

Open Set Recognition of Timber Species Using Deep Learning for Embedded Systems

M. Apolinario, D. Urcia, and S. Huaman, *Member, IEEE*

Abstract—Reliable and rapid identification of timber species is a very relevant issue for many countries in South America and especially for Peru, which is the second country with the largest extent of tropical forest, and that is because this issue is a necessity in order to develop an effective management of the forest resources, such as inspection and control of the timber commerce. Since current methods of identification are based on a closed set recognition approach, they are not reliable enough to be used in a practical application because scenarios of identification of timber species are by nature an open set recognition problem. For that reason, in this work we propose a convolutional neural network that has two main characteristics, being able to run in a real-time embedded system and being able to handle the open set recognition problem, that is, this model can discriminate between known and unknown species. In order to evaluate it, tests are performed in two timber species datasets and some experiments are developed in the embedded system Raspberry Pi3B+ to measure energy consumption. The results present high metrics, which means that it manages to discriminate the unknown species with accuracy and F1 score above 91% for two sets of images used. In addition to this, our proposed model obtain lower maximum power value (10-12%) and computational resource usage (5-13%) than a classical convolutional model and MobileNetsV2 measured on the Raspberry Pi3B+.

Index Terms—Timber species, Embedded system, Convolutional neural network, Open set recognition.

I. INTRODUCCIÓN

El sector maderero es importante económicamente para el mercado internacional, así como también por su impacto sobre el medio ambiente [1]. En Sudamérica, este sector cobra especial relevancia pues depende del bosque con mayor biodiversidad del mundo, la Amazonia, donde existe una amplia variedad de flora que en su mayoría provee recursos maderables. Debido a esta diversidad, la gestión efectiva de los recursos forestales se presenta como un desafío para las instituciones relacionadas con el sector. Un ejemplo de esto lo tenemos en Perú, el segundo país en Latinoamérica con la mayor extensión de bosques tropicales teniendo cerca de 73.3 millones de hectáreas cubiertas por bosques naturales [2], que cuenta con más de 20,000 especies de plantas de las cuales 5,509 son endémicas, y entre estas aproximadamente 450 especies son las que pueden proveer recursos maderables, y en este contexto, el elevado número de especies y recursos hace complicado que las instituciones gubernamentales puedan realizar un control y fiscalización efectivos sobre el comercio

de maderas, pues se necesita personal calificado que pueda realizar la identificación correcta de las especies maderables a fin de evitar el comercio ilegal de maderas o explotación de especies en peligro [3].

La problemática radica en la necesidad de realizar una identificación a la intemperie de especies maderables de forma rápida y confiable, de tal manera que incluso personal sin una amplia especialización pueda realizar tareas de fiscalización preliminares. En vista de ello, muchas propuestas en el área de visión computacional para la clasificación de especies han sido planteadas, en su mayoría usando métodos basados en máquinas de aprendizaje (*machine learning*), realizando una extracción de características y clasificación basado en SVM o *Multilayer Perceptron* [4]–[7], otras propuestas hacen uso de redes neuronales convolucionales (CNN) [8]–[10]; si bien todas han logrado resultados sobresalientes en la clasificación de especies maderables con resultados por encima del 97% de exactitud, aún tienen dos principales restricciones que limitan su uso en aplicaciones de campo o entornos no controlados. Tal vez la restricción más evidente, relacionada con la rapidez de identificación, es el elevado costo computacional que se requiere para ejecutar estos métodos, lo cual limita su aplicación en tiempo real y su implementación sobre sistemas portátiles, como los teléfonos inteligentes (*smartphones*), tabletas u otros sistemas embebidos. Estos dos aspectos son esenciales para realizar una identificación en campo donde por lo general las conexiones a Internet son limitadas, dificultando soluciones basadas en computación en nube (*cloud computing*). La segunda restricción menos evidente, pero de mayor importancia, está relacionada con la confiabilidad, dado que las propuestas actuales están pensadas para problemas de clasificación en conjuntos cerrados (*closed set recognition*) donde *a priori* se puede conocer el total de clases existentes. Sin embargo, el problema de identificación de especies para temas de fiscalización y control es por naturaleza un problema que implica reconocimiento en conjuntos de datos abiertos (*open set recognition*) [11], [12], pues si el método ha sido desarrollado para identificar una cantidad determinada de especies, entonces, cuando se presente una especie desconocida este debería reconocer que se ha presentado información de una especie nueva o diferente.

Debido a las dos restricciones mencionadas para los métodos actuales de identificación de especies maderables, en este trabajo se propone un método basado en una CNN de bajo costo computacional, para permitir que sea ejecutada sobre un sistema embebido en tiempo real, y la habilidad para poder reconocer entre información conocida y desconocida, elevando de esta forma su confiabilidad y permitiendo que sea aplicada

M. P. E. Apolinario, Instituto Nacional de Investigación y Capacitación de Telecomunicaciones (INICTEL-UNI), Universidad Nacional de Ingeniería, Lima, Perú. e-mail: mapolinario@uni.pe.

D. A. Urcia, INICTEL-UNI, Lima, Perú. e-mail: durcia@inictel-uni.edu.pe

S. G. Huaman, INICTEL-UNI, Lima, Perú. e-mail: shuaman@inictel-uni.edu.pe.

en la práctica. Así, la reducción del costo computacional se logra mediante el uso de bloques residuales invertidos, como se describen en [13], y el uso de capas convoluciones en profundidad (*depthwise*), mientras que el problema de reconocimiento en conjuntos de datos abiertos se aborda bajo un enfoque de agrupamientos (clusterización) en las proyecciones de la penúltima capa de la CNN, empleando algunos conceptos presentados en [11], [14]. Finalmente, el modelo propuesto se evalúa bajo las perspectivas de reconocimiento de conjunto abierto (*open set*) y conjunto cerrado (*closed set*) para contrastar su comportamiento, además se realiza una evaluación de la ejecución del modelo sobre un sistema embebido para mostrar el uso de recursos y el consumo energético.

