<u>49</u>
Figura 21. Versión de SQLiteStudio	<u>50</u>
Figura 22. Modelo de Datos	<u>52</u>
Figura 23. Diagrama de caso de uso general	<u>52</u>
Figura 24. Captura de la Huella Dactilar	<u>53</u>
Figura 25. Mostrar Información	<u>54</u>
Figura 26. Gráfica de tiempos	<u>60</u>
Figura 27. Interfaz	<u>61</u>

ÍNDICE DE TABLAS

Tabla 1: Comparación de tecnologías biométricas. Los datos están basados en la percepción de los autores del libro “Handbook of Fingerprint” [2].....	22
Tabla 2: Diez muestras	58
Tabla 3: Tabla de tiempos	59

Capítulo 1

Introducción

1.1. Antecedentes

Debido al avance tecnológico en la actualidad se están implementando sistemas para verificación de huellas dactilares tanto para instituciones del Estado como en empresas privadas en el ámbito forense y policial. Un ejemplo de esto se encuentra en el sector bancario, donde se utiliza la verificación de huellas dactilares como medida de seguridad en las transacciones financieras, brindando protección a los clientes contra posibles suplantaciones de identidad o robos. Además, en situaciones como la entrada de un delincuente a una sala de operaciones para someterse a múltiples cirugías estéticas que alteran significativamente su apariencia física, la identidad de cualquier persona puede ser confirmada por su huella dactilar que es única, invariable, inmutable. Del mismo modo, cuando una persona ha sufrido desfiguración debido a un accidente, es posible llevar a cabo una identificación inequívoca utilizando su huella dactilar, (Ruiz Rivera & Ruiz Lizama, 2021).

Cabe destacar que un punto importante, es el tiempo de recepción al practicar una búsqueda de la huella dactilar en la BD (Base de Datos) extensa o muy extensa. La búsqueda se realiza mediante diferentes métodos, teniendo en cuenta el tiempo de recepción, como también el costo consecuente, (Ruiz Rivera & Ruiz Lizama, 2021).

En la búsqueda que utiliza un algoritmo secuencial insume tiempo; lo cual origina un alto costo. Se sabe que las huellas dactilares poseen ciertas características, las cuales pueden ser agrupadas permitiendo su categorización; el enfoque de esta tesis radica en la búsqueda de huellas de personas en BD grandes, con la finalidad de disminuir costos y tiempo de procesamientos, (Ruiz Rivera & Ruiz Lizama, 2021).

La clasificación involucra la división de la BD en varios grupos empleando técnicas de ML (Machine Learning). Estos grupos se definen de acuerdo con su similitud en cuanto a sus características. Su clasificación tiene un impacto notable en la eficiencia de la identificación de huellas dactilares, mejorando tanto la precisión como la velocidad, aspectos críticos en este proceso, (Ruiz Rivera & Ruiz Lizama, 2021).

En la primera parte se clasifica las huellas por la aproximación de sus características, mediante ML, lo cual posibilita la creación de grupos de acuerdo con las características parecidas, el siguiente paso consiste en utilizar el algoritmo de búsqueda secuencial tal que cumpla la función de encontrar la huella buscada en un determinado grupo o segmento. En este caso, el utilizar ML permite formar o agrupar las clases de los más próximos, según sus características; dado que dentro de una de estas clases o grupos se encuentra la huella buscada, que será detectada e identificada por el algoritmo discreto utilizado, (Ruiz Rivera & Ruiz Lizama, 2021).

1.2. Definición del Problema

La delincuencia a nivel mundial ha ido incrementándose año a año, creando como necesidad el uso de huellas biométricas, para solucionar problemas de identificación de delincuentes o para reconocer personas siniestradas.

Crear tecnología biométrica no es sencillo, pero es necesario para resolver problemas de autenticar o validar la huella de una determinada persona. Las huellas biométricas demandan gran cantidad de almacenamiento, generando grandes bases de datos.

En la actualidad la búsqueda de huellas dactilares en una DB muy extensa se realiza utilizando un algoritmo discreto, que consiste en verificar cada uno de los datos que conforman dicha data, y buscar una huella de uno en uno en miles de estos datos, generando un tiempo de respuesta inmensamente largo, que a su vez refleja un costo muy alto. (Ruiz Rivera & Ruiz Lizama, 2021)

1.3. Objetivos

1.3.1. Principal

La presente tesis de investigación tiene como objetivo general proponer un método de búsqueda eficiente aplicando Machine Learning, para resolver el problema de identificación de huella dactilar a un menor costo y tiempo. (Ruiz Rivera & Ruiz Lizama, 2021)

1.3.2. Secundarios

- Evaluar modelos, métodos, técnicas y algoritmos existentes que permitan búsquedas de huellas dactilares en bases de datos grandes.
- Aplicar Inteligencia Artificial, a través de la disciplina a Machine Learning que permita segmentar la base de datos grande con el objetivo de tener una respuesta a menor tiempo y costo. (Ruiz Rivera & Ruiz Lizama, 2021)

1.4. Metodología

Como primer paso se revisa la literatura relacionada al tema investigación, encontrando artículos indexados de los últimos años respecto al tema. Para comprender y definir el tema de la investigación se consideró los artículos más relevantes encontrados en la búsqueda, (Ruiz Rivera & Ruiz Lizama, 2021).

Se aplica la técnica de Machine Learning que permite construir los grupos (segmentos) de acuerdo con los criterios de clasificación de huella dactilar.

El siguiente paso es la segmentación de la BD, donde las huellas dactilares serán agrupadas en cuanto a sus características más próximas. Luego se realiza la búsqueda de la huella en un solo segmento; esta propuesta, consigue que el tiempo de respuesta sea menor en el proceso de encontrar una determinada huella en una BD grande, (Ruiz Rivera & Ruiz Lizama, 2021).

1.5. Justificación

La contribución de este trabajo de investigación radica en la aplicación de la tecnología biométrica para buscar huellas dactilares de personas, donde su rostro no es fácil de reconocer, o en casos de algún accidente en que el reconocimiento de la persona se tiene que dar por su huella dactilar y donde la aplicación de algoritmos discretos usados para resolver el problema de búsqueda exhaustiva en una base de datos, demandan un alto costo de procesamiento.

El almacenamiento de huellas dactilares en cantidades muy grandes en una BD puede ocasionar un tiempo de respuesta considerable, lo cual se refleja en un alto costo de procesamiento al realizar la búsqueda de una huella dactilar, (Ruiz Rivera & Ruiz Lizama, 2021).

1.6. Alcances

El alcance de la investigación se enfoca en realizar el análisis del proceso de búsqueda d una huella dactilar en una BD grande; para resolver el problema de reconocimiento o autenticación de una persona a través de su huella dactilar, (Ruiz Rivera & Ruiz Lizama, 2021).

1.7. Hipótesis

La aplicación de un método de búsqueda eficiente para resolver el problema de búsqueda de huella dactilar mediante Machine Learning y un algoritmo de búsqueda

secuencial, permitirá una mejorara en el tiempo de respuesta, a un menor costo. (Ruiz Rivera & Ruiz Lizama, 2021)

1.8. Discusión

El enfoque descrito por (Dass & Jain, 2004) consta de cuatro etapas principales. La primera etapa implica la extracción del campo de orientación de una imagen de huella dactilar. Luego, en la segunda etapa, se generan curvas de flujo de orientación de campo (OFFC, por sus siglas en inglés) a partir de esta información. En la tercera etapa, cada OFFC se etiqueta en una de las cuatro clases: bucles izquierdos, bucles derechos, espirales y arcos. Finalmente, en la cuarta etapa, se realiza la clasificación de la huella dactilar en una de cuatro clases (grupos) utilizando enfoques matemáticos, particularmente, mediante vectores de orientación en la imagen de la huella digital y métodos de geometría diferencial, (Ruiz Rivera & Ruiz Lizama, 2021).

Según lo reportado por (Dass & Jain, 2004), se logró una tasa de precisión del 94.4% en la clasificación de huellas digitales, (Ruiz Rivera & Ruiz Lizama, 2021).

El propósito de este estudio es presentar un método eficiente de búsqueda que utiliza ML para segmentar huellas dactilares en una BD grande, agrupándolas de acuerdo con sus características individuales. El siguiente paso es aplicar un algoritmo de búsqueda secuencial en uno de los segmentos seleccionados, que contiene la huella que se necesita identificar. El objetivo del planteamiento es minimizar el tiempo de búsqueda, con un nivel de confianza del 95%. Además, la investigación demuestra que la búsqueda segmentada tiene mayor eficacia que la búsqueda secuencial en una BD extensa, y supera en eficiencia al método presentado por (Dass & Jain, 2004).

1.9. Propuesta de la Solución

La propuesta consiste en aplicar un método de búsqueda de huellas dactilares en una BD grande utilizando ML.

El objetivo es utilizar Machine Learning para clasificar primero las huellas por su característica más próximas en segmentos o clústeres. Luego vincular la huella ingresada al segmento que lo podría contener por su aproximación. Por último, aplicar un algoritmo discreto secuencial a este segmento determinado para poder encontrar la huella dactilar con un tiempo de respuesta menor.

1.10. Organización de la Tesis

Este trabajo está organizado de la siguiente manera:

En el capítulo I, se explica los antecedentes, la metodología, el problema y la solución propuesta, así como la contribución y el impacto que tendrá el aporte de la tesis en la comunidad.

En el capítulo II, se presenta una descripción general de los sistemas biométricos y su arquitectura para analizar la huella dactilar y se explica la clasificación de huellas según sus características. También se habla de método Clustering y de Machine Learning; es decir, se explica la información necesaria para comprender el tema.

En el capítulo III, se detalla los diferentes trabajos relacionados al almacenamiento de las huellas dactilares, la clasificación de huellas según sus características, búsqueda de huellas en base de datos grandes, entre otros; que servirán como base fundamental para la elaboración del planteamiento de este método.

En el Capítulo IV, se presenta el método propuesto, que mejore los resultados presentados en los métodos actuales. Se explica detalladamente cada uno de los procesos, así mismo también se expone como se realizará el agrupamiento de las huellas de acuerdo a sus características más cercanas utilizando ML, (Ruiz Rivera & Ruiz Lizama, 2021).

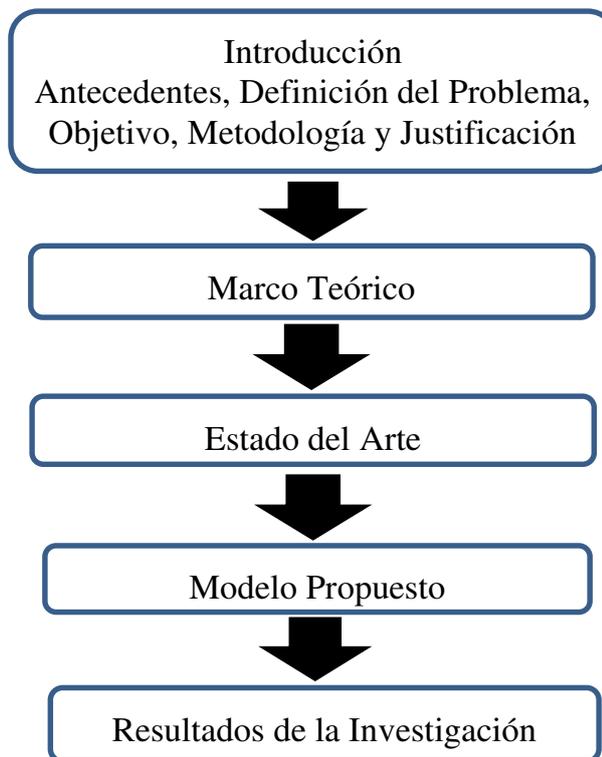
El Capítulo V, presenta dos aplicaciones de búsqueda de información de huellas, en una base de dato, la primera es una búsqueda secuencial en toda la base de datos, y la segunda es una búsqueda realizada en un determinado segmento que fue creada según su

característica más próximas de cada huella. Finalmente se explica el resultado de la prueba experimental. que comprobara el grado de eficiencia del método planteado.

En el Capítulo VI, se menciona las conclusiones del trabajo y las recomendaciones para trabajos de investigación futuros, relativos al tema presentado.

En la figura 1. se presenta la estructura completa del trabajo de investigación.

Figura 1. Estructura del Trabajo de Investigación



Fuente: (Elaboración propia)

Capítulo 2

Este capítulo explica en términos generales el sistema biométrico, su arquitectura. Describe la huella dactilar y sus características. También se habla de agrupamiento, de Algoritmo K-Means y de Machine Learning.

Marco Teórico

2.1. Biométrica

En la actualidad el concepto de biometría se utiliza para referirse al estudio de los métodos automáticos para el reconocimiento único de personas basado en uno o más rasgos conductuales o rasgos físicos intrínsecos. Por lo tanto, es bastante común escuchar el término “biometría informática”. De acuerdo con (Micolta López & Oltra Badenes, 2015) dicho término combina las diversas técnicas matemáticas, estadísticas y de inteligencia artificial para aplicar la autenticación e identificación de personas, principalmente en sistemas de seguridad informática.

Las técnicas biométricas se basan en la medida (directa o indirecta) y posterior análisis de uno o un conjunto de rasgos (estáticos y/o dinámicos) del individuo para

reconocerlo o verificar automáticamente su identidad. Según (Micolta López & Oltra Badenes, 2015), a este conjunto de rasgos también se le conoce como *indicador biométrico*.

