

Article

Optimizing 1001-Class Handwritten Digit Sequence Recognition for the Mexican Electoral Process Using Asymmetric ResNet-18 and CBAM

Miguel Angel Camargo-Rojas, Gabriel Sánchez-Pérez , José Portillo-Portillo * , Linda Karina Toscano-Medina * , Aldo Hernández-Suárez , Jesús Olivares-Mercado  and Héctor Manuel Pérez-Meana 

ESIME Culhuacan, Instituto Politécnico Nacional (National Polytechnic Institute), Mexico City 04440, Mexico; mcamargor1601@alumno.ipn.mx (M.A.C.-R); gasanchezp@ipn.mx (G.S.-P.); alhernandezsu@ipn.mx (A.H.-S.); jolivares@ipn.mx (J.O.-M.); hmperez@ipn.mx (H.M.P.-M.)

* Correspondence: jportillo@ipn.mx (J.P.-P.); ltoscano@ipn.mx (L.K.T.-M.)

Abstract

The automated digitization of handwritten electoral results is critical for ensuring transparency and speed in democratic processes. While recurrent sequence-to-sequence models (e.g., CRNN+CTC) achieve high accuracy, they inherently violate the strict latency constraints of high-throughput administrative environments. Conversely, standard lightweight CNNs exhibit suboptimal performance on the long-tail distribution of high-cardinality scenarios. To bridge this gap, this study reformulates sequence recognition into a latency-bound classification task. We propose a specialized Handwritten Digit Sequence Recognition (HDSR) framework for the Mexican Preliminary Election Results Program (PREP) based on a modified ResNet-18 architecture. The methodology introduces an asymmetric stride designed explicitly to preserve the 1:3 horizontal feature resolution of electoral tally sheets, integrating a lightweight Convolutional Block Attention Module (CBAM) in deep stages to refine classification across 1001 possible sequences. Leveraging a megadiverse dataset of 3.77 million real-world images, the model was trained using AdamW and label smoothing to mitigate human-induced label noise. Results demonstrate a global accuracy of 97.82% and a significant improvement in Macro-Precision (0.8878) for rare sequences. With an inference latency of 9.1 ms on standard CPU hardware, the proposed solution offers a scalable, high-confidence alternative that prioritizes spatial preservation and fail-controlled deployment.

Keywords: handwritten digit sequence recognition; asymmetric strides; Convolutional Block Attention Module (CBAM); long-tail distribution; high-cardinality classification; Preliminary Electoral Results Program (PREP)



Academic Editors: Honggang Chen and Chao Ren

Received: 14 May 2026

Revised: 4 July 2026

Accepted: 7 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The digitization of electoral processes is a cornerstone of modern democracy, ensuring transparency and the timely delivery of results. In Mexico, the National Electoral Institute (INE) [1] is the autonomous public institution responsible for organizing federal elections. A key component of this infrastructure is the Preliminary Electoral Results Program (PREP), a system designed to provide real-time preliminary information on Election Day by capturing data from *Actas de Escrutinio y Cómputo* (AEC, vote counting and tallying reports) [2].

The AEC forms, completed by polling station officials, are structured in rows where each entry typically sequences a political party logo, the vote count written in words, and the numerical digits. While each row contains redundant information, this study focuses exclusively on the automated recognition of the digit-based sequences (0–9). Within the PREP operation, these forms are digitized at Data Collection and Transmission Centers (CATD) into segmented Regions of Interest (ROIs). Under an inter-institutional collaboration agreement between the INE and the National Polytechnic Institute (IPN), this research utilizes these standardized ROIs to ensure the proposed system can operate in parallel with existing infrastructure, effectively preventing transcription bias and ensuring the integrity of the manual data entry process.

A significant challenge in automating this process lies in Handwritten Digit Sequence Recognition (HDSR). Existing HDSR approaches reported in the literature predominantly rely on unconstrained sequence-to-sequence modeling. While state-of-the-art paradigms such as Vision Transformers (ViTs) or Convolutional Recurrent Neural Networks with Connectionist Temporal Classification (CRNN+CTC) excel in these unconstrained scenarios, they introduce significant parameter overhead and decoding latency. Our proposed method offers a domain-specific alternative to these traditional HDSR frameworks by reformulating the problem as a latency-bound classification architecture. Given the strict operational requirement of maintaining inference latency below 10 ms on standard CPU hardware at the CATDs, conventional sequential models are currently prohibitive.

Consequently, this study isolates the problem as a purely convolutional, high-cardinality classification task. Our approach addresses a megadiverse dataset of 3.77 million ROI images categorized into 1001 distinct classes. Building upon our foundational evaluations of the PREP system records [3], we selected a modified ResNet-18 backbone [4]. In our prior comparative analyses, ResNet architectures demonstrated competitive representational capacity (achieving substantially higher Macro F1 scores), mitigating the representational collapse observed in ultra-lightweight models (e.g., MobileNetV3 [5], EfficientNet [6], and ShuffleNetV2 [7]) when confronted with extreme long-tail distributions.

By standardizing input images to a rectangular 64×192 pixel resolution, we introduce a strongly asymmetric stride within the feature extraction layers. This geometric preservation, synergized with the Convolutional Block Attention Module (CBAM) [8], maintains horizontal spatial information essential for distinguishing dense three-digit sequences without losing resolution due to uniform downsampling. Naturally, deploying this architecture on large-scale, real-world datasets involves inherent trade-offs. While the model maximizes global Accuracy, Macro-Precision is heavily influenced by the long-tail distribution and human-induced label noise. Under these adversarial conditions, our asymmetric approach prioritizes total error reduction, which is critical for high-throughput electoral systems.

To address the limitations of existing methodologies within this operational context, the specific adaptations of this study are summarized as follows:

- **Architectural Reformulation for Strict Latency Constraints:** Unlike traditional OCR and CRNN+CTC paradigms that rely on temporally sequential decoding, we approach bounded sequence recognition as a high-cardinality discrete classification problem (1001 classes). This shift effectively bypasses recurrent bottlenecks, supporting a sub-10 ms inference latency on standard CPU hardware (9.1 ms) while maintaining representational capacity.
- **Domain-Specific Geometric Preservation:** By engineering a progressive $S = (2, 1)$ stride configuration, the network is adapted to preserve the 1:3 horizontal aspect ratio of three-digit sequences, mitigating the morphological blending of adjacent characters that typically affects standard symmetric ResNet architectures.

- **Deep-Stage Attention for Long-Tail Extremes:** Rather than employing generic attention mechanisms across all layers, we isolate the Convolutional Block Attention Module (CBAM) to deep semantic stages. This focuses the network to filter background artifacts and human-induced noise, stabilizing Macro-Precision across the extreme long-tail distribution of the electoral dataset.

In line with transparency standards, the source code for this project is made publicly available, while access to the dataset remains restricted due to the aforementioned legal agreements with the INE.

2. Related Work

Modern deep learning algorithms have been successfully deployed across a wide spectrum of applications, ranging from complex pattern recognition in diverse scientific fields as in [9] to the adaptation of pre-trained models via transfer learning and fine-tuning as demonstrated in [10]. However, addressing the strict operational constraints of electoral digitization requires a departure from these generalized approaches. In alignment with the institutional requirements of the INE, this study prioritizes a custom training paradigm from scratch. This necessity drives our selection and structural modification of the ResNet-18 backbone, ensuring the architecture is explicitly optimized for handwritten tally sheets rather than inheriting biases from generic pre-training.