El resto del documento está organizado de la siguiente manera. La Sección II describe la metodología propuesta en detalle, analizando primero los procesos de adquisición de imágenes y las características de las mismas, para luego detallar los problemas existentes en las aplicaciones prácticas de métodos de reconocimiento de imágenes y, finalmente, presentando la arquitectura del modelo de red neuronal convolucional propuesto. Luego, en la Sección III se presentan los resultados obtenidos de la evaluación del modelo propuesto bajo los enfoques de *closed set* y *open set recognition*, además de un análisis del comportamiento del modelo propuesto cuando se ejecuta en un sistema embebido de limitados recursos computacionales. Finalmente, en la Sección IV se exponen las conclusiones del presente trabajo.

II. MATERIALES Y MÉTODOS

A. Especies Maderables

Las especies maderables para el presente estudio fueron escogidas teniendo como base la investigación realizada por SERFOR en su anuario de recursos forestales 2016, el cual describe y cuantifica el aprovechamiento de los recursos forestales maderables y no maderables, de los servicios asociados al sector, problemas y su potencial [15], además de la alta demanda que tienen dichas especies como parte de la actividad primaria en cada uno de los departamentos de Perú [2]. Entre los nombres comunes de algunas especies maderables se tienen: aguano masha, catahua amarilla, cedro, copaiba, lupuna blanca, tornillo y otros de cuyos productos transformados se abastecen los mercados, ver Tabla I. Entre las especies maderables con mayor volumen bruto de extracción están la lupuna, tornillo y cachimbo, ello motivó a que se estudiaran con mayor énfasis debido a su repercusión en la extracción de recursos forestales y alta demanda en el mercado local e internacional [1]. Además, la elección de las especies maderables restantes cuyos volúmenes de madera rolliza son inferiores a las mencionadas, responde a que poseen características macroscópicas que las diferencian principalmente en color, textura y opacidad, además del tamaño y forma de pequeños orificios microscópicos llamados miembros de vaso y la red celular que constituyen los parénquimas axiales, vistos en un corte transversal [16], [17].

TABLA I
ESPECIES MADERABLES CONSIDERADAS PARA LA IDENTIFICACIÓN

ID	Especie	Nombre común	Nro. imágenes
1	Machaerium inundatum	Aguano masha	24
2	Ormosia coccinea	Huayruro	14
3	Guazuma crinitia	Bolaina	29
4	Cariniana domestica	Cachimbo	31
5	Swietenia macrophylla	Caoba	24
6	Calycophyllum spruceanum	Capirona	26
7	Hura crepitans	Catahua amarilla	23
8	Cedrela odorata	Cedro	29
9	Cedrela fissilis	Cedro rojo	20
10	Copaifera officinalis	Copaiba	27
11	Amburana cearensis	Ishipingo	16
12	Calophyllum brasiliense	Lagarto caspi	10
13	Junglans neotropica	Nogal	24
14	Aspidosperma macrocarpon	Pumaquiro	13
15	Dipteryx micrantha	Shihuahuaco	15
16	Cedrelinga cateniformes	Tornillo	30

B. Conjunto de Imágenes y su Preprocesamiento

1) *Procedimiento de adquisición de imágenes:* las imágenes usadas en este trabajo pertenecen a dos conjuntos. El primero fue construido, y agrupa a las 16 especies maderables mostradas en la Tabla I, subdivididas en muestras de 10, 40 y 60 aumentos, las cuales fueron obtenidas con apoyo del Laboratorio de Anatomía e Identificación de Maderas de la Universidad Nacional Agraria de La Molina en Lima, Perú, al cual denominaremos como Conjunto de Imágenes de Maderas Peruanas (CIMP). Y el segundo, corresponde a Forest Species Database – Macroscopic (FSD-M) [18].

El protocolo adoptado para la adquisición de las muestras se compone de cuatro pasos. En el primer paso, se tomaron pedazos de la parte del tronco de cada una de las especies maderables seleccionadas. Luego, la madera es cortada de forma transversal al eje del tronco con la finalidad de apreciar las características anatómicas más distintivas de cada especie. En el tercer paso, se limpia la zona a fotografiar y se humedece ligeramente con un poco de agua para retirar impurezas. Finalmente, las imágenes son adquiridas usando una cámara compacta digital de la marca *Canon Powershot SX50* montada sobre el soporte de un microscopio con dos lentes: el primero, con resolución espacial de 145 píxeles por milímetro (R1) y el segundo, con 589 píxeles por milímetro (R2), ubicados a una distancia tal que el enfoque sea óptimo para la visualización de las características. Las imágenes resultantes son guardadas

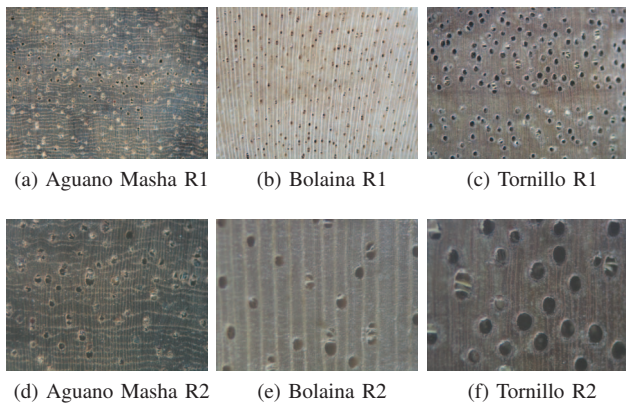


Fig. 1. Muestras a R1 y R2 adquiridas con una cámara Canon Powershot SX50.

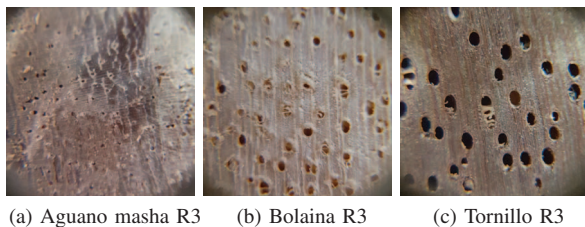


Fig. 2. Muestras a R3 adquiridas con la cámara de un smartphone Samsung Galaxy S6 Edge.

en formato JPG con una resolución en píxeles de 2592 x 1944. Cabe señalar que, para construir el tercer grupo de imágenes, correspondiente a una resolución espacial 243 píxeles por milímetro (R3), se usó un lente externo de fácil montaje sobre la cámara de un smartphone de la marca *Samsung* cuyo modelo es el *Galaxy S6 Edge* en lugar de la cámara compacta digital siguiendo el mismo protocolo, en este caso las imágenes fueron guardadas en el mismo formato a un tamaño de 5312 x 2988 píxeles.

