

Evaluation of end-to-end CNN models for palm vein recognition

José I. Santamaría*, Ruber Hernández-García[†], Ricardo J. Barrientos*, Francisco Manuel Castro[‡],
Julián Ramos-Cózar[‡], and Nicolás Guil[‡]

*Laboratorio LITRP, Depto. DCI, Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule, Talca, Chile

[†]Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Universidad Católica del Maule, Talca, Chile

[‡]Depto. Arquitectura de Computadores, ETSI Informática, Universidad de Málaga, Málaga, España

Email: rhernandez@ucm.cl

Abstract—In recent years, biometric systems have positioned themselves among the most widely used technologies for people recognition. In this context, palm vein patterns have received the attention of researchers due to their uniqueness, non-intrusion, and reliability. Currently, research on palm vein recognition based on deep learning is still very preliminary, most of the works are based on very deep models by using pre-trained models and transfer learning techniques. In this work, we evaluate end-to-end CNN models for palm vein recognition. The proposed method was implemented on seven public databases of palm vein images and two convolutional neural network architectures were evaluated: SingleNet, the proposed architecture of few convolutional layers, and a deeper architecture based on ResNet32. The experimental results demonstrate the superiority of the SingleNet model, outperforming the state-of-the-art results for the IITI, PUT, and FYO databases, achieving the same results on the Tongji and PolyU datasets, and obtaining a slightly lower performance for the VERA and CASIA databases. Comparing to the state-of-the-art approaches, our proposed method is computationally simpler than those that are based on very deep architectures and others that fuse hand-crafted and CNN extracted features.

Index Terms—Convolutional Neural Networks, Palm vein recognition, Biometrics.

I. INTRODUCCIÓN

En la sociedad moderna, la identificación y verificación de personas mediante sus rasgos biométricos son componentes esenciales de los gobiernos y las empresas privadas, contando con varias áreas de aplicación como el comercio electrónico, servicios bancarios y financieros, servicios hospitalarios, el internet de las cosas, entre muchas otras. Dentro de los sistemas biométricos la huella dactilar es la tecnología más utilizada por su aceptación universal y bajo coste, sin embargo, ésta cuenta con varios inconvenientes relacionados con la privacidad y seguridad. En la actualidad, y de hecho en el futuro inmediato, la pandemia de COVID-19 ha repercutido en la demanda de sistemas biométricos basados en contacto. Concretamente los sistemas de huellas dactilares se han visto afectados, puesto que el virus puede permanecer por un largo tiempo en las superficies. Así, las tecnologías biométricas sin contacto adquieren actualmente mayor relevancia.

En este contexto, la biometría basada en las venas de la palma de la mano es más segura y tiene mayor precisión que otros tipos de sistemas biométricos. Durante la última década, existen varios enfoques basados en el reconocimiento

de venas de la palma de la mano para la identificación de individuos [1]. Los enfoques más recientes se han centrado en las redes neuronales convolucionales (CNN), que son capaces de aprender representaciones de características más robustas. A pesar de ello, los enfoques basados en CNN presentan dos problemas muy importantes que hay que resolver para poder aplicarlos en escenarios del mundo real: (1) las bases de datos disponibles públicamente son limitadas; y (2) la mayoría de los trabajos se basan en modelos muy profundos centrados sobre todo en las técnicas de transferencia de aprendizaje usando modelos pre-entrenados.

El presente artículo propone la evaluación de arquitecturas CNN de principio a fin (end-to-end) para el desarrollo de un modelo de reconocimiento de personas por medio de venas de la palma que sea robusto, escalable y con una arquitectura poco profunda. Para esto se consideraron siete bases de datos públicas de venas de la palma existentes en la literatura, con el fin de establecer comparaciones a los trabajos del estado del arte. Se realizó una etapa de preprocesamiento de imágenes en donde, se extrajo el ROI, se aplicaron métodos de normalización, ecualización y nitidez, finalizando con la aumentación de los conjuntos de imágenes. Se utilizaron dos arquitecturas para realizar la evaluación: SingleNet una arquitectura de pocas capas convolucionales basada en AlexNet [2] y una arquitectura más profunda basada en ResNet32 [3]. Se entrenaron ambos modelos y se realizaron pruebas con el clasificador CNN y con un clasificador KNN usando el modelo entrenado como codificador de características.

De acuerdo a lo anterior, las principales contribuciones de la propuesta son las siguientes: (1) crear un modelo de principio a fin eficiente, el cual sea lo más simple posible desde el punto de vista de la arquitectura, lo cual posibilitaría ser desplegado en hardware de bajo costo como podrían ser arquitecturas paralelas y/o embebidas como NVIDIA Jetson; y (2) contribuir a extender las aplicaciones reales de los sistemas biométricos basados en venas de la palma de la mano.

El resto del artículo se compone de cuatro secciones. En primer lugar, se abordan los trabajos relacionados del estado del arte en la Sección II. En la Sección III, se presenta el modelo y metodología propuestos. Luego, se presentan los resultados experimentales obtenidos en la Sección IV y finalmente, se ofrecen las conclusiones en la Sección V.

II. TRABAJOS RELACIONADOS

En la Tabla I, se presenta un resumen de los diferentes enfoques abordados en las investigaciones basadas en CNN sobre reconocimiento de venas de la palma para la identificación de personas. Los resultados del desempeño representan la tasa de acierto extraída de los artículos originales. Los mejores resultados para cada base de datos se destacan en negrita.