2.1. Evolution of Sequence Recognition Architectures

Standard Convolutional Neural Networks (CNNs) have demonstrated remarkable effectiveness in image classification tasks by leveraging hierarchical spatial feature extraction. Early breakthroughs in multi-digit recognition, such as the direct transcription of the Street View House Numbers (SVHN) dataset [11], demonstrated that deep CNNs could successfully process bounded digit sequences. However, their broader application to unconstrained, variable-length sequence recognition presents fundamental limitations due to their strict requirement for fixed-length inputs and explicit character segmentation [12]. This segmentation-first paradigm introduces a critical bottleneck: inaccuracies in dividing text images cascade into overall recognition failures, particularly in scripts with variable morphological structures [13].

To overcome these limitations, the field converged on Convolutional Recurrent Neural Networks (CRNNs). By combining CNNs with Recurrent Neural Networks (RNNs) and Connectionist Temporal Classification (CTC) loss [14], CRNNs enabled end-to-end trainable pipelines that eliminated the need for explicit character-level alignment [15]. While this hybrid paradigm achieved strong performance across diverse unconstrained recognition benchmarks, the recurrent component introduces an inherent sequential processing bottleneck. Because RNNs process input sequences step by step, each time step depends on the previous hidden state, precluding parallelization across the temporal dimension. This sequential dependency limits computational throughput and significantly increases inference latency [16].

Recently, the limitations of recurrent architectures have motivated the adoption of Transformer-based models [17]. By leveraging self-attention mechanisms, Transformers capture global contextual dependencies across entire input sequences simultaneously, outperforming conventional CRNN+CTC models on both regular and irregular text recognition benchmarks [18]. However, despite their superior global context modeling, Transformer-based frameworks introduce substantial computational overhead. The self-attention mechanism requires computing pairwise token interactions, resulting in quadratic complexity with respect to sequence length. Consequently, deploying these state-of-the-art architectures in resource-constrained or high-throughput environments remains an open challenge,

driving current research toward lightweight attention alternatives that can balance representation capacity with inference efficiency.

2.2. Data Imbalance and Model Calibration

In real-world image classification tasks, datasets inherently exhibit long-tailed distributions, wherein a small number of head classes contain abundant samples while the majority of tail classes are severely underrepresented. This class imbalance fundamentally degrades the performance of conventional deep neural networks, which become biased toward learning majority class features while neglecting minority class representations [19]. As demonstrated across multiple complex domains, from medical imaging to remote sensing, standard cross-entropy loss trained on imbalanced data often leads to minority class collapse, wherein tail categories are systematically misclassified [20].

To counteract frequency bias, the research community has converged on two primary mitigation strategies: class-balanced loss functions and advanced data augmentation. Cost-sensitive approaches and formulations like Focal Loss address this issue by down-weighting well-classified instances and heavily penalizing errors on hard, underrepresented samples [21]. Concurrently, data augmentation strategies, including Mixup and semantic feature-space variations, dynamically adjust class representation during training to increase feature diversity for tail classes without the risk of naive overfitting [22].

Beyond class distribution challenges, deploying deep learning architectures in high-stakes environments requires strict predictive reliability. Modern neural networks trained with standard cross-entropy loss are well-documented to produce overconfident misclassifications, wherein predicted probability scores substantially exceed empirical accuracy [23]. This miscalibration is particularly dangerous in safety-critical domains where confidently wrong decisions can propagate downstream with severe consequences [24]. In the specific context of electoral document processing, automated Optical Character Recognition (OCR) of tally sheets significantly reduces human transcription errors and accelerates vote tabulation [25]. However, for critical infrastructure such as the Preliminary Electoral Results Program (PREP), silent failures are entirely unacceptable.

To mitigate the risk of overconfident errors, recent literature emphasizes the necessity of model calibration and selective classification. Regularization strategies such as Label Smoothing [26] prevent overconfidence during training by softening hard target distributions, while metrics like the Expected Calibration Error (ECE) formally quantify the alignment between predictive confidence and true accuracy. Integrating this calibrated uncertainty into human-in-the-loop pipelines enables the design of fail-controlled architectures, which proactively transform potential misclassifications into system rejections [27]. Within our proposed framework, this selective mechanism ensures that only highly confident predictions are automated, whereas ambiguous sequences on electoral records trigger systematic rejection for manual verification [28].

3. Materials and Methods

This section describes the comprehensive methodology employed to address the high-cardinality task of Handwritten Digit Sequence Recognition (HDSR) within the context of the Mexican Preliminary Electoral Results Program (PREP). The methodology is structured into four main components: first, we detail the characteristics of the megadiverse dataset and its pre-processing; second, we present the proposed architectural modifications, including the mathematical formulation of the asymmetric strides and CBAM integration; third, we describe the implementation details, including hardware, software, and the specific training protocol; and finally, we define the evaluation metrics used to validate the model's operational reliability.

3.1. Operational Context: The Manual Transcription Protocol

The current PREP infrastructure utilizes a redundant manual transcription protocol requiring double- or triple-blind data entry by geographically distributed clerks. While this decentralized consensus system prevents local collusion and ensures operational resilience, it significantly increases logistical complexity and processing latency. This inherent bottleneck underscores the critical need for a high-performance, automated HDSR system capable of matching human-level precision with minimal computational overhead.

3.2. Dataset Characterization and Pre-Processing

The performance of the proposed architecture is evaluated using a large-scale, megadiverse dataset provided by the National Electoral Institute (*Instituto Nacional Electoral*, INE) of Mexico. The images originate from the 2017–2018 Federal Electoral Process, encompassing a diverse range of documents from presidential, senatorial, and deputy elections. The complete dataset comprises a total of 3,776,996 handwritten digit sequence images extracted from the AEC. This corpus is partitioned into a training set of 2,936,082 samples and a dedicated test set of 840,914 samples, ensuring a robust evaluation of the model's generalization capabilities on unseen data across various types of electoral forms. A fundamental characteristic of this entire corpus is its extreme class imbalance, which stems directly from the natural distribution of real-world voting patterns. As illustrated by the training partition in Figure 1, the data exhibits a severe long-tail distribution, presenting a significant high-cardinality challenge for automated recognition.