Las 16 especies maderables escogidas fueron seccionadas en paralelepípedos, con tamaños variables para una sencilla visualización de sus principales características anatómicas. Las imágenes del corte transversal contienen mayor información de las especies maderables, por esta razón fueron usadas para construir el conjunto de imágenes. En las Fig. 1 y Fig. 2 se pueden observar algunas muestras de las especies maderables en las diferentes resoluciones espaciales usadas para la construcción de CIMP.

2) *Preprocesamiento de imágenes*: si bien esta etapa no forma parte del modelo de reconocimiento de las especies maderables, su aplicación al conjunto de imágenes de la cámara Canon Powershot SX50 fue necesaria. Por ejemplo, en la Fig. 3a, se observa una muestra adquirida con esa cámara a R2 que presenta líneas horizontales, apenas perceptibles, superpuestas a las características de la madera. Estas líneas son producidas por la cámara y no se pueden controlar mediante su configuración. A pesar de eso, las imágenes aún son útiles y pueden ser filtradas para atenuar ese patrón lineal conocido como ruido periódico; de esta manera, se logra un conjunto de imágenes similares a otras cámaras. Existen algunas her-

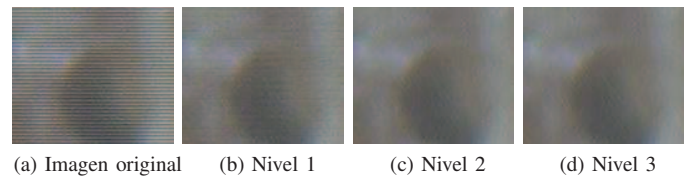


Fig. 3. Una de las muestras de capirona a R2 adquiridas con una cámara Canon Powershot SX50 y sus correspondientes niveles de TDW-2D.

ramientas que permiten atenuar el ruido en las imágenes; una de ellas, caracterizada por su alta eficiencia, es la que usa la Transformada Discreta Wavelet 2D (TDW-2D) que descompone una imagen, en sub bandas de aproximación y detalle en la etapa de análisis, donde usa funciones base conocidas como *wavelet* equivalentes a oscilaciones breves, realizando operaciones de producto escalar (convolución) con porciones de una imagen conocida utilizando el modelo matricial de color RGB, el mismo del conjunto de imágenes adquiridas. En el procesamiento de imágenes, uno de los usos de la TDW-2D es la atenuación de ruido en las mismas, a través de la manipulación de los coeficientes de detalle, los cuales junto a los coeficientes de aproximación descomponen una imagen [19]. Hay diversas bases de *wavelets* en el software Matlab para la aplicación de la TDW-2D, en este caso se utilizó una *Symlet* denominada *Sym4* del tipo ortogonal, la cual tiene menor asimetría que otras funciones y fase casi lineal, que permite la atenuación del cambio brusco del patrón no deseado del conjunto de imágenes construido. Se evaluaron tres niveles de análisis y umbral *soft*, cuyos resultados se visualizan en la Fig. 3. Luego, al atenuar el coeficiente de detalle, donde se concentra el ruido periódico, y sintetizar la coeficientes restantes, se obtiene la imagen filtrada. Sin embargo, al filtrar una imagen puede eliminarse información. Para minimizar la pérdida de esta, se evalúa el grado de des-correlación entre la imagen filtrada y la diferencia entre la imagen original y la filtrada, que corresponde al ruido periódico. Los valores obtenidos de los coeficientes de correlación fueron 0.0673, 0.0095 y 0.0082 para las imágenes resultantes del nivel 1, nivel 2 y nivel 3, respectivamente. Esos valores tienden a cero, lo cual indica des-correlación, así mismo, el nivel 2 presenta un valor mucho menor al nivel 1 y próximo al nivel 3, sin embargo, el nivel 3 demanda mayor costo computacional. Por estas razones, se usa el nivel 2 de descomposición de la TDW-2D. Para finalizar, el procedimiento se aplicó a todas las imágenes adquiridas con la cámara Canon Powershot SX50 con resolución espacial R1 y R2. Posteriormente, se hizo el redimensionamiento de todas las imágenes a R1, con el objetivo de uniformizar el conjunto de imágenes para la CNN.

C. Reconocimiento de Especies Maderables

En esta sección se abarca el tema de reconocimiento de especies maderables orientado hacia aplicaciones prácticas que puedan ser llevadas a cabo en el campo, en escenarios donde se necesitan herramientas tecnológicas para la fiscalización y control del comercio de maderas, por ello se toman en consideración dos de las principales limitaciones que presentan

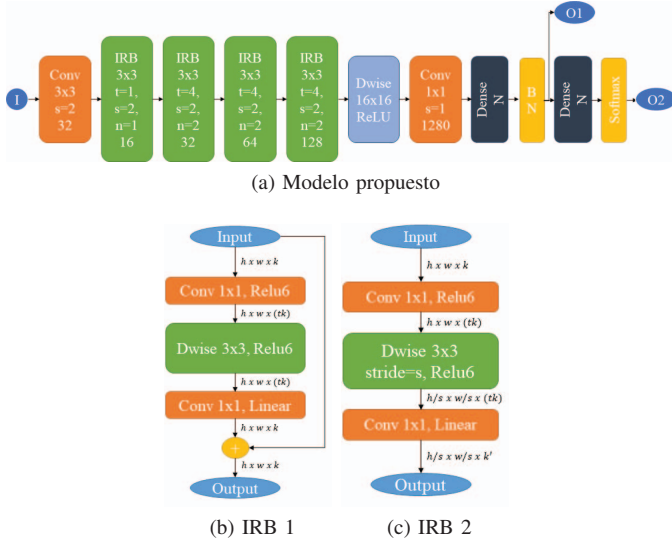


Fig. 4. (a) Modelo de la red neuronal propuesta, (b) Estructura de un bloque IRB con stride igual a 1 y (c) estructura de un bloque IRB con stride diferente de 1. Donde, h , w y k corresponde al alto, ancho y número de canales del mapa de características de entrada, respectivamente.

los modelos actuales: la rapidez y la confiabilidad. La primera está directamente relacionada con el costo computacional del modelo usado para el reconocimiento, por lo que en el caso de las CNN se busca optimizar las capas convoluciones. Mientras que la confiabilidad, en este escenario, se relaciona con la capacidad de un modelo para discriminar entre lo conocido y desconocido, lo cual se conoce como el problema de *open set* [11]. Con estas consideraciones, a continuación se describe la arquitectura de la red neuronal convolucional propuesta y los resultados de su entrenamiento.

