Handwritten Word Recognition for Low-Resource Languages: A CRNN-CTC Framework for Kirundi

***Niyifasha Patrick**

Anhui University of Technology, Ma'anshan, Anhui 243002, China

***Correspondence Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11070019

Received: 26 June 2026; Accepted: 01 July 2026; Published: 28 July 2026

ABSTRACT

Handwritten Text Recognition (HTR) has experienced remarkable progress with the development of deep learning techniques. However, most existing studies focus on high-resource languages for which large annotated datasets are readily available. In contrast, low-resource languages remain largely underrepresented in handwriting recognition research due to the scarcity of handwritten corpora, linguistic resources, and benchmark datasets. This paper presents a handwritten word recognition framework for Kirundi, a low-resource Bantu language spoken primarily in Burundi. The proposed system employs a Convolutional Recurrent Neural Network (CRNN) combined with Connectionist Temporal Classification (CTC) for end-to-end sequence recognition without explicit character segmentation.

To address severe data scarcity, a small handwritten Kirundi dataset consisting of manually collected word samples was constructed and annotated. Data augmentation techniques, including rotation, translation, Gaussian noise, Gaussian blur, and elastic distortion, were applied to increase sample diversity. In addition, synthetic handwritten-style data were generated to further expand the training set. Three experimental configurations were investigated: real handwritten data only, real data with augmentation, and real data combined with augmentation and synthetic handwritten-style images.

Experimental results demonstrate that synthetic data generation improved recognition performance and reduced Character Error Rate (CER) from 0.8354 to 0.7560, corresponding to an approximate relative improvement of 9.5%. Although exact word-level recognition remained difficult because of the extremely limited dataset size, the proposed framework successfully learned meaningful sequential patterns and produced increasingly structured Kirundi-like predictions. The study establishes an initial benchmark for Kirundi handwritten word recognition and highlights the potential of synthetic data generation for low-resource handwriting recognition tasks.

Keywords: Handwritten Word Recognition; Low-Resource Languages; Deep Learning; CRNN; CTC

INTRODUCTION

Handwritten Text Recognition (HTR) aims to automatically convert handwritten text images into machine-readable digital representations. It plays a crucial role in document digitization, archival preservation, postal automation, banking systems, historical manuscript analysis, and intelligent document processing. Over the past two decades, the field has evolved from traditional handcrafted feature-based methods toward deep learning approaches capable of learning complex visual and sequential representations directly from raw image data.

Early HTR systems relied heavily on manually engineered features and statistical classifiers. Such approaches generally followed a multi-stage pipeline consisting of preprocessing, segmentation, feature extraction, and classification. Although these methods achieved moderate success in constrained environments, their performance was highly dependent on the quality of feature engineering and character segmentation.

Variations in handwriting style, writing speed, document quality, and character overlap often led to significant performance degradation.

The emergence of deep learning substantially transformed handwriting recognition research. Convolutional Neural Networks (CNNs) enabled automatic extraction of hierarchical visual features from handwritten images, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks improved the modeling of sequential dependencies. The integration of CNNs and RNNs into Convolutional Recurrent Neural Networks (CRNNs) further advanced the field by enabling end-to-end sequence recognition. Combined with Connectionist Temporal Classification (CTC), CRNN architectures eliminate the need for explicit character-level segmentation and have demonstrated strong performance across multiple handwriting and scene text recognition benchmarks.

Despite these advances, existing HTR research remains concentrated on high-resource languages such as English, Chinese, Arabic, and French. These languages benefit from extensive handwritten datasets, large linguistic corpora, and well-established evaluation benchmarks. In contrast, many African languages remain significantly underrepresented in handwriting recognition studies. The absence of publicly available datasets and standardized benchmarks limits both research progress and practical deployment.

Kirundi is a Bantu language belonging to the Niger-Congo language family and is spoken by millions of people primarily in Burundi and neighboring regions. Although Kirundi plays an important role in education, administration, and everyday communication, digital language resources remain scarce. To the best of our knowledge, no publicly available handwritten Kirundi dataset currently exists, and very limited research has been conducted on Kirundi handwriting recognition.

The development of handwriting recognition systems for low-resource languages faces several challenges. First, collecting and annotating handwritten data is labor-intensive and time-consuming. Second, deep learning models typically require large amounts of labeled data to achieve satisfactory performance. Third, low-resource languages often lack auxiliary resources such as language models, dictionaries, and pre-trained representations that can support recognition tasks.

To address these challenges, recent studies have explored data augmentation and synthetic data generation techniques. Data augmentation increases dataset diversity through image transformations, while synthetic data generation creates additional training samples using handwriting-style fonts and image distortions. These approaches have demonstrated promising results in various low-resource computer vision tasks by improving model generalization and reducing overfitting.