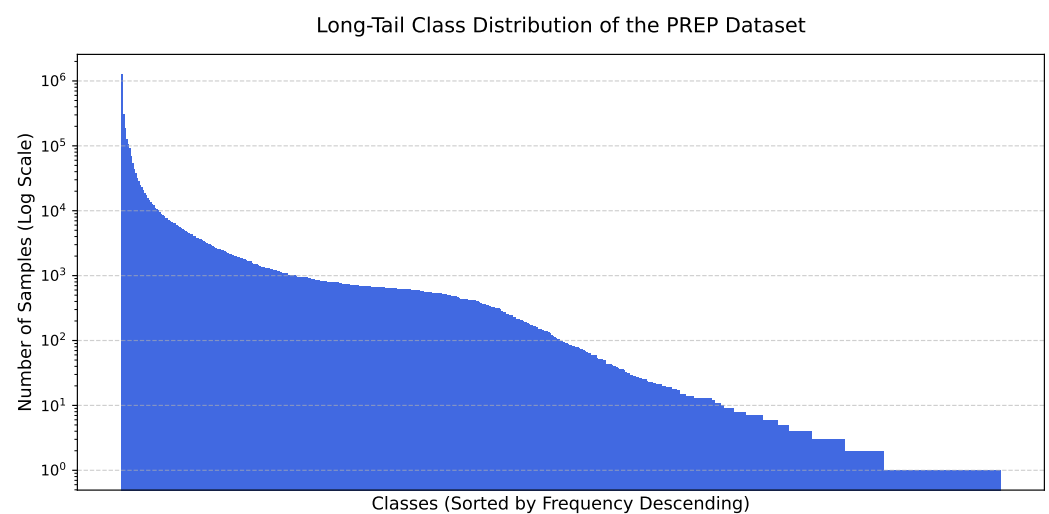


Figure 1. Long-tail class distribution of the PREP training set. Due to the high cardinality of the dataset (1001 possible sequence classes), individual labels are omitted from the x-axis for visual clarity. Instead, the classes are sorted by frequency in descending order, with the y-axis presented in a logarithmic scale. To contextualize the severity of the imbalance, the most frequent class corresponds to the sequence representing zero votes ('000') with 1,271,164 samples, while the extreme tail consists of several rare classes (e.g., '801') containing only a single sample. Furthermore, there are 420 theoretical classes with zero samples; this absence is not a sampling defect, but a direct reflection of the physical constraints of the Mexican electoral system (e.g., the statutory limit of ballots per polling station inherently precludes the occurrence of highly inflated vote sequences).

Given the extreme long-tail distribution of the 1001 classes, partitioning the dataset into a traditional three-way split (train/validation/test) was deliberately avoided. Creating an independent validation set would have severely compromised the representational capacity of the minority classes (e.g., classes with fewer than three samples), effectively forcing them into either zero-shot training scenarios or complete absence from the evalua-

tion metrics. For identical reasons, resampling techniques such as k-fold cross-validation were excluded; beyond the prohibitive computational cost of repeatedly retraining a deep architecture on 3.77 million images, stratified folding is mathematically impossible for tail classes containing only one or two samples. Consequently, a strict, mutually exclusive two-way split was maintained. The test set was rigorously isolated to prevent any potential data leakage and used exclusively for final performance benchmarking, while hyperparameter configuration was guided by established architectural baselines and empirical observation of the training loss convergence.

The task is framed as a high-cardinality classification problem consisting of 1001 distinct classes. These classes represent every possible integer sequence ranging from 0 to 999, with an additional “blank” category to account for Regions of Interest (ROIs) where no numerical data was recorded. A defining characteristic of this dataset is its inherent long-tail distribution; due to the nature of electoral voting patterns, lower numerical values appear with significantly higher frequency, while high-value classes are sparsely represented. Furthermore, as these images are captured under diverse operational conditions at Data Collection and Transmission Centers (*Centros de Acopio y Transmisión de Datos*, CATD), the dataset includes a non-negligible amount of label noise and significant variability in handwriting styles.

To standardize the input for the feature extraction backbone, a rigorous pre-processing pipeline was implemented. Each segmented ROI is resized to a fixed resolution of 64×192 pixels using bi-linear interpolation. This specific 1:3 aspect ratio was selected to preserve the integrity of the three-digit horizontal sequences, preventing the distortion of morphological features. Subsequent to resizing, pixel values are normalized to a range of $[0, 1]$ to facilitate gradient stability during the training phase. This standardized rectangular geometry serves as the primary justification for the proposed asymmetric stride modification, ensuring that horizontal spatial information is preserved throughout the downsampling process in the convolutional stages.

3.3. Proposed Architecture and Attention Mechanism

The proposed architecture utilizes a modified ResNet-18 as the feature extraction backbone. The model initiates with a preliminary feature extraction stem, adhering to the standard ResNet-18 layer sequence: a 7×7 convolutional layer with a symmetric stride $S = 2$, followed by Batch Normalization, ReLU activation, and a 3×3 MaxPooling operation. To preserve high-resolution spatial cues for the subsequent stages, the MaxPooling layer is configured with a stride of 1, maintaining the feature map resolution at 32×96 after the initial downsampling from the 64×192 input.

The core structural contribution is implemented within the four residual stages that follow the stem. In these stages, the downsampling strategy is re-engineered through a progressive stride configuration $S^{(i)} = (s_v^{(i)}, s_h^{(i)})$ for each stage $i \in \{1, 2, 3, 4\}$, as defined below:

- **Stage 1:** Utilizes an asymmetric stride $S^{(1)} = (2, 1)$, reducing vertical resolution to 16 pixels while preserving the full 96-pixel horizontal resolution to maintain character boundary integrity.
- **Stages 2 and 3:** Employ standard symmetric strides $S^{(2,3)} = (2, 2)$, gradually compressing both dimensions to 8×48 and 4×24 , respectively.
- **Stage 4:** Implements an aggressive symmetric stride $S^{(4)} = (4, 4)$ to force a rapid spatial collapse, facilitating high-level semantic abstraction before the final classification head.

The output dimensions (H_{out}, W_{out}) for any convolutional layer within these stages are determined by:

$$H_{out} = \left\lfloor \frac{H_{in} + 2p - k}{s_v^{(i)}} \right\rfloor + 1, \quad W_{out} = \left\lfloor \frac{W_{in} + 2p - k}{s_h^{(i)}} \right\rfloor + 1 \quad (1)$$

where k denotes the kernel size, p the padding, and $s^{(i)}$ the specific stride components for each stage.

Furthermore, to enhance the representation of long-tail classes, the Convolutional Block Attention Module (CBAM) is integrated into Stages 3 and 4. The attention mechanism refines the feature maps F through sequential channel and spatial attention gates:

$$F' = M_c(F) \otimes F; \quad F'' = M_s(F') \otimes F' \quad (2)$$

where the attention maps M_c (channel) and M_s (spatial) are computed as:

$$\begin{aligned} M_c(F) &= \sigma(W_1(\text{ReLU}(W_0(F_{avg}^c))) + W_2(\text{ReLU}(W_0(F_{max}^c)))) \\ M_s(F') &= \sigma(f^{7 \times 7}([\text{AvgPool}(F'); \text{MaxPool}(F')])) \end{aligned} \quad (3)$$

This hierarchical design ensures that horizontal resolution is prioritized in shallow layers for sequence integrity, while semantic attention is focused in deeper layers for classification robustness. The complete layer-by-layer configuration of the network, including the asymmetric downsampling operations and the precise location of the attention module, is detailed in Table 1.

Table 1. Detailed architectural specifications of the proposed Asymmetric ResNet-18 with CBAM (A-R18 + CBAM). Spatial dimensions are denoted as Height $\times$ Width. Downsampling occurs at the first convolutional layer of each stage using the specified stride.