La mayoría de las arquitecturas CNN utilizadas en la literatura son complejas y muy profundas. Cabe destacar que los enfoques reportados han mejorado los resultados de rendimiento en las bases de datos de venas de la palma, logrando un rendimiento de muy alta precisión e incluso el 100% en cuatro de las siete bases de datos públicas. Sin embargo, aún presentan limitaciones relevantes por resolver para poder aplicarlas en escenarios del mundo real, dentro de las cuales se destacan: (1) los conjuntos de datos son limitados, las bases de datos disponibles públicamente tienen un número pequeño de sujetos y un número limitado de muestras por clase, lo que restringe los datos de entrenamiento disponibles para tareas de reconocimiento de venas, y al mismo tiempo, reducen el rendimiento de generalización de los modelos CNN; (2) los modelos son muy profundos, aún existen grandes desafíos en el desarrollo de un modelo basado en CNN simple y factible para el reconocimiento de venas de la palma, ya que la mayoría de los enfoques del estado del arte se basan en arquitecturas muy profundas (p.e. VGG, ResNet, GoogLeNet, DenseNet, ResNeXt, MobileNet, entre otros) que requieren un proceso de entrenamiento muy lento y complicado con millones de parámetros para optimizar; y (3) el bajo costo y miniaturización de los dispositivos de implementación, lo cual está relacionado con el tema anterior, la ventaja más importante de los modelos más simples no es solo su entrenamiento en un tiempo razonable, sino también su despliegue en sistemas de bajos recursos.

III. METODOLOGÍA PROPUESTA

En esta sección, se presenta de forma detallada cada uno de los elementos utilizados para resolver el problema de estudio. Se entrega una descripción precisa de los aspectos necesarios para llevar a cabo la metodología propuesta.

A. Modelos CNN evaluados

Se seleccionaron dos arquitecturas CNN, con las cuales se desarrollaron los modelos utilizados para realizar los experimentos en cada una de las bases de datos de estudio. Primero se implementó una arquitectura de pocas capas convolucionales nombrada SingleNet y además se evaluó una arquitectura ResNet con 32 capas, ambas se describen a continuación.

1) *SingleNet*: La gran mayoría de las arquitecturas de la literatura son muy profundas, lo cual se ve reflejado en un gran número de capas convolucionales en la red y esto conlleva una considerable complejidad computacional y mucho tiempo de entrenamiento. La razón por la cual se seleccionó una arquitectura de pocas capas convolucionales fue para obtener un modelo computacional lo más simple posible con resultados

de aprendizaje comparable a los modelos más complejos del estado del arte.

Las especificaciones de la arquitectura implementada se presentan en la Tabla II, la cual se basa en el modelo AlexNet [2]. La red está compuesta por nueve capas (contando la entrada y salida), de ellas cinco capas convolucionales y tres capas densas (*fully-connected*). De acuerdo a experimentaciones realizadas para el refinamiento de la arquitectura, con el objetivo de reducir el sobreajuste de la red, fue necesario añadir operaciones de normalización (*BatchNorm*) a las capas convolucionales y operaciones de abandono (*dropout*) luego de las dos primeras capas *fully-connected*.

2) *ResNet*: La arquitectura ResNet modifica la estructura tradicional de la red neuronal convolucional incorporando módulos residuales que realizan una conexión de acceso directo. Los módulos residuales se apilan secuencialmente y la conexión de acceso directo reinyecta representaciones previas al flujo de entrada de la siguiente capa [3]. Esta arquitectura se tuvo en cuenta con el objetivo de agregar un grado mayor de complejidad en comparación a la arquitectura SingleNet propuesta, sin aumentar significativamente el número de parámetros ni la complejidad computacional del modelo.

La arquitectura Resnet permite de forma fácil variar la profundidad de la red, por lo que teniendo en cuenta la complejidad de la tarea de clasificación se optó por la variante ResNet-32, la cual se inspira en gran medida en la arquitectura de ResNet-34 [3]. La Tabla III resume las especificaciones de la arquitectura implementada. El modelo está compuesto por tres grupos de módulos residuales (ResStack), donde todas las capas cuentan con la misma cantidad y tamaño de filtros. La primera capa convolucional de cada grupo residual reduce los mapas de características a la mitad por medio de un valor de $stride = 2$. Finalmente, después del pipeline convolucional, se agrega una agrupación promedio (*AvgPooling*) antes de la capa densa de salida para la clasificación.

B. Bases de datos

Las bases de datos de venas de la palma de la mano son de gran interés para la comunidad científica, dadas las características y beneficios de este rasgo biométrico. En particular, los patrones de venas se visualizan en las imágenes como líneas oscuras que aparecen en la palma de la mano a través de la irradiación de luz en una longitud de onda cercana al infrarrojo (NIR), y solo tienen validez en el cuerpo vivo [4]. En la Tabla IV se presenta la información de las siete bases de datos públicas encontradas en la literatura, las cuales son utilizadas para la evaluación de los modelos propuestos. El número de muestras se expresa como muestras por número de sesiones de adquisición.

Debido a que las bases de datos disponibles públicamente tienen un número pequeño de sujetos y un número limitado de muestras por clase, se propone hacer un proceso de aumentación de muestras. La literatura aborda este problema utilizando diferentes técnicas que permiten ampliar el conjunto de datos existentes logrando reducir en gran medida el error

TABLA I: Resumen de los diferentes enfoques CNN para el reconocimiento de las venas de la palma.

