A Low-cost Ergonomic Device for Banknote Recognition Using Computer Vision and IoT

Nicolás Esleyder Caytuiro-Silva [1], Benjamín Maraza-Quispe [2,*], Eveling Gloria Castro-Gutierrez [1], Jackeline Melady Peña-Alejandro [1], Olga Melina Alejandro-Oviedo [2], Victor Hugo Rosas-Iman [2], Giuliana Feliciano-Yucra [2], Atilio Cesar Martinez-Lopez [2], and Roberto Carlos Pari-Viza [2]

[1] Faculty of Physical and Formal Sciences and Engineering, Universidad Católica de Santa María de Arequipa, Arequipa, Peru
[2] Faculty of Educational Sciences, Universidad Nacional de San Agustín de Arequipa, Arequipa, Peru
Email: nicolas.caytuiro@ucsm.edu.pe (N.E.C.-S.); bmaraza@unsa.edu.pe (B.M.-Q.); ecastrog@ucsm.edu.pe (E.G.C.-G.); 72525180@ucsm.edu.pe (J.M.P.-A.); oalejandro@unsa.edu.pe (O.M.A.-O.); vrosasi@unsa.edu.pe (V.H.R.-I.); gfeliciano@unsa.edu.pe (G.F.-Y.); amartinezl@unsa.edu.pe (A.C.M.-L.); rpariv@unsa.edu.pe (R.C.P.-V.)
*Corresponding author

***Abstract*—This study presents the design, implementation, and validation of a low-cost ergonomic assistive device for autonomous recognition of Peruvian banknotes using computer vision and Internet of Things (IoT) technologies. A dataset of 9054 images distributed across 16 classes was constructed to train and evaluate five pretrained convolutional Neural Network (CNN) architectures: Visual Geometry Group 16-layer network (VGG16), Visual Geometry Group 19-layer network (VGG19), 50-layer Residual Network (ResNet50), MobileNet version 2 (MobileNetV2), and Extreme Inception (Xception). Among these models, Xception achieved the best performance, with 99.8% test accuracy, macro-averaged precision of 0.997, macro-averaged recall of 0.997, and macro-averaged F1-Score of 0.997. Five-fold cross-validation confirmed the model's stability, yielding a mean accuracy of 99.68% with a standard deviation of 0.12 percentage points. Robustness tests under low-light conditions, partial occlusion, and folded banknote scenarios showed recognition performance above 96%. The selected model was deployed on a Raspberry Pi 4B+ IoT platform with audio feedback. Hardware benchmarking reported an average inference time of 3.82 ± 0.27 s and 3.2 h of battery autonomy. Usability evaluation with 44 visually impaired participants produced a System Usability Scale (SUS) score above 80 and high User Experience Questionnaire (UEQ) ratings. The economic feasibility analysis, including sensitivity scenarios, confirmed financial sustainability, with an Internal Rate of Return (IRR) of 11.85% and a Net Present Value (NPV) of 13,013.02 Peruvian soles (S/). Ethical approval was obtained, and informed consent procedures were completed. These findings demonstrate that the proposed assistive device is technically robust, socially relevant, and economically viable.**

***Keywords*—assistive technology, computer vision, Internet of Things (IoT), deep learning, banknote recognition, digital inclusion**

I. Introduction

According to the World Health Organization, there are approximately 285 million people with visual disabilities worldwide, of whom 39 million are blind, and the World Blind Union [1] estimates that 253 million have visual impairments, evidencing a global issue that significantly impacts quality of life. In Peru, nearly 160,000 people are blind and about 600,000 have visual impairments [2], a situation reflected in Arequipa where 3.4% of the population reports visual limitations even when using glasses [3]. These figures highlight the need for accessible devices that promote inclusion and autonomy for this vulnerable group, given that existing assistive technologies remain limited, expensive, or poorly ergonomic [4, 5]. Although proposals such as smart canes [6, 7] or vibrating glasses based on Arduino [8] have been developed, most focus on mobility and navigation, leaving gaps in other daily activities such as object and banknote recognition. In this context, this research is justified by its contribution to the design and implementation of an ergonomic and low-cost device based on computer vision and Internet of Things (IoT), specifically aimed at banknote recognition for visually impaired people in Arequipa, contributing not only to their economic and social independence but also to generating a replicable model in other scenarios, in line with the fundamental role technology must play in improving quality of life and social integration [9–11].

Historically, visually impaired people have relied on traditional tools such as the white cane, the braille devise, or guide dogs, symbols of autonomy that have accompanied this population for decades [12, 13]. However, in a globalized context, these devices are limited in addressing current demands for inclusion and autonomy, becoming insufficient and even obsolete in the face of new challenges of social integration [14]. Moreover, many modern assistive tools are expensive or difficult to use, leading to their abandonment [15].

Manuscript received January 29, 2026; revised March 5, 2026; accepted May 28, 2026; published August 19, 2026.

II. LITERATURE REVIEW

Various technological devices have been proposed to enhance autonomy for visually impaired individuals, particularly through low-cost and adaptable devices. Early developments focused primarily on mobility assistance. For instance, Ayala [16] and Avila [17] designed electronic canes integrating ultrasonic sensors and Arduino modules to detect obstacles and generate vibration and sound alerts. Similarly, Dávila [18] incorporated a Kinect V2 camera with user morphological analysis to personalize detection within a 360° environment. Although these devices represent significant advances, their scope remains centered on spatial orientation rather than broader functional autonomy.

Subsequent research expanded toward object identification. Ortiz [19] proposed a Radio-Frequency Identification (RFID)-based device that enables the recognition of product characteristics in shopping environments, illustrating the potential of assistive technologies to enhance daily interaction, although its applicability remains context-specific. Likewise, Peralta and Urmendiz [20] developed a cane-adaptable guidance device with multi-level height detection and mobile connectivity, reinforcing safety and integration with smart devices.

In the educational domain, Hernandez *et al.* [21] introduced a Braille-learning device equipped with a voice synthesizer to reduce acquisition time. Although valuable for inclusive education, such devices remain limited to academic contexts and do not address other dimensions of independence, such as financial transactions.

More recently, computer vision and deep learning have significantly expanded assistive capabilities. Shah *et al.* [22], Rusli *et al.* [23], and Mukhiddinov *et al.* [24] employed You Only Look Once version 4 (YOLOv4) and datasets such as Common Objects in Context (COCO) for real-time object detection, reporting high precision while also identifying ergonomic and user-interface limitations. In the specific context of banknote recognition, Awad *et al.* [25] and Suganya *et al.* [26] demonstrated the feasibility of Convolutional Neural Network (CNN)-based real-time classification devices, achieving high recognition accuracy and highlighting the practical applicability of deep learning in economic activities.

Complementary investigations emphasize persistent technical and ethical challenges. Morparia *et al.* [27] and Valipoor and de Antonio [28] discussed usability, portability, and privacy concerns as critical barriers to the adoption of computer vision-based assistive technologies. Meanwhile, Croce *et al.* [29] and Das *et al.* [30] explored indoor and outdoor navigation using deep learning approaches, whereas Portilla [31] presented an Artificial Intelligence (AI)-based robotic assistant, raising ethical considerations related to data management and human–machine interaction.

Research specifically addressing banknote recognition provides further methodological insights. Sufri *et al.* [32] demonstrated that image orientation and capture region significantly influence classification accuracy, even achieving 100% accuracy with traditional machine learning models such as Support Vector Machine (SVM) and Bayesian Classifier (BC). These findings underscore the importance of practical acquisition conditions and reinforce the need for ergonomic device design. Similarly, Castelar *et al.* [33] proposed hybrid approaches combining classical feature extraction methods, such as Scale-Invariant Feature Transform (SIFT) and Oriented FAST and Rotated BRIEF (ORB), with CNN architectures such as LempiraNet, surpassing 98% accuracy and enabling rapid adaptation to new currency designs.

