IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 July 2024, accepted 12 August 2024, date of publication 16 August 2024, date of current version 27 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444897

RESEARCH ARTICLE

Handwritten Amharic Word Recognition With Additive Attention Mechanism

RUCHIKA MALHOTRA[1] AND MARU TESFAYE ADDIS[2,3]

[1]Department of Software Engineering, Delhi Technological University, New Delhi 110042, India
[2]Department of Computer Science, Debre Tabor University, Debre Tabor 251272, Ethiopia
[3]Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India

Corresponding author: Maru Tesfaye Addis (marutcomp@gmail.com)

ABSTRACT Amharic stands as the second most widely spoken Semitic language globally, trailing only Arabic. As a result, it is a highly resourceful language for hardcopy-document collection, and automatic recognition technology is required. Recognizing handwritten Amharic words is a difficult task because of factors such as variations in individual handwriting for the same words, connectivity of two words without spaces, similarities in the shapes of alphabet characters, and noise from scanned images. The Amharic alphabet has a vast character set that adapted most of its characters from the most popular script called Ethiopic script, formerly known as Ge'ez script, and added unique characters. Despite significant advancements in optical character recognition (OCR) research, Amharic script recognition has received less attention. This study employs a deep learning (DL) approach with additive attention to recurrent neural networks (RNN) to achieve precise recognition of handwritten Amharic words. Seven convolutional neural networks (CNN) and two RNN, using a connectionist temporal classification (CTC) strategy, make up this recognition model architecture, enabling efficient recognition through sequential feature extraction. The study addressed deep learning data insufficiency by using augmentation techniques to increase datasets. The study used an original dataset of 12,047 handwritten Amharic words and an augmented dataset of 22,00 images. The developed model achieved an average character error rate (CER) of 2.84% and an average word error rate (WER) of 9.75% for the testing dataset. These results are indeed promising, offering a glimpse into the potential of this attention-based approach for handwritten Amharic word recognition. This research represents a significant step toward bridging the gap in OCR technology for Amharic script and showcases the transformative capabilities of DL in pattern recognition.

INDEX TERMS Amharic handwriting, CNN, deep learning, handwritten word recognition, pattern recognition.

I. INTRODUCTION

Amharic stands as the second most widely spoken Semitic language globally, trailing only Arabic [1], [2], [3]. This rich language boasts a vast community of over 100 million speakers; spread across various nations, including Ethiopia, which is the origin country [4], [5]. Notably, Amharic holds the prestigious position of being the official language of Ethiopia [4], [6], [7], [8], underscoring its significance and influence.

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

One of the remarkable aspects of Amharic is its written form [9], which draws upon the indigenous symbols of the Ge'ez language [10], predominantly employed in Ethiopia and Eritrea [11], [12], giving rise to the unique and ancient script known as the Ethiopic script [13]. The Amharic alphabet includes a diverse array of characters, encompassing regular letters, digits, and punctuation marks. Here, the Amharic language adapted Arabic numerals rather than the Ge'ez digit [14]. In total, this alphabet comprises a staggering 300 distinct character symbols, as visually depicted in **FIGURE 1**. There are 267 alphabetic characters to create Amharic words, as well as 10 Arabic numerals to display