Enfoque	Arquitectura CNN	Año	Base de Datos	Desempeño
Red neuronal convolucional profunda y clasificador SVM [4]	Inception ResNet v1	2018	Tongji	100%
Mejora y evaluación de la Transformada de Onda Discreta de CNN previamente entrenadas [5]	AlexNet	2019	PUT	93.92%
	VGG-16			90.83%
	VGG-19			92.27%
	GoogLeNet			76.92%
Red convolucional de codificador-decodificador y una red siamesa previamente entrenada que utiliza Triplet Loss [6]	U-net encoder-decoder + conv6-fc triplet Siamese	2019	CASIA	85.16%
			IITI	97.47%
			PolyU	98.78%
Modelo CNN de extremo a extremo con métodos de regularización [7]	regularized conv2+bnorm+fc2	2019	PolyU	99.90%
Transferencia del aprendizaje y el ajuste fino en los modelos de CNN [8]	conv6-fc2	2019	PUT	95.16%
	AlexNet			92.16%
	VGG-16			97.33%
	ResNet-50			99.83%
	SqueezeNet			91.66%
Modelo de CNN previamente entrenado con PCA y clasificador de bosque aleatorio [9]	AlexNet	2019	PolyU	100%
CNN pre entrenado y autocodificador convolucional densamente conectado [10]	DenseNet-161	2020	PolyU	99.69%
	ResNeXt-101			98.67%
Filtro Jerman y modelo CNN optimizado mediante optimización bayesiana [11]	3 blocks (n*conv+bnorm) +fc	2020	CASIA	99.40%
Fusión a nivel de puntuación de un potente descriptor BSIF basado en texturas y un clasificador CNN [12]	AlexNet	2020	CASIA	99.83%
			FYO	98.91%
			PUT	99.00%
			VERA	100%
			Tongji	100%
Evaluación de modelos CNN clásicos y recientes mediante el uso del aprendizaje por transferencia [13]	MobileNet v3	2021	PolyU	100%
	EfficientNet		Tongji	98.67%
			PolyU	100%
			Tongji	99.00%
	GhostNet		PolyU	99.60%
			Tongji	96.20%
			PolyU	98.57%
	ResNeSt-50		Tongji	94.92%

TABLA II: Especificaciones de la arquitectura SingleNet. El número de salidas de la capa Full03 se corresponde con el número de clases C .

Capas	Especificaciones (filtros / norm / pooling)
Input	128 × 128
Conv01	7 × 7 × 96 + BatchNorm + Max-Pooling: 2 × 2
Conv02	5 × 5 × 192 + BatchNorm + Max-Pooling: 2 × 2
Conv03	3 × 3 × 284 + BatchNorm + Max-Pooling: 2 × 2
Conv04	3 × 3 × 512 + BatchNorm + Max-Pooling: 2 × 2
Conv05	2 × 2 × 1024 + BatchNorm + Max-Pooling: 2 × 2
Full01	1024 + Dropout: 0.65
Full02	512 + Dropout: 0.65
Full03	C

TABLA III: Especificaciones de la arquitectura ResNet-32. El número de salidas de la capa Full01 se corresponde con el número de clases C .

Capas/Bloques	Especificaciones (tamaño × número de filtros)
Input	128 × 128
Conv2D-BN-ReLU	3 × 3 × 16 + BatchNorm
ResStack01	32 × 32 × 16
ResStack02	16 × 16 × 32
ResStack03	8 × 8 × 64
AvgPooling	8 × 8
Full01	C

y mejorar la precisión [20]. Entre las transformaciones aplicadas se encuentran rotación, traslación, acercamiento, cortes, perturbación de colores, entre otras. En la Fig. 1 se presenta

TABLA IV: Detalles de las bases de datos de venas de la palma disponibles en la literatura. Por razones de espacio en el resto del documento, las bases de datos CASIA-850 y PolyU-NIR son identificadas como CASIA y PolyU, respectivamente.

Base de Datos	Sujetos × Manos	Muestras	Imágenes Totales	Tamaño de la Imagen
Tongji [4]	300 × 2	10 × 2	12.000	800 × 600
IITI [14]	185 × 2	6	2.220	2592 × 1944
CASIA-850 [15]	100 × 2	3 × 2	1.200	768 × 576
VERA [16]	110 × 2	5 × 2	2.200	480 × 680
PUT [17]	50 × 2	4 × 3	1.200	1280 × 960
PolyU-NIR [18]	250 × 2	6 × 2	6.000	352 × 288
FYO [19]	160 × 2	1 × 2	640	800 × 600

el proceso de aumentación de muestras realizado. El proceso propuesto con este fin consistió en fijar una región de interés (ROI) y aumentar el conjunto de imágenes realizando traslaciones de 5 píxeles en los ejes de las abscisas y ordenadas, además de realizar rotaciones de 5 grados con respecto al centro de la palma de la mano, de esta manera fue posible aumentar cada una de las muestras en 87 nuevas imágenes. Como salida de este preprocesamiento se aplicaron por separado tres técnicas diferentes para mejorar las imágenes: (1) el algoritmo CLAHE [21] para mejorar el contraste por medio de la equalización del histograma; (2) un filtro de refinamiento (*sharpening*) por máscara de desenfoque [22] para resaltar los bordes de los patrones de venas; y (3) una normalización de

los niveles de grises. Las imágenes resultantes de la aplicación de las tres técnicas anteriores dieron origen a tres diferentes variantes de cada base de datos, las cuales fueron llamadas Clahe, Sharpen y Normal, respectivamente. Al realizar pruebas sobre cada una de las variantes, se pudo comprobar la influencia del preprocesamiento en el rendimiento de los modelos y así obtener para cada base de datos la de mejor desempeño.