Stage	Input Size	Output Size	Block Type	Kernel, Padding	Stride (s_v, s_h)	Channels
Input	64×192	64×192	–	–	–	3
Stem (Conv1)	64×192	32×96	Conv2D + BN + ReLU	$7 \times 7, p = 3$	(2, 2)	64
Stem (Pool)	32×96	32×96	MaxPool2D	$3 \times 3, p = 1$	(1, 1)	64
Stage 1	32×96	16×96	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 2$	$3 \times 3, p = 1$	(2, 1)	64
Stage 2	16×96	8×48	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 2$	$3 \times 3, p = 1$	(2, 2)	128
Stage 3	8×48	4×24	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 2 +$ CBAM	$3 \times 3, p = 1$	(2, 2)	256
Stage 4	4×24	1×6	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 2 +$ CBAM	$3 \times 3, p = 1$	(4, 4)	512
Classifier	1×6	1×1	AdaptiveAvgPool2D	–	–	512
Output	1×1	–	Fully Connected	–	–	1001

Note: Residual projection branches (1×1 Convolutions) are applied at the beginning of Stages 2, 3, and 4 to match channel dimensions. CBAM modules (Reduction Ratio $r = 16$) are sequentially applied at the terminal end of the residual blocks in Stages 3 and 4.

3.4. Implementation Details and Training Protocol

The proposed architecture was implemented using Python 3.14.4 [29] and PyTorch 2.12.0 [30] on a Linux workstation equipped with an Intel Core i9-9900K pro-

cessor (Intel Corporation, Santa Clara, CA, USA), 32 GiB of RAM, an NVIDIA RTX 2080 Ti GPU (NVIDIA Corporation, Santa Clara, CA, USA), and Manjaro Linux (kernel 6.18.32).

3.4.1. Data Augmentation Strategy

To enhance generalization across diverse handwriting styles while avoiding human bias in transformation selection, we implemented an automated RandAugment strategy [31]. By randomly selecting $N = 2$ operations per image from a curated policy space, RandAugment introduces sufficient stochasticity to prevent overfitting. The global distortion magnitude was empirically set to $M = 6$. The allowable operations space was strictly limited to photometric filters (e.g., AutoContrast, Brightness, Sharpness) and horizontal geometric transforms (ShearX, TranslateX). Vertical geometric operations were explicitly excluded to prevent semantic class ambiguity (e.g., confusing a handwritten ‘6’ with a ‘9’). Figure 2 displays a random sampling of ROIs after augmentation, highlighting the significant intra-class variability the architecture must accommodate.

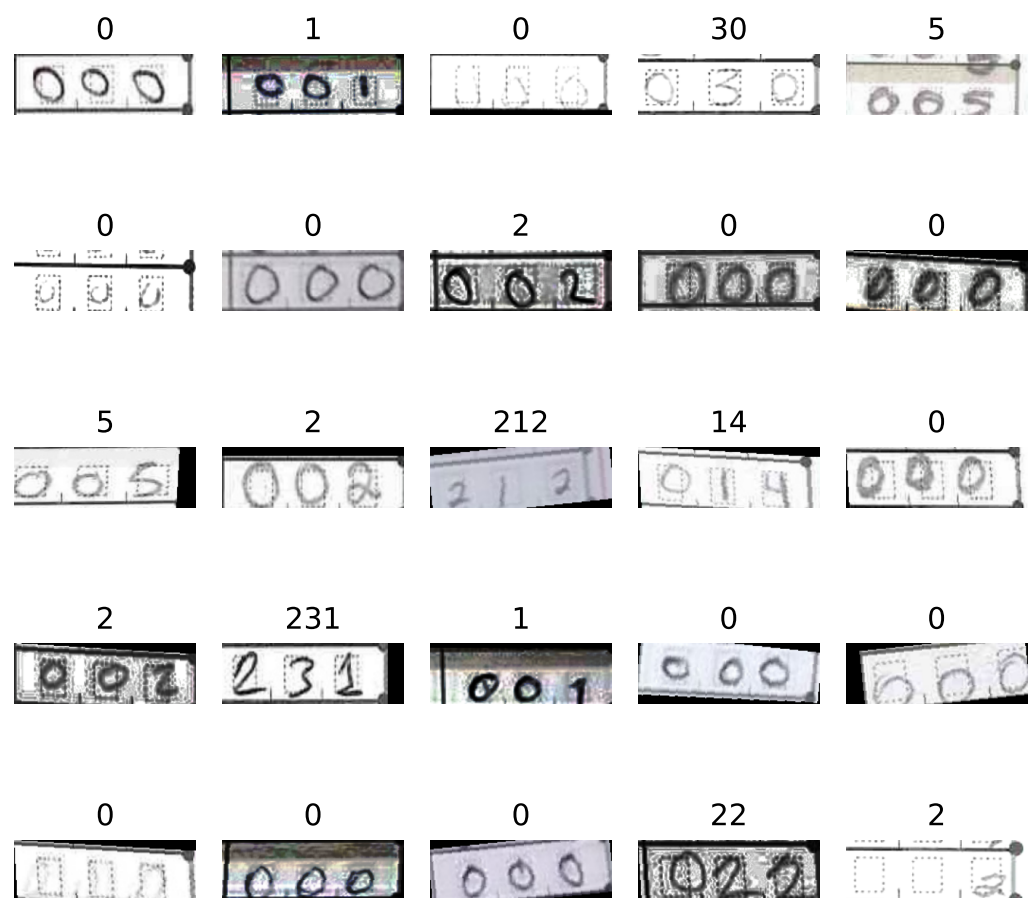


Figure 2. Visual sample of 25 Regions of Interest (ROIs) showcasing the megadiverse and challenging nature of the dataset. The preview illustrates inconsistent spatial distributions where digit sequences often overflow or shift beyond the pre-printed bounding box delimiters, coupled with extreme calligraphic variability.

3.4.2. Optimization and Regularization

Training was conducted for 10 epochs using the AdamW optimizer [32] with a weight decay of 0.01. The learning rate was modulated via a OneCycleLR scheduler, featuring a 10,000-iteration linear warmup to a peak of 2×10^{-3} , followed by cosine annealing towards a final value of 1×10^{-5} to stabilize deep residual convergence. A batch size of 64 was maintained across all experiments. Crucially, to handle the noisy labels inherent to

manual electoral annotations and to mitigate overconfidence on the long-tail distribution, Label Smoothing ($\alpha = 0.1$) was applied to the cross-entropy loss formulation.

3.5. Evaluation Metrics for Operational Reliability

The validation of the system is performed through a multi-metric approach. We prioritize Top-1 Accuracy as the primary indicator, as each digit sequence must be identified with absolute precision. Given the long-tail distribution, we also report Macro-averaged Precision and Macro-averaged F1-score to ensure reliability across all 1001 classes. Finally, computational efficiency is quantified via Inference Latency (ms) and GFLOPS.

4. Results

This section provides a quantitative assessment of the proposed Asymmetric ResNet-18 architecture, evaluating its performance on the 1001-class HDSR dataset. The experimental findings are organized into three primary areas: an ablation study to isolate architectural contributions, an analysis of classification metrics across the long-tail distribution, and an evaluation of computational efficiency in terms of throughput and latency. All results reported herein were obtained following the training protocol described in Section 3 and were validated using a test partition consisting of 840,914 samples to ensure statistical significance. The training phase utilized a megadiverse dataset containing 2,936,082 images, providing a robust and representative foundation for high-cardinality classification.

4.1. Ablation Study

To validate the architectural decisions behind the proposed model, we conducted a systematic ablation study. This analysis evaluates the transition from a standard symmetric ResNet-18 to the proposed asymmetric geometry, while also assessing the impact of CBAM integration across different stages and reduction ratios (r). The objective is to identify the configuration that provides the optimal trade-off between recognition quality and computational overhead, ensuring the architecture's suitability for high-throughput electoral data processing.