More advanced mathematical representations have also been explored. Huang and Gai [34] integrated the quaternion wavelet transform with deep convolutional networks to improve real-time banknote classification performance by leveraging multiscale and multiorientation features. These approaches highlight the increasing sophistication of classification architectures for currency recognition.

From an applied hardware perspective, Awad *et al.* [25] achieved 98.6% accuracy in Iraqi banknote recognition using a 19-layer CNN with integrated voice feedback, emphasizing accessibility for visually impaired users. Similarly, Fan *et al.* [35] deployed MobileNetV2 on a Raspberry Pi 3B+, achieving 96.4% accuracy for Ethiopian banknotes and demonstrating the feasibility of low-cost embedded implementations. At the Latin American level, Zapata [36] implemented a device for Colombian banknotes using Multilayer Perceptron (MLP) networks and Raspberry Pi, reaching 95% accuracy and confirming the viability of locally adapted devices.

Mobile-based solutions have also emerged. Khalil *et al.* [37] and Geerts *et al.* [38] developed a smartphone application using MobileNetV2 and transfer learning for United Arab Emirates (UAE) banknotes, achieving 70%–88% accuracy. Although its performance was lower than that of embedded CNN-based prototypes, the study demonstrates scalability and accessibility through widely available mobile platforms.

Overall, the literature indicates that computer vision and deep learning have matured sufficiently to support real-time banknote recognition devices. Nevertheless, recurring limitations persist: insufficient ergonomic integration, dependence on localized datasets, limited robustness validation, and partial consideration of accessibility factors. In response, the present study advances beyond algorithmic accuracy by integrating deep learning with an ergonomic, low-cost IoT device tailored to the local context of Arequipa. The objective is not only to achieve high classification performance but also to ensure usability, portability, and practical impact on the financial autonomy of visually impaired users.

III. MATERIALS AND METHODS

A. Study Design and Participants

This study followed an applied and experimental validation approach aimed at designing, implementing, and evaluating a low-cost assistive device for Peruvian banknote recognition. The study combined computer vision, deep learning, and embedded IoT deployment to support financial autonomy among visually impaired users.

The user validation stage involved visually impaired participants from the Unión de Ciegos de Arequipa Association. Participants were selected according to their availability and willingness to interact with the prototype under real-use conditions. The evaluation focused on usability, perceived usefulness, accessibility, and interaction with the device during banknote recognition tasks.

Fig. 1 presents representative examples of the captured Peruvian banknote images, illustrating front and back views, denomination variability, and relevant visual detail regions considered during dataset construction.

Fig. 1. Examples of captured Peruvian banknote images, including front, back, denomination, and visual detail regions.

B. Dataset Construction and Preprocessing

A dataset of Peruvian banknote images was constructed using banknotes from the S/10, S/20, S/50, and S/100 denominations, including both old and new families and front and back sides. Images were captured under controlled and real-world conditions to improve the representativeness of the dataset. The images were organized into class labels according to denomination, family, and side.

Before model training, the images were resized to match the input requirements of the selected CNN architectures. Data augmentation techniques, including rotation, brightness variation, and horizontal and vertical flipping, were applied only to the training set. This procedure was used to improve model robustness while avoiding data leakage into the validation and testing subsets.

Banknote images were captured under both controlled and real-world conditions. Training images were obtained using white or neutral backgrounds and included both the front and back sides of each banknote. Validation images were captured in everyday environments with varying lighting conditions, while the testing subset was selected independently from the available image set. The dataset was organized into 16 classes according to denomination, family, and side. A systematic labeling scheme was applied, and images were resized from 3072×4080 px to 854×480 px to reduce computational load while preserving visual features relevant for classification. The dataset was made available through IEEE Dataport [39] and Mendeley Data [40].

Additional details regarding the image acquisition device, camera specifications, and dataset distribution are provided in Appendix. Specifically, Table A1 reports the smartphone hardware used for banknote image capture, including processor, memory, storage, camera resolution, and operating system specifications. In addition, Fig. A1 presents the percentage distribution of Peruvian banknotes by denomination and family, allowing a clearer understanding of dataset composition and potential class imbalance considerations.

C. CNN Model Training and Validation

Five pretrained convolutional neural network architectures were evaluated: VGG16, VGG19, ResNet50, MobileNetV2, and Xception. Transfer learning was initially applied using pretrained weights, followed by fine-tuning of the upper layers to adapt the models to the Peruvian banknote recognition task.

The dataset was divided into mutually exclusive training, validation, and testing subsets. Model performance was assessed using accuracy, precision, recall, F1-Score, and confusion matrix analysis. In addition, a 5-fold cross-validation procedure was applied to evaluate model stability and reduce the risk of overfitting.

Five pretrained CNN architectures were evaluated: VGG16, VGG19, Xception, ResNet50, and MobileNetV2.

The models were trained in Python using TensorFlow/Keras in Google Colab Pro+ with GPU acceleration. The dataset was divided into training, validation, and testing subsets using a 70%, 15%, and 15% split, respectively. Data augmentation, including rotation, brightness adjustment, and horizontal and vertical flipping, was applied only to the training set. Transfer learning was first performed using pretrained weights, followed by fine-tuning to adapt the models to the Peruvian banknote recognition task.

The complete CNN training configuration is reported in Appendix. Table A2 summarizes the hyperparameters used for the evaluated pretrained architectures, including the input image size, optimizer, learning rate, batch size, number of epochs, fine-tuning strategy, loss function, and main evaluation metric. This information was included to improve the reproducibility of the experimental training procedure.

D. Prototype Evaluation Procedure

The best-performing CNN model was selected for deployment on the embedded platform. The prototype was evaluated in terms of recognition performance, embedded inference stability, response time, usability, user experience, and economic feasibility. The user evaluation was conducted through structured instruments, including usability and user experience questionnaires, complemented by direct observation during prototype interaction.

E. Ethical Considerations

The study was conducted in accordance with the principles of the Declaration of Helsinki. All participants provided informed consent before participating in the evaluation. Personal data were anonymized and securely stored. The research protocol was reviewed and approved by the corresponding institutional ethics committee.

IV. System Design and Implementation

A. System Requirements

To guide the design of the proposed IoT-based banknote recognition device, functional and non-functional requirements were defined. The functional requirements specify the main operations that the device must perform, including image capture, banknote recognition, and audio feedback generation. The non-functional requirements describe the expected quality attributes related to usability, reliability, portability, efficiency, and ergonomics.

As shown in Tables I and II, the device was designed to support a simple interaction sequence: the user activates the capture button, the system processes the banknote image, and the recognized denomination is communicated through audio feedback. These requirements guided the selection of the hardware components, the embedded software workflow, and the ergonomic enclosure design.

B. General System Architecture

The proposed system was designed as an integrated assistive device combining image acquisition, embedded computer vision processing, and audio feedback.

Fig. 2 shows the general architecture of the proposed IoT-based banknote recognition device. The system is organized into three main stages: data capture, processing, and output. In the data capture stage, the banknote image is acquired using a Raspberry Pi 5 MP camera integrated into the IoT device. In the processing stage, the captured image is analyzed on a Raspberry Pi 4B+ using the trained deep learning model [41]. Finally, in the output stage, the predicted denomination is communicated to the user through audio feedback using a speaker.

