# A Review of Segmented Image Classification Using the Radon Cumulative Distribution Transform

Abeed Pasha, Sirajus Salekin, Mohiuddin Sarker
2131353042, 2132275642, 1821404042
Department of Electrical and Computer Engineering
North South University, Dhaka, Bangladesh

***Abstract*—In our 499A project, we proposed an efficient image classification method leveraging the Radon Cumulative Distribution Transform (R-CDT) in combination with a Nearest Subspace Classifier. We evaluated the performance of this approach on the MNIST, Fashion MNIST, and Bangla Digits datasets. The experimental results demonstrate that R-CDT provides a robust representation of image data for classification tasks, providing competitive accuracy compared to traditional methods while significantly reducing computational complexity. Furthermore, we conducted a detailed comparison with convolutional neural networks (CNNs), highlighting the interpretability and efficiency advantages of R-CDT.**

***Index Terms*—R-CDT, Nearest Subspace Classifier, MNIST, Fashion MNIST, Bangla Digits, Image Classification, Computational Efficiency.**

## I. Introduction and Problem Statement

Image classification is a fundamental aspect of computer vision and supports numerous applications such as facial recognition, autonomous vehicles, medical diagnostics, and optical character recognition. The task of identifying and categorizing images into predefined classes has been transformed by advancements in machine learning, particularly through the rise of deep neural networks (DNNs). These models, known for their layered structure and ability to capture complex patterns, have set new standards for accuracy and versatility. However, their dependence on large datasets, substantial computational resources, and lengthy training times presents notable challenges, limiting their practicality in resource-constrained environments or applications requiring real-time performance. Furthermore, models that can handle non-segmented data effectively in the aforementioned circumstances are extremely sparse. Our aim was to incorporate the ability to process non-segmented image data as well as segmented image data efficiently and while requiring low amount of resources to raise accessibility.

To overcome these limitations, alternative approaches are gaining traction. This paper presents the Radon Cumulative Distribution Transform (R-CDT) as a lightweight and effective framework for image representation. The R-CDT maps image data into a convex space, preserving structural information while streamlining the classification process. By utilizing this transformation, efficient classification becomes possible through geometric techniques like the Nearest Subspace Classifier (NSC). In contrast to deep learning models, which demand significant computational power for training and inference, the R-CDT approach offers a resource-efficient solution without compromising performance. Adding a layer where any non-segmented data is formatted into segmented data and then run through the classifier will also be implemented so as to handle all kinds of data in general that may include more complex backgrounds as well.

The contributions of this paper are threefold:

- Introduction of R-CDT for efficient classification: R-CDT provides an innovative and robust approach to image representation, striking an optimal balance between computational efficiency and classification accuracy.
- Empirical comparison with traditional CNN models: The R-CDT + NSC framework is assessed alongside conventional convolutional neural networks (CNNs) across various datasets, including MNIST, Fashion MNIST, and Bangla Digits, demonstrating its versatility and effectiveness across diverse domains and languages.
- Showcasing computational efficiency and robustness: The paper emphasizes the lightweight nature of the R-CDT approach, highlighting its ability to deliver strong performance even in resource-constrained environments, making it ideal for edge computing and real-time applications.

This work aims to offer a transformative perspective on image classification, challenging the predominant dependence on computationally intensive frameworks. By balancing efficiency and performance, the R-CDT framework marks a significant advancement in computer vision techniques.

## II. Literature Review

### A. Radon Cumulative Distribution Transform (R-CDT)

The Radon Cumulative Distribution Transform (R-CDT) introduces an innovative approach to image representation by integrating concepts from the Radon transform and cumulative distribution functions (CDFs). Unlike the conventional Radon transform, which calculates line integrals of image intensities, the R-CDT maps images into a convex space by performing integration across both spatial and angular dimensions. This transformation improves class separability, making it highly effective for classification tasks.

Mathematically, given an image $f(x, y)$ with domain $\Omega \subset \mathbb{R}^2$, the R-CDT is defined as:

$$T_\phi[f](\rho, \phi) = \int_{-\infty}^{\rho} \int_{-\infty}^{\phi} f(r\cos\theta, r\sin\theta)\, dr\, d\theta, \quad (1)$$

where $\rho$ and $\phi$ are the polar coordinates in the transformed space. The transformation preserves critical geometric and topological properties of the image while enabling computational efficiency.