Según (Sánchez Calle, 2008) la medición de los rasgos estáticos hace referencia a la anatomía del usuario como las huellas digitales, la imagen facial, la geometría de la mano, los patrones del iris, la retina, etc. Mientras que los rasgos dinámicos miden características del comportamiento dinámico del usuario como patrones de voz, escritura o firma manuscrita, la cadencia del paso, los gestos, etc.

Debido a lo anterior y de acuerdo con (Micolta López & Oltra Badenes, 2015) un sistema de gestión de la identidad biométrica debe ser capaz de:

1. Realizar un proceso de identificación de manera rápida, precisa, y con recursos no muy elevados.
2. No ser oneroso para sus usuarios y ser aceptado por la población a la que está dirigido.
3. Ser lo suficientemente robusto como para evitar cualquier tipo de fraude o falsificación.

2.2. Sistemas biométricos

Es un sistema automatizado que hace uso de la tecnología para realizar tareas de biometría. Es decir, un sistema que puede hacer labores de reconocimiento en función a las características físicas o de comportamiento de un ser humano, (Ruiz Rivera & Ruiz Lizama, 2021).

Para solucionar este tipo de problemas se desarrollan técnicas con base en pequeños datos biométricos que garantizan la identidad del individuo. Según (Micolta López & Oltra Badenes, 2015), estos datos biométricos están organizados en dos tipos: uno: Biometría fisiológica, tomando como base las partes de cuerpo humano, tal como lo especifica (Fernandez, 2008), las huellas son tomadas de los dedos de las manos; el iris, la retina, tomadas de los ojos; la voz, y el rostro. Dos: la Biometría conductual: basada

en las acciones de un individuo; como por ejemplo la firma de una persona cualquiera, (Ruiz Rivera & Ruiz Lizama, 2021).

2.2.1. Identificación Biométrica

De acuerdo con (Fernandez, 2008), el profesor Lakhmi Jain define la identificación biométrica como el procedimiento automático que vincula la identidad de una persona a través del uso de ciertas características físicas o del comportamiento particular de esa persona, (Ruiz Rivera & Ruiz Lizama, 2021).

2.2.2. Características de un indicador biométrico

De acuerdo con (Fernandez, 2008) un indicador biométrico es alguna característica con la cual se puede realizar biometría, siendo las siguientes, (Ruiz Rivera & Ruiz Lizama, 2021):

- Universalidad: el rasgo biométrico existe para todas las personas.
- Unicidad: el rasgo identifica unívocamente a cada persona.
- Permanencia: el rasgo se mantiene invariable con el tiempo a corto plazo.
- Inmutabilidad: el rasgo se mantiene invariable con el tiempo a largo plazo (durante toda la vida).
- Mensurabilidad: el rasgo es apto para ser caracterizado cuantitativamente.
- Rendimiento: el rasgo permite el reconocimiento del individuo con rapidez, robustez y precisión.
- Aceptabilidad: el rasgo debe ser aceptado por la mayoría de la población.
- Invulnerabilidad: el rasgo permite la robustez del sistema frente a los métodos de acceso fraudulentos.

La tabla 1 presenta las diversas tecnologías biométricas según los grados de confianza (Alto, Medio, Bajo) de las propiedades anteriormente descritas de acuerdo con (Maltoni, Maio, Jain, & Prabhakar, 2003).

Tabla 1: Comparación de tecnologías biométricas. Fundamentados en la apreciación de los autores del libro “Handbook of Fingerprint”

Identificador biométrico	Universalidad	Unicidad	Permanencia	Mensurabilidad	Rendimiento	Aceptabilidad	Invulnerabilidad
	Alto	Alto	Alto	Bajo	Alto	Bajo	Bajo
Oreja	Medio	Medio	Alto	Medio	Medio	Alto	Medio
Cara	Alto	Bajo	Medio	Alto	Bajo	Alto	Alto
Termograma facial	Alto	Alto	Bajo	Alto	Medio	Alto	Bajo
Huella dactilar	Medio	Alto	Alto	Medio	Alto	Medio	Medio
Modo de andar	Medio	Bajo	Bajo	Alto	Bajo	Alto	Medio
Geometría de la mano	Medio	Medio	Medio	Alto	Medio	Medio	Medio
Venas de la mano	Medio	Medio	Medio	Medio	Medio	Medio	Bajo
Iris	Alto	Alto	Alto	Medio	Alto	Bajo	Bajo
Pulsación de teclado	Bajo	Bajo	Bajo	Medio	Bajo	Medio	Medio
Olor	Alto	Alto	Alto	Bajo	Bajo	Medio	Bajo
Retina	Alto	Alto	Medio	Bajo	Alto	Bajo	Bajo
Firma	Bajo	Bajo	Bajo	Alto	Bajo	Alto	Alto
Voz	Medio	Bajo	Bajo	Medio	Bajo	Alto	Alto

Fuente: (Maltoni, Maio, Jain, & Prabhakar, 2003)

2.3. Huella dactilar

Representa la apariencia superficial tomada de la epidermis de un dedo, esto según la descripción de (Persto S.A. de C.V., 2020). Dicha marca dactilar está compuesta por crestas papilares y se forma durante la etapa fetal de un individuo. Se destaca que, a menos

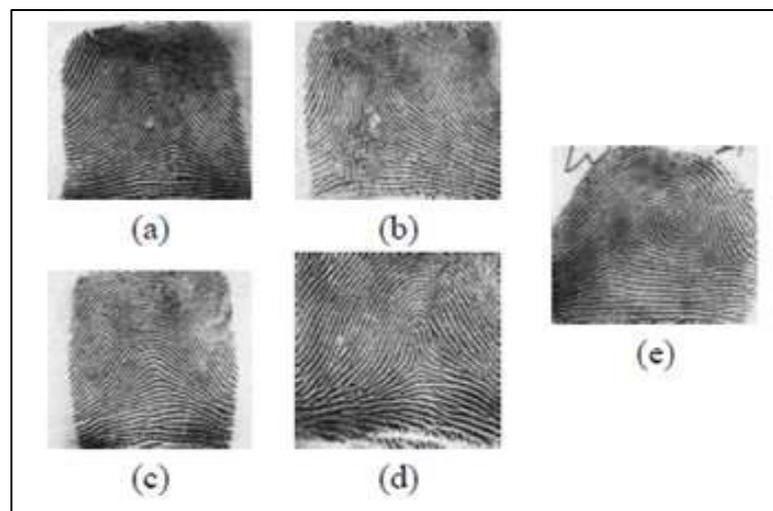
que experimente lesiones o daños graves, la huella dactilar permanece inalterable a lo largo de toda la vida (Villamizar R., 1994).

2.3.1. Características de la huella dactilar

El primer sistema de calificación fue propuesto por Francis Galton en 1892 estableciendo que la huella dactilar es individual y permanente, señalando que posee “puntos finos”, (Persto S.A. de C.V., 2020).

(Dass & Jain, 2004) tomaron como referencia las cinco categorías principales de clasificación de huellas digitales propuestas por Henry. Estas categorías incluyen (a) arco izquierdo, (b) arco derecho, (c) arco, (d) arco de carpa y (e) circular o espiral. Esta clasificación se basó en el estándar NIST4, tal como se presenta en la figura 2.

Figura 2. Agrupación de huellas dactilares en BD del Instituto Nacional de Estándares y Tecnología (NIST4)



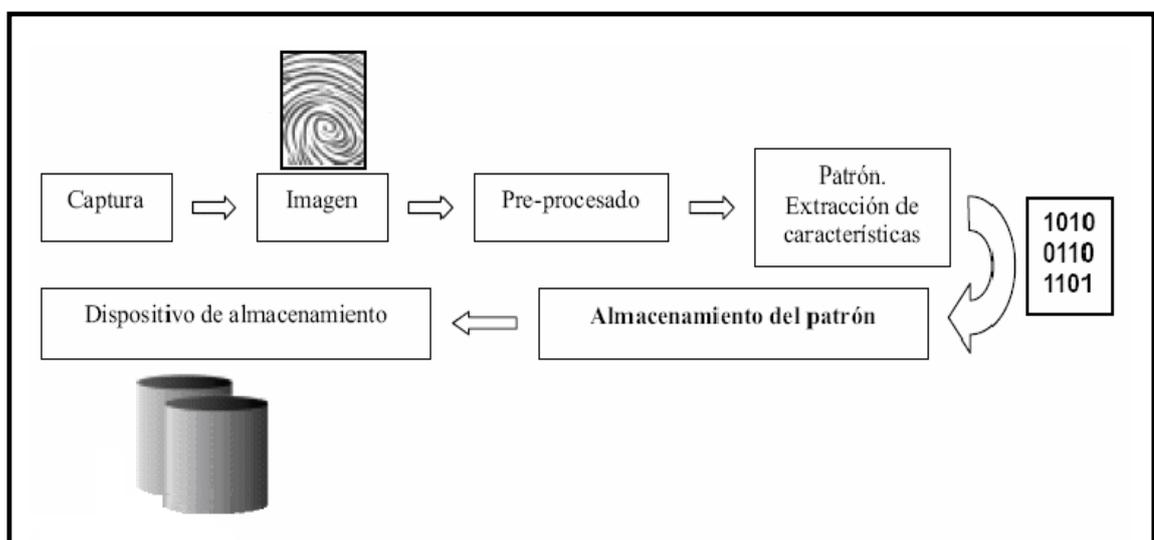
Fuente: (Dass & Jain, 2004)

2.3.2. Componentes básicos de los sistemas biométricos basados en huella dactilar

Los sistemas biométricos basados en huella dactilar proponen inicialmente un mecanismo denominado inscripción o reclutamiento para realizar la carga de las huellas

dactilares en la base de datos, para esto se recoge una imagen o un registro; a partir de aquí el proceso desarrollado para la extracción de características y codificación de los datos depende de los algoritmos utilizados por los fabricantes, estos fabricantes llaman a estas como plantillas y pueden ser almacenadas para su posterior utilización en el sistema. Este proceso se puede observar en la figura 3, en donde se especifica como entrada una imagen de huella dactilar y luego continua el proceso anteriormente mencionado (Coomonte Belmonte, 2006).

Figura 3. *Proceso de registro de las imágenes de huella dactilar*



Fuente: (Coomonte Belmonte, 2006)

Se debe tener en cuenta que los sistemas biométricos en general funcionan de dos maneras, en modo de verificación y en modo de identificación.

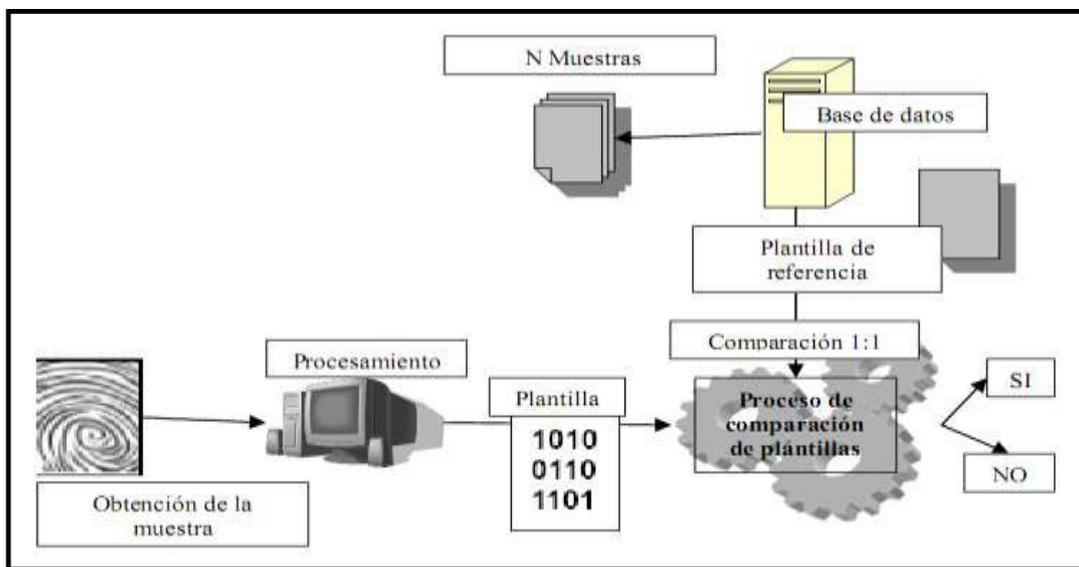
2.3.3 Verificación biométrica de huella dactilar

El funcionamiento del sistema biométrico basado en huella dactilar en modo de verificación o autenticación tiene como finalidad corroborar si la identidad de la persona corresponde con la imagen dactilar que esta presenta. Por ejemplo, si un usuario le dice al sistema que es el trabajador número 1234, el sistema buscará la plantilla (huella dactilar) relacionada con ese usuario comparándola con la que obtiene en el momento del

acceso, realizar la comparación y luego emitir una respuesta de aceptación o rechazo. Es por esto por lo que este modo de funcionamiento también es denominado comparación 1 a 1, ya que el sistema obtiene la huella de la persona y busca una huella en la base de datos para realizar la comparación (Coomonte Belmonte, 2006).

Los componentes básicos que comprenden estos sistemas, se muestran en la figura 4, (Coomonte Belmonte, 2006). Además, se debe precisar que este tipo de funcionamiento corresponde a los sistemas de verificación de huellas dactilares.