This paper investigates the feasibility of handwritten word recognition for Kirundi using a CRNN-CTC architecture under extremely limited data conditions. A small handwritten dataset collected from native speakers serves as the foundation for the experiments. Augmentation and synthetic handwritten-style data generation are employed to mitigate data scarcity. Three experimental configurations are evaluated and compared using Character Error Rate (CER), Word Error Rate (WER), and Word Accuracy Rate (WAR).

The main contributions of this work are summarized as follows:

1. Construction of an initial Kirundi handwritten word dataset collected from native speakers.
2. Development of a CRNN-CTC framework for Kirundi handwritten word recognition.
3. Investigation of augmentation and synthetic data generation techniques for low-resource handwriting recognition.
4. Experimental analysis of recognition behavior under extremely limited training conditions.
5. Establishment of an initial benchmark for future Kirundi handwriting recognition research.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

LITERATURE REVIEW

Traditional Handwritten Text Recognition

Handwritten Text Recognition (HTR) has been extensively studied for several decades. Early approaches relied on handcrafted feature extraction methods combined with statistical classification techniques. Traditional systems typically followed a pipeline consisting of preprocessing, segmentation, feature extraction, and classification. Common feature descriptors included zoning features, projection profiles, contour-based features, and Histogram of Oriented Gradients (HOG). Classification was often performed using Hidden Markov Models (HMMs), Support Vector Machines (SVMs), or k-Nearest Neighbor (k-NN) algorithms.

Although these approaches achieved reasonable performance on constrained datasets, they exhibited several limitations. The effectiveness of handcrafted features depended heavily on domain expertise, and segmentation errors frequently propagated throughout the recognition pipeline. Furthermore, variations in handwriting style, stroke thickness, writing speed, and document quality often reduced recognition accuracy. These limitations motivated the development of learning-based methods capable of automatically extracting discriminative representations from raw image data.

Deep Learning-Based Handwritten Recognition

The introduction of deep learning significantly transformed handwritten recognition research. Convolutional Neural Networks (CNNs) demonstrated remarkable capability in automatically learning hierarchical visual features from image data. Unlike handcrafted approaches, CNNs learn feature representations directly from training samples, reducing the need for manual feature engineering.

Subsequent developments incorporated recurrent neural networks to model sequential dependencies within handwritten text. Long Short-Term Memory (LSTM) networks addressed the vanishing gradient problem and enabled the modeling of long-range contextual information. Bidirectional LSTM (BiLSTM) architectures further improved performance by processing sequences in both forward and backward directions.

The integration of CNNs and recurrent architectures led to the development of Convolutional Recurrent Neural Networks (CRNNs). CRNN models combine visual feature extraction and sequence modeling within a unified framework, making them particularly suitable for handwriting recognition tasks. CRNN-based systems have demonstrated strong performance in both handwritten and scene text recognition applications.

Connectionist Temporal Classification

A major challenge in handwriting recognition is the alignment between image sequences and character labels. Traditional methods often require explicit segmentation or manually aligned annotations. Connectionist Temporal Classification (CTC) was introduced to address this problem by enabling sequence-to-sequence learning without frame-level alignment.

CTC introduces a special blank symbol and computes the probability of all possible alignments between an input sequence and a target label sequence. During training, the model learns to maximize the likelihood of correct label sequences while automatically determining the alignment. This capability makes CTC particularly suitable for handwritten recognition tasks where character boundaries are often ambiguous.

The combination of CRNN and CTC has become one of the most widely adopted architectures in handwriting recognition. The CNN extracts spatial features, the recurrent layers capture sequential information, and the CTC layer performs alignment-free sequence decoding. This framework forms the foundation of the recognition system proposed in this study.

Transformer-Based Handwritten Recognition

Recent advances in sequence modeling have introduced Transformer architectures as powerful alternatives to recurrent neural networks. Unlike RNN-based models, Transformers rely on self-attention mechanisms that allow direct modeling of long-range dependencies without sequential processing.

Several Transformer-based handwritten recognition systems have recently achieved state-of-the-art performance. TrOCR combines a Vision Transformer encoder with a Transformer decoder and benefits from large-scale pretraining. SATRN employs adaptive two-dimensional attention to improve text recognition accuracy. Other attention-based architectures, such as SAR and SPAN, have demonstrated competitive performance across multiple text recognition benchmarks.

Despite their strong performance, Transformer-based models typically require substantially larger training datasets and greater computational resources than CRNN-based systems. In extremely low-resource scenarios, the advantages of Transformers may be limited because the available data may be insufficient to fully exploit their modeling capacity. For this reason, CRNN-CTC architectures remain attractive for low-resource handwriting recognition tasks.