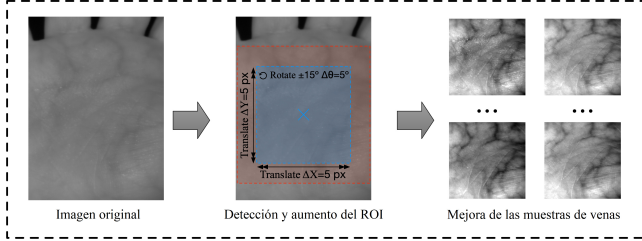


Fig. 1: Proceso de aumentación de muestras del ROI.

C. Diseño experimental

1) *Selección de conjuntos de entrenamiento y prueba:* Para lograr seleccionar la cantidad de datos para entrenamiento y prueba, fue necesario explorar y cruzar datos de los diferentes artículos en donde hicieron uso de estas bases de datos en redes neuronales convolucionales. A partir esto fue posible establecer la cantidad de datos para entrenamiento y prueba que se utilizaron para desarrollar y evaluar los modelos. Esta selección se realizó de acuerdo con la frecuencia con que estos se utilizaron en los diferentes artículos con el objetivo de establecer comparaciones equivalentes, en la Tabla V se resume dicha información.

TABLA V: Cantidad de muestras originales seleccionadas para los procesos de entrenamiento y prueba.

Base de Datos	Muestras de entrenamiento (%)	Muestras de prueba (%)	Total de muestras por clase
Tongji	16 (80%)	4 (20%)	20
ITI	3 (50%)	3 (50%)	6
CASIA	4 (80%)	2 (20%)	6
VERA	8 (80%)	2 (20%)	10
PUT	9 (80%)	3 (20%)	12
PolyU	7 (60%)	5 (40%)	12
FYO	1 (50%)	1 (50%)	2

2) *Pruebas de rendimiento realizadas:* Después del entrenamiento de ambos modelos, se realizaron pruebas de reconocimiento de individuos en cada base de datos estudiada. Para esto, se emplearon los modelos CNN obtenidos como clasificador y además se utilizaron como codificadores de características junto a un clasificador KNN.

En todos los casos los datos utilizados corresponden al conjunto de prueba establecidos con anterioridad para cada una de las bases de datos. En las pruebas con KNN, este conjunto de datos fue separado para ser parte de la galería y como muestras de consulta. Se realizaron varias pruebas, primero se utilizó solo una imagen para la galería y todas las demás para consulta, posteriormente se fue aumentando la cantidad de imágenes para la galería y de igual forma se consideraron todas las restantes para el conjunto de prueba, esta información

se resume en la Tabla VI. En el caso particular de la base de datos FYO, la cual cuenta con solo una imagen desconocida para el modelo, se utiliza la imagen de entrenamiento para la galería y la restante para la prueba. Además, para las imágenes de la galería se evaluó el uso o no del proceso de aumentación de muestras.

TABLA VI: Especificaciones de las particiones realizadas en las pruebas con el clasificador KNN. La cantidad de muestras considera las imágenes originales sin aumentación para cada individuo.

Base de Datos	Cantidad de muestras Galería:Prueba
Tongji	1:3; 2:2 y 3:1
ITI	1:2 y 2:1
CASIA	1:1
VERA	1:1
PUT	1:2 y 2:1
PolyU	1:4; 2:3 y 3:2
FYO	1:1

Con respecto a las pruebas del clasificador KNN se utilizó el método *KNeighborsClassifier* de la biblioteca *sklearn* del lenguaje de programación Python. Para la ejecución de los experimentos, se utilizó el modelo con mejores resultados de los entrenamientos, al cual se le eliminaron las 5 últimas capas (correspondientes al clasificador), utilizando la salida de la última capa convolucional como descriptor de características. En cada modelo los descriptores de características están conformados por vectores *float* de longitud 1024 en SingleNet y 256 en ResNet. A partir de estas características se implementó el KNN teniendo en cuenta diferentes valores de $k = [1, 3, 5]$ y usando la distancia euclidiana como métrica de similitud entre las codificaciones de características de las muestras en la galería y las de prueba.

3) *Métricas de rendimiento:* Las métricas consideradas para evaluar el entrenamiento de los modelos propuestos corresponden a las tasas de exactitud de la clasificación (*accuracy*) y la pérdida (loss) obtenidas a lo largo de las épocas de entrenamiento. Por otra parte, respecto a las pruebas de reconocimiento se identifican las mejores tasas de exactitud obtenidas por los modelos y el clasificador KNN.

IV. RESULTADOS EXPERIMENTALES

En esta sección se presentan los resultados obtenidos a partir de la ejecución de los experimentos mencionados en la Sección III-C. En los experimentos, se utilizaron dos GPU NVIDIA GeForce GTX 1080Ti basadas en la arquitectura Pascal sobre un servidor dedicado con sistema operativo Debian GNU/Linux 10 (buster) (kernel 4.19.0-17-amd64 x86_64), compuesto por dos CPU Intel Xeon Gold 6140, con 128 GB de memoria RAM. Los modelos propuestos fueron desarrollados en el lenguaje de programación Python (v.3.7.3), haciendo uso de las bibliotecas de OpenCV (v.4.1) Tensorflow (v.1.15.0), Keras (v.2.3.1) y Scikit-learn (v.0.24).