As observed in Table 2, while the standard symmetric baseline (Vanilla ResNet-18) offers the lowest computational cost (0.4470 GFLOPS), it exhibits inferior performance on minority classes, as evidenced by its lower Macro F1 score (0.8736). Conversely, introducing the asymmetric geometry alongside aggressive attention mechanisms across all stages yields conflicting metrics. For instance, the configuration with CBAM in Stages 1–4 ($r = 8$) exhibits an accuracy paradox typical of extreme long-tail distributions: while Top-1 Accuracy peaks at 98.35%, the Macro Precision and Macro F1 scores collapse to approximately 0.72. This indicates that the network over-optimizes for the high-frequency majority classes at the severe expense of the tail.

Even when the reduction ratio is adjusted (Stages 1–4, $r = 16$) to achieve a marginally higher Macro F1 (0.8889) compared to our selected configuration (0.8826), it violates our theoretical premise of semantic abstraction. Applying attention mechanisms to early convolutional stages forces the network to allocate parameter weight to low-level morphological noise, increasing the risk of overfitting to background artifacts in degraded ballots.

Therefore, the selected configuration (Asymmetric Stride, CBAM in Stages 3–4, $r = 16$) represents the Pareto-optimal solution for this specific operational environment. It successfully mitigates the representational collapse of tail classes, significantly outperforming the Vanilla baseline in sequence integrity, while strictly confining attention to deep semantic features. Moreover, its nominal computational increase (0.7197 GFLOPS) comfortably maintains an inference latency of 9.1 ms, remaining effectively below the strict 10 ms threshold

on standard CPU hardware. This ensures robust long-tail generalization without violating high-throughput deployment constraints.

Table 2. Comprehensive ablation study of the proposed architecture for 1001-class HDSR. Results evaluate the performance shift from Standard (Symmetric) to Proposed (Asymmetric) geometries.

Configuration	Ratio (r)	Params (M)	GFLOPS	Top-1 Acc.	Macro Prec.	Macro F1
Standard Geometry (Symmetric Stride)						
Vanilla ResNet-18 (Baseline)	–	11.69	0.4470	97.78%	0.8779	0.8736
CBAM in Stages 1–4	16	11.78	0.4477	97.75%	0.8931	0.8917
CBAM in Stages 3–4	16	11.77	0.4473	97.70%	0.8776	0.8767
Proposed Geometry (Asymmetric Stride)						
Pure Asymmetric	–	11.69	0.7195	97.78%	0.8627	0.8631
CBAM in Stages 1–4	8	11.87	0.7206	98.35%	0.7212	0.7198
CBAM in Stages 1–4	16	11.78	0.7204	97.87%	0.8956	0.8889
CBAM in Stages 3–4	8	11.85	0.7199	98.34%	0.7174	0.7158
CBAM in Stages 3–4 *	16	11.78	0.7197	97.82%	0.8878	0.8826
CBAM in Stages 3–4	32	11.74	0.7196	98.33%	0.7233	0.7222

* Proposed configuration (Ours). Performance evaluated on 840,914 test samples. **Bold values** indicate the best performance in each metric across all configurations.

4.2. Complexity and Operational Trade-Offs

The relationship between model complexity, inference latency, and predictive accuracy is visualized in the bubble chart (Figure 3) and the Pareto frontier (Figure 4).

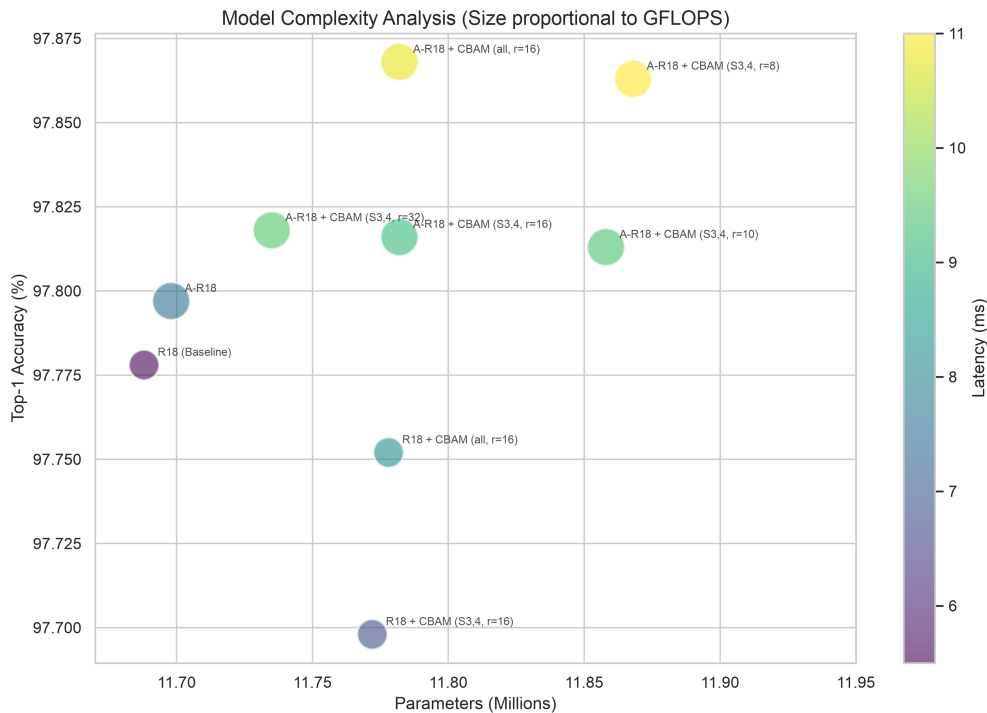


Figure 3. Model complexity analysis. Bubble size is proportional to GFLOPS, while color represents CPU inference latency. The proposed A-R18 + CBAM (S3,4, r = 16) configuration achieves a strategic balance between parameter count and high-resolution feature preservation.

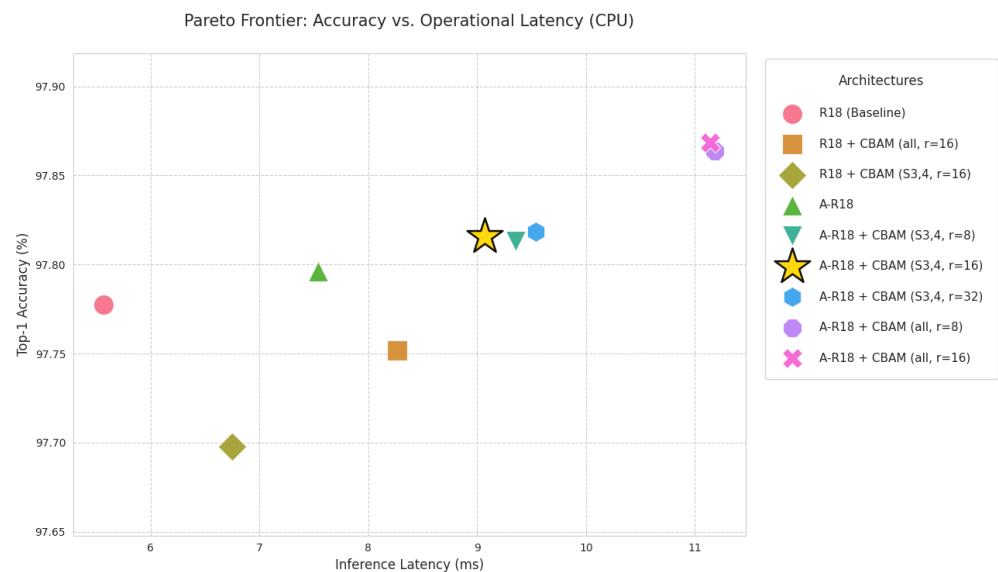


Figure 4. Pareto frontier of Top-1 Accuracy vs. Inference Latency. The proposed architecture (A-R18 + CBAM, S3,4, $r = 16$) resides in the optimal region, offering a significant accuracy boost over the baseline while maintaining latency below the 10ms threshold required for real-time CATD operations.