TABLE I. Functional Requirements of The Device

Code	Name	Purpose	Input	Output
RF01	Banknote Capture	Capture images with HD camera	Physical button	Banknote image
RF02	Banknote Recognition	Classification of value with CNNs	Banknote image	Predicted class
RF03	Audio Output	Generate audio feedback	Predicted class	User audio

TABLE II. Non-Functional Requirements

Criterion	Description
Usability	Intuitive interaction, accessible button layout, and clear audio feedback.
Reliability	Stable recognition performance under controlled and real-world capture conditions.
Portability	Compact structure suitable for handheld or chest-level use.
Efficiency	Acceptable response time and battery autonomy for assistive use.
Ergonomics	Lightweight, resistant, and user-oriented enclosure adapted to daily use.

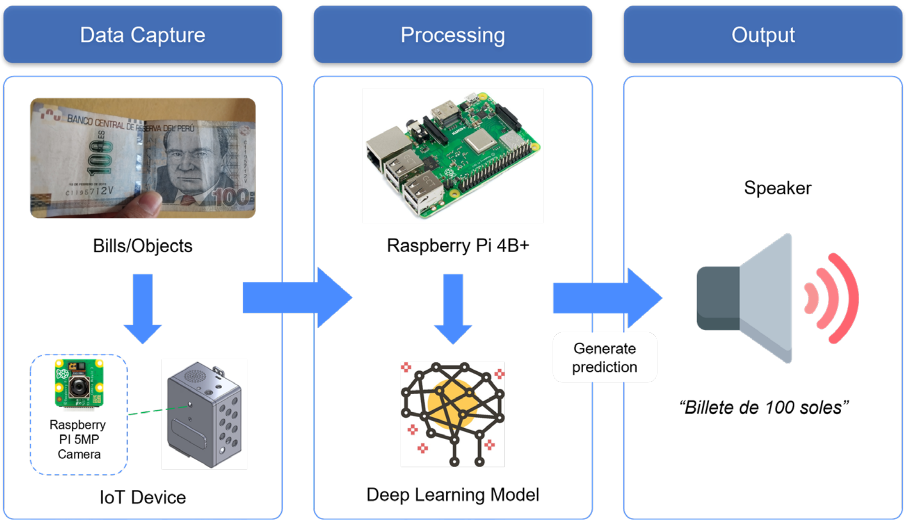

Fig. 2. General architecture of the proposed device.

As shown in Fig. 2, the proposed architecture follows a sequential and user-centered workflow. The user activates the device to capture the banknote image, the embedded platform preprocesses and classifies the image using the trained CNN model, and the result is transformed into an audio message. This architecture allows the device to operate as an autonomous assistive system without requiring visual interaction from the user.

C. Software Workflow for Dataset Construction and Model Training

Fig. 3 illustrates the software workflow followed for dataset construction and CNN model training. The process begins with the acquisition of Peruvian banknote images under different capture conditions. Then, the images are selected, labeled, and preprocessed to ensure consistency and compatibility with the input requirements of the evaluated CNN architectures. After dataset preparation, several pretrained CNN models are trained using transfer learning and fine-tuning. The best-performing model is finally exported as model weights for deployment on the embedded IoT platform.

As shown in Fig. 3, the software workflow connects the dataset preparation stage with the model training stage. This organization ensures that the banknote images are systematically processed before training and that the final CNN model is obtained through a structured sequence of selection, labeling, preprocessing, training, validation, and weight export. The exported model weights are then used in the embedded recognition system described in the following subsection.

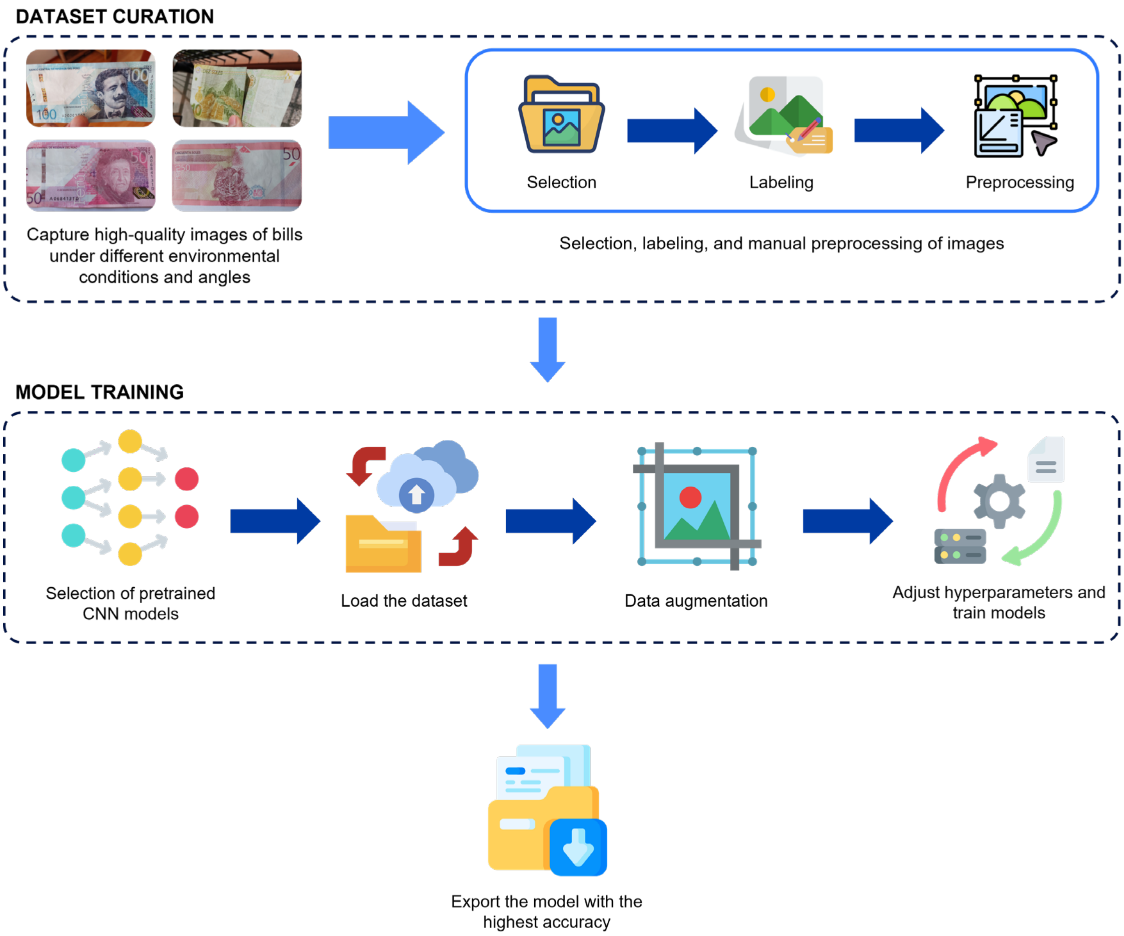

Fig. 3. Software workflow for dataset construction and model training.

D. Embedded Hardware and Inference Workflowe

Fig. 4 presents the embedded hardware and inference workflow of the proposed banknote recognition device. After the best-performing CNN model is exported, it is integrated into the Raspberry Pi 4B+ platform. During operation, the user places the banknote in front of the Raspberry Pi camera, and the captured image is processed by the embedded system. The trained model is then loaded to generate the predicted denomination. Finally, the prediction result is converted into an audio message using the Adafruit MAX98357 audio amplifier and a speaker.

As shown in Fig. 4, the embedded workflow connects the trained deep learning model with the physical components of the device. The Raspberry Pi 4B+ performs image preprocessing, model loading, and prediction generation, while the audio module communicates the recognized denomination to the user. This integration enables the prototype to operate as an autonomous assistive device without requiring an external computer during real-time use.