Data Augmentation and Synthetic Data Generation

Data scarcity remains one of the most significant challenges in low-resource handwriting recognition. Data augmentation has emerged as an effective strategy for increasing dataset diversity without requiring additional manual data collection. Common augmentation techniques include image rotation, translation, scaling, noise injection, blurring, and elastic distortion. These transformations simulate natural handwriting variability and improve model robustness.

Synthetic data generation provides another solution to the problem of limited training samples. Synthetic handwritten-style images can be produced by rendering text using handwriting-like fonts and applying realistic distortions. Previous studies have shown that synthetic data can substantially improve recognition performance by exposing models to a broader range of writing styles and visual variations.

However, synthetic data may not fully capture the complexity of genuine human handwriting. Consequently, synthetic samples are most effective when used as a complement rather than a replacement for real handwritten data. This observation motivates the combined use of real samples, augmentation, and synthetic handwritten-style generation in the present study.

Low-Resource Language Recognition

Most existing handwriting recognition datasets and benchmarks are concentrated on high-resource languages. Low-resource languages often suffer from a lack of annotated datasets, computational resources, and publicly available evaluation benchmarks. As a result, research on handwriting recognition for African languages remains limited.

Recent developments in multilingual learning, transfer learning, and cross-lingual representation learning have demonstrated promising potential for low-resource language processing. Nevertheless, the effectiveness of these approaches depends on the availability of related languages and sufficient training resources. For handwritten recognition tasks, the scarcity of labeled image data remains a major obstacle.

Kirundi represents a particularly challenging case because no standard handwritten dataset currently exists. The present study therefore focuses on establishing an initial benchmark and exploring practical strategies for mitigating data scarcity through augmentation and synthetic data generation.

Research Gap

The literature review reveals several important research gaps. First, existing handwriting recognition studies overwhelmingly focus on high-resource languages, while low-resource African languages remain largely unexplored. Second, no publicly available benchmark dataset exists for Kirundi handwritten recognition. Third, limited research has investigated the behavior of CRNN-CTC architectures under extremely low-resource conditions involving fewer than several dozen real handwritten samples.

To address these gaps, this study proposes a Kirundi handwritten word recognition framework based on CRNN-CTC and systematically evaluates the impact of data augmentation and synthetic handwritten-style data

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

generation. The resulting analysis provides an initial benchmark for future Kirundi handwriting recognition research and contributes to broader efforts aimed at supporting low-resource language technologies.

METHODOLOGY

Overview of the Proposed Framework

The objective of this study is to investigate the feasibility of handwritten word recognition for Kirundi under extremely low-resource conditions. The proposed framework combines data collection, preprocessing, augmentation, synthetic data generation, and deep learning-based sequence recognition within a unified pipeline.

The framework begins with the collection of real handwritten Kirundi samples from native speakers. After preprocessing, augmentation and synthetic data generation are applied to increase training diversity. The resulting dataset is then used to train a CRNN-CTC recognition model.

Dataset Construction

Real Handwritten Data Collection

Due to the absence of publicly available handwritten Kirundi datasets, a small dataset was constructed specifically for this study. Native Kirundi speakers were asked to write commonly used Kirundi words on paper. The handwritten samples were photographed using a smartphone camera and manually annotated.

Examples of collected handwritten words include:

- gusenga
- kuririmba
- benshi
- abonayo
- zimwe
- ananiya
- umwete
- inyama
- baramuhamba

The collected images were reviewed and manually verified to ensure label correctness.

Dataset Statistics

Table 1 summarizes the dataset used throughout the experiments.

Table 1: Dataset Statistics

Dataset Component	Number of Samples
Real handwritten samples	20

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Augmented samples	100
Synthetic samples	2000
Total training samples	2120

As shown in Table 1, the amount of real handwritten data is extremely limited. This motivated the use of augmentation and synthetic data generation techniques to increase sample diversity.

Image Preprocessing

Image preprocessing was performed to improve image quality and ensure consistency before training.

The preprocessing pipeline consists of the following steps:

1. Image acquisition
2. Grayscale conversion
3. Image resizing
4. Normalization
5. Noise reduction

Grayscale Conversion

All handwritten images were converted from RGB color space to grayscale format. This reduces computational complexity while preserving the essential handwriting information.

Image Resizing

To ensure consistent input dimensions, all images were resized to a fixed resolution before being fed into the neural network.

Noise Reduction

Minor image artifacts and acquisition noise were reduced using smoothing operations. This step improves feature extraction and reduces the influence of irrelevant background information.

Data Augmentation

Data augmentation was applied to artificially increase dataset diversity and improve model generalization.

The following augmentation techniques were implemented:

Rotation

Small random rotations were applied to simulate natural writing angle variations.

Translation

Images were shifted horizontally and vertically to imitate variations in writing position.