As shown in Figure 4, while the all-stages attention variants achieve slightly higher Top-1 accuracy (up to 98.35%), they incur a latency penalty of over 11 ms. In contrast, the proposed A-R18 + CBAM (S3,4, $r = 16$) maintains a competitive accuracy of 97.87% with a latency of only 9.13 ms. This represents the optimal operational point for the Mexican electoral infrastructure, where high-throughput processing is as critical as individual sequence precision.

4.3. Statistical Reliability and Calibration

To validate the model's suitability for legal and electoral deployment, we evaluated its calibration through a Reliability Diagram and the Expected Calibration Error (ECE), as shown in Figure 5.

The resulting ECE of 0.0957 is remarkably low for a high-cardinality task involving 1001 classes. The alignment between the model's confidence and its actual accuracy (following the diagonal) demonstrates that the integration of Label Smoothing prevents over-confident misclassifications. This calibration is vital for the PREP system, as it allows for the automatic flagging of low-confidence sequences for manual human verification, thereby maintaining the overall integrity of the electoral results.

In summary, the experimental results confirm that the proposed asymmetric stride geometry, combined with selective attention in deep residual stages, successfully addresses the challenges of large-scale HDSR. By outperforming the symmetric ResNet-18 baseline in both precision and reliability while remaining within strict computational constraints, the A-R18 + CBAM architecture proves to be an effective approach for the automated transcription of handwritten electoral reports.

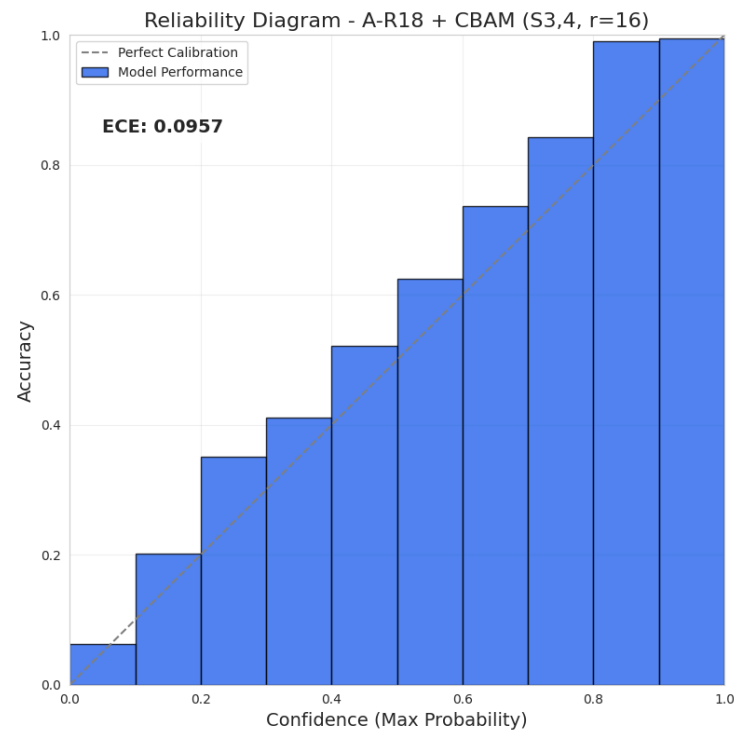


Figure 5. Reliability Diagram for the proposed A-R18 + CBAM (S3,4, r = 16) model. The low ECE of 0.0957 indicates that the model's confidence scores are well-calibrated, ensuring that predicted probabilities accurately reflect the empirical likelihood of correct classification.

5. Discussion

The experimental results demonstrate that the strategic integration of asymmetric strides and attention mechanisms significantly enhances the recognition of handwritten digit sequences. In this section, we interpret these findings by analyzing how the proposed geometry preserves spatial integrity compared to symmetric baselines. Furthermore, we discuss the model's resilience against the inherent challenges of the INE dataset, such as high-cardinality imbalance and label noise. Finally, we evaluate the feasibility of deploying this architecture within the CATD infrastructure, considering the critical trade-off between predictive precision and operational latency during electoral processes.

5.1. Theoretical Basis of Asymmetric Geometry and Receptive Field

The core advantage of the proposed A-R18 architecture lies in its ability to explicitly preserve horizontal spatial resolution through targeted receptive field scaling. Given that the input ROIs are normalized to a strictly rectangular 64×192 resolution (a 1:3 aspect ratio), the choice of stride fundamentally dictates the spatial decay of the feature maps. A standard symmetric downsampling configuration $S = (2, 2)$ aggressively reduces both dimensions, causing the receptive fields of adjacent horizontal characters to overlap prematurely, leading to morphological blending in deep layers. Conversely, a vertical-preserving stride $S = (1, 2)$ would exacerbate this horizontal compression, destroying the sequence integrity.

By implementing an asymmetric stride $S = (2, 1)$ in Stage 1, the architecture theoretically halves the redundant vertical background space while retaining 100% of the horizontal resolution. This theoretical premise was corroborated by preliminary exploratory experiments, which revealed that a $S = (1, 2)$ stride resulted in severe sequence corruption and representational collapse. Consequently, our formal ablation study (Table 2) focuses exclusively on the viable geometries, empirically demonstrating that the transition from the standard $S = (2, 2)$ to the proposed $S = (2, 1)$ configuration acts as the primary driver for the improvement in Macro Precision.

5.2. Robustness Against Label Noise and High-Cardinality Imbalance

Classifying 1001 distinct sequences presents a significant challenge due to the extreme long-tail distribution of electoral data. Beyond the severe class imbalance, the dataset is affected by two distinct types of noise: morphological noise (e.g., degraded ink, overlapping strokes) and label noise (e.g., incorrect ground-truth annotations caused by operator fatigue during the manual data entry phase). While the integration of CBAM in the deeper stages (S3 and S4) addresses morphological noise by selectively attending to valid structural features, label noise poses a unique optimization challenge, as it can force the network to memorize incorrect targets.

To explicitly mitigate the impact of human-induced label noise, the architecture employs Label Smoothing ($\alpha = 0.1$). By softening the hard targets during the cross-entropy calculation, this regularization prevents the model from assigning absolute certainty to any single class, thereby reducing the network's susceptibility to overfitting on flawed ground truths. The effectiveness of this combined approach is quantitatively validated by the reliability analysis presented in Figure 5, demonstrating an exceptionally well-calibrated ECE of 0.0957. For the INE's operational workflow, this ensures that the system does not produce over-confident misclassifications when encountering incorrectly labeled samples. Instead, it maintains well-calibrated probabilities that allow for reliable human-in-the-loop verification, ultimately reducing the risk of automated transcription errors in the final vote count.