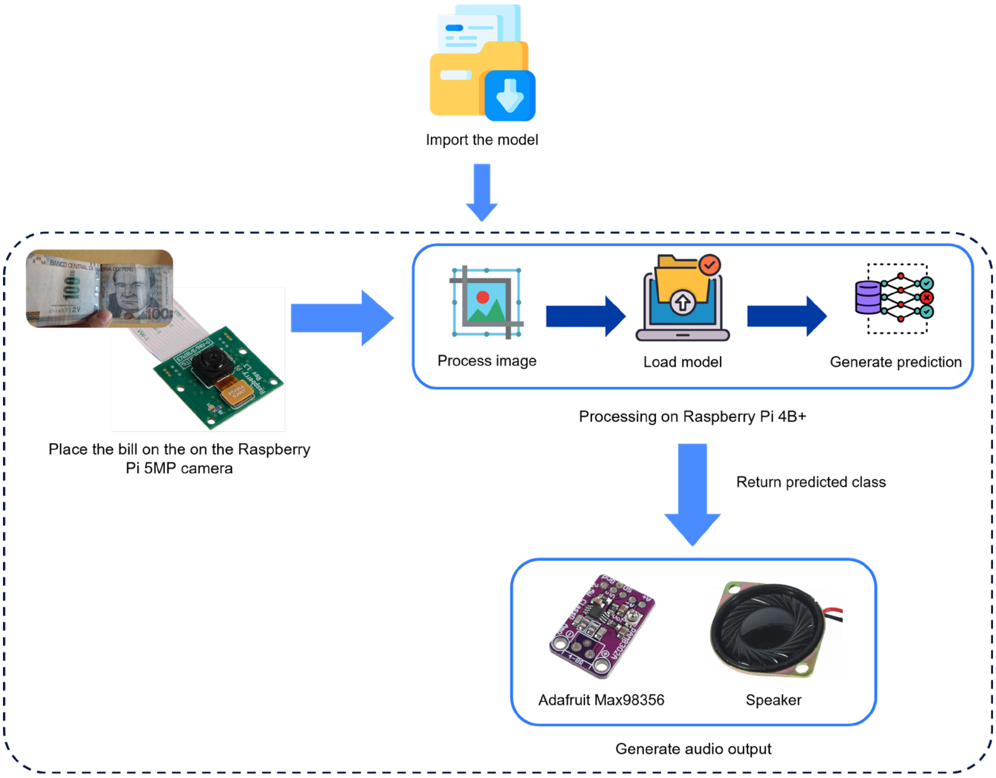

Fig. 4. Embedded hardware and inference workflow of the proposed IoT-based banknote recognition device.

E. *Prototype Enclosure and Ergonomic Design*

Fig. 5 shows the physical design and final implementation of the proposed assistive device. The left side presents the CAD-based enclosure design, while the right side shows the 3D-printed prototype during handling. The enclosure was designed to protect the internal electronic components, facilitate portability, and support intuitive interaction by visually impaired users. The case includes openings for the camera, speaker, control buttons, ventilation, and charging access.

As shown in Fig. 5, the prototype integrates the main hardware components within a compact enclosure suitable for daily use. The design considers ergonomic aspects such as grip, button accessibility, speaker positioning, and portability. These characteristics are relevant because the device is intended to be used independently by visually impaired users in real-world financial interaction contexts.

Fig. 5. CAD design and final 3D-printed prototype of the assistive device.

F. *Operational Workflow*

Fig. 6 presents the operational workflow of the proposed banknote recognition device during real-time use. The process begins when the user presses the capture button. Then, the camera captures the image of the banknote and sends it to the embedded processing unit. The system preprocesses the image and applies the trained CNN model to identify the denomination. If the banknote is correctly recognized, the device generates an audio message with the predicted denomination. If the recognition is not successful, the user is asked to repeat the capture process.

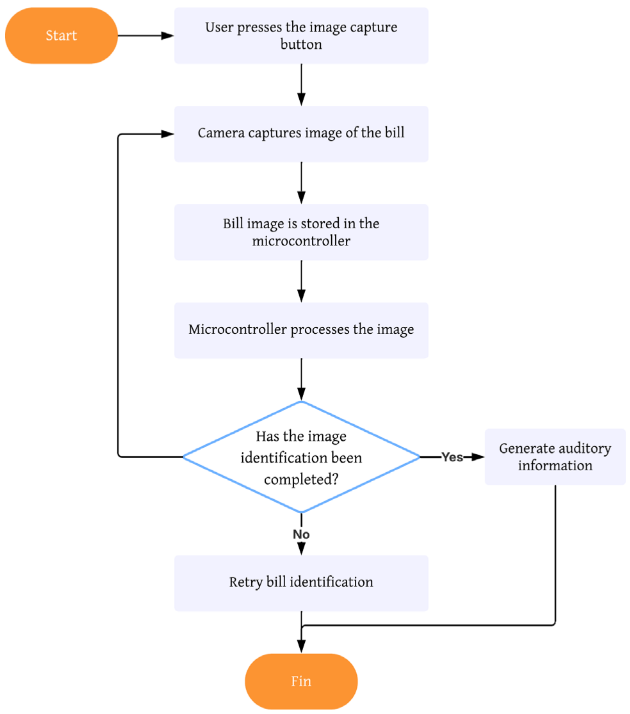

Fig. 6. Operational workflow of the proposed banknote recognition device.

As shown in Fig. 6, the proposed workflow was designed to be simple and accessible for visually impaired users. The interaction only requires pressing a physical button and listening to the audio feedback generated by the device. This sequence reduces visual dependency and allows the user to repeat the capture process when the banknote is not correctly identified.

V. Results and Discussion

A. Dataset and Model Performance

The dataset included images of Peruvian banknotes organized according to denomination, family, and side. The classification task was formulated as a multi-class recognition problem. Among the evaluated CNN architectures, Xception achieved the best overall performance, outperforming VGG16, VGG19, ResNet50, and MobileNetV2 in terms of classification accuracy and macro-averaged evaluation metrics.

The performance analysis showed that the Xception model provided the most reliable balance between accuracy and generalization. Precision, recall, and F1-Score values were consistently high across classes, indicating that the model was able to distinguish between denominations and banknote families with minimal misclassification. Cross-validation further confirmed the stability of the model across different data partitions.

Among the five evaluated CNN architectures, Xception achieved the best performance, with 99.7% accuracy during transfer learning and 99.8% after fine-tuning. Therefore, this model was selected as the baseline architecture for deployment on the embedded IoT prototype.

TABLE III. Total Image Distribution per Denomination

Denomination	New Family	Old Family	Total
10	728	1025	1753
20	894	1162	2056
50	1190	1831	3021
100	939	1285	2224
Total	3751	5303	9054

As shown in Table III, the dataset contains 3751 images from the new banknote family and 5303 images from the old banknote family, resulting in a total of 9054 images. The S/50 denomination had the highest representation, with 3021 images, whereas the S/10 denomination had the lowest representation, with 1753 images. Although the distribution is not completely uniform, all denominations and families were represented in the dataset, allowing the CNN models to learn visual patterns from both banknote families.

In addition, Fig. A1 in Appendix A shows the percentage distribution of Peruvian banknotes in the dataset used to train the computer vision model. A higher representation was observed for 50-soles banknotes from the old family (20%), whereas the lowest representation corresponded to 10-soles banknotes from the new family (8%). This relatively balanced distribution contributed to dataset diversity; however, the over-representation of certain categories may introduce training bias. Therefore, future versions of the dataset should consider balancing techniques to ensure uniform recognition performance across all denominations and banknote families.