A. Entrenamiento de los modelos CNN

El entrenamiento de los modelos CNN propuestos SingleNet y ResNet, se realizó sobre cada una de las siete bases de datos

de estudio, considerando las imágenes originales y aumentadas con las variantes de preprocesamiento obtenidas Clahe, Normal y Sharpen. Cabe señalar que las manos izquierdas y derechas de los individuos se consideraron como sujetos diferentes para todas las bases de datos, por lo cual, el número de clases aumentó al doble.

En cuanto al entrenamiento de las bases de datos sin aumentación, la cantidad de muestras para validación se estableció en 20%, a diferencia de las imágenes con aumentación en donde se utilizó un 10% como conjunto de validación. Esto se debió a que al separar las muestras no fue posible conservar las mismas proporciones de clases para los conjuntos de entrenamiento y validación.

Para la mayoría de las bases de datos el tamaño de la capa de entrada fue de 128×128 píxeles (como se muestra en las tablas II y III). Sin embargo, para el caso particular de la base de datos Tongji fue necesario establecerla en 64×64 píxeles, debido a que la cantidad de memoria del servidor no fue suficiente para almacenar la red y los datos de entrenamiento en su totalidad.

Con relación a los hiperparámetros de entrenamiento se establecieron los siguientes:

- Épocas de entrenamiento: 100 con un tamaño de lote $batch_size = 32$, más 25 épocas adicionales en las que se usaron todos los datos de entrenamiento disponibles sin considerar conjunto de validación, adicionalmente se configuró una parada temprana (*EarlyStopping*) para evitar el sobreentrenamiento;
- Tasa de abandono: (*dropout*): 65% para SingleNet y 45% para ResNet;
- Tasa de aprendizaje inicial: 0.01, se configuró una función de reducción de la tasa de aprendizaje (*ReduceLRon-Plateau*) con un factor igual a 0.2 para reducir la tasa de aprendizaje cuando la función de pérdida se estanque;
- Función de optimización: *SGD* (con $momentum = 0.9$ y $weight\ decay = 1 \times 10^{-5}$) para SingleNet y *Adam* para ResNet;
- Función de pérdida: *categorical_crossentropy* para ambos modelos.

B. Resultados para el modelo SingleNet

Respecto a los resultados de entrenamiento del modelo SingleNet sobre los datos sin aumentación se identificaron los siguientes hallazgos: (1) los modelos obtenidos para las bases de datos CASIA y IITI presentaron sobreajuste *overfitting*; (2) para las bases de datos PolyU, Tongji y FYO se obtuvieron buenos resultados de aprendizaje con altas tasas de precisión y bajas tasas de pérdida; y (3) se observó que en las bases de datos PUT y VERA se logran buenos resultados gracias al escalonamiento de la tasa de aprendizaje. Por otro lado, el comportamiento del entrenamiento considerando el proceso de aumentación de datos mostró buenos resultados para todas las bases de datos, identificando un excelente aprendizaje con altas tasas de precisión y bajas tasas de pérdida, tanto para el conjunto de entrenamiento y validación. El comportamiento

de entrenamiento para el modelo SingleNet se presenta en la Tabla VII.

Por otra parte, los resultados obtenidos por el clasificador del modelo SingleNet para el conjunto de datos de prueba se presentan en la Tabla VIII. A partir de estos es posible identificar las siguientes características: (1) las mejores tasas de acierto fueron obtenidas en su totalidad por el entrenamiento realizado con aumentación de muestras, observándose las diferencias más significativas en las bases de datos PUT, CASIA y IITI, por ese orden; (2) la mejor variante de preprocesamiento para el entrenamiento sin aumentación está representada Sharpen; mientras que para el entrenamiento con aumentación, la mejor variante fue Clahe mayoritariamente; y (3) el mejor rendimiento se obtuvo en la base de datos PolyU con una tasa de acierto del 100% por medio de la variante Clahe y la aumentación de muestras. En general, los resultados obtenidos por este modelo son comparables al estado del arte, donde se utilizan modelos CNN preentrenados y de mayor profundidad.

C. Resultados para el modelo ResNet

En relación al comportamiento del entrenamiento del modelo ResNet usando datos sin aumentación de muestras se pudo identificar lo siguiente: (1) los modelos obtenidos para las bases de datos CASIA, IITI, PUT, PolyU y FYO presentaron sobreajuste; (2) en las bases de datos VERA y Tongji se obtuvieron buenos resultados y se logra identificar una excelente capacidad de generalización; y (3) nuevamente el escalonamiento de la tasa de aprendizaje fue fundamental en la optimización del modelo. En cuanto al entrenamiento con aumentación de datos, el modelo ResNet obtuvo buenos resultados para todas las bases de datos identificando un excelente aprendizaje con altas tasas de precisión y bajas tasas de pérdida. El comportamiento de entrenamiento para el modelo ResNet se presenta en la Tabla IX.