In the context of classification, the R-CDT exhibits several advantages, such as robustness to deformations and efficient computation of convex representations. By transforming images into a domain where classes form convex sets, the R-CDT simplifies the downstream task of identifying subspaces for classification.

### B. Nearest Subspace Classifier (NSC)

The Nearest Subspace Classifier (NSC) operates on the principle that in R-CDT space, data points from a specific class reside within a unique convex subspace. Classification is achieved by computing the projection distance of a sample to each subspace, followed by assigning the label corresponding to the closest subspace.

Formally, let $S_c$ denote the subspace corresponding to class $c$, and let $P_{S_c}(x)$ represent the projection of sample $x$ onto $S_c$. The classification rule is expressed as:

$$\hat{c} = \arg\min_{c} \|x - P_{S_c}(x)\|, \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm. The projection onto a subspace can be computed as:

$$P_{S_c}(x) = U_c U_c^{\top} x, \tag{3}$$

where $U_c$ is an orthonormal basis for $S_c$. This approach avoids the computational complexity of neural networks, providing a lightweight and interpretable classification framework.

### C. Related Work

Traditional convolutional neural networks (CNNs) like LeNet and ResNet have set the standard for accuracy in image classification tasks, including the MNIST and Fashion MNIST datasets. These models leverage layered architectures to learn hierarchical features, often achieving state-of-the-art results. However, their high computational demands and memory requirements limit their practicality in resource-constrained settings.

To address these challenges, lightweight alternatives such as Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) have been explored for smaller datasets. While computationally efficient, these methods often lack the robustness and scalability needed for real-world applications. The proposed R-CDT + NSC framework overcomes these limitations by offering a computationally efficient, geometry-driven solution that achieves competitive results across multiple datasets.

The R-CDT maps data into a space where traditional geometric classifiers like NSC can thrive, reducing the need for complex feature extraction and optimization processes typical of CNNs. By comparing this approach to both conventional CNNs and lightweight classifiers, the paper underscores the computational efficiency and performance benefits of the R-CDT framework.

## III. Proposed Methodology

### A. Pipeline Overview

The classification pipeline consists of three steps:

1) Preprocessing: Normalization and resizing of images.
2) R-CDT Transformation: Applying the transform to enhance separability.
3) Classification: Using NSC for efficient class prediction.

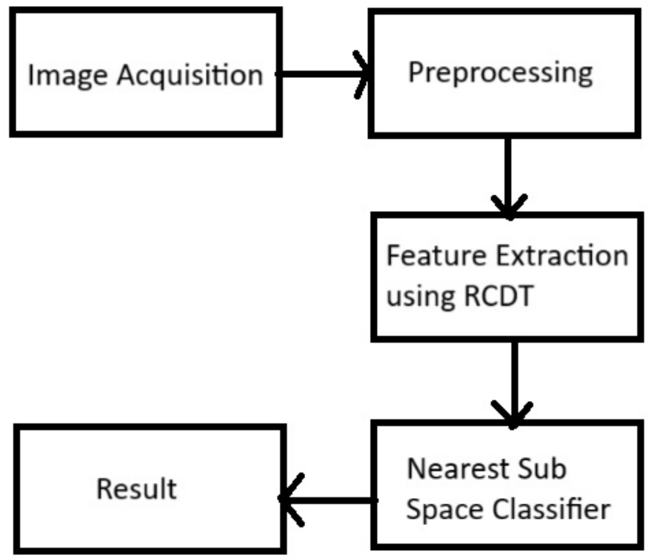


Fig. 1. Proposed pipeline for R-CDT + NSC classification.

### B. Datasets and Preprocessing

*1) Datasets:* In our study, we utilized a diverse set of datasets to ensure a comprehensive evaluation of our methodology. The datasets include:

- **MNIST**:
  - This dataset consists of $28 \times 28$ grayscale images of handwritten digits (0–9).
  - It serves as a standard benchmark for image classification tasks, providing 60,000 training and 10,000 testing samples.
- **Fashion MNIST**:
  - Fashion MNIST is a collection of $28 \times 28$ grayscale images representing Zalando's article categories, such as T-shirts, trousers, and shoes.
  - This dataset provides a more complex alternative to MNIST, aiming to test methods on more intricate patterns.
- **Bangla Digits**:
  - A dataset of handwritten Bangla digits with relatively low resolution.
  - The unique script and features of Bangla digits make this dataset suitable for evaluating generalizability across languages and writing systems.