Table IV presents the training time and classification accuracy obtained by each evaluated CNN architecture during the transfer learning and fine-tuning stages. This comparison allowed us to identify the model with the best balance between training performance and classification accuracy for subsequent deployment on the embedded IoT platform.

TABLE IV. Training time and Accuracy of the Models

Model	TL Time	FT Time	TL Accuracy (%)	FT Accuracy (%)
VGG16	1h 47m	1h 18m	50.9	66.2
VGG19	1h 10m	40m	39.7	67.7
Xception	3h 12m	1h 50m	99.7	99.8
ResNet50	1h 40m	1h 40m	53.3	54.6
MobileNetV2	1h 57m	1h 42m	63.8	63.8

As shown in Table IV, Xception achieved the highest accuracy in both training stages, reaching 99.7% during transfer learning and 99.8% after fine-tuning. Although Xception required the longest training time, its classification performance was substantially higher than that of the other evaluated architectures. VGG19 showed a relevant improvement after fine-tuning, increasing from 39.7% to 67.7%, while MobileNetV2 maintained stable performance across both stages. Based on these results, Xception was selected as the final model for deployment in the proposed IoT-based banknote recognition device.

After identifying Xception as the best-performing architecture, a class-wise evaluation was conducted to analyze its recognition performance across the different banknote categories. Precision, recall, and F1-Score were calculated for each denomination and banknote family.

TABLE V. CLASS-WISE PERFORMANCE METRICS OF THE XCEPTION MODEL

Class	Precision	Recall	F1-Score
10 old	0.997	0.996	0.996
10 new	0.995	0.996	0.995
20 old	0.998	0.997	0.997
20 new	0.997	0.998	0.997
50 old	0.999	0.999	0.999
50 new	0.998	0.997	0.997
100 old	0.999	0.998	0.998
100 new	0.998	0.999	0.998
Macro Avg	0.997	0.997	0.997

Table V presents the detailed performance metrics of the Xception model.

As shown in Table V, the Xception model achieved consistently high performance across all banknote classes. The macro-average precision, recall, and F1-Score reached 0.997, indicating balanced classification performance across denominations and banknote families. The lowest F1-Score was observed for the S/10 new family class, although its value remained high at 0.995. These results suggest that the model was able to distinguish the visual characteristics of both old and new Peruvian banknote families with minimal classification errors.

To determine whether the differences in classification accuracy among the evaluated CNN architectures were statistically significant, a one-way ANOVA test was performed. The analysis compared the mean accuracy values obtained by VGG16, VGG19, Xception, ResNet50, and MobileNetV2. Table VI presents the ANOVA results for the comparison of CNN architecture performance.

TABLE VI. ONE-WAY ANOVA RESULTS FOR CNN ARCHITECTURE COMPARISON

Source of Variation	Sum of Squares	df	Mean Square	F	*p*-value	η^2
Between groups	4.812	4	1.203	38.72	<0.001	0.886
Within groups	0.621	20	0.031	-	-	-
Total	5.433	24	-	-	-	-

As shown in Table VI, the one-way ANOVA revealed statistically significant differences in classification accuracy among the evaluated CNN architectures, $F(4, 20) = 38.72$, $p < 0.001$. The effect size was large ($\eta^2 = 0.886$), indicating that the type of CNN architecture had a substantial influence on classification performance. These results support the selection of Xception as the most suitable model for deployment in the proposed IoT-based banknote recognition device.

B. Embedded Hardware Performance

The selected model was deployed on a Raspberry Pi-based embedded platform to evaluate its feasibility for assistive use. The system was assessed in terms of response time, inference stability, memory usage, and energy consumption. Although embedded inference required more time than GPU-based processing, the response time remained acceptable for real-world assistive interaction, since users typically require a short period to position and stabilize the banknote before capture.

The hardware evaluation confirmed that the proposed device can operate within the computational constraints of a low-cost embedded platform. The results also suggest that future optimization through TensorFlow Lite conversion and model quantization could further reduce latency, memory footprint, and power consumption.

The embedded implementation was evaluated on a Raspberry Pi 4B+ platform. The device achieved an average end-to-end response time of 3.82 ± 0.27 s, including image capture, preprocessing, CNN inference, and audio feedback. Average RAM usage during inference was 1.2 GB, average Central Processing Unit (CPU) utilization was 62%, and average power consumption was 4.8 W. Using a 5 V, 5000 mAh portable battery, the estimated continuous autonomy was approximately 3.2 h. These results indicate that the prototype is technically feasible for assistive use, although future optimization through TensorFlow Lite conversion and model quantization may further reduce latency and energy consumption.

TABLE VII. ROBUSTNESS EVALUATION UNDER ENVIRONMENTAL AND PHYSICAL VARIATIONS

Condition	Accuracy (%)
Normal lighting	99.8
Low lighting	97.9
Partial occlusion	96.4
Folded banknotes	97.2

As shown in Table VII, the Xception model maintained high recognition accuracy under all evaluated conditions. The highest performance was obtained under normal lighting, with 99.8% accuracy, while the lowest accuracy was observed under partial occlusion, with 96.4%. Despite this reduction, the model remained above 96% accuracy in all scenarios, indicating adequate robustness for real-world assistive use. These results suggest that the proposed device can maintain reliable performance when banknotes are captured under moderate environmental or physical variations.

A more detailed analysis of the embedded inference process is provided in Table A3. The proposed device achieved an average end-to-end response time of 3.82 ± 0.27 s, including preprocessing, Convolutional Neural

Network (CNN) inference, and audio feedback generation. The Xception model required 2.45 s for inference on the Raspberry Pi 4B+ platform, while the audio feedback process required 0.46 s. Resource utilization remained within the operational limits of the embedded platform, with an average Random-Access Memory (RAM) usage of 1.2 GB, peak RAM usage of 1.35 GB, and average CPU utilization of 62%.

Regarding energy consumption, the device required an average of 4.8 W during normal operation, with a peak consumption of 5.3 W during inference and 3.1 W in idle mode. Using a 5 V, 5000 mAh portable battery, the estimated continuous operating time was approximately 3.2 h, while intermittent real-world use could extend autonomy to 5–6 h depending on the frequency of interaction. These results confirm that the prototype is technically feasible for assistive use, although future optimization through TensorFlow Lite conversion and post-training quantization may further reduce latency and power consumption.

After selecting Xception as the final recognition model, its deployment was evaluated on the Raspberry Pi 4B+ platform to determine whether the proposed device could operate under embedded hardware constraints. The evaluation considered response time, model inference time, memory usage, power consumption, and estimated battery autonomy. Table VIII summarizes the key embedded hardware performance metrics.

TABLE VIII. KEY EMBEDDED HARDWARE PERFORMANCE METRICS

Parameter	Value
Average response time	3.82 s
Model inference time	2.45 s
RAM usage	1.2 GB
Average power consumption	4.8 W
Estimated battery autonomy	3.2 h

As shown in Table VIII, the proposed device achieved an average end-to-end response time of 3.82 s, including image capture, preprocessing, CNN inference, and audio feedback generation. The Xception model required 2.45 s for inference on the Raspberry Pi 4B+ platform. During operation, the system used approximately 1.2 GB of RAM and consumed an average of 4.8 W. Considering a 5 V, 5000 mAh portable battery, the estimated continuous battery autonomy was approximately 3.2 h. These results indicate that the prototype is technically feasible for assistive use, although future optimization strategies such as TensorFlow Lite conversion and model quantization may further reduce latency and energy consumption.