Figura 4. Esquema del proceso de verificación de huella dactilar

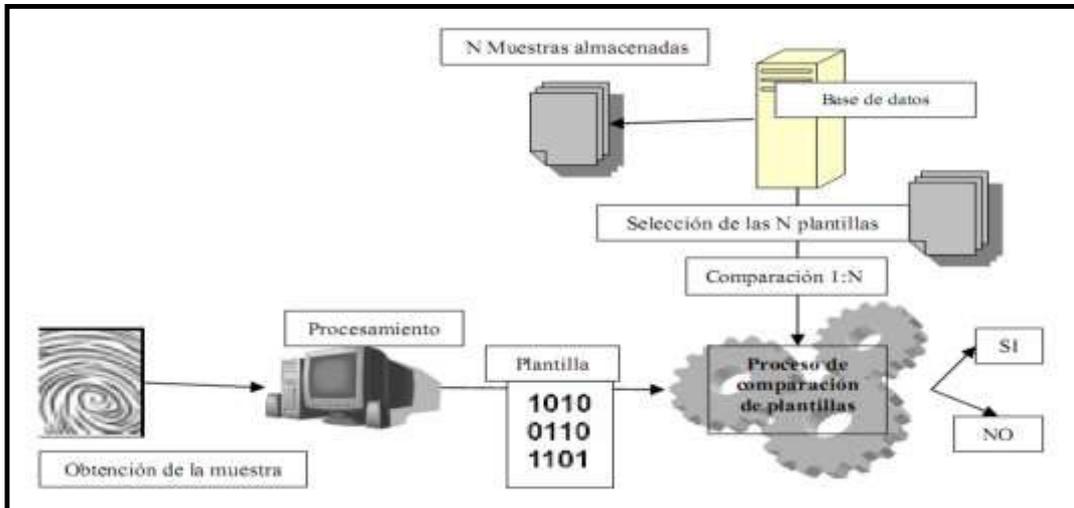


Fuente: (Coomonte Belmonte, 2006)

2.3.4 Identificación biométrica de la huella dactilar

El funcionamiento de un sistema biométrico basado en huella dactilar en modo de identificación tiene como finalidad averiguar la identidad de una persona a través de sus huellas dactilares. Por ejemplo, estos sistemas son muy utilizados por la policía para poder obtener la identidad de un ciudadano. Este modo de funcionamiento también se le denomina comparación 1 a N, ya que el sistema obtiene la huella de la persona y compara la huella contra todos los registros. Los componentes básicos que comprenden estos sistemas se pueden apreciar en la figura 5, (Coomonte Belmonte, 2006).

Figura 5: Esquema del proceso de identificación de huella dactilar



Fuente: (Coomonte Belmonte, 2006)

El modo de funcionamiento de identificación del sistema presenta dos variaciones para este caso:

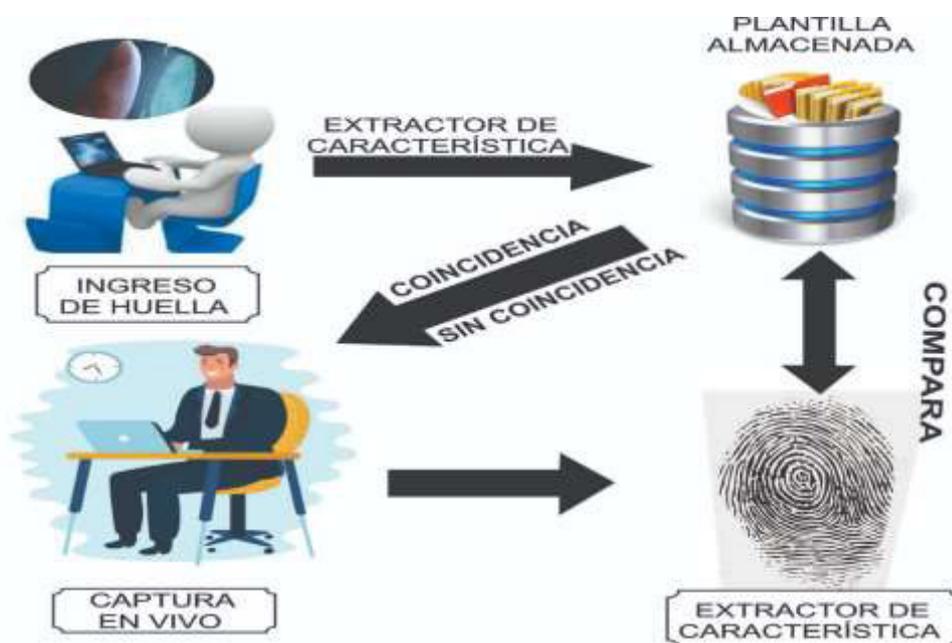
- **Sistemas de identificación positiva de huella dactilar:** Esta variante tiene como objetivo el brindar acceso a instalaciones o recursos para las personas, buscando si la persona se encuentra registrada.
- **Sistemas de identificación negativa de huella dactilar:** El sistema de identificación negativa, se usa para asegurar que una persona que accede al sistema ya no esté presente en la base de datos, en el caso de que alguien trate de acceder al sistema de manera múltiple. Así el sistema detecta si la persona que intenta hacer uso del sistema está o no contenida en la base de datos de usuarios.

2.4. Procesos del sistema biométrico

El procedimiento inicia con la lectura de una huella dactilar en un dispositivo de reconocimiento de huellas, mediante un escáner óptico de huellas. Este dispositivo transforma esta información al formato digital. A continuación, las huellas digitalizadas se almacena en la BD.

Luego, se procede a capturar la huella en vivo que se busca identificar. Se extraen sus características y se lleva a cabo una búsqueda en la base de datos donde previamente se han almacenado otras huellas dactilares. Si se encuentran coincidencias entre la imagen de la huella en vivo y las huellas almacenadas, se puede verificar la identidad asociada a esa imagen, lo que resulta en una respuesta positiva para la búsqueda. En caso de que no se encuentren coincidencias entre la imagen digitalizada de la huella y las de la base de datos, se genera un resultado negativo en la búsqueda, como se muestra en la figura 6.

Figura 6: La arquitectura de los procesos de un sistema biométrico



Fuente: Elaboración propia

2.5. Método de Clustering

El clustering encuentra diversas aplicaciones en el campo de la informática. Esto incluye la compresión de imágenes, como se mencionó en el trabajo de (Scheunders, 1997), y la digitalización de la voz, como se investigó en el estudio de (Makhoul, Roucos, & Gish, 1985). Además, se utiliza en la recuperación de información relacionada, como señalan (Bathia & Deogun, 1998). También se aplica en la minería de datos para buscar

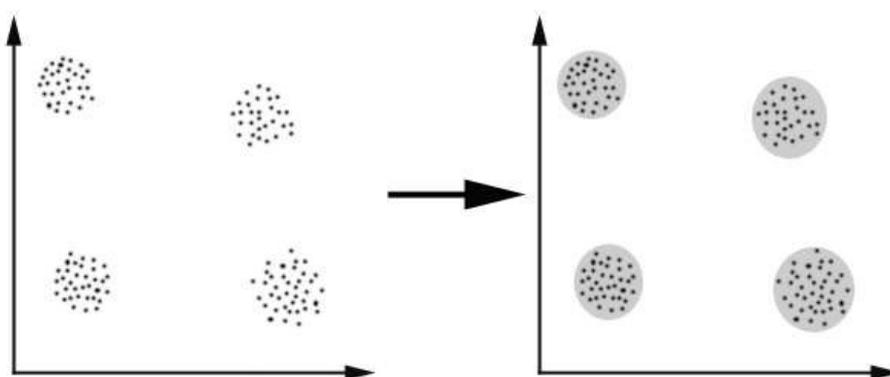
grupos con características de interés específicas, como el descubrimiento de nuevos segmentos de clientes para mejorar los servicios, tal como se planteó en la investigación de (Fayyad, Piatetsky-Shapiro, & Padhraic, 1996).

Los métodos de clustering varían en la forma en que agrupan los elementos. Aquellos que crean clusters basados en una partición del conjunto de objetos se denominan métodos de Hard-clustering (Kearns, Mansour, & Ng, 1997), y uno de los más conocidos entre ellos es el algoritmo K-Means, que fue introducido por (Forgy, 1965) y desarrollado posteriormente por (MacQueen, 1967).

(Maggi, Matteucci, & Zanero, 2008), definen al Clustering como el proceso de organizar objetos en grupos y cuyos elementos sean similares de alguna manera. Un cluster es una colección de objetos “similares” entre ellos y que no lo son con los objetos de otros clústeres

Un criterio de similitud bastante usado es la distancia: dos o más objetos pertenecen al mismo cluster si están “cerca” de acuerdo con una distancia dada, a esto se le llama distance – based clustering. Otra clase de clustering es el conceptual clustering, en donde dos o más objetos pertenecen al mismo clúster si uno de estos objetos define un concepto común para todos los demás objetos, ver figura 7. (Bonarini, Bonacina, & Matteucci, 2001).

Figura 7: *Clustering*



Fuente: (Bonarini, Bonacina, & Matteucci, 2001)

2.6. Algoritmo K-Means

El algoritmo K-Means, que fue presentado por (Forgy, 1965) y posteriormente desarrollado por (MacQueen, 1967), se utiliza ampliamente en aplicaciones de clustering. La idea fundamental detrás de este algoritmo es comenzar con K centros iniciales y formar grupos asignando objetos de X a los centros más cercanos, luego, se recalculan los centros. Si los nuevos centros no difieren de los centros anteriores, el algoritmo concluye. En caso contrario, el proceso de asignación con nuevos centros se repite hasta que no haya cambios en los centros o se cumpla algún nuevo criterio de detención que implique pocas reasignaciones de objetos.

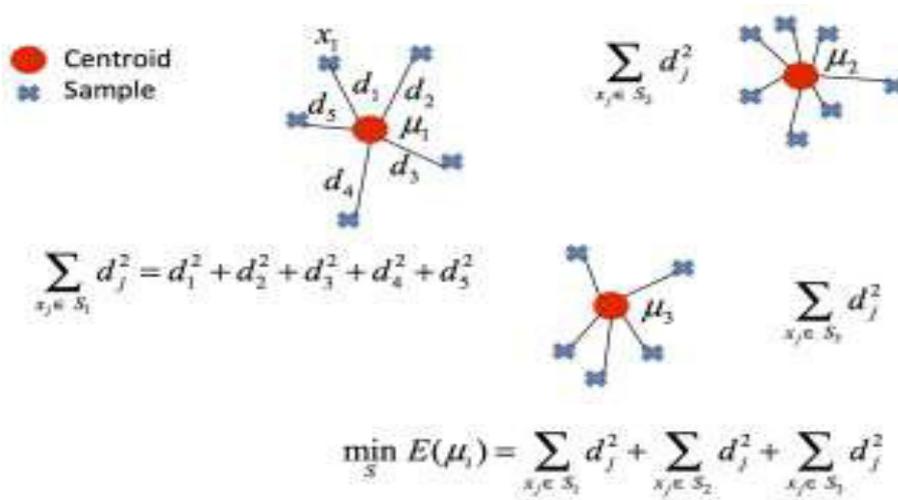
2.7. Machine Learning

ML posee algoritmos que proporciona a las maquinas la capacidad de identificar patrones.

En el presente estudio se utiliza el algoritmo K-Means o K Medias, que pertenece a la categoría de algoritmos utilizados para la segmentación y agrupación de datos. Este método fue propuesto inicialmente en 1957 por el matemático Stuart P. Lloyd, aunque su publicación oficial se realizó en 1982 en un artículo titulado "Least Squares Quantization in PCM". A lo largo de las décadas siguientes, el algoritmo K-Means ha experimentado múltiples optimizaciones hasta llegar a las implementaciones más recientes, (Ruiz Rivera & Ruiz Lizama, 2021).