In addition, the ablation study systematically justifies the integration depth of the CBAM module. Shallow convolutional layers (Stages 1 and 2) primarily act as edge and gradient detectors; applying attention mechanisms at this level often results in the network erroneously amplifying high-frequency morphological noise or background artifacts. In contrast, deeper layers (Stages 3 and 4) encapsulate high-level semantic abstractions. By restricting CBAM to these terminal stages, the spatial and channel attention gates can efficiently filter class-specific strokes (e.g., distinguishing the loops of an '8' from a '3') without wasting computational parameters on low-level feature maps. The empirical results in Table 2 confirm this theoretical premise, showing that integrating CBAM exclusively in Stages 3 and 4 provides the optimal balance, avoiding the latency penalty and potential overfitting observed when attention is applied across all stages.

5.3. Operational Efficiency and Deployment Feasibility

Electoral processes require a delicate balance between predictive accuracy and processing speed. The practical utility of the proposed architecture is best understood when scaling its metrics to a national deployment scenario. During a Mexican federal election, the PREP infrastructure must process millions of AEC forms within a highly constrained 24-h window. The traditional double- or triple-blind manual transcription protocol, while secure, is bottlenecked by human fatigue and logistical overhead.

As demonstrated by the Pareto frontier (Figure 4), the proposed A-R18 + CBAM ($r = 16$) configuration offers a strategic operational balance. By maintaining an inference latency of 9.13 ms on standard CPU hardware, a single computational thread achieves a throughput of approximately 109 Regions of Interest (ROIs) per second. In a decentralized Data Collection and Transmission Center (CATD) utilizing standard multi-core servers, this translates to the capacity of processing tens of thousands of records per minute, substantively reducing the digital processing bottleneck.

Furthermore, coupling the 97.82% Top-1 Accuracy with the well-calibrated confidence mechanism (ECE of 0.0957) enables a highly efficient workflow shift. Instead of relying entirely on manual data entry for every document, the system can safely automate the highest-confidence predictions. The fail-controlled pipeline systematically rejects ambigu-

ous or degraded sequences (triggering manual review), while tabulating the confident predictions. This hybrid human-in-the-loop paradigm optimizes resource allocation and minimizes the risk of fatigue-induced transcription errors during the critical hours of election night.

5.4. Future Work

To address scalability, future research will transition to Connectionist Temporal Classification (CTC) for dynamic sequence decoding. Building upon the established constraint that non-recurrent structures are imperative for maintaining the 10 ms budget, the next phase will implement CTC utilizing Fully Convolutional Networks (FCNs), Gated FCNs (GFCNs), and 2D State Space Models (e.g., VMamba or Vim). By adapting Vision Mamba architectures, we aim to leverage linear-time sequence modeling over spatial feature maps without the latency bottlenecks of traditional recurrences. Within this updated framework, traditional Convolutional Recurrent Neural Networks (CRNNs) equipped with BiLSTM or BiGRU modules will be deployed exclusively as latency baselines to benchmark the computational superiority of 2D SSMs and FCNs under the CTC paradigm. Additionally, to optimize the training pipeline against dataset noise, we will introduce a self-labeling strategy. By enforcing a strict confidence threshold (≥ 0.95), this mechanism will automatically identify and correct potentially mislabeled ground-truth samples. Concurrently, to overcome the data scarcity inherent to the long-tail distribution, we will explore the integration of Generative AI techniques as an offline data augmentation framework. Finally, while the current methodology focuses exclusively on numerical digits (Arabic numerals), future iterations will expand the recognition pipeline to process the alphabetical representation of the vote counts (i.e., numbers written in words). Integrating this capability will enable automated cross-verification between the numerical and handwritten text fields of the tally sheets, providing an additional layer of redundancy to definitively fortify the integrity of the digitized electoral records.

6. Conclusions

This study addressed the critical challenge of unconstrained handwritten digit sequence recognition within hardware-constrained environments, specifically targeting the strict operational requirements of electoral document digitization.

Primary Findings. Our experimental results demonstrate that formulating bounded multi-digit recognition (up to three digits) as a high-cardinality classification task (1001 classes) effectively circumvents the sequential bottlenecks inherent to recurrent architectures. By augmenting a purely convolutional backbone with a Convolutional Block Attention Module (CBAM), the proposed framework adapts to morphological variations while reliably maintaining inference latency below the critical 10 ms threshold on standard CPU hardware (9.1 ms).

Limitations and Failure Analysis. While the high-cardinality formulation satisfies strict deployment budgets, it introduces inherent limitations. The discrete 1001-class mapping restricts vocabulary scalability, limits generalizability to unbounded sequence lengths, and creates an artificial performance ceiling due to the severe long-tail distribution of real-world electoral data. Furthermore, an analysis of specific failure cases reveals vulnerabilities to extreme morphological noise. The model struggles with severe calligraphic ambiguity (e.g., highly cursive '4's mimicking '9's), heavy strike-throughs where polling officials manually corrected a sequence, and structural artifacts such as ink bleed-through from the reverse side of the tally sheet. In these specific failure modes, the lack of character-level alignment prevents partial sequence recovery, forcing the system to either misclassify the entire region or trigger a low-confidence rejection for human verification.

Future Work. To address scalability, future research will transition to Connectionist Temporal Classification (CTC) for dynamic sequence decoding. Building upon the established constraint that non-recurrent structures are imperative for maintaining the 10 ms budget, the next phase will implement CTC utilizing Fully Convolutional Networks (FCNs), Gated FCNs (GFCNs), and 2D State Space Models (e.g., VMamba or Vim). By adapting Vision Mamba architectures, we aim to leverage linear-time sequence modeling over spatial feature maps without the latency bottlenecks of traditional recurrences. Within this updated framework, traditional Convolutional Recurrent Neural Networks (CRNNs) equipped with BiLSTM or BiGRU modules will be deployed exclusively as latency baselines to benchmark the computational superiority of 2D SSMs and FCNs under the CTC paradigm. Additionally, to optimize the training pipeline against dataset noise, we will introduce a self-labeling strategy. By enforcing a strict confidence threshold (≥ 0.95), this mechanism will automatically identify and correct potentially mislabeled ground-truth samples. Concurrently, to overcome the data scarcity inherent to the long-tail distribution, we will explore the integration of Generative AI techniques as an offline data augmentation framework. This will allow the synthesis of high-fidelity artificial samples for the missing tail classes, thereby balancing the training distribution for future architectures without impacting online inference performance.

Author Contributions: Conceptualization, M.A.C.-R., G.S.-P., J.P.-P., L.K.T.-M., A.H.-S., J.O.-M. and H.M.P.-M.; methodology, M.A.C.-R., G.S.-P. and J.P.-P.; software, M.A.C.-R.; validation, M.A.C.-R., G.S.-P., J.P.-P., L.K.T.-M., A.H.-S., J.O.-M. and H.M.P.-M.; formal analysis, M.A.C.-R., J.P.-P. and L.K.T.-M.; investigation, M.A.C.-R. and G.S.-P.; resources, J.O.-M. and H.M.P.-M.; data curation, M.A.C.-R.; writing—original draft preparation, M.A.C.-R.; writing—review and editing, M.A.C.-R., G.S.-P., J.P.-P., L.K.T.-M., A.H.-S., J.O.-M. and H.M.P.-M.; visualization, M.A.C.-R.; supervision, A.H.-S., J.O.-M. and H.M.P.-M.; project administration, L.K.T.-M. and A.H.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. The data that support the findings of this study were obtained from the National Electoral Institute (Instituto Nacional Electoral, INE, Mexico) under an inter-institutional collaboration agreement and are not publicly available due to legal and institutional restrictions. The code used for the experiments is available at <https://github.com/mikecr96/A-R18-CBAM> (accessed on 6 July 2026).