A more detailed breakdown of the embedded inference process, resource utilization, and power consumption is presented in Appendix. Table A3 extends the hardware performance analysis by reporting preprocessing time, model inference time, audio output delay, average and peak RAM usage, CPU utilization, average and peak power consumption, idle power consumption, and estimated battery autonomy under continuous and intermittent use. These complementary results support the technical feasibility of the prototype for assistive use and identify optimization opportunities for future versions.

C. Usability and User Experience Evaluation

The user evaluation showed favorable acceptance of the prototype among visually impaired participants. Usability results indicated that participants perceived the device as accessible, understandable, and useful for supporting banknote recognition. User experience results also reflected positive perceptions regarding attractiveness, novelty, and practical value.

These findings are important because assistive technologies must not only achieve technical accuracy but also be usable and acceptable for the target population. The integration of audio feedback, physical interaction, and ergonomic design contributed to the practical relevance of the proposed system.

The usability and user experience evaluation was conducted with 44 visually impaired participants. The UEQ results showed favorable perceptions of the device, particularly in attractiveness, novelty, and expectation fulfillment. In addition, the SUS score was above 80/100, indicating high perceived usability. These findings suggest that the proposed assistive device was positively accepted by the target users and that its audio feedback, ergonomic design, and simple interaction mechanism support its practical use in real-world contexts.

D. Economic Feasibility

The economic analysis showed that the device can be produced at a relatively low cost while maintaining financial feasibility. The estimated production cost, sale price, profit margin, net present value, and internal rate of return indicate that the system could be sustainable if scaled under appropriate production and distribution conditions. This strengthens the social and practical relevance of the proposal, particularly for users who require affordable assistive technologies.

Table IX shows that the assistance device is economically viable, since with a production cost of S/ 269.95 and a selling price of S/ 320.00, a profit margin of 10.02% is achieved, maintaining both accessibility and sustainability. Moreover, the Net Present Value (NPV) of S/ 13,013.02 confirms long-term profitability, while the Internal Rate of Return (IRR) of 11.85%, higher than the considered discount rate, reinforces the financial feasibility of the project, positioning it as an inclusive, low-cost devise with expansion potential to benefit people with visual impairments.

TABLE IX. COST/BENEFIT ANALYSIS OF THE DEVICE

Concept	Value
Production cost (S/)	269.95
Sale price (S/)	320.00
Profit margin	10.02%
NPV (S/)	13,013.02
IRR (%)	11.85

E. Comparative Analysis with Previous Studies

Compared with previous banknote recognition systems, the proposed device contributes not only through high classification performance but also through the integration of embedded deployment, ergonomic design, real-user validation, and economic feasibility analysis. While prior studies have mainly focused on algorithmic performance, this study provides a more comprehensive validation by considering technical, usability, and deployment-related dimensions.

Table X presents a comparative analysis of banknote recognition devices reported in previous studies. Prior works generally achieved high accuracy values, particularly those based on deep learning approaches. However, several studies do not report key deployment indicators such as dataset size, inference time, embedded hardware performance, or validation with real users. In contrast, the present study achieved 99.8% accuracy using a dataset of 9054 images, with an average embedded response time of 3.82 s on a Raspberry Pi 4B+ platform. Furthermore, unlike most previous studies, the proposed device was validated with 44 visually impaired users, strengthening the evidence of its practical applicability.

To complement the quantitative comparison presented in Table IV, Fig. 7 illustrates the accuracy achieved by each CNN architecture during the transfer learning and fine-tuning stages. This visual comparison allows a clearer identification of the best-performing model for the Peruvian banknote recognition task.

TABLE X. COMPARATIVE ANALYSIS OF BANKNOTE RECOGNITION DEVICES ACROSS DIFFERENT COUNTRIES

Study	Country	Accuracy (%)	Dataset Size	Inference Time	Hardware	Real Users
Awad *et al.* [25]	Iraq	98.6	Not reported	Not reported	PC	No
Fan *et al.* [35]	Ethiopia	96.4	Not reported	Not reported	Raspberry Pi	No
Sufri *et al.* [32]	Malaysia	97.2	~1500 images	Not reported	PC	No
Huang and Gai [34]	China	98.9	Not reported	Not reported	PC	No
Khalil *et al.* [37]	UAE	98.3	Not reported	Real-time (mobile)	Mobile device	No
Meshram *et al.* [42]	India	97.8	~3000 images	Not reported	PC	No
Silva *et al.* [39, 41]	Peru	-	3500 images	-	-	No
Present Study	Peru	99.8	9054 images	3.82 s	Raspberry Pi	Yes (44)

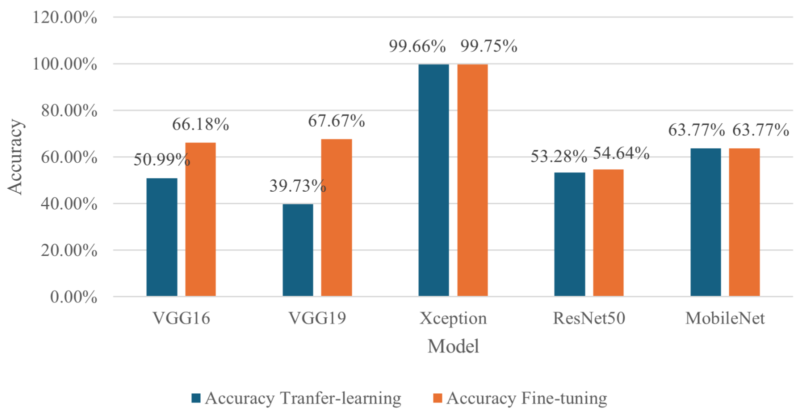

Fig. 7. Comparison of accuracy between transfer learning and fine-tuning.

As shown in Fig. 7, Xception achieved the highest accuracy in both stages, reaching 99.66% during transfer learning and 99.75% after fine-tuning. This performance was substantially higher than that of VGG16, VGG19, ResNet50, and MobileNetV2. Although fine-tuning improved the performance of most architectures, Xception remained the most accurate model. Therefore, Xception was selected as the final CNN architecture for deployment in the proposed IoT-based banknote recognition device.

F. Discussion

The results demonstrate that deep learning models can be effectively adapted to the recognition of Peruvian banknotes when supported by a locally constructed dataset and appropriate preprocessing strategies. The superior performance of the Xception architecture suggests that depthwise separable convolutions are suitable for extracting discriminative visual features from banknote images.

From an implementation perspective, the proposed system advances beyond purely algorithmic approaches by integrating computer vision into a portable IoT-based assistive device. This is particularly relevant for visually impaired users, for whom accessibility, audio feedback, portability, and simplicity of interaction are essential design requirements.

However, the results should be interpreted considering the specific scope of the study. The device was designed for Peruvian banknotes and does not currently detect counterfeit, severely damaged, or highly worn banknotes. In addition, although the embedded implementation is functional, further optimization is required to reduce inference latency and extend battery autonomy.

G. Limitations and Future Work

The main limitations of the study are related to currency scope, hardware optimization, and real-world variability. The current version recognizes selected Peruvian banknote denominations and does not include counterfeit detection. Furthermore, although the prototype is functional, future versions should reduce device size, improve portability, and optimize embedded inference.

Future work will focus on converting the trained model to TensorFlow Lite, applying post-training quantization, reducing computational load, and improving battery autonomy. Additional studies should also evaluate the device with larger and more diverse user groups, as well as under broader real-world conditions involving different lighting, worn banknotes, and extended daily use.