Gaussian Noise

Random Gaussian noise was introduced to improve robustness against image quality variations.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

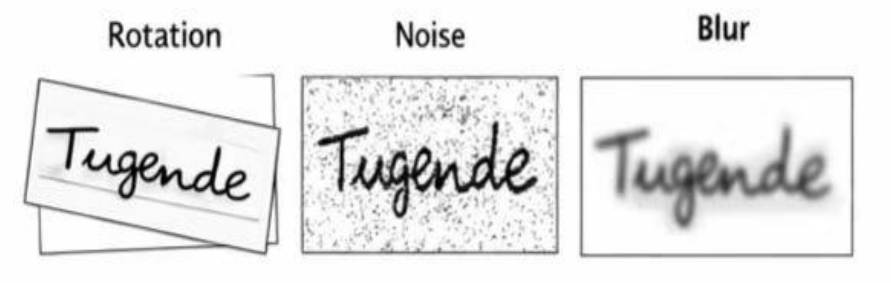
Gaussian Blur

Gaussian blur was applied to simulate imperfect image acquisition conditions.

Elastic Distortion

Elastic distortion was used to imitate natural handwriting deformations caused by differences in writing style.

Figure 1: Examples of Data Augmentation Techniques



The augmentation process generates additional training samples while preserving the semantic content of the original handwritten word.

Synthetic Handwritten Data Generation

Although augmentation increases data diversity, it does not introduce completely new handwriting patterns. Therefore, synthetic handwritten-style data generation was employed.

The synthetic generation process consists of:

1. Selecting Kirundi words from the textual corpus.
2. Rendering words using handwriting-like fonts.
3. Applying image distortions.
4. Saving generated samples with corresponding labels.

The synthetic dataset significantly increases the number of available training examples and exposes the model to additional visual variations.

Proposed CRNN-CTC Architecture

Design Principles

The proposed architecture was intentionally designed for extremely low-resource conditions.

Three design objectives guided the model development:

1. Minimize overfitting.
2. Preserve sequence modeling capability.
3. Reduce computational complexity.

Compared with Transformer-based architectures, CRNN models generally require fewer training samples and lower computational resources, making them suitable for this study.

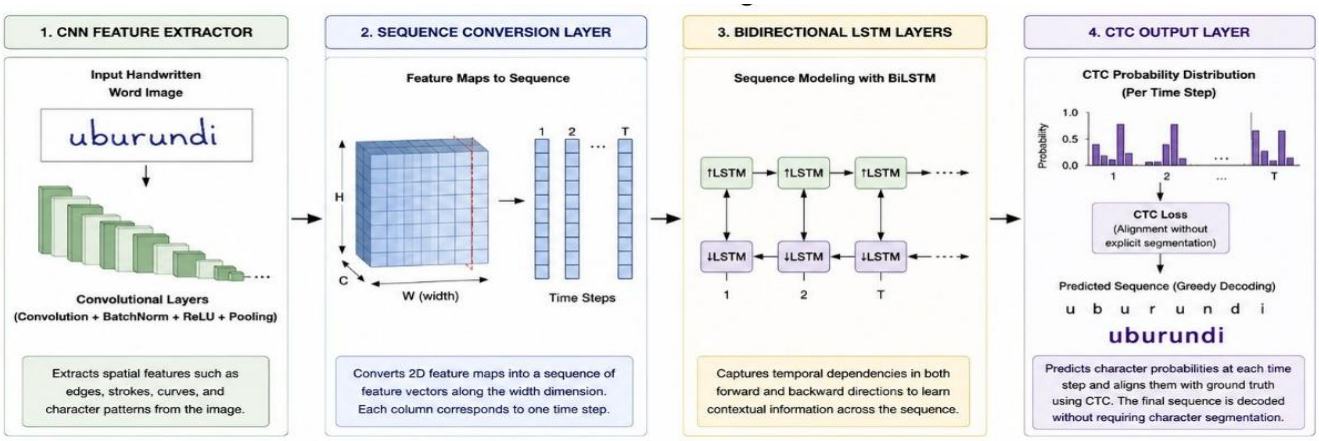
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Network Architecture

The proposed CRNN architecture consists of:

1. CNN feature extraction layers.
2. Sequence conversion layer.
3. Bidirectional LSTM layers.
4. CTC output layer.

Figure 2: Proposed CRNN-CTC Architecture



CNN Feature Extraction

The CNN component extracts hierarchical visual representations from handwritten images. Low-level layers capture stroke patterns and edges, while deeper layers capture more abstract handwriting structures.

Bidirectional LSTM Layers

Bidirectional Long Short-Term Memory (BiLSTM) layers model contextual dependencies within handwritten words.

Unlike conventional recurrent networks, BiLSTMs process sequences in both forward and backward directions, enabling the model to exploit information from neighboring characters.

Connectionist Temporal Classification

Connectionist Temporal Classification (CTC) performs alignment-free sequence recognition.