0	ሀ	15	ሐ	30	ሠ	45	ሯ	60	ሾ	75	ብ	90	ቴ	105	ኃ	120	ኛ	135	ኪ
1	ሁ	16	ሑ	31	ሡ	46	ሰ	61	ሿ	76	ቦ	91	ት	106	ኄ	121	ኜ	136	ካ
2	ሂ	17	ሒ	32	ሢ	47	ሱ	62	ቀ	77	ቧ	92	ቶ	107	ኅ	122	ኝ	137	ኬ
3	ሃ	18	ሓ	33	ሣ	48	ሲ	63	ቁ	78	ቨ	93	ቷ	108	ኆ	123	ኞ	138	ክ
4	ሄ	19	ሔ	34	ሤ	49	ሳ	64	ቂ	79	ቩ	94	ቸ	109	ነ	124	ኟ	139	ኮ
5	ህ	20	ሕ	35	ሥ	50	ሴ	65	ቃ	80	ቪ	95	ቹ	110	ኑ	125	አ	140	ኳ
6	ሆ	21	ሖ	36	ሦ	51	ስ	66	ቄ	81	ቫ	96	ቺ	111	ኒ	126	ኡ	141	ኸ
7	ለ	22	መ	37	ሧ	52	ሶ	67	ቅ	82	ቬ	97	ቻ	112	ና	127	ኢ	142	ኹ
8	ሉ	23	ሙ	38	ረ	53	ሷ	68	ቆ	83	ቭ	98	ቼ	113	ኔ	128	ኣ	143	ኺ
9	ሊ	24	ሚ	39	ሩ	54	ሸ	69	ቋ	84	ቮ	99	ች	114	ን	129	ኤ	144	ኻ
10	ላ	25	ማ	40	ሪ	55	ሹ	70	በ	85	ቯ	100	ቾ	115	ኖ	130	እ	145	ኼ
11	ሌ	26	ሜ	41	ራ	56	ሺ	71	ቡ	86	ተ	101	ቿ	116	ኗ	131	ኦ	146	ኽ
12	ል	27	ም	42	ሬ	57	ሻ	72	ቢ	87	ቱ	102	ኀ	117	ኘ	132	ኧ	147	ኾ
13	ሎ	28	ሞ	43	ር	58	ሼ	73	ባ	88	ቲ	103	ኁ	118	ኙ	133	ከ	148	ዀ
14	ሏ	29	ሟ	44	ሮ	59	ሽ	74	ቤ	89	ታ	104	ኂ	119	ኚ	134	ኩ	149	ዃ

150	ወ	165	ዘ	180	ዧ	195	ዷ	210	ጎ	225	ጬ	240	ጻ	255	ፋ	270	<	285	1
151	ዉ	166	ዙ	181	የ	196	ጀ	211	ጐ	226	ጭ	241	ጼ	256	ፌ	271	(	286	2
152	ዊ	167	ዚ	182	ዩ	197	ጁ	212	ጓ	227	ጮ	242	ጽ	257	ፍ	272	«	287	3
153	ዋ	168	ዛ	183	ዪ	198	ጂ	213	ጠ	228	ጯ	243	ጾ	258	ፎ	273	፣	288	4
154	ዌ	169	ዜ	184	ያ	199	ጃ	214	ጡ	229	ጰ	244	ጿ	259	ፏ	274	%	289	5
155	ው	170	ዝ	185	ዬ	200	ጄ	215	ጢ	230	ጱ	245	ፀ	260	ፐ	275	»	290	6
156	ዎ	171	ዞ	186	ይ	201	ጅ	216	ጣ	231	ጲ	246	ፁ	261	ፑ	276	)	291	7
157	ዏ	172	ዟ	187	ዮ	202	ጆ	217	ጤ	232	ጳ	247	ፂ	262	ፒ	277	>	292	8
158	ዐ	173	ዠ	188	ደ	203	ጇ	218	ጥ	233	ጴ	248	ፃ	263	ፓ	278	.	293	9
159	ዑ	174	ዡ	189	ዱ	204	ገ	219	ጦ	234	ጵ	249	ፄ	264	ፔ	279	+	294	፡
160	ዒ	175	ዢ	190	ዲ	205	ጉ	220	ጧ	235	ጶ	250	ፅ	265	ፕ	280	፤	295	፧
161	ዓ	176	ዣ	191	ዳ	206	ጊ	221	ጨ	236	ጷ	251	ፆ	266	ፖ	281	–	296	…
162	ዔ	177	ዤ	192	ዴ	207	ጋ	222	ጩ	237	ጸ	252	ፈ	267	ፗ	282	።	297	*
163	ዕ	178	ዥ	193	ድ	208	ጌ	223	ጪ	238	ጹ	253	ፉ	268	!	283	/	298	#
164	ዖ	179	ዦ	194	ዶ	209	ግ	224	ጫ	239	ጺ	254	ፊ	269	፦	284	0	299	?

FIGURE 1. **Amharic alphabetic symbols and their index values were used in this study [7].**

numerical values. The Amharic writing system uses the remaining 23 characters as punctuation marks.

Currently, numerous federal and regional organizations in Ethiopia are actively seeking to reduce paper usage and transition to digital workflow [12]. Institutions spanning hospitals, bank offices, primary school libraries, and locality offices are inundated with many handwritten documents [13]. Consequently, it is critical to convert these handwritten resources into digital formats for streamlined management. The escalating demand for effective OCR technology to meet this imperative is evident, as OCR serves as a pivotal component of any text recognition task [15].