VI. CONCLUSIONS

This study presented the design, implementation, and validation of a low-cost IoT-based assistive device for Peruvian banknote recognition using deep learning on an embedded platform. A computer vision model was trained and evaluated using a locally constructed dataset of Peruvian banknotes, and the best-performing architecture was integrated into a Raspberry Pi-based prototype with audio feedback.

The results demonstrated that the proposed device can accurately recognize Peruvian banknotes while offering a practical and accessible interaction mechanism for visually impaired users. The integration of CNN-based classification, embedded processing, ergonomic design, and audio output supports the feasibility of the system as an assistive technology for promoting financial autonomy.

In addition to technical performance, the study included usability evaluation and economic feasibility analysis, strengthening the applied relevance of the proposal. Nevertheless, limitations remain regarding counterfeit detection, recognition of severely damaged banknotes, embedded inference latency, and battery autonomy. Future work will focus on model optimization, hardware miniaturization, broader user validation, and the incorporation of additional recognition capabilities.

APPENDIX A. IMAGE CAPTURE AND DATASET PREPARATION DETAILS

TABLE A1. CAPTURE EQUIPMENT SPECIFICATIONS

Feature	Description
Processor	Snapdragon 685
RAM	8 GB LPDDR4X
Storage	128 GB
Front Camera	13 MP f/2.45
Rear Cameras	50 MP f/1.8 (main), 8 MP f/2.2 (wide), 2 MP f/2.4 (macro)
Operating System	Android 12 + MIUI 14

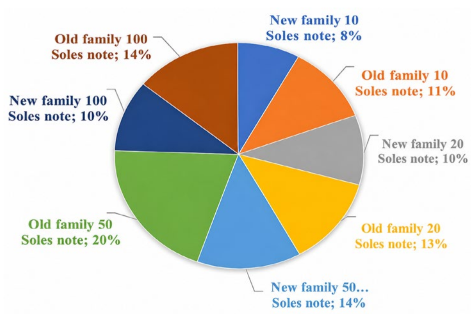

Fig. A1. Percentage distribution per denomination.

APPENDIX B. CNN TRAINING CONFIGURATION

TABLE A2. TRAINING CONFIGURATION AND HYPERPARAMETERS OF THE EVALUATED CNN MODELS

Parameter	Value
Input Shape	224 × 224 × 3
Optimizer	Adam (lr = 0.0001)
Batch Size	64
Epochs	100
Fine-tuning	Top 2 layers
Loss Function	Categorical Crossentropy
Main Metric	Accuracy

APPENDIX C. EMBEDDED HARDWARE AND ENERGY CONSUMPTION DETAILS

TABLE A3. DETAILED EMBEDDED INFERENCE, RESOURCE UTILIZATION, AND POWER CONSUMPTION ANALYSIS

Parameter	Value	Description
Preprocessing time	0.91 s	Time required for image acquisition, resizing, normalization, and preparation before inference.
Model inference time	2.45 s	Time required by the Xception model to classify the captured banknote image.
Audio output delay	0.46 s	Time required to generate and play the audio feedback after classification.
Average response time	3.82 ± 0.27 s	End-to-end time from image capture to audio feedback generation.
RAM usage during inference	1.2 GB	Average memory consumption during the execution of the recognition model.
Peak RAM usage	1.35 GB	Maximum memory consumption observed during continuous operation.
Average CPU usage	62%	Average processor utilization during embedded inference.
Average power consumption	4.8 W	Mean power consumption during normal device operation.
Peak power consumption	5.3 W	Maximum power consumption observed during CNN inference.
Idle power consumption	3.1 W	Power consumption when the device is active but not performing inference.
Estimated continuous battery autonomy	3.2 h	Estimated operating time using a 5 V, 5000 mAh portable battery under continuous use.
Estimated intermittent battery autonomy	5–6 h	Expected operating time under real-world intermittent use conditions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

B.M.-Q. conceived and designed the study, coordinated the research process, and led the writing of the manuscript. O.M.A.-O. and R.C.P.-V. contributed to the theoretical framework, methodological design, and critical revision of the manuscript. R.C.P.-V. and V.H.R.-I. participated in data collection, data organization, and preliminary analysis. G.F.-Y. and A.C.M.-L. supported the interpretation of results and contributed to the discussion section. N.E.C.-S. assisted in statistical analysis and visualization of results, while E.G.C.-G. and J.M.P.-A. contributed to the technical validation and review of the study's findings. All authors reviewed, approved, and agreed to the final version of the manuscript.

REFERENCES

[1] World Blind Union. (2023). World Blind Union. [Online]. Available: https://worldblindunion.org/es/

[2] Ministry of Health of Peru. (2014). In Peru, around 160,000 people are blind and some 600,000 suffer from some form of visual impairment. [Online]. Available: https://www.gob.pe/institucion/minsa/noticias/31145-en-el-peru-cerca-de-160-000-personas-son-invidentes-y-unas-600-000-sufren-de-alguna-discapacidad-visual

[3] National Institute of Statistics and Informatics. (2014). First specialized national survey on disability 2012. [Online]. Available: https://proyectos.inei.gob.pe/microdatos/

[4] C. A. G. Gavilánez and H. X. H. Orbea, "Development of an electronic prototype to guide blind people in the collection of objects through audio and computer vision," Bachelor's Thesis, Electronic Engineering, Universidad Politécnica Salesiana, 2019.

[5] R. V. Salazar and A. C. Mesa, "Assistive devices for mobility in people with visual disabilities: A literature review," *Revista Politécnica*, vol. 15, no. 28, pp. 107–116, 2019. doi: 10.33571/rpolitec.v15n28a1

[6] C. R. L. González, "Proposal for the design of an electronic cane for blind people to improve the quality of their daily mobility," Bachelor's Thesis, Industrial Engineering, Universidad Continental, Arequipa, Peru, 2018.

[7] R. Fernandez, "Smart geolocating sensory cane to support the mobility of blind people in the regional organization of the blind of Peru–Chiclayo," Ph.D. Thesis, Universidad Católica Santo Toribio de Mogrovejo, 2021.

[8] S. Vargas, "Prototype of glasses for obstacle detection through vibrations for blind people based on Arduino technology," Ph.D. Thesis, Computer Science, 2020.

[9] A. Hernández, "Assistance devise for visually impaired people based on object detection with deep learning," M.S. Thesis, University of Chile, 2022.

[10] I. Soderback, I*nternational Handbook of Occupational Therapy Interventions*, 2nd ed. Springer, 2015.

[11] M. Beheshti *et al.*, "A smart service devise for spatial intelligence and onboard navigation for individuals with visual impairment (vis4ion Thailand): Study protocol of a randomized controlled trial of visually impaired students at the Ratchasuda College, Thailand," *Trials*, vol. 24, no. 1, 2023. doi: 10.1186/s13063-023-07173-8

[12] National Organization of Spanish Blind People, *Visual Disability and Personal Autonomy*, 1st ed., 2011. (in Spanish)

[13] J. I. García and D. P. Cañadas, "The education of blind and visually impaired people: Historical vision of an inclusion process," in *Proc. El Largo Camino Hacia una Educación Inclusiva: La Educación Especial y Social del Siglo XIX a Nuestros díAs: XV Coloquio de Historia de la Educación*, vol. 1, pp. 453–454, 2009. (in Spanish)

[14] M. E. P. Parea and C. S. Añapa, "Contribution in the use of ICTs for visually impaired people through the implementation of the JAWS program and screen magnifier in cybercafes of the Esmeraldas parish, during 2011, and evaluation of the results of the pilot project," M.S. Thesis, Special Education, Universidad Politécnica Salesiana, 2013.