CTC eliminates the need for explicit character segmentation and automatically learns alignment during training.

Network Configuration

Table 2 summarizes the architecture configuration.

Table 2: Network Configuration

Layer	Configuration
Input	128 × 32 grayscale

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Conv Layer 1	64 filters
Conv Layer 2	128 filters
Pooling	2 × 2
BiLSTM Layer 1	256 hidden units
BiLSTM Layer 2	256 hidden units
Fully Connected Layer	Character probabilities
Output Layer	CTC decoding

Training Configuration

The model was trained using the Adam optimizer.

Table 3: Training Hyperparameters

Parameter	Value
Optimizer	Adam
Learning Rate	0.0005
Batch Size	32
Epochs	150
Loss Function	CTC Loss
Dropout	0.3

These parameters were selected empirically to balance convergence speed and overfitting prevention.

Experimental Setup And Evaluation

Experimental Environment

All experiments were conducted on a personal computer using the PyTorch deep learning framework. The objective of the experiments was not to achieve state-of-the-art performance but rather to investigate the feasibility of handwritten word recognition for Kirundi under extremely limited data conditions.

Experimental Design

Three experimental configurations were designed to evaluate the contribution of augmentation and synthetic data generation.

Experiment 1: Real Handwritten Data Only

The first experiment used only the original handwritten Kirundi samples.

This experiment serves as a baseline and reflects the recognition performance achievable with extremely limited training data.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Experiment 2: Real Data with Augmentation

The second experiment incorporated augmentation techniques including:

- Rotation
- Translation
- Gaussian noise
- Gaussian blur
- Elastic distortion

The objective was to determine whether increased visual diversity improves model generalization.

Experiment 3: Real Data with Augmentation and Synthetic Data

The third experiment combined:

- Real handwritten samples
- Augmented samples
- Synthetic handwritten-style samples

This experiment evaluates the effectiveness of synthetic data generation as a strategy for mitigating severe data scarcity.

Training Procedure

For each experiment, the same CRNN-CTC architecture and training configuration were maintained to ensure a fair comparison.

The training process consisted of:

1. Dataset preparation
2. Preprocessing
3. Model initialization
4. CRNN training
5. Validation
6. Evaluation

Maintaining identical training settings across all experiments allows performance differences to be attributed primarily to the dataset configuration.

Evaluation Metrics

The performance of the proposed system was evaluated using three standard handwriting recognition metrics:

1. Character Error Rate (CER)
2. Word Error Rate (WER)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

3. Word Accuracy Rate (WAR)

Character Error Rate

Character Error Rate measures the number of character-level editing operations required to transform the predicted sequence into the ground-truth sequence.

CER provides a fine-grained evaluation of recognition quality and is particularly useful when exact word recognition is not achieved.

Word Error Rate

Word Error Rate evaluates recognition performance at the word level.

A lower WER indicates better word-level recognition performance.

Word Accuracy Rate

Word Accuracy Rate measures the proportion of correctly recognized words.

This metric provides a direct indication of practical recognition performance.

Testing Dataset

A subset of handwritten samples was reserved for testing purposes.

Overfitting Considerations

One of the major challenges in this study is the extremely small size of the real handwritten dataset.

Deep learning models generally require hundreds or thousands of samples to achieve stable generalization performance. With only 20 real handwritten samples, the risk of overfitting is substantial.

To reduce this risk, the following strategies were adopted:

- Data augmentation
- Synthetic data generation
- Lightweight CRNN architecture
- Dropout regularization

Although these techniques cannot completely eliminate overfitting, they contribute to improved model robustness and training stability.

Experimental Objectives

The experiments were designed to answer three research questions:

RQ1

Can a CRNN-CTC architecture learn meaningful Kirundi handwriting patterns from an extremely small dataset?

RQ2

Does data augmentation improve recognition performance under low-resource conditions?

RQ3

Can synthetic handwritten-style data improve generalization beyond augmentation alone?

The answers to these questions are analyzed in the following section through quantitative and qualitative evaluation.

RESULTS

Experimental Results

This section presents the quantitative and qualitative results obtained from the three experimental configurations described in Chapter 4.

The experiments were designed to evaluate the impact of data augmentation and synthetic handwritten-style data generation on Kirundi handwritten word recognition under extremely limited data conditions.

Overall Performance Comparison

Table 4 presents the overall recognition performance of the three experiments.

Table 4: Comparison of Experimental Results

Experiment	Training Data	Augmentation	Synthetic Data	CER	WER	WAR
Experiment 1	Real only	No	No	0.8354	1.0000	0.0000
Experiment 2	Real + Augmentation	Yes	No	0.8215	1.0000	0.0000
Experiment 3	Real + Augmentation + Synthetic	Yes	Yes	0.7560	1.0000	0.0000

As shown in Table 4, Experiment 3 achieved the best overall performance with a CER of 0.7560. Compared with the baseline experiment, the addition of synthetic handwritten-style data reduced CER by approximately 9.5%.