1) *Arquitectura de la red neuronal convolucional*: la red neuronal propuesta está diseñada usando pocos parámetros, por ende requiere un reducido número de operaciones matemáticas en la etapa de inferencia, sin que ello implique una pérdida significativa de exactitud; adicionalmente dicha red neuronal también puede discriminar, con cierto grado de eficiencia, entre imágenes de entrada de clases conocidas y desconocidas.

Para que el modelo propuesto realice la menor cantidad de operaciones por cada inferencia, se diseñó usando capas convoluciones del tipo *depthwise*, las cuales requieren 10 veces menos operaciones respecto de las capas convoluciones convencionales, por lo que se optó por hacer uso de los bloques *Inverted Residual Bottlenecks* (IRB) para la etapa inicial de extracción de características, cuyas estructuras se muestran en la Fig. 4b y Fig. 4c [13]. Adicionalmente, para que el modelo propuesto tuviera una respuesta hacia las imágenes de clases desconocidas, se realizó un análisis de las proyecciones de las imágenes de entrada en la penúltima capa *fully-connected* siguiendo un enfoque de clusterización para discriminar entre lo conocido y desconocido, empleando conceptos como los propuestos en [14], [20].

En la Fig. 4a se muestra un diagrama de la arquitectura del modelo propuesto, como se puede observar, existen una entrada (I) y dos salidas (O1 y O2) cada una con su propia función de costo. En la entrada del modelo ingresan imágenes

de dimensión $512 \times 512 \times 3$ normalizadas en el rango de 0 a 1. En la primera salida tenemos un vector de dimensión N , correspondiente al número de clases conocidas en el CIMP, que representa la proyección de las imágenes de entrada en un espacio N -dimensional. Mientras, la segunda salida es un vector también de dimensión N , generado con la función *softmax*, que indica las probabilidades que la imagen de entrada pertenezca a una de las clases. Además, en el diagrama se puede observar que la primera capa, denominada como Conv, es un bloque convolucional convencional con 32 *kernels* de dimensión 3×3 con un *stride* (s) igual a dos con una salida lineal seguida por una capa interna de *Batch Normalization* (BN) con una función de activación ReLU en la salida. La labor de este primer bloque es extraer las características de bajo nivel de las imágenes a la vez que se reducen sus dimensiones espaciales, lo sigue un bloque IRB con 16 canales de salida con $s = 2$, un factor de expansión $t = 1$ y $n = 1$, lo cual indica que solo contiene un bloque IRB 2, seguido por tres bloques IRB con $t = 4$, $s = 2$, $n = 2$ (un bloque IRB 2 seguido de un bloque IRB 1) y número de canales de salidas de 32, 64 y 128, respectivamente; todos los bloques IRB tienen una función de activación lineal en la salida. La siguiente capa es una convolución del tipo *depthwise*, denominada como “Dwise”, con un kernel de dimensión 16×16 y función de activación lineal usando *valid padding*, es decir que reduce las dimensiones del tensor de entrada; a continuación se encuentra un bloque convolucional convencional con 1280 *kernels* de dimensión 1×1 con una salida lineal seguida por una capa BN con función de activación ReLU, en este punto el vector de salida tiene una dimensión de $1 \times 1 \times 1280$ por lo que se reforma en un vector de dimensión 1280 que ingresa a las capas *fully connected*.

La primera capa *fully connected*, denominada como Dense, tiene N neuronas con función de activación lineal cuya salida ingresa a una capa BN para obtener la primera salida de la red (O1). El objetivo es obtener un espacio vectorial de menor dimensión a partir del espacio de mayor dimensión formado por la información de las imágenes, luego los vectores de salida se agrupan utilizando el criterio de mínima distancia entre el centroide y los vectores que formaran el cluster, así mismo, se maximiza la distancia entre los centroides de los posibles clústeres de diferentes clases, como se describe en [20]. Por último, la siguiente capa Dense tiene también N neuronas y su propósito es clasificar a cada una de las imágenes en las N clases conocidas iniciales usando una función de activación *softmax* en su salida para determinar las probabilidades de cada clase.

2) *Entrenamiento de la red neuronal*: para entrenar el modelo propuesto se usaron dos funciones de costo, una para cada vector de salida, las cuales fueron ponderadas para calcular una función de costo global. Para la primera salida, se implementó la función de costo *ii_loss* en *tensorflow*, como fue propuesta en [20], ya que esta tiene la particularidad de minimizar la distancia entre elementos de una misma clase y maximizar la distancia entre elementos de clases diferentes, agrupando de esta forma las imágenes en clústeres, los cuales pueden ser usados para calcular un valor umbral a partir del cual puede determinarse si una nueva imagen

pertenece a una de las clases conocidas o, en su defecto, una clase desconocida. Luego, para la segunda salida, se usó la combinación de la función de costo *categorical cross entropy* (cce), la cual es comúnmente usada en problemas de clasificación multiclase pues compara la similitud entre dos distribuciones de probabilidad, una correspondiente a la etiqueta verdadera de la clase, codificada en *one hot*, y la otra proveniente de una capa *softmax*. Con estas funciones de costo el modelo propuesto buscará agrupar las imágenes en un espacio N-dimensional y luego clasificarlas. Finalmente, para el proceso de *backpropagation* se adicionan las dos funciones de costo con un peso de uno para *ii_loss* y un peso de 10 para *categorical cross entropy*, ya que esta combinación de pesos dio los mejores resultados para el entrenamiento.

Para comparar el desempeño obtenido por el modelo propuesto se implementan dos modelos adicionales, el primero con una red neuronal convolucional convencional semejante a las propuestas en [8]–[10] y el segundo con una versión completa de MobilenetV2 [13], a los que denominamos ModelC1 y ModelC2 respectivamente, en ambos casos se modifican las últimas capas para adecuarlo al número de clases del conjunto de imágenes (CIMP). Posterior al entrenamiento se evalúan los tres modelos bajo los enfoques de *closed set* y *open set recognition* como se detalla en la Sección III-A.