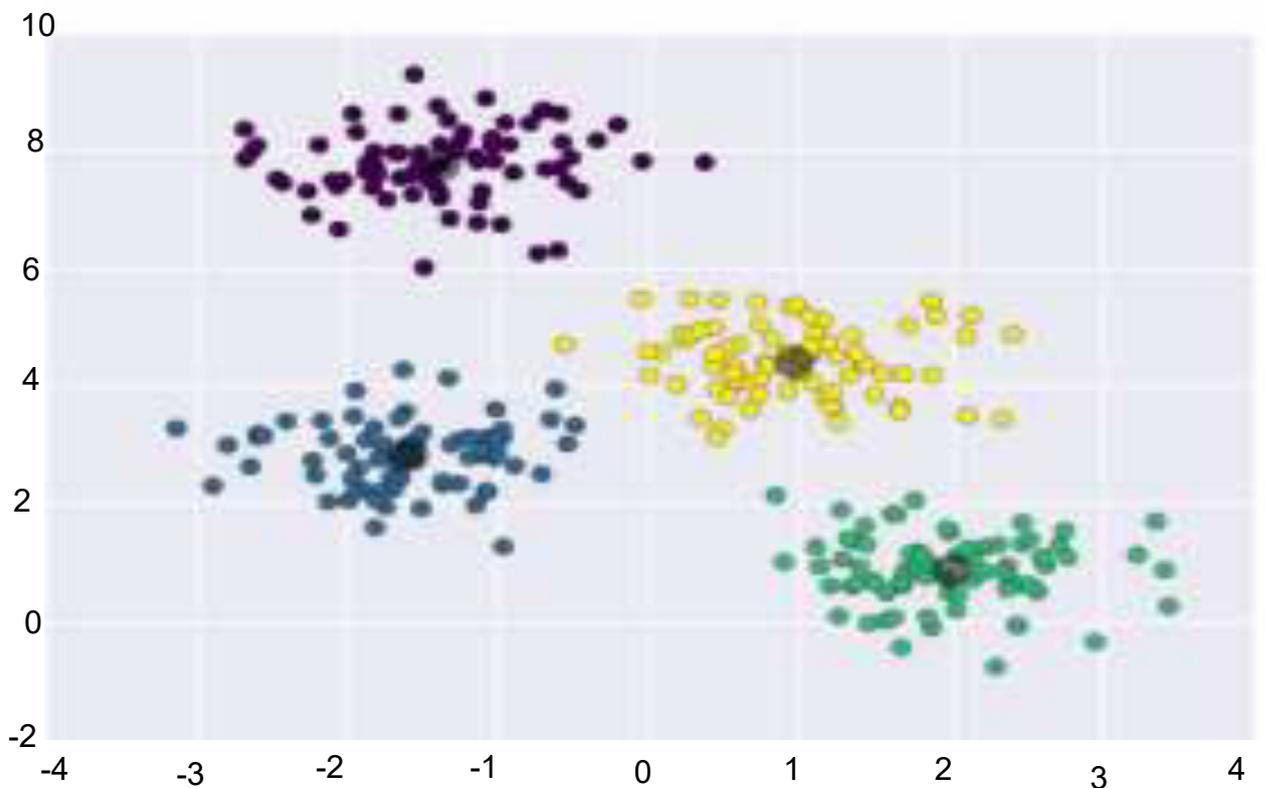
La esencia del algoritmo K-Means consiste en: tomar un conjunto de puntos con el propósito de obtener su segmentación. Estos puntos son seleccionados de manera aleatoria, tomando k centroides; siendo k un valor arbitrario que elige el usuario. A continuación, los clusters se forman con el criterio de la mínima distancia entre cada punto y su respectivo centroides, tal como puede verse en la figura 8.

figura 8: Algoritmo K-Means



Fuente: Tomado de: https://www.unioviedo.es/compmum/laboratorios_py/new/kmeans.html
 (Fecha de visita 15 de julio de 2021)

Figura 9: Agrupación de datos bajo criterio del algoritmo K-Means



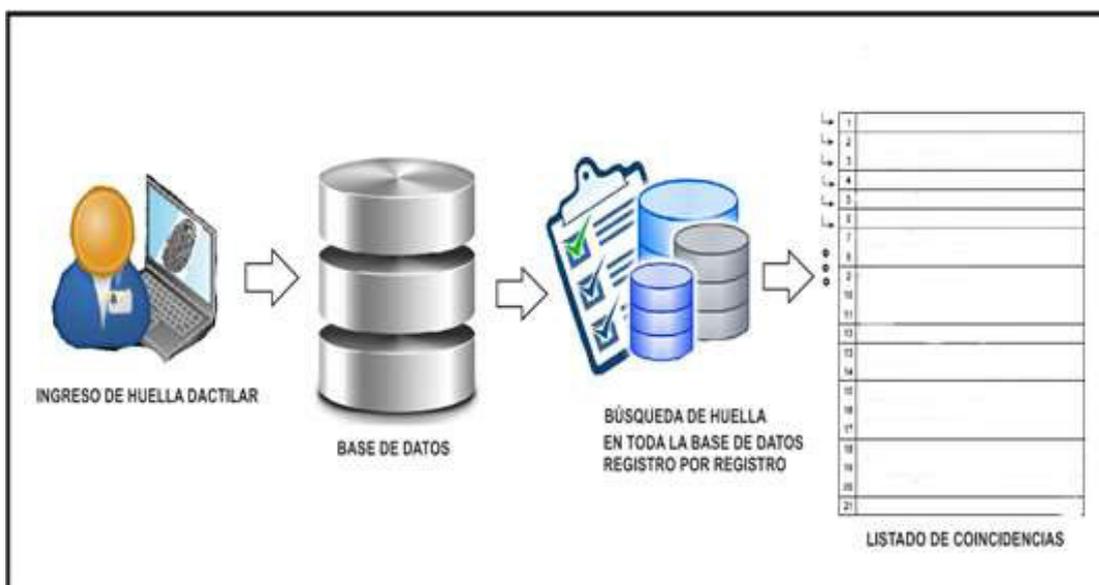
Fuente: Tomado de <https://jakevdp.github.io/PythonDataScienceHandbook/05.11-k-means.html> (Fecha de visita 20 de julio de 2021)

Obtenidos los k-clusters se recalcula los centroides tomando la media de los conjuntos de puntos dentro de cada cluster, (ver figura 9). El algoritmo es iterativo por ello se aplica a todos los centroides que convergen. Se entiende por convergencia que los clústeres formados se mantienen constantes en iteraciones posteriores.

2.8. Método de Búsqueda Secuencial

La Identificación de huellas dactilares automática es una de las tecnologías biométricas más importantes. Comparar huellas de manera secuencial en una gran base de datos, con un sistema de indexación consume mucho tiempo y trabajo para encontrar una coincidencia como respuesta. Como se observa en la figura 10, ingresa la huella a identificar a través de un dispositivo, luego ingresa a la base de datos donde se realiza la búsqueda secuencial huella por huella, dado que la búsqueda insume tiempo, este proceso se puede relacionar en términos de valor tiempo-costo.

Figura 10. *Método de Búsqueda de Huellas Dactilares*

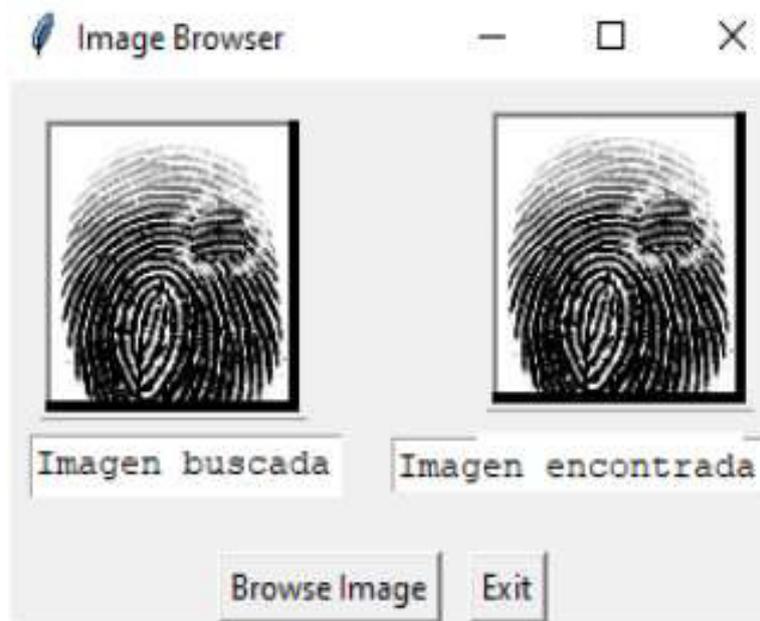


Fuente: Elaboración propia

2.8.1. Interfaz gráfica de búsqueda secuencial

Cuando se pone en ejecución el Sistema de búsqueda secuencial de huella digital, se muestra al usuario una ventana mediante la cual se puede realizar la búsqueda. Se hace click en el botón **Image Browser** (examinar imagen, es decir cargar la imagen a buscar) que se encuentra en la parte inferior de dicha ventana. A continuación, aparece una ventana emergente que permitirá buscar y seleccionar la imagen de la huella que se busca en la base de datos, ver figura 11.

Figura 11: *Interfaz gráfica de usuario*



Fuente: Elaboración propia

Si la imagen que a buscar se encuentra en la base de datos, entonces se mostrará tanto la imagen buscada como la imagen encontrada; lo cual corresponde a lo solicitado.

En la siguiente sección se presenta parte del código en lenguaje Python que es utilizado para conectar con la base de datos y las funciones de búsqueda. También se explica acerca de la búsqueda secuencial que realiza en SQLite cuando mediante un *query* aplicado a una determinada imagen de huella digital.

2.8.2. Funcionamiento interno de Sistema de Búsqueda Secuencial

Para conectarse a la base de datos, se utiliza el siguiente código (Ver figura 12):

Figura 12: Código para conectarse a la base de datos

```
import sqlite3

def readBlobData(matriz_de_huella):
    try:
        sqliteConnection=
sqlite3.connect('C:/Users/User/Desktop/FingerPrintDatabase.db)
        cursor = sqliteConnections.cursor()
        print("Conectado a SQLite")
        sql_fetch_blobl_query = """SELECT * FROM FP_BSTRING WHERE
BLOBSTRING = ?"""
        cursor.execute(sql_fetch_blob_query, (matriz_de_huella,))

        if len(record) <= 0:
            cursor.close()
    except sqlite3.Error as error:
        print("No se ha logrado leer la data desde la tabla de SQLite",
error)
    finally:
        if sqliteConnection:
            sqliteConnection.close()
```

Fuente: Elaboración propia

De acuerdo con la figura 12, en primer lugar, se importa el módulo `sqlite3` que provee todos los mecanismos necesarios para conectarse a la base de datos. Luego se define la función cuyo nombre es `readBlobData` siendo el parámetro una matriz producto del cambio de representación de una imagen de entrada; es decir de una imagen a una matriz de ceros y unos. Dentro de la función se crea un objeto con la ruta de la base de datos y con este un objeto de la clase `cursor`, el cual es una abstracción pensada para el recorrido sobre el conjunto de datos. Luego, se tiene el `query` almacenado en una variable de tipo `string` y se ejecuta ese `query` mediante la función `execute`. Esta función recibe un `string` con la consulta y una tupla que contiene la matriz de la imagen. Adicionalmente, se comprueba el resultado de la búsqueda y se informa del éxito o fracaso de la búsqueda, finalmente se cierra la conexión.

2.8.3. Búsqueda en tablas sin índices

La mayoría de las tablas en SQLite consisten en cero o más filas con una clave entera única (el id de fila) seguida por el contenido o dato. Las filas son lógicamente almacenadas en orden ascendente según el id de fila mencionado. Como ejemplo, en la

Capítulo 3

Estado del arte

Este capítulo tiene el propósito de describir la revisión de la literatura de los temas relacionados a la investigación, Dicha información servirá como base fundamental para proponer el modelo de búsqueda en una BD extensa de huellas dactilares.

3.1. Clasificación de huellas dactilares usando curvas de flujo de campo de orientación

(Dass & Jain, 2004), llevaron a cabo una revisión de diversas corrientes de enfoque relacionadas con clasificación de huellas dactilares. Adicionalmente, notaron que los métodos híbridos desarrollados anteriormente no habían sido sometidos a pruebas en BD grandes. En contraste con otros investigadores que habían empleado cuatro categorías sin obtener resultados efectivos en la clasificación, los autores optaron por proponer el uso de cinco categorías con la expectativa de lograr mejores resultados.

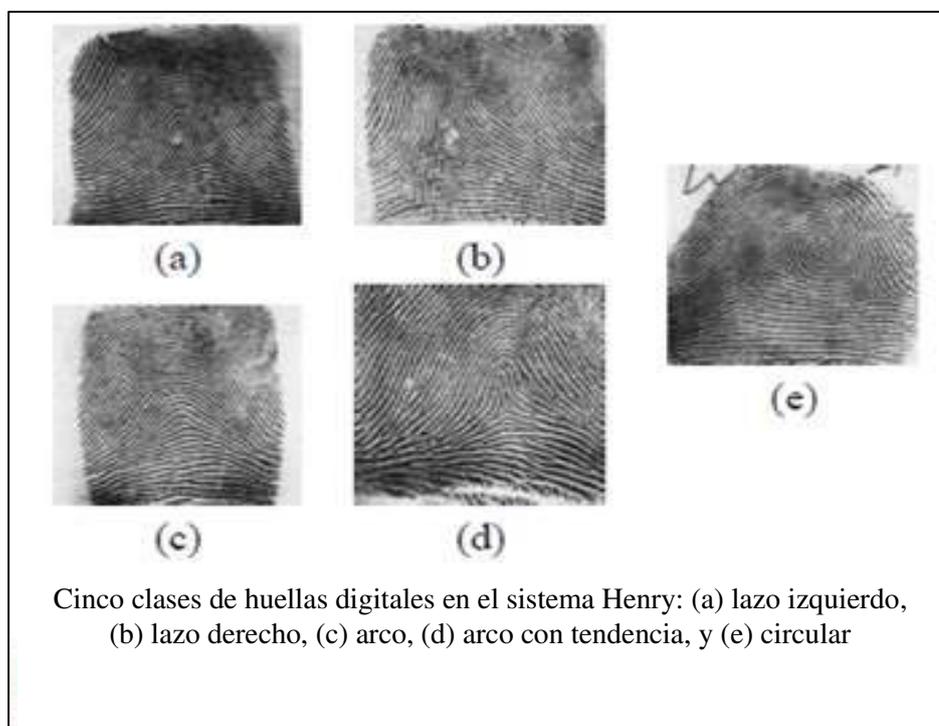
El enfoque presentado por (Dass & Jain, 2004) consta de cuatro etapas fundamentales. En primer lugar, se realiza la extracción del campo de orientación a partir de la imagen de la huella dactilar proporcionada. En segundo lugar, se generan las curvas de flujo de orientación de campo (OFFCs). En la tercera etapa, se lleva a cabo la asignación de cada OFFC a una de las cuatro categorías, que incluyen bucles izquierdos, bucles derechos, espirales y arcos. Esto implica realizar la clasificación de la imagen de

la huella en una de cuatro categorías, basado en principios matemáticos, tomando el valor de orientación de la imagen de la huella como un vector.