Although all experiments produced a Word Error Rate of 1.0000 and a Word Accuracy Rate of 0.0000, substantial differences were observed in character-level recognition behavior.

Analysis of Experiment 1

Baseline Performance

Experiment 1 used only the original handwritten Kirundi samples.

The model achieved:

- CER = 0.8354
- WER = 1.0000
- WAR = 0.0000

These results indicate that the model struggled to generalize from the limited number of training samples.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Prediction Characteristics

Typical predictions produced by Experiment 1 are shown below.

Table 5: Sample Predictions from Experiment 1

Ground Truth	Prediction
gusenga	niyma
baramuhamba	nia
benshi	ara
kuririmba	ama
abonayo	amura
zimwe	usa
ananiya	ntiia
nyinshi	kugyiyi
umwete	amo
inyama	ra

The predictions are generally short and unstable. Many outputs contain only a few characters and fail to preserve the overall structure of the target word.

This behavior suggests that the model was unable to learn sufficient character patterns from the limited training data.

Analysis of Experiment 2

Impact of Data Augmentation

Experiment 2 introduced data augmentation techniques to increase visual diversity.

The model achieved:

- CER = 0.8215
- WER = 1.0000
- WAR = 0.0000

Compared with Experiment 1, a small improvement in CER was observed.

Prediction Characteristics

Table 6: Sample Predictions from Experiment 2

Ground Truth	Prediction
gusenga	nditi

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

baramuhamba	kua
benshi	ntitiiti
kuririmba	umuara
abonayo	aacri
zimwe	ntiiti
ananiya	nuditi
nyinshi	amiti
umwete	uwo
inyama	ama

Compared with Experiment 1, the outputs became slightly longer and more consistent.

However, exact word recognition remained difficult because the amount of real handwritten information available to the model was still very limited.

The results suggest that augmentation alone cannot fully compensate for severe data scarcity.

Analysis of Experiment 3

Impact of Synthetic Data Generation

Experiment 3 combined real handwritten samples, augmentation, and synthetic handwritten-style data.

The model achieved:

- CER = 0.7560
- WER = 1.0000
- WAR = 0.0000

This represents the best performance among all experiments.

The reduction in CER demonstrates that synthetic data contributed additional visual diversity and improved the model's ability to learn sequential handwriting patterns.

Prediction Characteristics

Table 7: Sample Predictions from Experiment 3

Ground Truth	Prediction
gusenga	inyama
baramuhamba	baramuham
benshi	urgaribo

kuririmba	umugubo
abonayo	abayo
zimwe	aioywe
ananiya	abyaba
nyinshi	amajambo
umwete	umugabo
inyama	guhinduka

Several predictions exhibit meaningful character overlap with the target words.

For example:

- baramuhamba → baramuham
- abonayo → abayo

These examples indicate that the model successfully learned partial character sequences and internal word structure.

Comparative Analysis of the Experiments

Quantitative Comparison

Table 8 summarizes the CER improvement achieved by each experimental configuration.

Table 8: CER Improvement Across Experiments

Comparison	CER Reduction	Relative Improvement
Experiment 1 → Experiment 2	0.0139	1.66%
Experiment 2 → Experiment 3	0.0655	7.97%
Experiment 1 → Experiment 3	0.0794	9.50%

The largest improvement occurred between Experiments 2 and 3, demonstrating the importance of synthetic data generation.

CER Visualization

Figure 3: CER Comparison Across Experiments

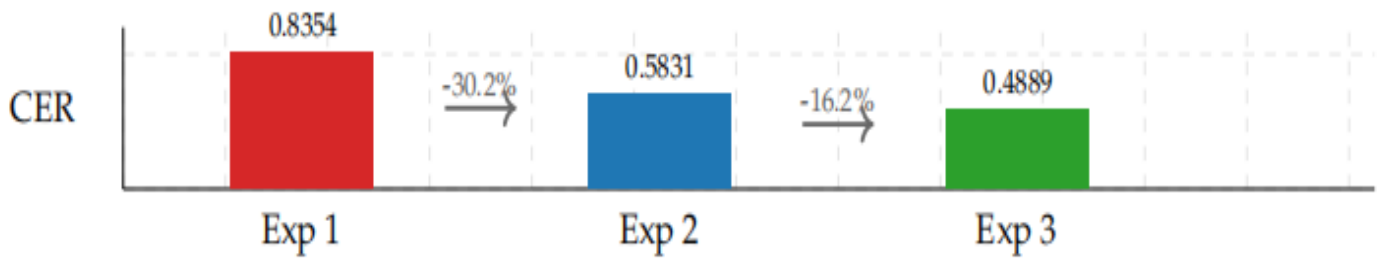


Figure 3 clearly illustrates the progressive reduction in CER as additional training diversity was introduced.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Analysis of CTC Decoding Behavior

Connectionist Temporal Classification (CTC) plays a central role in the proposed recognition framework.