Todos los modelos antes mencionados fueron implementados en Python 3.6 usando Tensorflow en un servidor con un procesador Intel Xeon E5-2620 a 2.1 GHz y dos GPUs NVIDIA GeForce GTX 1080. El entrenamiento se realizó usando únicamente CIMP, el cual es descrito en la Sección II-A; posterior al entrenamiento se evalúan los modelos bajo la perspectiva *closed set* usando CIMP; y adicionalmente, para la evaluación en un escenario *open set*, se realiza una simulación, mediante la selección de 11 especies de las presentes en FSD-M, las cuales no están presentes en CIMP, de tal forma que las especies conocidas son las que pertenecen a CIMP y las otras 11 de FSD-M son las desconocidas; el análisis de estos resultados se muestran en la Sección III-A.

Luego, el total de imágenes de las especies presentes en CIMP se dividió aleatoriamente en los conjuntos de entrenamiento, validación y test a razón de 70%, 15% y 15% respectivamente. En seguida, se realizó recortes aleatorios sobre las imágenes originales a fin de obtener imágenes de menor tamaño que pudieran representar correctamente la especie, en este sentido las secciones de recorte tienen una dimensión de 512x512, este valor fue escogido basado en una extrapolación de experiencias previas con diferentes tamaños de imágenes de madera [21]. Por último, los modelos fueron entrenados durante 300 épocas usando además técnicas de *data augmentation* como *flips* y rotaciones aleatorias lo que brinda un efecto de regularización sobre los modelos.

Al finalizar el entrenamiento el modelo propuesto obtuvo una exactitud de 95.76 % en el conjunto de validación, mientras que los modelos de comparación, ModelC1 y ModelC2, obtuvieron 91.88 % y 96.78 %, respectivamente. Estos valores de exactitud indican que los modelos fueron entrenados correctamente y son capaces de generalizar para información no vista durante el entrenamiento, además, a pesar de que el modelo propuesto tiene solo una fracción del total de

parámetros de los otros dos modelos muestra un desempeño similar.

III. RESULTADOS

En esta sección, se analiza el desempeño del modelo propuesto y los otros dos adicionales. Primero, evaluando su desempeño en el reconocimiento bajo las perspectivas de *closed set* y *open set*. Luego, se evalúan los rendimientos de cada uno de los modelos en su ejecución sobre un sistema embebido.

A. Evaluación de los Modelos en el Reconocimiento de Especies Maderables

El modelo aquí propuesto tiene como característica brindar una aproximación al problema de *open set* en el reconocimiento de especies maderables, sin embargo, ya que los modelos presentes en la literatura no consideran esta problemática se optó en un primer paso por comparar los modelos bajo una perspectiva de *closed set*, ya que es el entorno para el que fueron diseñados, y luego contrastar este desempeño bajo una perspectiva de *open set*.

1) *Desempeño bajo una perspectiva de closed set* : en esta sección se evalúa únicamente el comportamiento de los modelos en la clasificación de imágenes sobre las 16 especies de CIMP. Para evaluar el desempeño por cada clase se usa la matriz de confusión, la cual da una idea global del comportamiento de los clasificadores.

Las matrices de confusión se muestran en Fig. 5, en estas se puede notar una correcta clasificación de las imágenes,

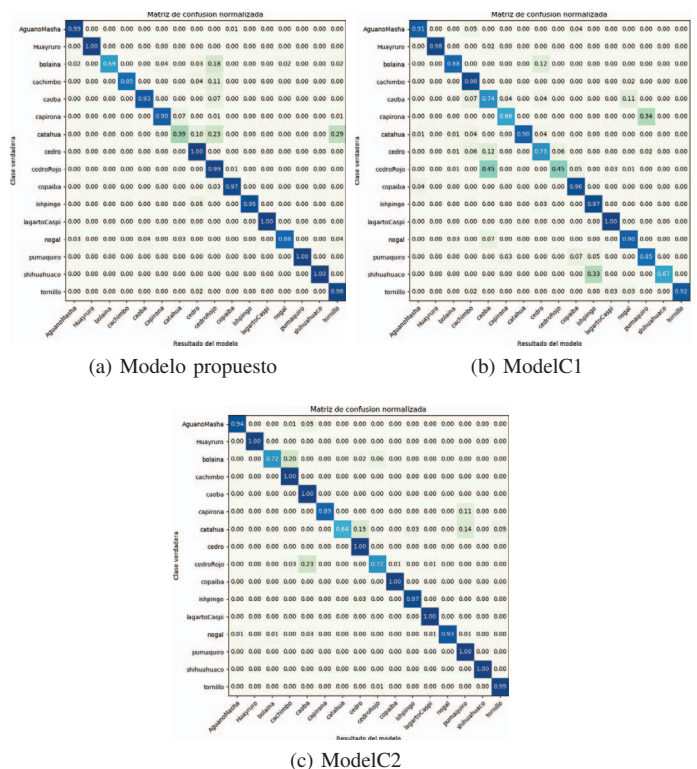


Fig. 5. Matrices de confusión para el conjunto de prueba de CIMP con los modelos: (a) modelo propuesto, (b) ModelC1 y (c) ModelC2.

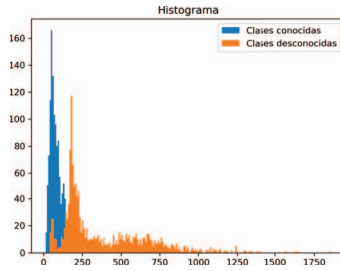


Fig. 6. Histograma de las distancias mínimas medidas entre las proyecciones de imágenes de entrada y los clústeres más cercanos a ellas. En azul se muestra la distribución para un total de 1220 imágenes de clases conocidas de CIMP y en naranja para las 1430 imágenes de clases desconocidas pertenecientes a FSD-M.

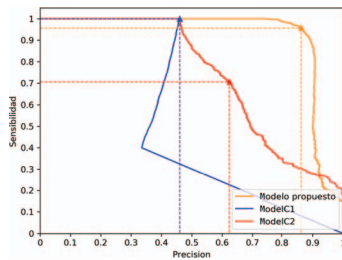


Fig. 7. Curvas de precisión y sensibilidad de los tres modelos calculada para diferentes valores umbral. Los triángulos sobre cada una de las curvas señalan los resultados para el umbral óptimo.

además de una similitud en su comportamiento, ya que para determinadas especies los modelos clasifican correctamente, mientras que para otras especies se equivocan con cierto grado de error. De esto podemos concluir que el modelo propuesto, a pesar de tener menos parámetros que los otros dos modelos, puede realizar correctamente tareas de clasificación en conjuntos de datos cerrados con una exactitud superior a 90 %, calculado a partir de la matriz de confusión.