La Tabla X resume los resultados obtenidos por el clasificador del modelo ResNet para el conjunto de datos de prueba, de los cuales es posible identificar las siguientes características: (1) las mejores tasas de acierto fueron obtenidas en su totalidad por el entrenamiento realizado con aumentación de imágenes, donde las diferencias más marcadas estuvieron dadas en los conjuntos de datos PUT, FYO, CASIA, IITI y PolyU, en ese orden, lo que es natural dado que un modelo más complejo requiere de más muestras para su entrenamiento; (2) la mejor variante de preprocesamiento identificada para las pruebas es Clahe, tanto usando aumentación como sin ésta; y (3) en la base de datos Tongji se obtuvo el mejor desempeño por medio del entrenamiento con aumentación de imágenes y la variante Clahe con un 99.96% de tasa de acierto.

Al comparar los resultados de ambos modelos evaluados, se puede comprobar que el modelo SingleNet obtuvo mejores tasas de acierto para cinco de las siete bases de datos estudiadas (FYO, IITI, PolyU, PUT, VERA), con resultados superiores al 99% en todas; mientras que ResNet fue mejor en las dos restantes (CASIA, Tongji). Este comportamiento se atribuye a la complejidad de estas últimas bases de datos, en

TABLA VII: Resultados del entrenamiento del modelo SingleNet para cada base de datos estudiada.

Base de datos	Aumentación de muestras	Mejor variante	Épocas	Tiempo	Entrenamiento		Validación	
					Precisión	Pérdida	Precisión	Pérdida
CASIA	No	normal	68	44s	99.17	0,0698	78.75	1,0394
	Sí	clahe	42	30m	99.99	0,0191	100	0,0186
FYO	No	normal	72	2m 33s	99.22	0,0424	100	0,0176
	Sí	normal	32	4h 37m	99.99	0,0195	100	0,0192
IITI	No	sharpen	109	1m 24s	99.80	0,053	88.29	0,5827
	Sí	clahe	41	41m	99.99	0,0176	100	0,017
PolyU	No	clahe	54	2m 9s	99.43	0,0412	100	0,0172
	Sí	clahe	32	2h 8m	99.94	0,0219	100	0,0201
PUT	No	normal	88	1m 0s	98.44	0,0699	100	0,0186
	Sí	clahe	25	20m	99.99	0,0173	100	0,0169
Tongji	No	clahe	54	5m 18s	99.74	0,0263	100	0,0181
	Sí	sharpen	124	8h 36m	100	0,0113	100	0,0112
VERA	No	sharpen	57	1m 12s	99.66	0,0379	100	0,017
	Sí	normal	68	1h 45m	99.99	0,0135	100	0,0133

TABLA VIII: Resultados obtenidos por el clasificador del modelo SingleNet para el conjunto de datos de prueba.

Base de Datos	Mejor Variante	Sin Aumentación	Mejor Variante	Con Aumentación
CASIA	normal	78.50%	clahe	97.75%
FYO	sharpen	97.06%	clahe	99.97%
IITI	sharpen	86.40%	clahe	98.38%
PolyU	clahe	96.68%	clahe	100%
PUT	normal	79.00%	clahe	99.00%
Tongji	clahe	99.62%	sharpen	99.87%
VERA	sharpen	96.36%	normal	99.77%

las que se presenta mayor variabilidad inter-clase por lo que es más efectivo un modelo más robusto como ResNet.

D. Resultados con clasificador KNN

Los resultados obtenidos por el clasificador KNN para ambos modelos se presentan en la Tabla XI, a partir de los cuales se puede notar los siguientes elementos: (1) los mejores resultados fueron obtenidos mayoritariamente para el modelo SingleNet, lo que demuestra nuevamente su alto poder de generalización; (2) las mejores tasas de acierto fueron obtenidas considerando la aumentación de imágenes en las muestras de la galería; (3) la mejor variante identificada para la clasificación sin aumentación y con aumentación es mayoritariamente Clahe; (4) en las bases de datos PUT, Tongji y PolyU se obtuvo el mejor rendimiento, alcanzando el 100% cuando se utilizaron una cantidad superior de imágenes de galería respecto a las de prueba; (5) el número de K que logró obtener las mejores tasas de acierto para la clasificación fue $k=1$; y (6) al usar el modelo CNN como codificador de características y un clasificador KNN, los resultados de reconocimiento mejoraron en todas las bases de datos excepto en la VERA donde fue solo ligeramente inferior al clasificador SingleNet. Este último punto es muy significativo porque demuestra la efectividad de los modelos propuestos como codificadores de características y su alta efectividad de generalización, lo que posibilita el uso de clasificadores de similitud como el KNN para extender el modelo de reconocimiento sin necesidad de un reentrenamiento. No obstante, esto deberá ser comprobado con más detalle en trabajos futuros mediante la evaluación de pruebas cruzadas entre bases de datos.

E. Comparación contra el estado del arte

La Tabla XII muestra una comparación de los mejores resultados del estado del arte obtenidos por modelos CNN y los mejores resultados de los modelos propuestos, para cada una de las bases de datos estudiadas. A partir de estos, se puede observar que para las bases de datos FYO, IITI y PUT el modelo SingleNet junto al clasificador KNN superan los resultados del estado del arte. Por otra parte, para las bases de datos Tongji y PolyU se logra igualar los métodos de referencia, aunque en nuestro caso con arquitecturas de menor complejidad. Con respecto a las bases de datos CASIA y VERA, aunque no se obtiene el mejor resultado del estado del arte no se aprecian diferencias significativas, igualmente con una red de más baja profundidad.