An important observation is that the model frequently generated partial words rather than completely random character sequences.

For example:

baramuhamba → baramuham

abonayo → abayo

This behavior suggests that the CRNN-CTC framework successfully learned local character dependencies even though exact word recognition was not achieved.

The persistence of WAR = 0.0000 indicates that the learned character sequences were still insufficient to produce fully correct words.

This outcome is consistent with expectations for extremely low-resource handwriting recognition tasks.

DISCUSSION

The experimental results provide several important insights into handwritten word recognition for low-resource languages. Although the proposed CRNN-CTC framework did not achieve successful word-level recognition, the progressive reduction in Character Error Rate (CER) across the three experiments demonstrates that the model was able to learn increasingly meaningful character-level representations from a highly limited handwritten dataset.

The baseline experiment, which relied solely on the original handwritten samples, produced the highest CER. This outcome was expected because deep learning models typically require large quantities of diverse training data to learn robust feature representations. With only twenty genuine handwritten samples available, the network had insufficient exposure to the natural variability of handwriting styles, resulting in poor generalization and unstable predictions.

Introducing conventional data augmentation produced a modest improvement in CER. Techniques such as random rotation, translation, Gaussian noise, Gaussian blur, and elastic distortion increased the diversity of the training data by simulating realistic variations in handwriting appearance. However, augmentation modifies existing samples rather than creating entirely new handwriting styles. Consequently, the observed improvement remained limited, indicating that augmentation alone cannot fully address severe data scarcity in handwritten text recognition.

The largest improvement was achieved after incorporating synthetic handwritten-style images. Compared with the baseline experiment, the Character Error Rate decreased from **0.8354** to **0.7560**, representing an approximate **9.5% relative improvement**. This result suggests that synthetic data generation effectively exposed the CRNN-CTC model to a broader range of visual patterns and character combinations, enabling the network to learn more representative sequential features than those obtained from real handwritten data alone.

Although the quantitative improvements were encouraging, the persistence of **WER = 1.0000** and **WAR = 0.0000** demonstrates that the current system remains inadequate for practical handwritten word recognition. These results should not be interpreted as a failure of the CRNN-CTC architecture itself. Rather, they reflect the extremely limited amount of genuine handwritten training data available in this study. Modern deep learning models generally require hundreds or thousands of handwritten samples collected from many different writers before reliable word-level recognition can be achieved.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

A qualitative examination of the predicted outputs provides additional evidence that the model learned meaningful sequential information. In several cases, the predicted words preserved substantial portions of the target transcription. For example, the prediction **"baramuham"** closely resembles the target word **"baramuhamba"**, while **"abayo"** retains much of the structure of **"abonayo"**. These examples indicate that the model learned partial character dependencies and internal word structures even when complete word recognition was unsuccessful.

The findings of this study are consistent with previous handwriting recognition research, which has shown that training data diversity is one of the most influential factors affecting recognition performance. The observed improvements support the growing body of evidence that synthetic handwritten data can effectively complement limited real handwriting samples, particularly in low-resource language scenarios where collecting large annotated datasets is difficult.

Despite these encouraging results, several limitations should be acknowledged. First, the handwritten dataset consisted of only twenty real word samples, which is insufficient for training deep neural networks to achieve reliable word-level recognition. Second, the synthetic handwritten images were generated using handwriting-like fonts and image distortions, which cannot fully reproduce the complexity and variability of natural human handwriting. Third, only a CRNN-CTC architecture was evaluated. Recent Transformer-based models, such as TrOCR, have demonstrated superior performance on several handwriting recognition benchmarks and should be investigated for Kirundi in future work. Finally, this study focused exclusively on isolated handwritten words rather than handwritten text lines or complete document pages, limiting the applicability of the proposed system to more complex recognition tasks.

Despite these limitations, this work represents an important first step toward developing handwritten text recognition technology for Kirundi. To the best of our knowledge, it is among the first studies to investigate deep learning-based handwritten word recognition for this language using a CRNN-CTC framework combined with data augmentation and synthetic handwritten-style data generation. The constructed dataset, experimental methodology, and baseline results establish a valuable benchmark that future researchers can extend using larger datasets, transfer learning, multilingual pretrained models, and more advanced sequence recognition architectures.