2) *Desempeño bajo una perspectiva de open set:* como se mencionó anteriormente, la evaluación se realiza mediante la simulación de un escenario *open set* donde imágenes de las clases conocidas son las que están presentes en el conjunto de *test* de CIMP mientras que las clases desconocidas son imágenes aleatorias de once especies presentes en FSD-M. El modelo propuesto está pensado para proyectar las imágenes de entrada sobre un espacio que permita agruparlas y delimitar las especies conocidas y desconocidas usando un umbral de distancia. Para ilustrar esta característica se calculó la distancia mínima entre el vector $O1$ y los centroides para las imágenes de test, como se describe en la Sección II-C1. Esto se muestra en la Fig. 6, donde se puede observar que las distribuciones de clases conocidas y desconocidas pueden ser separadas fácilmente mediante la selección de una distancia umbral.

De los histogramas de las distancias, vistos en Fig. 6, asumimos que sus funciones de densidad de probabilidad se asemejan a gaussianas. Entonces, la distancia umbral será la media más dos veces la desviación estándar de la clase conocida, que en este caso es igual a 153.109. Sin embargo, este valor podría no ser el óptimo. Por ello, para hallar el valor óptimo se establece un conjunto de valores umbrales entre 0 y 350 para los cuales se trazan las curvas de la precisión

TABLA II
MÉTRICAS OBTENIDAS BAJO UN ENFOQUE OPEN SET CON CIMP

Métricas	Modelo propuesto	ModelC1	ModelC2
Exactitud	91.622 %	45.962 %	66.452 %
Precisión	88.741 %	45.996 %	61.897 %
Sensibilidad	93.688 %	99.836 %	70.573 %
F1 score	91.114 %	62.978 %	65.951 %

y sensibilidad, como se muestra en Fig. 7. En la curva del modelo propuesto, se observa que los valores máximos de precisión y sensibilidad se obtienen con un valor umbral igual a 143.889. Es importante señalar que los dos valores umbral son próximos, pero en la práctica se usará el calculado inicialmente, pues utiliza únicamente la información de las clases conocidas que es la que *a priori* se puede conocer.

Para comparar el modelo propuesto con los otros dos, se establece una probabilidad umbral a la cual los modelos, ModelC1 y ModelC2, van a determinar si una clase pertenece o no a una de sus clases, por lo que en caso de no pertenecer a ninguna de sus clases será una clase desconocida. En este escenario el problema se reduce a un problema de clasificación binario donde un valor cercano a cero representa a la clase desconocida y un valor cercano a uno a las clases conocidas. Siguiendo un procedimiento similar al usado para el modelo propuesto, se calculan los valores de precisión y sensibilidad para diferentes probabilidades umbral, como se muestra en Fig. 7, donde los umbrales óptimos para ModelC1 y ModelC2 son 0.292 y 0.997 respectivamente. A partir de la tendencia mostrada en la Fig. 7 se puede concluir que ModelC1 presenta un pésimo desempeño, por otro lado, ModelC2 tiene un mejor comportamiento para discernir entre conocido y desconocido, sin embargo esto solo se logra si *a priori* se tiene información de las clases desconocidas, lo cual en la práctica no siempre es posible. Una vez determinados los umbrales óptimos de los tres modelos, se calcularon la exactitud y *F1 score*, de esta manera, si comparamos todas las métricas, como se observa en la Tabla II, el modelo propuesto presenta el mejor comportamiento.

B. Evaluación del Modelo Propuesto Sobre el Sistema Embebido

1) *Costo computacional:* la información sobre el uso de los recursos de un modelo dado es importante para poder estimar tanto la velocidad de ejecución como los requerimientos en hardware, esta información está relacionada con la cantidad de parámetros y la arquitectura de cada modelo. El modelo propuesto tiene una cantidad menor de parámetros comparada con ModelC1 y ModelC2, como se observa en la Tabla III, esto nos permite intuir que el modelo propuesto hace uso de menos

TABLA III
USO DE LOS RECURSOS DEL SISTEMA EMBEBIDO

Variable	Modelo propuesto	ModelC1	ModelC2
CPU	54.313 %	59.782 %	67.077 %
RAM	257.903 Mb	262.292 Mb	258.322 Mb
Nro. parámetros	0.49 M	4.97 M	3.59 M

recursos. Para ver este efecto sobre el sistema embebido, se realizaron mediciones del uso de memoria RAM y CPU cada 0.5 segundos mientras el sistema realizaba 100 identificaciones de manera continua. Se hizo la comparativa entre los tres modelos de redes neuronales entrenados: el modelo propuesto, ModelC1 y ModelC2, cuyos resultados se muestran en la Tabla III.

2) *Consumo energético*: el consumo de energía está relacionado directamente al tiempo de autonomía que puede tener un dispositivo, por tanto es fundamental para el caso de los dispositivos portátiles. Para el cálculo del consumo se usaron como variables la corriente y tensión, además del tiempo de ejecución de las inferencias de cada CNN entrenada, para finalmente obtener el valor de la potencia consumida del sistema, antes y durante el proceso de inferencia. Las mediciones de corriente y tensión se realizaron usando un multímetro digital *Keysight 34470A* sobre un *Raspberry Pi3B+* sin ningún periférico conectado, de acuerdo al diagrama de conexión mostrado en Fig. 8. Para facilitar la comprensión de los datos obtenidos, se han clasificado en dos grupos: reposo (cuando se ha encendido el sistema embebido para la ejecución de su propio sistema operativo) y durante la ejecución (cuando se está ejecutando una de las CNN).

Todos los experimentos fueron realizados sobre el sistema embebido *Raspberry Pi3B+* el cual posee un microprocesador *Broadcom BCM2837B0* y GPU *Cortex A53* (ARMv8), con arquitectura de 64-bit SoC a 1.4 GHz. Las CNN sometidas a las pruebas de energía fueron el modelo propuesto, ModelC1 y ModelC2, cada una ejecutó 100 inferencias para muestras aleatorias durante el tiempo requerido de ejecución. Los datos obtenidos de corriente y tensión se adquirieron con una frecuencia de muestreo de 430 ms, repitiéndose cuatro veces cada prueba para obtener así promedios de potencia de cada tipo de CNN.