En general, se logran obtener muy buenos resultados, con un 100% de tasa de reconocimiento en cuatro de los siete conjuntos de datos evaluados (FYO, PolyU, PUT y Tongji) y solo una por debajo del 99% (IITI con 98.92%). Se debe destacar que los rendimientos de clasificación obtenidos son cercanos o superiores al estado del arte, lo que se alcanza con una arquitectura de poca profundidad y un clasificador KNN. De esta manera, el modelo propuesto es un método de poca complejidad, a diferencia de lo propuesto por otros autores de la literatura que utilizan redes y modelos de mayor profundidad o que se fusionan con métodos de extracción de características (*hand-crafted*).

V. CONCLUSIONES

En este artículo, se evaluaron dos arquitecturas de redes neuronales convolucionales con diferentes características, para la identificación de personas mediante venas de la palma. Se utilizó una arquitectura de pocas capas convolucionales denominada SingleNet y una arquitectura ResNet32. El método propuesto se evaluó en siete bases de datos públicas de venas de la palma representativas del estado del arte, las cuales fueron preprocesadas utilizando un procedimiento propio de aumentación de muestras y tres técnicas de mejora de imagen. Ambas modelos implementados fueron entrenados de extremo a extremo en cada base de datos. Las pruebas de rendimiento para el reconocimiento de personas se realizaron con el

TABLA IX: Resultados del entrenamiento del modelo ResNet para cada base de datos estudiada.

Base de datos	Aumentación de muestras	Mejor variante	Épocas	Tiempo	Entrenamiento		Validación	
					Precisión	Pérdida	Precisión	Pérdida
CASIA	No	clahe	47	2m 41s	100	0,1847	81.25	1,0886
	Sí	clahe	83	6h 3m	100	0,0002	100	0,0002
FYO	No	sharpen	64	13m 38s	96.34	0,1784	59.67	0.1408
	Sí	normal	52	15h 47m	100	0,0001	100	0,0001
IITI	No	clahe	86	6m 29s	100	0,1845	78.38	1,1421
	Sí	clahe	41	4h 4m	99.98	0,003	100	0,0033
PolyU	No	normal	86	19m 16s	97.34	0,1569	93.86	0,3786
	Sí	clahe	71	23h 0m	100	0,0001	100	0,0001
PUT	No	clahe	34	2m 22s	97.33	0,3445	67.22	2,0544
	Sí	clahe	70	5h 46m	100	0,0001	100	0,0001
Tongji	No	clahe	41	26m 22s	100	0,0303	100	0,0303
	Sí	clahe	76	1d 7h 28m	100	0,0001	100	0,0001
VERA	No	normal	77	9m 20s	100	0,0689	100	0,0694
	Sí	sharpen	70	11h 9m	100	0,0001	100	0,0001

TABLA X: Resultados obtenidos por el clasificador del modelo ResNet para el conjunto de datos de prueba.

Base de Datos	Mejor Variante	Sin Aumentación	Mejor Variante	Con Aumentación
CASIA	clahe	80.25%	clahe	98.75%
FYO	sharpen	78.78%	clahe	97.12%
IITI	clahe	81.17%	clahe	97.93%
PolyU	normal	88.40%	clahe	99.92%
PUT	clahe	74.67%	clahe	96.33%
Tongji	clahe	97.96%	clahe	99.96%
VERA	normal	96.59%	sharpen	99.32%

clasificador CNN y además con un clasificador KNN usando el modelo entrenado como codificador de características.

Los resultados experimentales evidenciaron que utilizar el clasificador KNN para el modelo SingleNet con aumentación de datos logra igualar e incluso superar el rendimiento de los trabajos reportados en la literatura. Para las bases de datos FYO, IITI y PUT se superaron los resultados de referencia, en la Tongji y PolyU se iguala su efectividad, mientras que para la VERA y CASIA se obtienen tasas de reconocimiento ligeramente inferiores. No obstante, se destacan las ventajas del modelo SingleNet, el cual disminuye el consumo de los recursos de hardware y el tiempo de entrenamiento en todas las bases de datos. A partir de la evaluación se comprobó que el enfoque propuesto consigue excelentes resultados de generalización.

Finalmente, como trabajos futuros con vistas a extender las pruebas realizadas se propone: (1) realizar nuevas pruebas de entrenamiento disminuyendo el tamaño de entrada de las imágenes para analizar si existen diferencias al usar imágenes de menor resolución; (2) extender el estudio del modelo propuesto para el reconocimiento de rasgos biométricos suaves como el sexo, la edad, el color de la piel, entre otras, los cuales tienen influencia en la visualización de los patrones de vena; y (3) desarrollar un dispositivo móvil sin contacto para la identificación de personas por medio de imágenes de venas de la palma, basado en el método propuesto y utilizando arquitecturas embebidas de bajo consumo como NVIDIA Jetson.