De acuerdo con lo señalado por (Dass & Jain, 2004), el método empleado les permitió lograr una tasa de precisión del 94.4% en la clasificación de huellas digitales. Además, plantean la posibilidad de incluir en futuros estudios la detección de las áreas de origen de los bucles más pequeños. Además, planean seguir utilizando la clasificación propuesta por Henry, que incluye categorías como arco, arco con tendencia, lazo derecho, lazo izquierdo y circular, como se detalla en el trabajo de (Henry E. R., 1900), como base para su investigación, como se muestra en la figura 16.

También refiere que su enfoque toma en cuenta la BD NIST4 (Instituto Nacional de Estándares y Tecnología) y combina elementos de tipo estructural, sintáctico y matemático, tal como se describe en el trabajo de (Watson & Wilson, 1992).

Figura 16. *Clasificación de las huellas digitales*



Fuente: (Dass & Jain, 2004)

3.2. Una técnica eficiente de clasificación y búsqueda de huellas dactilares

Los autores M H Bhuyan and D K Bhattacharyya realizan una revisión a la literatura de huellas dactilares, centrándose en la precisión y la velocidad, sabiendo que son puntos críticos para la búsqueda (Bhuvan & Bhattacharyya, 2009).

El enfoque de Ballan and Ayhan Sakarya's technique, basado en técnicas para clasificar huellas dactilares, aplica la técnica de extraer puntos singulares (deltas y núcleos) de las huellas dactilares basado en histograma direccionales. Se encuentra la dirección de la imagen direccional chequeando la orientación de pixeles individuales, calculando el histograma direccional usando un solapamiento de bloques en la imagen direccional, y clasifica las huellas dactilares en clases Wirbel (circular espiral y dos lasos iguales) o las clases lasso (arco, arco tendido, lazo derecho o lazo izquierdo) la complejidad de la técnica es el orden del número de pixeles de la imagen de la huella. Sin embargo, toma mucho tiempo para clasificar la imagen (Bhuvan & Bhattacharyya, 2009).

En Jain, Prabhakar and Hong's Multi-channellapproach, el método puede llegar a ser más exacto mientras clasifica las imágenes de huellas dactilares, las compara con una contra parte. En este caso las huellas dactilares se clasifican en cinco categorías: Verticilo, lasso derecho, laso izquierdo, arco y arco tendido. El algoritmo usa una representación novel (huella código) y se basa en un clasificador de dos etapas para hacer la clasificación eficaz. El clasificador de dos etapas utiliza un k- vecino más cercano en su primera etapa y un conjunto de clasificadores de redes neuronales en su segunda etapa para clasificar un vector de característica en una de las cinco clases de huellas digitales. Este algoritmo adolece del requisito de que la región de interés sea la correcta, lo que requiere la detección precisa del punto central de la imagen de la huella. En caso contrario, el algoritmo se puede considerar muy eficaz (Bhuvan & Bhattacharyya, 2009).

En Cho, Kim, Bae et al's approach plantea un algoritmo eficaz de clasificación de huellas dactilares usando la información relacionada con los puntos principales. El algoritmo detecta puntos principales de las imágenes candidatas y analiza la zona cercana al área de cada uno de los núcleos candidatos, en este análisis de núcleos si se encuentra un punto central falso debido al ruido se elimina. Usando esta información, se puede realizar la clasificación. Sin embargo, puede encontrarse difícil eliminar los puntos

singulares falsos, los cuales han sido usados para la clasificación. Esto demanda una mayor sofisticación del método para eliminar aquellos puntos núcleos falsos (Bhuvan & Bhattacharyya, 2009).

Yao, Marcialis, Pontil et al's reportan un nuevo algoritmo de clasificación de huellas dactilares que consta de dos enfoques de aprendizaje automático: vector de soporte de máquinas (SVMs) y redes neuronales recurrentes (RNNs). RNNs son entrenadas en una representación estructurada de la imagen de huellas dactilar. También se utilizan para extraer un conjunto de características distribuidas las cuales se pueden integrar en la SVMs y se combinan con un código de corrección de error de la imagen que, a diferencia de los sistemas anteriores, también puede explotar la información contenida en imágenes de huellas dactilares ambiguas (Bhuvan & Bhattacharyya, 2009).

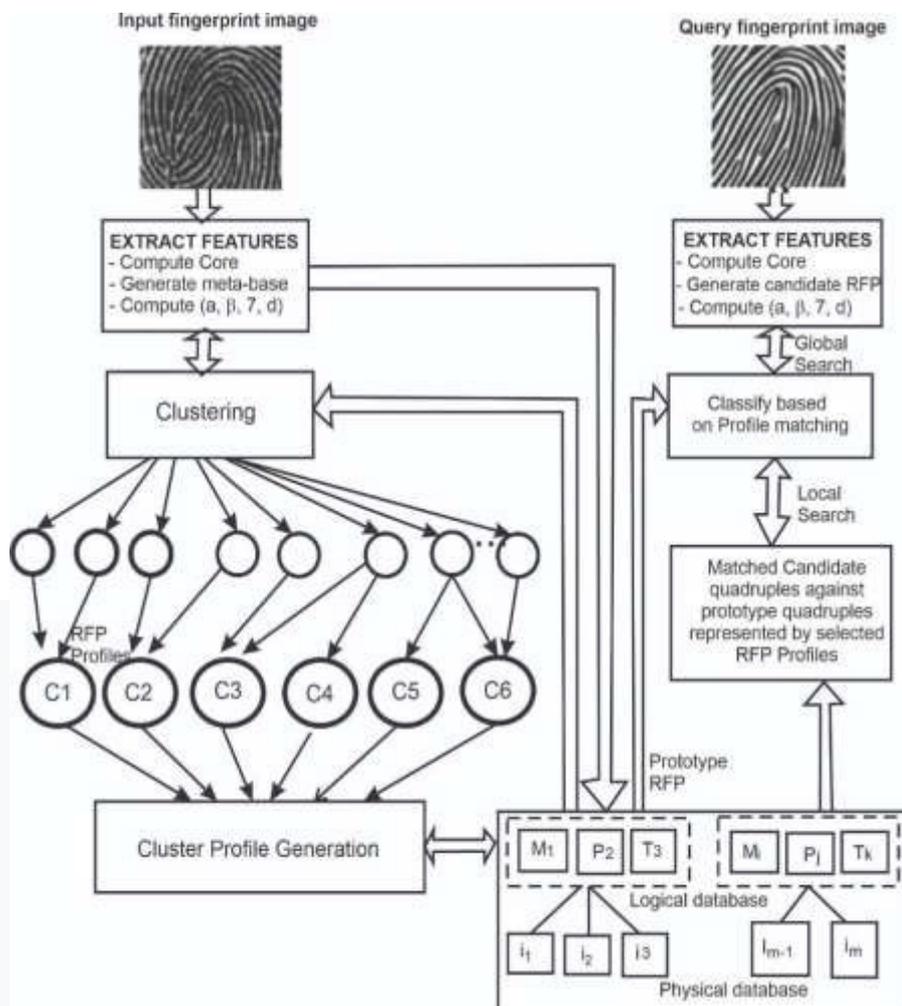
Shah and Sastry's approach clasifican imagenes de huellas dactilares en una de las cinco clases: arco, arco cubier5to, lasso izquierdo lasso derecho y espiral, este algoritmo útil en función de un vector de poco dimensione, obteniendo a partir de la salida de un detector de realimentación basado en la línea. Su línea detectora, es un sistema dinámico de cooperación que da líneas de orientación y conserva múltiples orientaciones en los puntos donde las líneas con orientación diferente se encuentran. Aquí se utilizan tres tipos de clasificadores: máquinas de clasificador de soporte, maquinas clasificador del vecino más cercano y clasificador de redes neuronales. El detector funciona únicamente en imagines binarias (Bhuvan & Bhattacharyya, 2009).

De acuerdo con (Bhuvan & Bhattacharyya, 2009), para la clasificación de imágenes de huellas se debe tomar en cuenta las características; estas dependen de la calidad de la imagen, la representación de datos de dichas imágenes, el modelo de procesamiento de estas y la evaluación hecha a partir de la extracción de las características.

La evaluación en tiempo real de una imagen está en función a la calidad de la imagen, por otro lado, imágenes de baja calidad requieren un mayor tratamiento. En ese sentido El tratamiento de una imagen (Bhuvan & Bhattacharyya, 2009), destacan que es importante tener en cuenta a calidad de las imágenes.

El objetivo de utilizar agrupamiento jerárquico facilita la búsqueda en base de datos grandes. De acuerdo con (Bhuvan & Bhattacharyya, 2009) el agrupamiento jerárquico que proponen, toma en cuenta el código de flujo de orientación de cresta. AL ingresar la imagen en consulta, se extraen sus características, luego se inicia el proceso de búsqueda. Este proceso funciona en dos fases: La primera se encarga de identificar a que clase pertenece la huella a consultar; la segunda se encarga de realizar una búsqueda local de la huella a consultar, en la clase identificada o seleccionada. Como muestra en la Figura 17.

Figura 17: *Query based search over clustered fingerprint*



Fuente: (Bhuvan & Bhattacharyya, 2009)

Clasificar las huellas dactilares de un individuo en cuatro o cinco tipos, admite un alto grado de precisión del 80 y 90 por ciento, según la mayoría de los métodos de clasificación. El método propuesto alcanza un nivel de precisión del 97%, al clasificar las huellas en seis clases, mostrando una mejora respecto a los anteriores. (Bhuvan & Bhattacharyya, 2009).

3.3. Un sistema de emparejamiento en tiempo real para grandes bases de datos de huellas dactilares.

La técnica que menciona (Ratha, Chen, & Jain, 1996) es más simple, esto consiste en colocar las imágenes en la base de datos como un texto plano, ésta es una secuencia de caracteres con la información de cada pixel. El principal inconveniente de este método es que la escena de descripción puede ser diferente en momentos diferentes para el mismo, dependiendo del contexto de la consulta.

Los autores muestran motivación cuando observan el gran espacio que ocupan estos datos al ser almacenados y plantean reducirlos usando técnicas de extracción de características. Una base de datos de huellas digitales se caracteriza por un gran número de registros (del orden de millones). El tamaño de la base de datos del FBI ha crecido de más de 0,8 millones de huellas dactilares tarjetas (10 huellas dactilares por tarjeta) en 1924, y a más de 114 millones de tarjetas de huellas dactilares en 1994. El requisito de almacenamiento para una colección tan grande de imágenes se requiere de 1.140 terabytes sin compresión. Igualmente, la clase de consulta que se anticipa que este sistema utilice difiere significativamente de las consultas comunes en otros ámbitos de aplicaciones de bases de datos de imágenes.

Según (Ratha, Chen, & Jain, 1996) la representación de un objeto puede cambiar según su orientación, las condiciones de iluminación circundante y el dispositivo de captura utilizado. La información de una imagen tiene mayor dimensionalidad que una información textual. En un conjunto de archivos digitales existen tres aspectos que son: a) toma de datos, b) administración de almacenamiento y c) método de búsqueda y consulta.

Con respecto a la proposición de (Douglas Colgado, 1993) se realiza una división de las imágenes de acuerdo a las características que poseen, para posteriormente almacenarlos en una base de datos. De este modo, para poder comprobar que la huella digital pertenece a una determinada persona, solo será necesario obtener la imagen de su huella y verificarlos según sus características en la base de datos.

3.4. Categorización de las Imágenes de Huellas Dactilares Usando la Técnica de Redes Neuronales

Según Neto, H.V.; Borges, D.L.; plantean que existen varios enfoques para la detección y clasificación de las huellas por patrones, pero esto, todavía es una cadena por completar. Por otra parte, la opción de utilizar una gran base de datos que compara y busca coincidencias particulares para ser igualadas a la huella dactilar, sería muy efectiva si existe una subdivisión de características que reduce significativamente el tiempo de búsqueda creando una clasificación. Para (Henry E. R., 1900) la clasificación debe contar con precisión en arco, cubierto arco, lazo izquierdo, derecho y bucle espiral con la precisión posición y el tipo de los puntos singulares tales como los núcleos y deltas. Según esta teoría o planteamiento existen dos tipos de características la de alto nivel que observa las crestas y los surcos y la de bajo nivel básicamente clasificadas por sus líneas especiales en las crestas y surcos observando las particulares de éstas con puntos característicos de la captura de la huella.

Las características detalladas como el núcleo y punto delta son puntos especiales. El punto central es el punto más arriba en la cresta más interna y el punto delta es la unión de líneas. Estos puntos son casi invariables y muy utilizados por los procesos de clasificación.

Al realizar la captura en vivo de la huella ingresada en muchos casos puede estar dañada debido a heridas mal sanadas o vacíos entre líneas por el deterioro del tiempo o causado también por químicos; al tener esta realidad muchas veces es difícil encontrar el punto delta como marca la clasificación y esta tarea de relacionar huella con un individuo es una tarea complicada de dar solución. Asimismo, se utiliza la base de datos de NIST para obtener posiciones y características de la imagen de la huella, sin embargo, suceden

problemas por la calidad de imagen que puede dar resultados falsos en la clasificación automática de coincidencias.