Las mediciones de corriente y potencia para cada CNN se hicieron sobre dos bases de tiempo: t_1 de 170 segundos y t_2 de 420 segundos debido a que el modelo propuesto y ModelC1 ejecutaban las inferencias en un tiempo mucho menor a su similar ModelC2. Luego, los cálculos se realizaron sólo en las etapas de carga (C) e inferencia (I) de las CNN, y no en el tramo final de la ejecución del algoritmo, obteniéndose los resultados mostrados en Tabla IV.

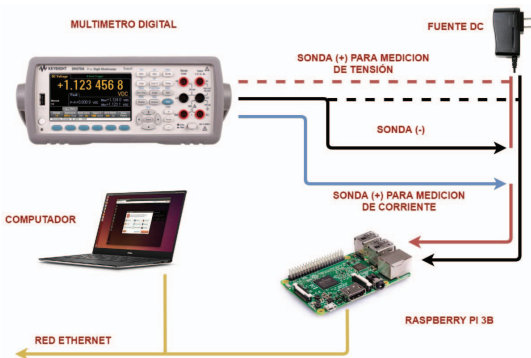


Fig. 8. Diagrama general para la adquisición de los datos y el cálculo del consumo energético antes y durante la ejecución de los algoritmos.

TABLA IV
INDICADORES ESTADÍSTICOS DE POTENCIA Y CORRIENTE DE LAS CNN

Métricas	Modelo propuesto		ModelC1		ModelC2		Raspberry reposo
	C	I	C	I	C	I	
Potencia min (W)	1.6175	1.3625	1.5800	1.3700	1.6950	1.6950	1.4160
Potencia max (W)	2.1075	2.9975	2.2750	3.3650	2.7650	3.4400	2.0810
Potencia promedio (mWh)	0.4801	0.6883	0.5151	0.6841	0.5479	0.7204	0.4225
Corriente min (A)	0.3212	0.2715	0.2725	0.2725	0.3175	0.3200	0.2642
Corriente max (A)	0.4187	0.5960	0.6700	0.4900	0.5175	0.6475	0.3879
Corriente promedio (mAh)	0.0954	0.1368	0.1025	0.1361	0.1026	0.1361	0.0787
Intervalo de medición (s)	25.33	94.86	15.81	117.64	142.8	221.68	-
Tiempo total (s)	170		170		420		170

ModelC1 generó consumo de corriente acumulada promedio igual a 0.1361 mAh en la etapa de inferencia, al igual que ModelC2, sin embargo esta última fue la que más inconvenientes ocasionó y se percibió un sobrecalentamiento de la placa del sistema embebido, además de realizar las inferencias en un tiempo igual a 221.68 s que es más de dos veces el tiempo que empleó el modelo propuesto y aproximadamente el doble del tiempo empleado por ModelC1. Además, registró un máximo de corriente de 0.6475 mA y potencia máxima de 3.44 llegando incluso a calentar excesivamente alguno de los componentes usados en las mediciones.

El consumo energético más eficiente corresponde al modelo propuesto que obtuvo una potencia acumulada promedio de 1.1684 mWh, frente a 1.1992 mWh de ModelC1 y 1.2683 mWh de ModelC2, en ambas etapas, siendo ModelC2 la que más potencia consume en total. Como se observa en la Tabla IV, el modelo propuesto ofrece un rendimiento energético menor en comparación a ModelC1 y ModelC2, así, se logra una reducción de la potencia máxima entre 10% y 12% y por consiguiente una mayor duración de la batería del *Raspberry Pi3B+* u otro sistema embebido donde se ejecute, además de garantizar un mejor desempeño de los componentes electrónicos.

IV. CONCLUSIONES

En este trabajo, se ha propuesto y analizado un modelo, basado en redes neuronales convolucionales para la identificación de especies maderables con un enfoque de reconocimiento de un conjunto abierto y de costo computacional y energético limitado. El enfoque del reconocimiento del conjunto abierto (open set recognition) es un criterio introducido en este trabajo para lidiar con la dificultad de la identificación de especies maderables en condiciones de campo, donde se presenten casos de especies maderables para las cuales no se entrenó el modelo propuesto. Los resultados del reconocimiento de las clases conocidas y no conocidas presentan métricas elevadas, con exactitud y *F1 score* arriba del 91% y mayor a los modelos usados para comparación. La sensibilidad alta es superada solo por aquella de ModelC1. Con respecto al costo computacional se puede notar que el modelo propuesto tiene siete veces menos parámetros que los dos modelos empleados para la comparación, ello se ve reflejado en un menor uso de los recursos del sistema embebido de prueba, CPU y RAM (5% y 13%), así como de un menor valor de potencia máxima (10% y 12%). Por otro lado, el modelo propuesto tiene un tiempo de ejecución de inferencia menor, en comparación a los otros dos modelos (19% y 57%). La evaluación de los resultados muestra que el modelo propuesto

tiene un buen desempeño bajo las perspectivas de *closed* y *open set recognition*, en comparación a otros dos modelos, y puede ser implementado en un sistema embebido de manera que no consuma gran parte sus recursos computacionales ni la dotación de energía, como la suministrada por una batería.

AGRADECIMIENTOS

Los autores desean agradecer al Laboratorio de Anatomía e Identificación de Maderas de la Universidad Nacional Agraria La Molina por facilitar las muestras de maderas y agradecer al Instituto Nacional de Investigación y Capacitación de Telecomunicaciones (INICTEL-UNI) por el soporte técnico y financiero brindado para el desarrollo de este trabajo.