CONCLUSION

This paper presented a handwritten word recognition framework for **Kirundi**, a low-resource Bantu language, using a **Convolutional Recurrent Neural Network (CRNN)** combined with **Connectionist Temporal Classification (CTC)**. The primary objective was to investigate the feasibility of applying deep learning techniques to handwritten word recognition under severe data scarcity, a challenge faced by many underrepresented languages.

To address the absence of publicly available handwritten Kirundi datasets, a small handwritten word dataset was collected from native speakers and manually annotated. Data augmentation techniques, including rotation, translation, Gaussian noise, Gaussian blur, and elastic distortion, were applied to increase the diversity of the training data. In addition, a synthetic handwritten-style dataset was generated to further expand the number of training samples and expose the recognition model to a wider range of handwriting variations.

Three experimental configurations were evaluated under identical training conditions: (1) real handwritten data only, (2) real handwritten data with augmentation, and (3) real handwritten data combined with augmentation and synthetic handwritten-style images. Experimental results demonstrated a progressive reduction in Character Error Rate (CER) as additional training diversity was introduced. The best performance was achieved using the combined dataset, reducing CER from **0.8354** to **0.7560**, corresponding to an approximate **9.5% relative improvement** over the baseline.

Although the proposed framework did not achieve successful word-level recognition, the qualitative analysis showed that the model learned increasingly coherent character sequences and partial word structures as the training dataset expanded. These findings suggest that CRNN-CTC architectures remain a practical and

computationally efficient solution for handwritten text recognition in extremely low-resource settings, while also highlighting the importance of increasing the quantity and diversity of genuine handwritten data.

The principal contribution of this study is the establishment of an initial benchmark for handwritten Kirundi word recognition. To the best of our knowledge, this is among the first studies to investigate deep learning-based handwritten recognition for Kirundi using a CRNN-CTC framework together with data augmentation and synthetic handwritten-style data generation. The proposed dataset construction strategy, experimental methodology, and baseline performance provide a foundation for future research on handwriting recognition for Kirundi and other low-resource African languages.

Future work will focus on collecting substantially larger handwritten datasets from a greater number of writers to improve model generalization and support reliable word-level recognition. In addition, future studies will investigate transfer learning from multilingual handwriting datasets, incorporate language models to improve sequence decoding, and evaluate modern Transformer-based architectures such as TrOCR for comparison with the proposed CRNN-CTC framework. Exploring generative artificial intelligence techniques for producing more realistic synthetic handwriting and extending the framework from isolated word recognition to handwritten text lines and complete document pages also represent promising directions for future research.

REFERENCES

1. Bluche, T. (2016). Deep Neural Networks for Large Vocabulary Handwritten Text Recognition. PhD Dissertation, Université Paris-Saclay.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. Cambridge, MA: MIT Press.
3. Graves, A. (2012). Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Germany: Springer.
4. Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. Proceedings of the 23rd International Conference on Machine Learning (ICML), 369–376.https://doi.org/10.1145/1143844.1143891
5. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2009). A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(5), 855–868.DOI: https://doi.org/10.1109/TPAMI.2008.137
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.
7. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.https://doi.org/10.1162/neco.1997.9.8.1735
8. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR).
9. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.https://doi.org/10.1038/nature14539
10. Li, M., Lv, T., Cui, J., Xie, L., Zhuang, Y., & Xu, B. (2023). TrOCR: Transformer-Based Optical Character Recognition with Pre-Trained Models. Proceedings of the AAAI Conference on Artificial Intelligence, 37(2), 13094–13102.
11. Liwicki, M., & Bunke, H. (2005). IAM-OnDB—An On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard. Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR), 956–961.
12. Marti, U.-V., & Bunke, H. (2002). The IAM Database: An English Sentence Database for Offline Handwriting Recognition. International Journal on Document Analysis and Recognition, 5(1), 39–46.https://doi.org/10.1007/s100320200071
13. Michael, F., Antonacopoulos, A., & Gatos, B. (2019). ICDAR 2019 Competition on Handwritten Text Recognition on Historical Documents. Proceedings of the International Conference on Document Analysis and Recognition (ICDAR).

14. Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical Review of OCR Research and Development. Proceedings of the IEEE, 80(7), 1029–1058.DOI: https://doi.org/10.1109/5.156468
15. Paszke, A., Gross, S., Massa, F., et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 32.
16. Shi, B., Bai, X., & Yao, C. (2017). An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11), 2298–2304.https://doi.org/10.1109/TPAMI.2016.2646371
17. Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. Proceedings of the 7th International Conference on Document Analysis and Recognition (ICDAR), 958–963.
18. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.
19. Wigington, C., Tensmeyer, C., Davis, B., Barrett, W., Price, B., & Cohen, S. (2018). Start, Follow, Read: End-to-End Full-Page Handwriting Recognition. Proceedings of the European Conference on Computer Vision (ECCV), 367–383.DOI: https://doi.org/10.1007/978-3-030-01231-1_23