La clasificación de la huella dactilar es una ventana a un panorama más complejo de lo ya analizado, ya que se cuenta con el método de clasificación por características que es la más conocida y utilizada. El algoritmo funciona muy bien en la búsqueda de similitudes de características de puntos, núcleo y demás; sin embargo al observar una huella poco visible surgen las complicaciones para hallar al individuo por la búsqueda de huella dactilar. El uso de éstas redes en la búsqueda de la huella dactilar reduce significativamente errores además de construir información contextual, siguiendo una serie de patrones de imágenes volviendo a reconstruir partes dañadas, si es necesario; al ser completo y seguro a la tolerancia a fallas, se espera la obtención de resultados óptimos en esta búsqueda del individuo con la huella dactilar.

Capítulo 4

Desarrollo de la Propuesta

Este capítulo ilustra el funcionamiento de una búsqueda secuencial y describe de manera detallada el método propuesto. Realiza una explicación del uso de Machine Learning para la segmentación. También muestra los diferentes diagramas del método, los casos de uso y finalmente el análisis de los resultados.

4.1. Método propuesto

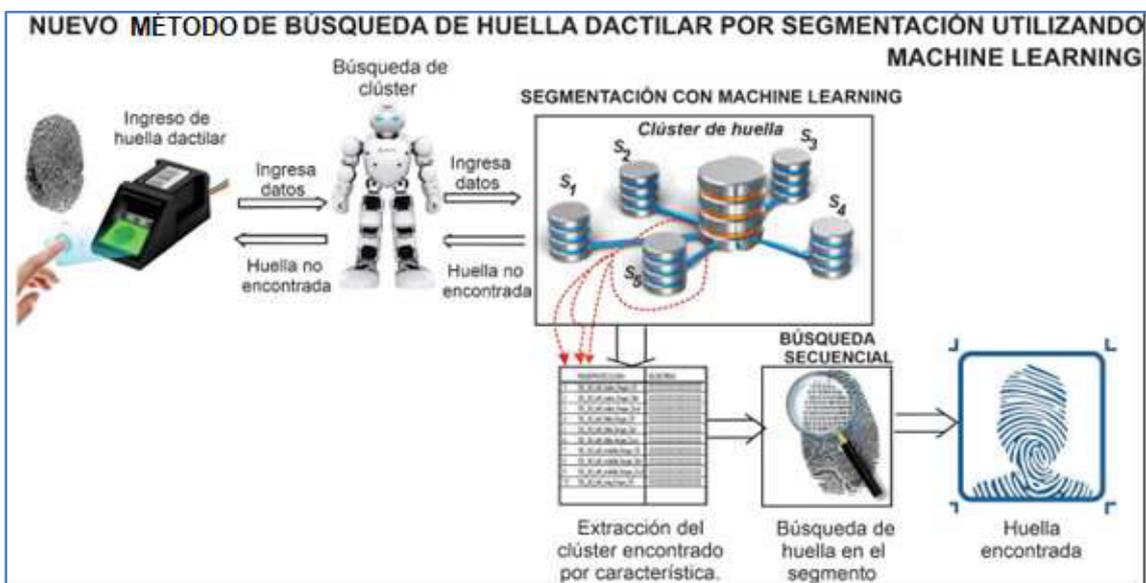
EL uso de los sistemas biométricos en la última década ha sido muy utilizado, para resolver el problema de reconocimiento de patrones de huellas dactilares. Existen propuestas para las soluciones de búsquedas en base de datos muy grandes, y también, existen nuevas propuestas, con el objetivo de acortar el tiempo de respuesta al realizar la búsqueda, (Ruiz Rivera & Ruiz Lizama, 2021).

Es importante tener en cuenta, el método que se utiliza para realizar la búsqueda en una gran base de datos. Desarrollar una búsqueda en una base de datos muy grande de manera secuencial o en su totalidad, demanda un tiempo de respuesta muy alto al identificar o encontrar la huella buscada. Por ello se plantea utilizar la clasificación de huellas por sus características más próximas, generando así agrupaciones o segmentos

que serán almacenadas en la base de datos, reduciendo el tiempo de búsqueda al realizar la búsqueda en un determinado segmento y no en toda la base de datos, (Ruiz Rivera & Ruiz Lizama, 2021).

El modelo propuesto implica la división de la base de datos original en fragmentos o segmentos según las características específicas de cada huella dactilar, permitiendo así que la búsqueda se lleve a cabo en determinado segmento y no requerirá trabajar con toda la base de datos; obteniendo así un resultado inmediato de identificación de la persona, tal como se muestra en la figura 18.

Figura 18. *Búsqueda de la Huella Dactilar, según su Clasificación en Segmentos utilizando Machine Learning*



Fuente: Elaboración propia

Inicialmente a través del lector se captura la huella dactilar de la persona a identificar. En la base de datos ya entrenada (figura 19) el sistema se encarga de obtener el segmento al cual pertenece dicha huella, para continuar con la búsqueda. La finalidad de este método es obtener un tiempo de respuesta corto, algo que es muy importante cuando se necesita tomar una decisión inmediata

La base de datos que guarda y almacena las huellas dactilares es de gran tamaño, y llevar a cabo una búsqueda secuencial, incluso cuando está indexada, requiere un período considerable de tiempo para proporcionar una respuesta. Estas huellas dactilares

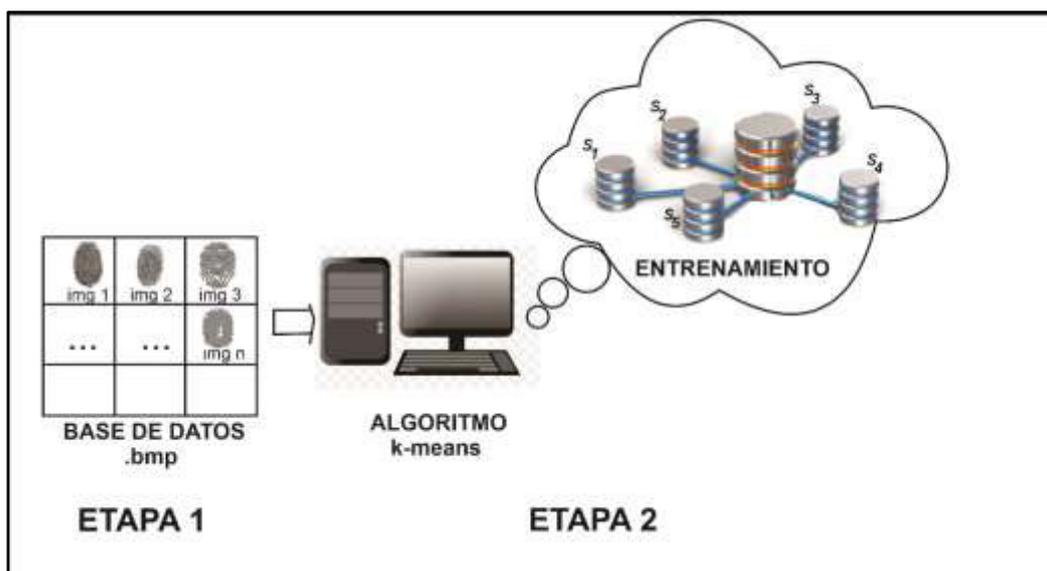
poseen características distintivas que permiten su clasificación en segmentos específicos. El siguiente paso consiste en efectuar la búsqueda solo en el segmento seleccionado, donde se tiene la esperanza de encontrar la huella que se busca. La segmentación se realiza mediante ML, en tanto el algoritmo K-Means, se encarga de agrupar las huellas en segmentos en función de las características que el algoritmo identifica como las más similares. Utilizando este método, dividimos la base de datos en cinco clusters que al final son llamados segmentos, (Ruiz Rivera & Ruiz Lizama, 2021).

Cuando se introduce la huella que se desea buscar, se determina en qué número de segmento pertenece. Una vez conocido el segmento, se hace la búsqueda en dicho segmento aplicando un algoritmo secuencial. Esto implica que la búsqueda no se realiza en toda la BD sino en el segmento identificado. El proceso permite identificar de manera rápida la huella buscada en un tiempo menor.

4.1.1. Fase de entrenamiento del método

A través de algoritmo K-Means se clasifica las imágenes de la base de datos en cinco grupos de acuerdo con sus características (etapa 1), luego en función a las características comunes entre una y otra huella crea los segmentos (etapa 2), ver figura 19.

Figura 19. Fase de entrenamiento del método, utilizando algoritmo K-Means



Fuente: Elaboración propia

Clasificada en tres etapas:

Etapas 1

Para esta etapa, se requiere una base de datos tipo SQL o una colección de imágenes almacenadas en una carpeta, en nuestro caso se utiliza una colección de imágenes de tipo BMP, que será utilizada para el entrenamiento y testeo, se utilizó un total de 539 imágenes donde 80% tomado de forma aleatoria fue utilizado para el entrenamiento y lo restante para el testeo o prueba.

Etapas 2

En esta etapa se utiliza el lenguaje Python como agente principal para programar el modelo, dentro de las librerías usadas, están numpy, pandas, keras y scikitlearn. Después de tener el código (algoritmo) listo, pasamos a la fase de entrenamiento donde utilizaremos el 80% de las imágenes para entrenar al modelo cuyo objetivo principal es buscar características presentes en las imágenes, la cual se pueda generar 5 segmentos donde se ubicarán las imágenes de acuerdo a sus similitudes más próximas encontradas por el método. Luego de entrenar el método se pasa al testeo, donde se observa la buena clasificación del método.

Etapas 3

En esta etapa el modelo ya está entrenado y listo para predecir en que segmento se encuentra una imagen nueva o existente en la base de datos. El modelo predice en que segmento debe colocarse y/o está presente para hacer la respectiva búsqueda secuencial.

4.2. Implementación en Python

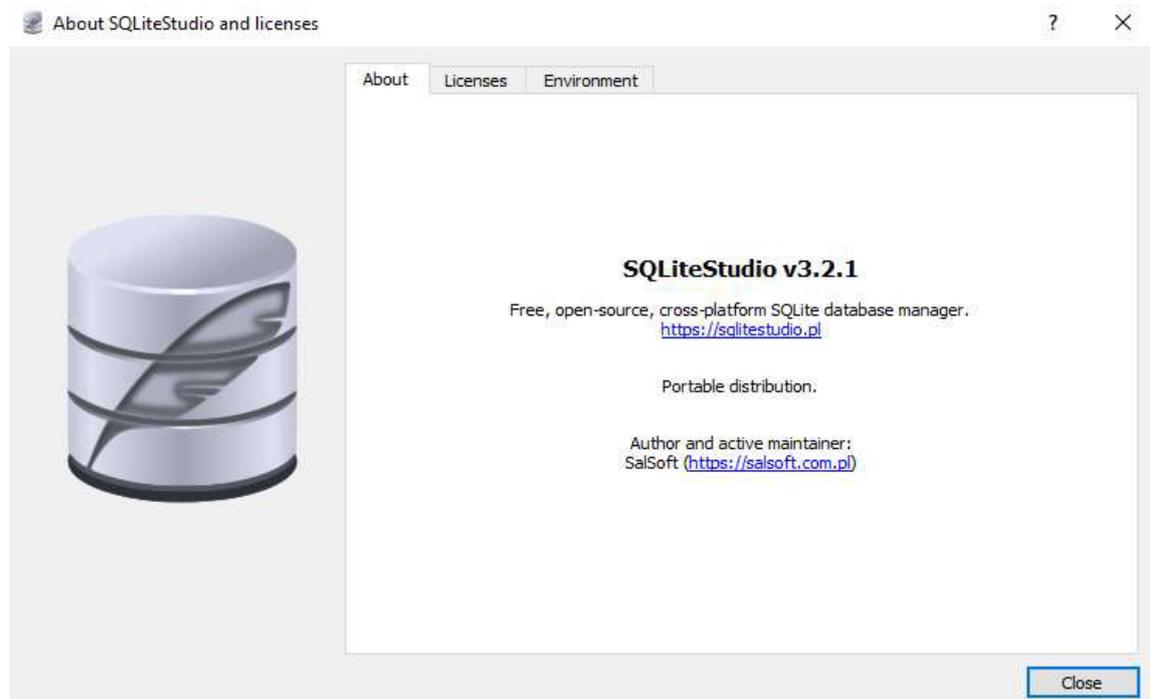
Para el presente proyecto se utilizó el lenguaje de programación Python, el cual es uno de los más utilizados en la actualidad para proyectos basados en Machine Learning. Las librerías necesarias utilizadas, son las siguientes:

- **Tkinter:** Es una librería que proporciona a las aplicaciones de Python una interfaz de usuario sencillo de programar. Además, es un conjunto de herramientas GUI de Tcl/Tk (Tcl: Tool Command Language), proporcionando una amplia gama de usos,

incluyendo aplicaciones web, de escritorio, redes, administración, pruebas y muchos más.