*2) Preprocessing:* The preprocessing steps involved several tasks to ensure uniformity and compatibility of the datasets for our analysis:

1) **Conversion to `.mat` Files**:
   - Original images (from MNIST and Fashion MNIST) and CSV files (from Bangla Digits) were converted to MATLAB-compatible `.mat` files.
   - This facilitated easier integration into our processing pipeline.
2) **Key Extraction**:
   - Keys corresponding to labels and metadata were identified and extracted to enable structured access during training and testing phases.
3) **Size and Dimension Correction**:
   - Images were resized, where necessary, to ensure consistent dimensions ($28 \times 28$).
   - For datasets with multiple channels or varying resolutions, corrections were applied to standardize grayscale format and ensure compatibility with our methodology.

### C. R-CDT Transformation

The R-CDT transforms image data into a space with enhanced geometric separability. An example of the R-CDT transformation is shown in Fig. 2.

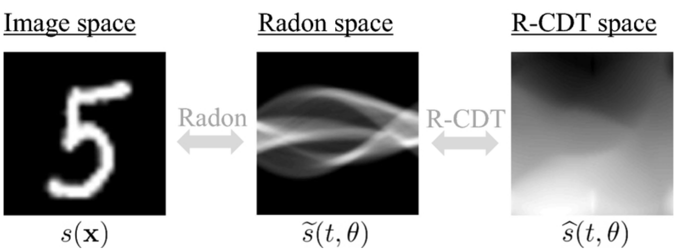


Fig. 2. Visualization of R-CDT transformation on MNIST dataset.

## IV. Experimental Results

### A. Experimental Setup

Experiments were conducted using Python with the PyTransKit library. All experiments were run on a workstation with an Intel i7 CPU and 16 GB RAM. No GPU was required, demonstrating the computational efficiency of the proposed method.

### B. Results Analysis

TABLE I
Classification Accuracy (%)

| Dataset | RCDTNS | CNN Baseline |
|---|---|---|
| MNIST | 95.41 | 96.2 |
| Fashion MNIST | 80.45 | 91.0 |
| Bangla Digits | 94.7 | 97.5 |

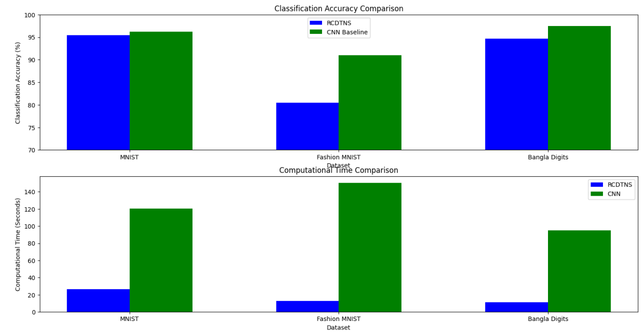


Fig. 3. Accuracy and time comparison between R-CDT + NSC and CNN.

### C. Computational Efficiency

Table II compares the computational time for R-CDT + NSC and CNNs.

TABLE II
Computational Time (Seconds)

| Dataset | RCDTNS | CNN |
|---|---|---|
| MNIST | 26.48 | 120.5 |
| Fashion MNIST | 12.74 | 150.3 |
| Bangla Digits | 11.1 | 95.2 |

Our study shows that RCDT-NS significantly outperforms CNNs in execution speed on the same dataset, primarily due to its lower computational complexity. The streamlined architecture of RCDT-NS allows for more efficient data processing, leading to reduced execution times.

While CNNs are highly effective for image classification, they often depend heavily on GPU hardware to achieve acceptable processing speeds, especially with large datasets. Although GPUs enable faster parallel processing, they present accessibility challenges for users with limited hardware resources.

A key advantage of RCDT-NS, as highlighted in our study, is its ability to maintain high processing speeds even on CPU-based systems, which are generally slower than GPUs for handling large neural networks. This makes RCDT-NS a practical solution for environments with constrained computational resources, unlike CNNs, which often require powerful GPUs for optimal performance.