[15] A. D. M. Delgado, "White cane to prevent obstacles," Bachelor's Thesis, Control and Automation Engineering, Instituto Politécnico Nacional, 2012.

[16] P. J. R. López, "Analysis, design and development of a comprehensive support system for mobility and interactive access to Internet information resources for blind people, based on mobile technologies," Bachelor's Thesis, Systems Engineering, Universidad Politécnica Salesiana, 2015.

[17] J. L. M. Moya and G. J. Á. Reasco, "Construction of an electronic cane for blind people in La Maná Canton," Bachelor's Thesis, Computer Engineering and Computational Systems, Universidad Técnica de Cotopaxi, 2017.

[18] B. F. D. Torres, "VIX—Smart cane for visually impaired people," Bachelor's Thesis, Communication Design, Universidad San Francisco de Quito, 2016.

[19] L. E. F. Ortiz Fernandez, "Object identification system for blind people using RFID technology," *Ingenius, Revista De Ciencia Y Tecnologia*, vol. 8, pp. 38–44, 2012.

[20] M. A. P. Mosquera and J. V. U. Terreros, "Assistance and guidance system for blind people," Bachelor's Thesis, Universidad Autónoma de Occidente, Santiago de Cali, Colombia, 2014.

[21] C. Hernandez, L. Pedraza, and D. López, "Technological device for optimizing Braille learning time in blind people," *Revista de Salud Pública,* vol. 13, pp. 1351–1358, 2011. (in Spanish)

[22] H. Shah, M. Amin, K. Dadwani, N. Desai, and A. Chatiwala, "Real-time object detection devise with voice feedback for the blind people," in *Smart Trends in Computing and Communications*, Y.-D. Zhang, T. Senjyu, C. So-In, and A. Joshi, Eds. Singapore: Springer, 2023, pp. 683–690.

[23] S. Rusli, S. Liawatimena, and A. Trisetyarso, "Context aware navigation devise for assisting blind people," *ICIC Express Letters*, vol. 16, pp. 1351–1358, 2022. doi: 10.24507/icicel.16.12.1351

[24] M. Mukhiddinov, A. Abdusalomov, and J. Cho, "Automatic fire detection and notification devise based on improved YOLOv4 for the blind and visually impaired," *Sensors*, vol. 22, no. 9, 3307, 2022. doi: 10.3390/s22093307

[25] S. R. Awad, B. T. Sharef, A. M. Salih, and F. L. Malallah, "Deep learning-based Iraqi banknotes classification devise for blind people," *Eastern-European Journal of Enterprise Technologies*, vol. 1, pp. 31–38, 2022. doi: 10.15587/1729-4061.2022.248642

[26] K. Suganya, G. Prabaharan, A. Prathik, N. Manjunathan, and R. Vijayabharathi, "Designing a deep learning approach to categorize rupee notes and assist the visually impaired members to recognize different denominations," *AIP Conference Proceedings*, vol. 3449, 020269, 2026. doi: 10.1063/5.0301127

[27] K. Morparia, A. Kanabar, S. Bhadoriya, and P. Kadam, "Face recognition with audio output: An aid for the visually impaired," in *Proc. 2022 4th Int. Conf. on Inventive Research in Computing Applications (ICIRCA)*, 2022, pp. 1519–1524. doi: 10.1109/ICIRCA54612.2022.9985550

[28] M. M. Valipoor and A. de Antonio, "Recent trends in computer vision-driven scene understanding for VI/blind users: A deviseatic mapping," *Universal Access in the Information Society*, vol. 22, no. 3, pp. 983–1005, 2022. doi: 10.1007/s10209-022-00868-w

[29] D. Croce, L. Giarrè, F. Pascucci, I. Tinnirello, G. E. Galioto, D. Garlisi, and A. Lo Valvo, "An indoor and outdoor navigation devise for visually impaired people," *IEEE Access*, vol. 7, pp. 170406–170418, 2019. doi: 10.1109/ACCESS.2019.2955046

[30] U. Das, V. Namboodiri, and H. He, "Pathlookup: A deep learning-based framework to assist visually impaired in outdoor wayfinding," in *Proc. 2021 IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2021, pp. 111–116. doi: 10.1109/PerComWorkshops51409.2021.9431007

[31] G. Portilla. (2022). The CSIC develops a robot-dog to guide dependent or disabled people. [Online]. Available: https://www.csic.es/

[32] N. A. J. Sufri, N. A. Rahmad, N. F. Ghazali, N. Shahar, and M. A. As'ari, "Vision based devise for banknote recognition using different machine learning and deep learning approach," in *Proc. 2019 IEEE 10th Control and Devise Graduate Research Colloquium (ICSGRC)*, 2019, pp. 5–8. doi: 10.1109/ICSGRC.2019.8837068

[33] S. Castelar, L. A. Banegas, D. A. Mendoza, J. C. Soto, and K. Davila, "Automated Honduran banknote image classification using machine learning," in *Proc. 2022 IEEE 40th Central America and Panama Convention (CONCAPAN)*, 2022, pp. 1–6. doi: 10.1109/CONCAPAN48024.2022.9997703

[34] X. Huang and S. Gai, "Banknote classification based on convolutional neural network in quaternion wavelet domain," *IEEE Access*, vol. 8, pp. 162141–162148, 2020. doi: 10.1109/ACCESS.2020.3021181

[35] C.-B. Fan, D. T. Aseffa, H. Kalla, and S. Mishra, "Ethiopian banknote recognition using convolutional neural network and its prototype development using embedded platform," *Journal of Sensors*, vol. 2022, 4505089, 2022. doi: 10.1155/2022/4505089

[36] K. J. T. Zapata, "Banknote recognition devise for visually impaired people using computer vision," Bachelor's Thesis, Mechatronics Engineering, Universidad Escuela de Ingeniería de Antioquia, Envigado, 2018.

[37] A. Khalil, M. Yaghi, T. Basmaji, M. Faizal, Z. Farhan, A. Ali, and M. Ghazal, "Mobile deep classification of UAE banknotes for the visually challenged," in *Proc. 2022 9th Int. Conf. on Future Internet of Things and Cloud (FiCloud)*, 2022, pp. 321–325. doi: 10.1109/FiCloud57274.2022.00053

[38] P. Geerts, J. Eijsink, A. Moser, P. ter Horst, C. Boersma, and M. Postma, "Rationale and development of an e-health application to deliver patient-centered care during treatment for recently diagnosed multiple myeloma patients: Pilot study of the MM e-coach," *Pilot and Feasibility Studies*, vol. 9, no. 1, 85, 2023.

[39] N. E. C. Silva, J. M. P. Alejandro, and E. G. Castro Gutierrez. (September 2023). Dataset of Peruvian banknotes. *IEEE Dataport*. [Online]. Available: https://ieee-dataport.org/documents/dataset-peruvian-banknotes

[40] K. C. Shahira and A. Lijiya, "Towards assisting the visually impaired: A review on techniques for decoding the visual data from chart images," *IEEE Access*, vol. 9, pp. 52926–52943, 2021. doi: 10.1109/ACCESS.2021.3069205

[41] N. E. C. Silva, J. M. P. Alejandro, and E. G. C. Gutierrez. (September 2023). Peruvian banknotes. *Mendeley Data*. [Online]. Available: https://data.mendeley.com/datasets/22jmt8xhpn/1

[42] V. Meshram, K. Patil, and V. Meshram, "Evaluation of top pretrained models using transfer learning on banknote dataset with quality parameter," *Ingénierie des Systèmes d'Information*, vol. 28, no. 3, pp. 693–701, 2023. doi: 10.18280/isi.280319