- **Numpy:** es una biblioteca para el lenguaje de programación Python que proporciona el fundamento para realizar vectores y matrices multidimensionales, además de una extensa recopilación de funciones matemáticas para poder operar con ellas.
- **Tensorflow:** es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos.
- **Keras:** es una biblioteca de Redes Neuronales de Código Abierto escrita en Python. Es capaz de ejecutarse sobre TensorFlow, Microsoft Cognitive Toolkit o Theano. Está especialmente diseñada para posibilitar la experimentación en más o menos poco tiempo con redes de Aprendizaje Profundo. Sus fuertes se centran en ser amigable para el usuario, modular y extensible.
- **Scikit learn:** también llamada sklearn, es un conjunto de rutinas escritas en Python para hacer análisis predictivo, que incluyen clasificadores, algoritmos de clusterización, etc. Está basada en NumPy, SciPy y matplotlib, de forma que es fácil reaprovechar el código que use estas librerías.
- **OpenCV:** es una biblioteca libre de visión artificial originalmente desarrollada por Intel. OpenCV significa Open Computer Vision (Visión Artificial Abierta). Desde que apareció su primera versión alfa en el mes de enero de 1999, se ha utilizado en una gran cantidad de aplicaciones, y hasta 2020 se la sigue mencionando como la biblioteca más popular de visión artificial. Detección de movimiento, reconocimiento de objetos, reconstrucción 3D a partir de imágenes, son sólo algunos ejemplos de aplicaciones de OpenCV.
- **Sqlite3:** SQLite es una base de datos SQL autónoma basada en archivos. SQLite viene incluido con Python y se puede usar en cualquiera de sus aplicaciones de Python sin tener que instalar ningún software adicional.

Figura 21. *Versión de SQLiteStudio*



Fuente: Elaboración propia

4.3.2. Verificando

Para verificar que tiene instalado lo necesario para ejecutar el programa, digite `pip list` desde la consola. Si desea consultar la versión de SQLite Studio de su entorno de trabajo, una vez abierto dicho programa, dé click en la pestaña **help** en la cinta de opciones, luego en **About**. Por otro lado, este sistema incluye un archivo con extensión **.db**, el cual contiene 7,000 huellas digitales almacenadas con el tipo de dato BLOB. El nombre del archivo es **FingerPrintDatabase.db**.

4.4. Modelamiento

4.4.1. Usuarios a quien va dirigido:

Conociendo el funcionamiento del sistema que se propone, se debe de identificar que personas estarán involucradas con el sistema. Se identificaron 2 tipos de usuarios las cuales son el usuario que pondrá su huella dactilar en el detector de huellas y el personal encargado del manejo de la aplicación que tiene como entrada a la huella dactilar del usuario.

a) Persona:

Este tipo de usuario será la persona que será identificada mediante su huella dactilar.

Funciones actuales:

Los usuarios en la actualidad tienen las funciones:

- Al poner su huella dactilar en el lector de huella digital, el tiempo de espera para reconocer sus datos en un sistema es prolongado.
- Existe un desconocimiento de los sistemas de huella digital, puesto que los usuarios no usan con mucha frecuencia este tipo de tecnologías.

Funciones con la implementación del sistema:

El usuario principalmente tendrá una función de esperar la información que el sistema devuelva mediante la aplicación y el lector de huella digital, ahora con el sistema ese tiempo de espera será menor al que es actualmente.

b) Operador

Este tipo de usuario será el encargado de captar las huellas dactilares de las personas mediante el lector de huella y mandar esa información a la aplicación, luego de ello podrá informar el resultado de la aplicación que devuelva.

Funciones actuales:

- Cuando un usuario solicita su información, el usuario administrativo busca en una aplicación mediante el DNI o algún identificador de esa persona, esto hace engorroso el proceso de búsqueda o identificación.
- Si la búsqueda es de un proceso forense, la búsqueda se hace aún más engorrosa, porque se debe de identificar mediante alguna seña del cuerpo o mediante algún rasgo físico o del rostro.

Funciones con la implementación del sistema:

Una vez implementado el sistema el usuario administrativo podrá identificar al usuario mediante su huella dactilar de una manera rápida y eficiente, ya que la huella es única e irrepetible por cada persona.

c) Administrador:

El administrador tendrá las funciones:

- Gestionar las personas, añadir, editar y eliminar las personas que tenga el sistema.
- Gestionar el acceso a los operadores al sistema de identificación biométrico.
- Gestionar las búsquedas realizadas.

4.4.2. Modelo de Datos

En la figura 22 se presenta el Modelo de Datos, que implementa la propuesta.

Figura 22. Modelo de Datos

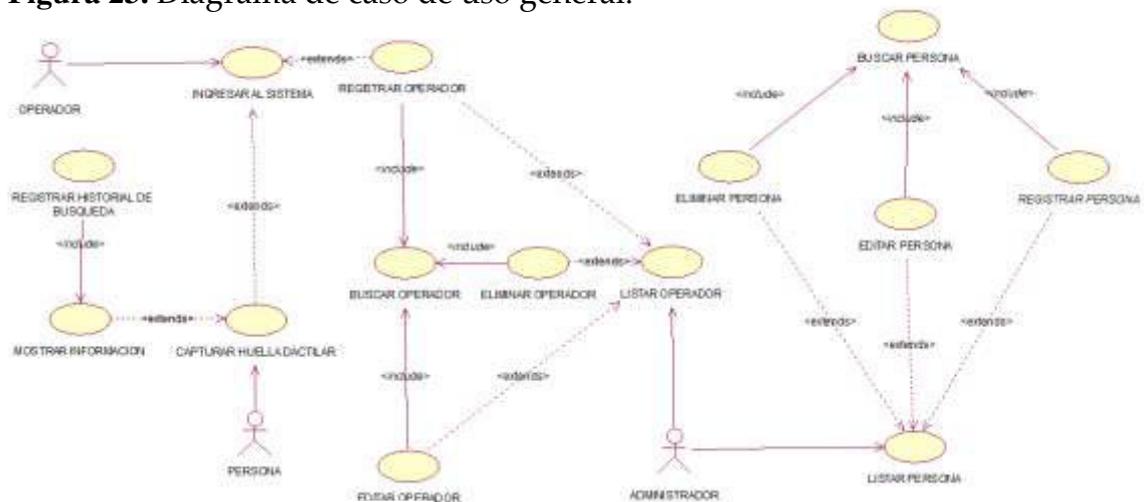


Fuente: Elaboración propia

4.4.3. Caso de Uso

En la figura 23 se presenta el diagrama de caso de uso general.

Figura 23. Diagrama de caso de uso general.



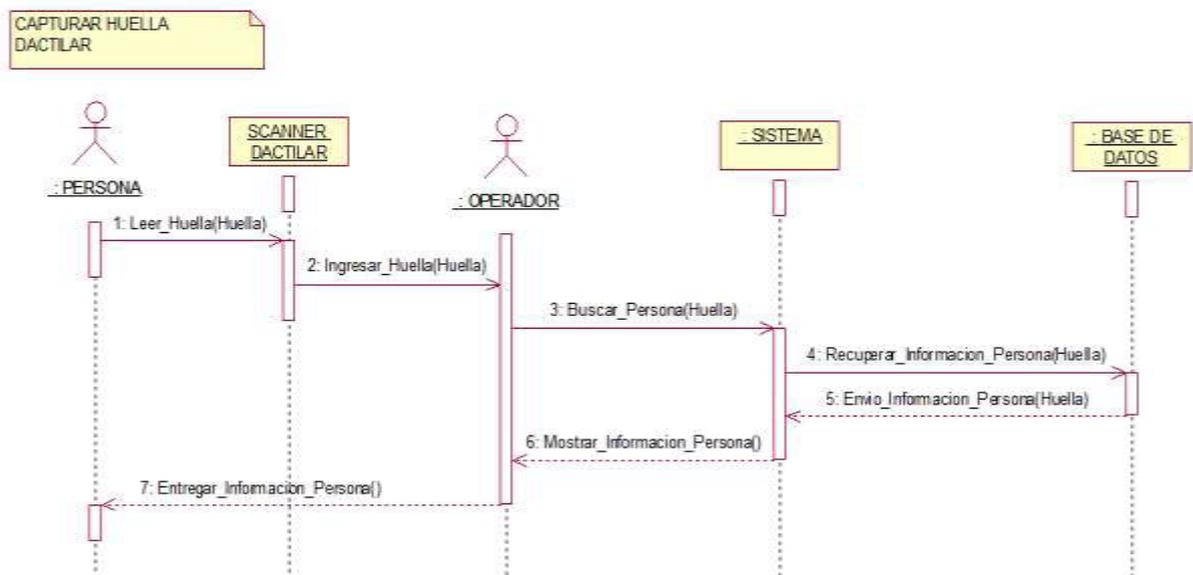
Fuente: Elaboración propia

4.4.4. Diagrama de Secuencia

En la figura 24 se presenta el diagrama de secuencia de la Captura de la Huella Dactilar.

En este caso la persona hará uso del scanner para obtener su información mediante su huella digital, el sistema web internamente realizará la búsqueda para obtener la información en el menor tiempo posible.

Figura 24. Captura de la Huella Dactilar

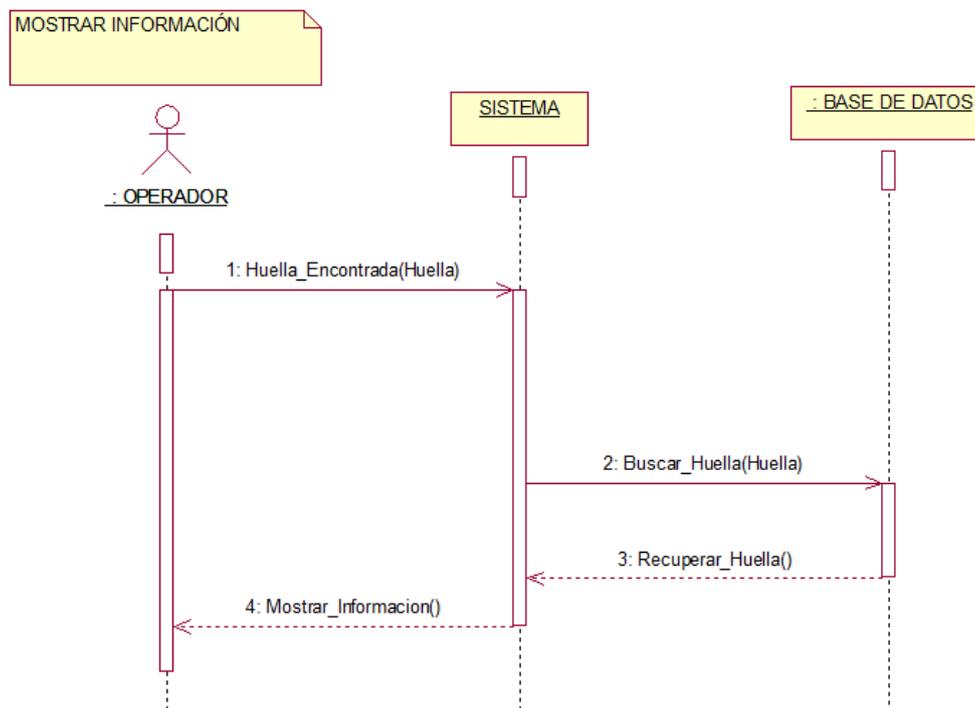


Fuente: Elaboración propia

En la figura 25 se presenta el diagrama que nos permite Mostrar la Información.

En este caso operador mostrará la información a la persona o la entidad que proporcionó la huella dactilar de la persona a buscar.

Figura 25. Mostrar Información



Fuente: Elaboración propia

4.5. Funciones y módulos

▪ Búsqueda secuencial no segmentada

Como se muestra a continuación el script que efectúa la búsqueda secuencial de una imagen en una BD que aún no está segmentada, el objetivo principal de este procedimiento es conseguir el tiempo que tarda el algoritmo en hallar la imagen buscada.

```
import time
import sqlite3
def busqueda_sec(bin):
    inicio = time.perf_counter()
    try:
        sqliteConnection = sqlite3.connect('nocluster.db')
        cursor = sqliteConnection.cursor()
        print("Connected to SQLite")

        sql_fetch_blob_query = """SELECT * from
```

```

                                nocluster where imagen = ?
                                LIMIT 1"""
                                cursor.execute(sql_fetch_blob_query,
                                                (bin,))
                                record = cursor.fetchall()
                                cursor.close()
except sqlite3.Error as error:
    print("Failed to read blob data from sqlite table",
          error)
# finally:
# if sqliteConnection:
# sqliteConnection.close()
# print("sqlite connection is closed")
    tiempo = time.perf_counter() - inicio
    dir_img = 'D:/Python/milhuellas/k-images/'
            +record[0][0]+' .bmp'
            #print(tiempo)
return tiempo, dir_img

```

▪ Predicción y búsqueda

Se tiene un modelo previamente entrenado cuyo objetivo es predecir el cluster en el que se ubica la imagen que se desea buscar. La ejecución del algoritmo de búsqueda en dicho cluster registra el tiempo total que toma el proceso, el cual puede verse en la figura 27, (Ruiz Rivera & Ruiz Lizama, 2021).

```

import time
from joblib import load
import sqlite3
cluster= load('2domodelo')
def busqueda_cluster(features,bin):
    inicio = time.perf_counter()
    clus = cluster.predict(features)[0] #clus = 0,1,2,3
    try:
        sqliteConnection = sqlite3.connect('cluster.db') |
        cursor = sqliteConnection.cursor()
        print("Connected to SQLite")
        sql_fetch_blob_query = """SELECT * from cluster{0} where imagen =
                                ? LIMIT 1""".format(clus+1)
        cursor.execute(sql_fetch_blob_query, (bin,))
        record = cursor.fetchall()
        cursor.close()
    except sqlite3.Error as error:
        print("Failed to read blob data from sqlite table", error)

# finally:
# if sqliteConnection:
# sqliteConnection.close()
# print("sqlite connection is closed")
    tiempo = time.perf_counter() - inicio
    dir_img = 'D:/Python/milhuellas/k-images/' +record[0][0]+' .bmp'
    #print(tiempo)
return tiempo, dir_img

```

- Funciones adicionales

Se utiliza una red neuronal que apoya al procesamiento de imágenes y a la identificación del objeto, (Ruiz Rivera & Ruiz Lizama, 2021).

```
def image_feature(imagen):
    model = InceptionV3(weights='imagenet', include_top = False)
    features = []
    img=image.load_img(imagen, target_size=(224,224))
    x = img_to_array(img)
    x = np.expand_dims(x,axis = 0)
    x = preprocess_input(x)
    feat=model.predict(x)
    feat=feat.flatten()
    features.append(feat)
    return features
def salir():
    shutil.rmtree('Temporal')
    exit()
```

Capítulo 5

Resultados de la Investigación

5. 1. Resultados Obtenidos

Análisis de tiempos

Luego de implementar el algoritmo mediante el programa, es pertinente determinar el tiempo de procesamiento de una búsqueda, con fines de comparación. para ello se realizó 2 pruebas diferentes:

1. Determinar el tiempo de búsqueda de una misma imagen 10 veces, para determinar el margen de fluctuación del tiempo de procesamiento
2. Determinar el tiempo necesario para buscar 50 imágenes seleccionadas al azar, sin repetición, con el propósito de evaluar los tiempos en relación con la posición de la imagen en la BD no segmentada.