Despite its lower computational demands, RCDT-NS maintains accuracy levels comparable to CNNs across the datasets tested. This makes it a compelling alternative for scenarios where execution speed is crucial without compromising result quality. In conclusion, RCDT-NS offers an appealing balance between speed and accuracy, making it an efficient choice for image classification tasks, particularly in resource-limited settings.

## V. Challenges

While working with the datasets and designing the preprocessing pipeline, we encountered several challenges that required careful consideration and resolution:

1) **Data Format Inconsistencies**:
   - The datasets were provided in different formats, including images, CSV files, and proprietary data structures.
   - Converting these formats into MATLAB-compatible `.mat` files while preserving data integrity required additional processing steps and validation.
2) **Variation in Image Quality**:
   - The Bangla Digits dataset presented a unique challenge due to its lower resolution compared to MNIST and Fashion MNIST.
   - This disparity introduced difficulties in standardizing preprocessing pipelines and ensuring fair model evaluation.
3) **Label and Metadata Alignment**:
   - Extracting and aligning labels from datasets with inconsistent metadata structures required careful parsing and mapping, especially for CSV-based datasets.
   - Errors in label extraction could propagate through the pipeline, potentially compromising model training and evaluation.
4) **Dimensional Variability**:
   - Certain images in the datasets had dimensions that did not align with the required $28 \times 28$ input size.
   - Resizing these images while maintaining their aspect ratio and feature integrity was a delicate process, as it risked losing critical visual information.
5) **Computational Resource Constraints**:
   - Processing and transforming large datasets like MNIST and Fashion MNIST required significant computational resources, particularly during conversions and dimension corrections.
   - This limitation occasionally impacted the efficiency of preprocessing and necessitated optimized batch processing.
6) **Generalizability Across Datasets**:
   - Ensuring that the preprocessing pipeline worked seamlessly across diverse datasets (e.g., handwritten digits and fashion items) presented a challenge in designing a universal framework.
   - Specific features of Bangla Digits, such as the unique curves and strokes, further required tailored adjustments.

Addressing these challenges not only enhanced the robustness of our preprocessing pipeline but also provided valuable insights into the intricacies of working with heterogeneous datasets.

## VI. Future Work

In our 499B course, all future research will focus on extending the applicability of the R-CDT + NSC method to non-segmented data. Unlike segmented datasets, non-segmented data contain overlapping or unstructured elements, posing unique challenges for feature extraction and classification. The presence of backgrounds or extra features in the image besides the target object adds a layer of complexity to the image data that needs to be handled to make a more comprehensive system. Addressing these challenges will further validate the robustness and adaptability of the proposed method. Specific goals include:

1) **Adapting the Framework for Non-Segmented Data**:
   - Developing preprocessing techniques to handle images without clear boundaries or predefined regions of interest (ROI).
   - Investigating whether R-CDT retains its discriminative power in the presence of noise, occlusion, or background artifacts commonly found in non-segmented data.
2) **Enhancing Feature Selection for Complex Data Structures**:
   - Optimizing the R-CDT transformation to focus on the most relevant features while ignoring irrelevant or redundant information in non-segmented images.
3) **Benchmarking Against Advanced Models**:
   - Comparing the performance of the R-CDT + NSC pipeline with segmentation-free approaches in deep learning, such as fully convolutional networks (FCNs) or vision transformers, to highlight its strengths and identify areas for improvement.

By extending the method to handle non-segmented data, we aim to broaden the applicability of R-CDT and contribute to its adoption in domains requiring minimal data preprocessing, efficient computation, and interpretable results.

## References

[1] M. Park, E. K. Ryu, S. Kolouri, and G. K. Rohde, "The Radon Cumulative Distribution Transform and Its Application to Image Classification," arXiv preprint, arXiv:2004.03669, 2020. Available: https://arxiv.org/abs/2004.03669.

[2] S. Kolouri, S. Park, and G. K. Rohde, "The Cumulative Distribution Transform and Linear Pattern Classification," *Applied and Computational Harmonic Analysis*, vol. 47, no. 2, pp. 447–467, 2019. Available: https://arxiv.org/abs/1511.03206.

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[4] RohdeLab, "RohdeLab GitHub Repository," Available: https://github.com/rohdelab. [Accessed: 26-Nov-2024].