REFERENCIAS

- [1] Organización de las Naciones Unidas para la Alimentación y la Agricultura – FAO “La industria de la madera en el Perú,” Organización de las Naciones Unidas para la Alimentación y la Agricultura y el Instituto Tecnológico de la Producción (ITP) - CITEMadera, Lima, 2018. [Online]. Disponible en: <http://www.fao.org/3/I8335ES/I8335es.pdf>
- [2] Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore, “COMBATING BIOPIRACY – THE PERUVIAN EXPERIENCE”, World Intellectual Property Organization, 2007. [Online]. Disponible en: https://www.wipo.int/edocs/mdocs/tk/en/wipo_grtkf_ic_11/wipo_grtkf_ic_11_8_a.pdf
- [3] Gerald Koch, Volker Haag, Immo Heinz, Hans-Georg Richter and Uwe Schmitt, “Control of Internationally Traded Timber - The Role of Macroscopic and Microscopic Wood Identification against Illegal Logging,” *Journal Forensic Research*, 2015, 6:6.
- [4] P. R. Cavalin, M. N. Kapp and L. S. Oliveira, “Multi-scale texture recognition systems with reduced cost: A case study on forest species,” 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 1316-1321.
- [5] C. Affonso, A. L. Debiase, F. H. Antunes and A. C. Ponce de Leon Ferreira de Carvalho. “Deep learning for biological image classification,” *Expert Systems With Applications*. Vol. 85, 2017, pp. 114-122.
- [6] A. R. Yadav, M. L. Dewal, R. S. Anand and S. Gupta, “Classification of hardwood species using ANN classifier”, 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Jodhpur, 2013, pp. 1-5.
- [7] B. Sugianto et al., “Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier”, 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 337-341.
- [8] L. G. Hafemann, L. S. Oliveira and P. Cavalin, “Forest Species Recognition Using Deep Convolutional Neural Networks,” 2014 22nd International Conference on Pattern Recognition, Stockholm, 2014, pp. 1103-1107.
- [9] G. Figueroa-Mata, E. Mata-Montero, J. C. Valverde-Otarola and D. Arias-Aguilar, “Using Deep Convolutional Networks for Species Identification of Xylotheque Samples” 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), San Carlos, 2018, pp. 1-9.
- [10] Ravindran, Prabu et al. “Classification of CITES-Listed and Other Neotropical Meliaceae Wood Images Using Convolutional Neural Networks” *Plant Methods* 14 (2018): 25. PMC. Web. 19 Apr. 2018.
- [11] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota and T. E. Boulton, “Toward Open Set Recognition,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757-1772, July 2013.
- [12] T. Boulton, S. Cruz, A. Dhamija, M. Günther, J. Henrydoss, W. J. Scheirer, “Learning and the Unknown: Surveying Steps Toward Open World Recognition,” AAAI Conference on Artificial Intelligence (AAAI 2019), Senior Member Track, January 2019.
- [13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 4510-4520.

- [14] A. Bendale and T. E. Boulton, “Towards Open Set Deep Networks,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 1563-1572.
- [15] Servicio Nacional Forestal y de Fauna Silvestre – SERFOR, Ministerio de Agricultura y Riego, República del Perú. Anuario forestal y de fauna silvestre 2016 [Online]. Disponible en <https://www.serfor.gob.pe/wp-content/uploads/2018/05/Anuario-2016.pdf>
- [16] Confederación Peruana de la Madera - CPM (2008). Compendio de información técnica de 32 especies forestales. Tomo I. Lima, Perú. Centro de Innovación Tecnológica de la Madera-CITEMadera/ Ministerio de la Producción.
- [17] Confederación Peruana de la Madera - CPM (2008). Compendio de información técnica de 32 especies forestales. Tomo II. Lima, Perú. Centro de Innovación Tecnológica de la Madera-CITEMadera/ Ministerio de la Producción.
- [18] P. L. Paula Filho, L. S. Oliveira, S. Nigkoski, A. S. Britto Jr, Forest Species Recognition using Macroscopic Images, *Machine Vision and Applications*, 25(4):1019-1031, 2014.
- [19] M. Misiti, Y. Misiti, G. Oppenheim and J.M. Poggi, *Wavelet Toolbox 4 User's Guide*, The Matworks Inc, 1997-2009.
- [20] Mehadi Hassen and Philip K. Chan. Learning a Neural-network-based Representation for Open Set Recognition, 2018; arXiv:1802.04365.
- [21] M. P. E. Apolinario Lainez, S. G. Huamán Bustamante and G. C. Orellana, “Deep Learning Applied to Identification of Commercial Timber Species from Peru,” 2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Lima, 2018, pp. 1-4.



Marco Paul Enrique Apolinario Lainez Obtuvo su bachiller en ciencias con mención en Ingeniería Electrónica de la Universidad Nacional de Ingeniería (UNI) en Perú en el 2017. Desde 2018 es Asistente de investigación y desarrollo tecnológico en el área de Procesamiento de Señales e Imágenes en el Instituto Nacional de Investigación y Capacitación de Telecomunicaciones (INICTEL-UNI). Entre 2017 y 2018 fue practicante en la Dirección de Investigación y Desarrollo Tecnológico de INICTEL-UNI, y a inicios del 2017 fue practicante en el Área de Investigación, Desarrollo e Innovación del Radio Observatorio de Jicamarca (ROJ) en Perú.



Daniel Augusto Urcia Paredes culminó sus estudios de pregrado de la carrera de Ingeniería Electrónica en la Universidad Nacional de Ingeniería (UNI), Lima, Perú en el año 2018. Desde 2018 se viene desempeñando como Asistente de Investigación en el área de Procesamiento de Señales en el Instituto Nacional de Investigación y Capacitación de Telecomunicaciones (INICTEL-UNI). Su interés actual incluye Desarrollo de Prototipos Electrónicos, Análisis de Imagen, Sistemas Embebidos y Programación de bajo y alto nivel.



Samuel Gustavo Huaman Bustamante obtuvo su doctorado en Ingeniería Eléctrica de la Pontificia Universidad Católica de Río de Janeiro (PUC-Río), en Brasil en el 2012. Maestría en Ingeniería Eléctrica en la PUC-Río en 2008. Bachillerato en Ingeniería Electrónica de la Universidad Nacional de Ingeniería (UNI), en Perú, en 1999. Desde 2016 es investigador en Instituto Nacional de Investigación y Capacitación de Telecomunicaciones (INICTEL-UNI) y, desde 2013, es profesor en unidad de posgrado de la Facultad de Ingeniería Eléctrica y Electrónica de la UNI. De 2013 a 2016 fue especialista electrónico en el Servicio Nacional de Meteorología e Hidrología del Perú. De 2007 a 2012 fue investigador en el Laboratorio de Inteligencia Computacional Aplicada de la PUC-RIO, en Brasil.