Desarrollo

Primero se determina la fluctuación en los tiempos de búsqueda, para poder visualizar el margen de error en los tiempos de búsqueda de la media correspondiente. Para eso, se ha ejecutado la búsqueda de un mismo elemento 10 veces. A continuación, en la tabla 2, se muestran los resultados obtenidos:

Tabla 2: Diez muestras

T. Seg.	T. Seg.
0.002906	0.000593
0.002537	0.000563
0.002604	0.000563
0.002631	0.000526
0.002538	0.000516
0.002396	0.000557
0.003113	0.000445
0.002258	0.000503
0.002609	0.000513
0.002917	0.000561

Los promedios de tiempo de búsqueda son, respectivamente, 0.002651s y 0.000534s respectivamente, con desviaciones estándar correspondientes de 0.000258s y 0.000042s. La desviación estándar permite cuantificar la propensidad de un dato a desviarse de la media. A partir de esto, podemos identificar el margen de error de los datos obtenidos con respecto a los valores ideales.

Entonces, para los cálculos siguientes, podemos considerar (a un intervalo de confianza de 95%) un margen de error en los tiempos obtenidos de $\pm 0.000160s$ y $\pm 0.000026s$ para las búsquedas Segmentada y Secuencial respectivamente.

Luego de ejecutar los algoritmos expuestos, se tienen los resultados de la tabla 3:

Tabla 3. *Tabla de tiempos.*

Nro.	Nro. Huella	Tiempo Segmentada (Seg)	Tiempo Secuencial (Seg)	Nro.	Nro. Huella	Tiempo Segmentada (Seg)	Tiempo Secuencial (Seg)
1	531	0.004379	0.008464	26	268	0.002802	0.004809
2	528	0.004228	0.008589	27	264	0.009219	0.008393
3	517	0.004777	0.008337	28	261	0.003089	0.005612
4	513	0.004227	0.008667	29	259	0.003477	0.006325
5	506	0.003872	0.008317	30	247	0.007648	0.008603
6	490	0.003828	0.008067	31	237	0.007601	0.009455
7	486	0.005438	0.008955	32	236	0.009113	0.005527
8	483	0.005157	0.010635	33	233	0.003252	0.006325
9	480	0.004468	0.008513	34	222	0.003704	0.005572
10	472	0.006281	0.015658	35	211	0.008135	0.010234
11	444	0.002972	0.007607	36	179	0.003113	0.003579
12	442	0.003441	0.007469	37	173	0.002846	0.002438
13	437	0.003763	0.007042	38	163	0.003166	0.002587
14	428	0.004522	0.016173	39	156	0.002703	0.002495
15	408	0.004283	0.006484	40	153	0.002625	0.002485
16	397	0.004105	0.008492	41	144	0.002265	0.002865
17	392	0.004123	0.007803	42	138	0.002307	0.002889
18	390	0.004025	0.006753	43	121	0.002772	0.001929
19	377	0.003137	0.006483	44	91	0.005962	0.003999
20	356	0.010098	0.014196	45	80	0.002847	0.001563
21	337	0.003842	0.008346	46	78	0.002946	0.001216
22	316	0.008645	0.013955	47	66	0.002495	0.001253
23	312	0.003817	0.006923	48	64	0.002835	0.001167
24	309	0.003928	0.005397	49	42	0.002091	0.001048
25	272	0.003292	0.006387	50	9	0.003076	0.000588

Fuente: Elaboración propia

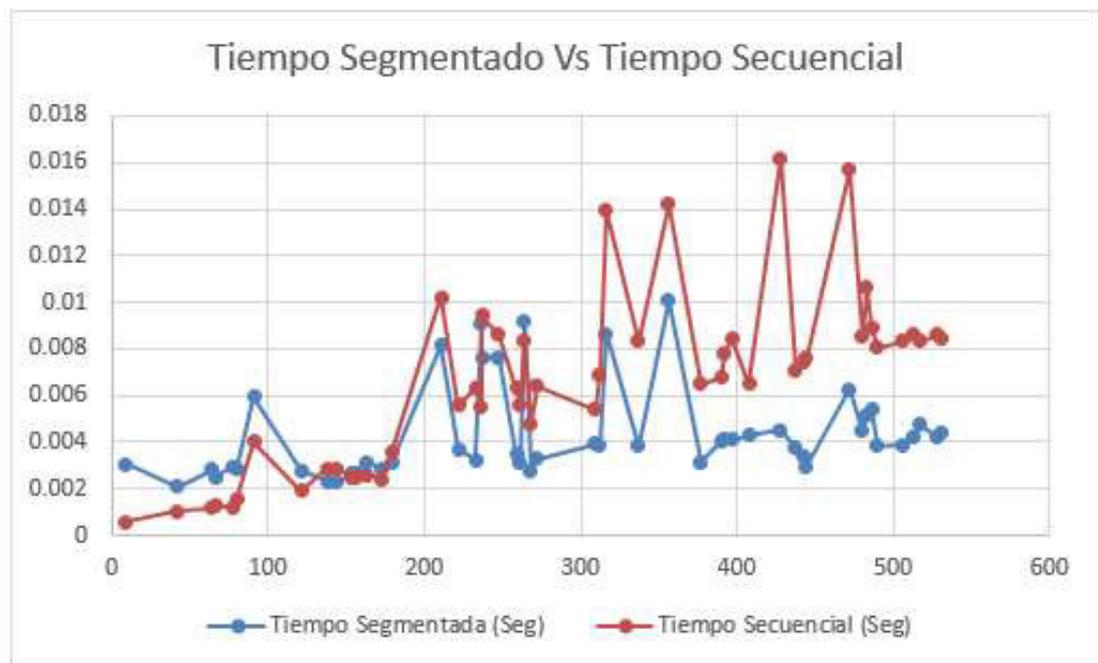
La primera columna describe la posición de la imagen en la base de datos, mientras que las siguientes dos columnas muestran el tiempo de búsqueda en segundos de la imagen usando los algoritmos segmentado y secuencial, respectivamente.

En conclusión, el experimento desarrollado permite confirmar que el algoritmo de búsqueda segmentada presenta varias grandes ventajas sobre el algoritmo secuencial clásico. Esto principalmente debido a los dos siguientes motivos:

- La búsqueda secuencial en un cluster es evidentemente considerablemente menor en promedio al tiempo de búsqueda en la base de datos completa
- El tiempo de evaluación de una imagen en el modelo de clasificación es lo suficientemente breve como para no exceder al tiempo de búsqueda del algoritmo secuencial

Tal como se aprecia en la Figura 26, la intersección de las rectas T. Secuencial versus T. Segmentada, se da en el punto 112.5. El diagrama de regresión muestra que para una BD de 537 imágenes, la búsqueda secuencial toma más tiempo a partir de la posición 113 en la BD no segmentada, (Ruiz Rivera & Ruiz Lizama, 2021).

Figura 26. Gráfica de tiempos



Fuente: Elaboración propia

5.2. Interfaz

Presentamos una interfaz donde con el botón abrir, seleccionamos la huella a buscar; luego presionamos el botón buscar para que realice la búsqueda de la huella a encontrar, finalmente nos muestra como salida los resultados de la búsqueda secuencial como también de la búsqueda clauterizada, como muestra la figura 27.

Figura 27.



Fuente: Elaboración propia

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

La presente tesis titulada “*Método de búsqueda eficiente para resolver el problema de identificación de huella dactilar aplicando machine learning*” concluye que, en concordancia con el objetivo principal del proyecto; el cual es reducir el tiempo de búsqueda al intentar identificar a una persona en cualquier tipo de evento, resulta más eficiente el uso de ML debido a que permite dividir una BD en grupos de huellas de acuerdo a sus características. Luego se emplea la técnica de búsqueda en secuencia en el segmento seleccionado, esto simplifica la localización de la huella dactilar que corresponde a un individuo.

El planteamiento presentado no realiza la búsqueda de una huella en toda una BD, puesto que hacerlo demanda mucho tiempo. Por ello la propuesta es realizar una búsqueda focalizada, en solo uno de los segmentos que fue clasificado de acuerdo con sus características; esto permite reducir el tiempo de localización de la huella lo que significa una reducción del 55.6%.

Recomendaciones

En el contexto de la investigación se recomienda realizar búsqueda segmentada en BD que posean gran cantidad de documentos, y luego manteniendo el enfoque planteado en el artículo de investigación. Para llevar a cabo la segmentación, se aconseja la utilización de algoritmos genéticos, que son técnicas adaptables y aplicables para abordar desafíos de búsqueda y optimización. El objetivo primordial es mejorar la eficiencia temporal de los resultados finales.

Los aportes en este ámbito relacionado a las huellas dactilares solucionan varias dificultades al momento de identificar la identidad de la persona hallada en cualquier tipo de siniestro.

Referencias

- Bathia, S., & Deogun, J. (1998). Conceptual Clustering in Information Retrieval. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 28(3), 427 - 436.
- Bhuvan, M., & Bhattacharyya, D. (2009). An Effective Fingerprint Classification and Search Method. *IJCSNS International Journal of Computer Science and Network Security*, 39 - 48.
- Coomonte Belmonte, Rafael. 2009. *Sistema de reconocimiento de personas mediante su patrón de iris basado en la Transformada Wavelet*. Proyecto Fin de Carrera, E. T. S. I. Telecomunicaciones – U. P. M., mayo 2006.
<https://docplayer.es/156992138-Proyecto-fin-de-carrera.html>
- Dass, S., & Jain, A. (December de 2004). Fingerprint Classification Using Orientation Field Flow Curves In *ICVGIP*. 650 - 655.
- Douglas H., D. (Nov. de 1993). Enhancement and Feature Purification of Fingerprint Images. *Pattern and Recognition*, 26(11), 1661 – 1671.
- Fayyad, U., Piatetsky-Shapiro, G., & Padhraic, S. (1996). From data Mining to Knowledge Discovery in *Databases*. *American Association for Artificial Intelligence*, 37 - 54.
- Fernandez, F. A. (2008). *Biometric sample quality and its application to multimodal authentication systems*. Madrid: Universidad Politecnica de Madrid.
- Forgy, E. (1965). Cluster analysis of multivariate data: Efficiency vs. Interpretability of classifications. *Biometrics*, 21, 768.
- Henry, E. R. (1900). *Classification and Users of Fingerprints*.

- Kearns, M., Mansour, Y., & Ng, A. (1997). An information-theoretic analysis of hard and soft assignment methods for clustering. (K. M. Publishers, Ed.) *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 282 - 293.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281 - 297.
- Makhoul, J., Roucos, S., & Gish, H. (1985). Vector quantization in speech coding. *Proceedings of the IEEE*, 73, 1551 - 1558.
- Maltoni, D., Maio, D., Jain, A., & Prabhakar, S. (2003). Multimodal biometric systems. *Handbook of fingerprint recognition*, 233 - 255.
- Micolta López JF; Oltra Badenes, RF. (2015). Gestión de la identidad biométrica en las organizaciones. 3C TIC, cuadernos de desarrollo aplicados a las TIC. 4(1):57-72. <http://hdl.handle.net/10251/94985>
- Persto S.A. de C.V. (n.d.). (15 de junio de 2020). Verificación de identidad biométrica. Obtenido de <http://www.persto.com/>
- Ratha, N. K., Karu, K., Chen, S., & Jain, A. (1996). A real-time matching system for large fingerprint databases. *IEEE transactions on pattern analysis and machine intelligence*, 18(8), 793 - 813.
- Ruiz, M., & Ruiz, E. (2021). *Método de búsqueda eficiente para resolver el problema de identificación de huella dactilar aplicando Machine Learning*. Revista Industrial Data 24(2): 293-317 (2021)
- Sanchez Calle, Angel, *Aplicaciones de la Visión Artificial y Biometría Informática*, Universidad Rey Juan Carlos, Madrid 2005
- Scheunders, P. (1997). A comparison of clustering algorithms applied to color image quantization. *Pattern Recognition Letters*, 18(11-13), 1379 - 1384.

Villamizar, J. A. (1994). Procesamiento y clasificación de huellas dactilares. *Lecturas Matemáticas*, 15, 149 -165.

Watson C.I.:Wilson, C. (March de 1992). Nits 4, special database. National Institute of Standars and Technology.