REFERENCIAS

- [1] W. Wu, S. Elliott, S. Lin, S. Sun, and Y. Tang, "Review of palm vein recognition," *IET Biometrics*, vol. 9, no. 1, pp. 1–10, 2019.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [4] L. Zhang, Z. Cheng, Y. Shen, and D. Wang, "Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset," *Symmetry*, vol. 10, no. 4, p. 78, 2018.
- [5] M. Wulandari, D. Gunawan *et al.*, "On the Performance of Pretrained CNN Aimed at Palm Vein Recognition Application," in *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 2019, pp. 1–6.
- [6] D. Thapar, G. Jaswal, A. Nigam, and V. Kanhangad, "PVSNet: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features," in *2019 IEEE 5th international conference on identity, security, and behavior analysis (ISBA)*. IEEE, 2019, pp. 1–8.
- [7] L. Lu, R. Fu, L. Yuan, W. Chen, Y. Liu *et al.*, "Palm vein recognition based on end-to-end convolutional neural network," *Journal of Southern Medical University*, vol. 39, no. 2, pp. 207–214, 2019.
- [8] S. Lefkovits, L. Lefkovits, and L. Szilágyi, "Applications of different CNN architectures for palm vein identification," in *International Conference on Modeling Decisions for Artificial Intelligence*. Springer, 2019, pp. 295–306.
- [9] Y. Lisha, L. Mengying, L. Yaqin, Y. Feng, and H. Jing, "Palm vein classification based on deep neural network and random forest," *Laser Optoelectron. Prog.*, vol. 56, no. 10, 2019.
- [10] R. S. Kuzu, E. Maiorana, and P. Campisi, "Vein-Based Biometric Verification Using Densely-Connected Convolutional Autoencoder," *IEEE Signal Processing Letters*, vol. 27, pp. 1869–1873, 2020.
- [11] M. I. Obayya, M. El-Ghandour, and F. Alrowais, "Contactless Palm Vein Authentication Using Deep Learning With Bayesian Optimization," *IEEE Access*, vol. 9, pp. 1940–1957, 2020.
- [12] F. O. Babalola, Y. Bitirim, and Ö. Toygar, "Palm vein recognition through fusion of texture-based and cnn-based methods," *Signal, Image and Video Processing*, vol. 15, no. 3, pp. 459–466, 2021.
- [13] W. Jia, J. Gao, W. Xia, Y. Zhao, H. Min, and J.-T. Lu, "A performance evaluation of classic convolutional neural networks for 2D and 3D palmprint and palm vein recognition," *International Journal of Automation and Computing*, vol. 18, no. 1, pp. 18–44, 2021.
- [14] S. Bhilare, G. Jaswal, V. Kanhangad, and A. Nigam, "Single-sensor hand-vein multimodal biometric recognition using multiscale deep pyramidal approach," *Machine Vision and Applications*, vol. 29, no. 8, pp. 1269–1286, 2018.
- [15] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 281–284.

TABLA XI: Resultados obtenidos por el clasificador KNN. Nomenclatura: G=cantidad de imágenes para galería; P=cantidad de imágenes de prueba; y K=número de elementos más cercanos.

Base de Datos	G	P	Mejor Modelo y Variante	K	Sin Aumentación	Mejor Modelo y Variante	K	Con Aumentación
CASIA	1	1	ResNet Clahe	1	86.50%	SingleNet Clahe	1	99.00%
FYO	1	1	SingleNet Normal	1	97.78%	SingleNet Normal	1	100%
IITI	2	1	SingleNet Sharpen	1	90.00%	SingleNet Clahe	1,3	98.92%
PolyU	3	2	SingleNet Clahe	1	99.60%	SingleNet Clahe	1	100%
PUT	2	1	ResNet Clahe	1	87.00%	SingleNet Clahe	1,3,5	100%
Tongji	3	1	SingleNet Clahe	1	100%	SingleNet/Resnet Clahe	3,5	100%
VERA	1	1	SingleNet Sharpen	1	96.36%	SingleNet Sharpen	1,3,5	99.55%

TABLA XII: Comparación de los mejores resultados de los modelos propuestos y los modelos basados en CNN encontrados en la literatura.

Base de Datos	Mejor Modelo del Estado del Arte	Resultado del Estado del Arte	Mejor Modelo Obtenido	Resultado Obtenido
CASIA	AlexNet [12]	99.83%	Clasificador KNN modelo SingleNet (clahé)	99.00%
FYO	AlexNet [12]	98.91%	Clasificador KNN modelo SingleNet (normal)	100%
IITI	U-net encoder-decoder + conv6-fc triplet Siamese [6]	97.47%	Clasificador KNN modelo SingleNet (clahé)	98.92%
PolyU	AlexNet [9] MobileNet v3 y EfficientNet [13]	100%	Clasificador SingleNet, ResNet y KNN para ambos modelos (clahé)	100%
PUT	ResNet-50 [8]	99.83%	Clasificador KNN modelo SingleNet (clahé)	100%
Tongji	AlexNet [12] Inception ResNet v1 [4]	100%	Clasificador KNN modelo SingleNet (clahé)	100%
VERA	AlexNet [12]	100%	Clasificador SingleNet (normal)	99.77%

- [16] P. Tome and S. Marcel, "Palm vein database and experimental framework for reproducible research," in *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2015, pp. 1–7.
- [17] R. Kabaciński and M. Kowalski, "Vein pattern database and benchmark results," *Electronics Letters*, vol. 47, no. 20, pp. 1127–1128, 2011.
- [18] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," *IEEE transactions on instrumentation and measurement*, vol. 59, no. 2, pp. 480–490, 2009.
- [19] Ö. Toygar, F. O. Babalola, and Y. Bitirim, "FYO: a novel multimodal vein database with palmar, dorsal and wrist biometrics," *IEEE Access*, vol. 8, pp. 82461–82470, 2020.
- [20] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation," *Expert Systems with Applications*, vol. 95, pp. 43–56, 2018.
- [21] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Chapter VIII.5, Graphics Gems IV*, P. S. Heckbert, Ed. Academic Press Professional, Inc., 1994, pp. 474–485.
- [22] J. Chaki and N. Dey, *A beginner's guide to image preprocessing techniques*. CRC Press, 2018.