Acknowledgments: The authors wish to thank the Instituto Politécnico Nacional (IPN) for the administrative and financial support provided through the Multidisciplinary Project 20260201. Additionally, M.A. Camargo-Rojas acknowledges the financial support provided by the Secretariat of Science, Humanities, Technology and Innovation (SECIHTI) for his doctoral studies.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Instituto Nacional Electoral. Acerca del INE. 2026. Available online: <https://www.ine.mx/sobre-el-ine/> (accessed on 21 June 2026).
2. Instituto Nacional Electoral. Programa de Resultados Electorales Preliminares (PREP). 2026. Available online: <https://www.ine.mx/voto-y-elecciones/prep/> (accessed on 21 June 2026).
3. Camargo-Rojas, M.A.; Sánchez-Pérez, G.; Portillo-Portillo, J.; Toscano-Medina, L.K.; Hernández-Suárez, A.; Olivares-Mercado, J.; Pérez-Meana, H.M.; García Villalba, L.J. Towards Electoral Digitization: Automatic Classification of Handwritten Numbers in PREP System Records. *Eng. Proc.* **2026**, *123*, 35. [CrossRef]

4. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
5. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
6. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; Volume 97, pp. 6105–6114.
7. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. *arXiv* **2018**, arXiv:1807.11164.
8. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
9. Liu, Y.; Li, H.; Yao, L.; Xu, Y.; Dai, D.; Wang, Z. Lithium-ion batteries state-of-health estimation using degradation information and segmented voltage signals. *Process Saf. Environ. Prot.* **2026**, *212*, 108919. [[CrossRef](#)]
10. Zhao, H.; Wang, B.; Fu, Y.; Li, N.; Gao, Z. Fixed-time adaptive fault-tolerant control of a multi-mode VTOL UAV with variable prescribed performance boundaries under random disturbances. *ISA Trans.* **2026**, *175*, 126–138. [[CrossRef](#)] [[PubMed](#)]
11. Goodfellow, I.J.; Bulatov, Y.; Ibarz, J.; Arnoud, S.; Shet, V. Multi-digit number recognition from street view imagery using deep convolutional neural networks. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
12. Xiao, D.; Xu, X.; Shan, L.; Liu, T.; Li, X.; Zhang, Y. Attention-Based Handwritten Chinese Recognition for Power Grid Maintenance Documents. *Sci. Prog.* **2025**, *108*. [[CrossRef](#)] [[PubMed](#)]
13. Jain, A.; Gupta, J.; Khandelwal, S.; Kaur, S. Vehicle License Plate Recognition. *Fusion Pract. Appl.* **2021**, *4*, 15–21. [[CrossRef](#)]
14. Graves, A.; Fernández, S.; Gomez, F.; Schmidhuber, J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, Pittsburgh, PA, USA, 25–29 June 2006; ACM: New York, NY, USA, 2006; pp. 369–376.
15. Akoushideh, A.; Rashtehroudi, A.R.; Shahbahrami, A. Persian/Arabic Scene Text Recognition With Convolutional Recurrent Neural Network. *IET Smart Cities* **2025**, *7*, e70001. [[CrossRef](#)]
16. Wang, J.; Hu, X. Convolutional Neural Networks With Gated Recurrent Connections. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3421–3435. [[CrossRef](#)] [[PubMed](#)]
17. Selvam, P.; Tavera Romero, C.A.; Alharbi, M.; Mehbodniya, A.; Webber, J.; Sengan, S. A Transformer-Based Framework for Scene Text Recognition. *IEEE Access* **2022**, *10*, 100895–100910. [[CrossRef](#)]
18. Pal, U.; Halder, A.; Shivakumara, P.; Blumenstein, M. A Comprehensive Review on Text Detection and Recognition in Scene Images. *Artif. Intell. Appl.* **2024**, *2*, 229–249. [[CrossRef](#)]
19. Song, Y.; Li, M.; Wang, B. Long-Tailed Visual Recognition via Improved Cross-Window Self-Attention and TrivialAugment. *IEEE Access* **2023**, *11*, 49601–49610. [[CrossRef](#)]
20. Liao, K.; Pang, J.; Zheng, Y.; Wang, K.; Lin, G.; Tan, Y. Multi-Expert Dynamic Gating and Feature Decoupling Algorithm for Long-Tail Image Classification. *Concurr. Comput. Pract. Exp.* **2025**, *37*, e70287. [[CrossRef](#)]
21. Yang, Q.; Feng, X.; Tan, S. One-Stage Long-Tailed Object Detection Based on Adaptive Class Suppression Equalized Focal Loss. *IEEE Trans. Electr. Electron. Eng.* **2025**, *20*, 1424–1435. [[CrossRef](#)]
22. Zheng, X.; Bai, X. A Mixup-based Margin Aware and Calibration Model for Imbalanced in Soil Classification. *Acad. J. Comput. Inf. Sci.* **2024**, *7*, 12–18. [[CrossRef](#)]
23. Pearce, J.; Meger, D. Adaptive Confidence Calibration. Proceedings of the Canadian Conference on Artificial Intelligence, Toronto, ON, Canada, 30 May–3 June 2022. Available online: <https://caiac.pubpub.org/pub/camnr5ix> (accessed on 6 July 2026).
24. Zhang, X.; Wang, T.; Ma, L.; Mahadevan, S. Reliability engineering, risk management, and trustworthiness assurance for AI systems. *J. Reliab. Sci. Eng.* **2025**, *1*, 022001. [[CrossRef](#)]
25. Oyibo, K.U.; Louis, J.D.; Gilbert, J.E. iOCR: Informed optical character recognition for election ballot tallies. *arXiv* **2022**, arXiv:2208.00865.
26. Müller, R.; Kornblith, S.; Hinton, G.E. When does label smoothing help? In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.
27. Khokhar, F.; Zoppi, T.; Ceccarelli, A.; Montecchi, L.; Bondavalli, A. Fail-Controlled Classifiers: A Swiss-Army Knife Toward Trustworthy Systems. *Softw. Pract. Exp.* **2025**, *56*, 239–259. [[CrossRef](#)]
28. Mena, J.; Pujol, O.; Vitrià, J. Uncertainty-Based Rejection Wrappers for Black-Box Classifiers. *IEEE Access* **2020**, *8*, 101721–101746. [[CrossRef](#)]
29. Python Software Foundation. *Python Language Reference*, Ver. 3.14.4; Python Software Foundation: Wolfboro Falls, NH, USA, 2026.
30. PyTorch Team. *PyTorch*, Version 2.12.0; PyTorch Foundation: San Francisco, CA, USA, 2026.

31. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. RandAugment: Practical automated data augmentation with a reduced search space. *arXiv* **2019**, arXiv:1909.13719.
32. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.