OCR operates as a transformative system, capable of converting both printed and handwritten text from hardcopy documents into formats comprehensible by computers [16]. Its overarching goal is to convert these documents into electronic counterparts, making them easier to manage within computer systems. The implementation of an OCR system, which is the central problem and popular in pattern recognition [17], [18], offers numerous advantages to organizations. It bolsters document management operations, enhances data accessibility, and expedites the retrieval of information, collectively contributing to more efficient organizational processes.

Recognizing handwritten Amharic words is not an effortless task with OCR technology. This is due to factors such as individual writing style variation for the same words [9], connectivity of two words without space, and small shape differences among the letters in the alphabet [19], [20]. Noise from scanned images and vast character sets from the alphabets contribute to this difficulty. So, identifying handwritten Amharic words in image documents presents a unique OCR challenge.

Despite the research on pattern recognition for popular scripts, Amharic OCR has received less attention than other language OCR studies [2], [5], [21], [22]. Languages such as Latin, Chinese, and Arabic scripts [23] were studied well with different approaches. Previous research [1], [2], [24], [25] on Amharic-related script OCR has mostly been based on printed texts. But the case of handwritten Amharic texts has not been studied much [21], [26] because there aren't many public research datasets [12], [26]. It is rare to find publicly available Amharic handwritten text datasets, which are typically required for deep learning algorithms [23]. So, developing a DL model and preparing the required dataset is a must in the handwritten Amharic word recognition area to fill the gaps.

Conventional transcription and analysis methods often require manual involvement and can be prone to errors. Consequently, there is a demand for automated and effective methodologies capable of precisely recognizing Amharic words. Such techniques would be critical in preserving and comprehending handwritten documents, as well as addressing the challenges associated with traditional methods. Recent advances in artificial intelligence, specifically the use of DL techniques, have greatly improved pattern recognition [12]. They have proven to be more effective than traditional machine learning (ML) methods [27].

Nevertheless, the effectiveness of these DL techniques is contingent upon having substantial labeled data, a resource that can prove challenging and expensive to acquire in numerous instances. In response to this constraint, image augmentation has surfaced as a potent strategy. This method increases both the diversity and volume of training datasets, enhancing the model's ability to generalize and improving its resilience [12].

This study makes significant contributions to the study of Amharic OCR, offering the following key advancements:

1) Expanding the handwritten Amharic word dataset: A crucial contribution lies in the expansion of the handwritten Amharic word dataset from 12,047 to 34,047 images. This substantial increment was achieved through meticulous data augmentation techniques applied to the original dataset of 12,047 handwritten Amharic word images prepared by Abdurahma et al. [7].
2) Development of a comprehensive recognition model: The study introduces a comprehensive handwritten word recognition model developed using the TensorFlow framework. In this model, the best parts of CNN layers for automatically extracting features, bidirectional gated recurrent unit (BGRU) networks for

sequencing, and CTC loss functions [28] are all put together. Additionally, this architecture applies an additive attention mechanism to select the focused region of the image. This holistic architecture empowers the model to autonomously extract pertinent features while harnessing the temporal context of the image. This integrated approach streamlines the recognition process and enhances the model's performance.

3) Performance evaluation of the recognition model: The study conducts rigorous evaluations of the developed handwritten Amharic word recognition model. After conducting various experiments, including the selection of hyper parameters value, the model's performance is evaluated. The proposed method of handwriting Amharic word recognition appears to be effective based on these results. These preliminary findings signify a promising stride forward in the domain of OCR for handwritten Amharic script recognition.

These contributions collectively represent significant strides toward advancing the capabilities of OCR technology for the Amharic script, illustrating the promise of the proposed approach in recognizing handwritten Amharic words with enhanced precision and efficacy.

Throughout this paper, the following sections are divided into four parts: Section II presents an overview of relevant research related to handwritten Amharic word recognition. Section III describes the proposed approach, including the DL architecture and training procedures. In Section IV, we examine the experimental setup and present the evaluation results. Lastly, Section V provides a summary of this study.

II. LITERATURE REVIEW

Exploration in the realm of Amharic OCR remains an underdeveloped area, both within ML and DL techniques. This holds particularly true when addressing handwritten Amharic words. Within this context, this section explores the details of the Amharic alphabet and its writing system, exploring diverse methodologies employed for the recognition of handwritten Amharic words. Past research has predominantly leaned towards traditional ML methods, while contemporary studies showcase an increasing trend toward leveraging DL and ensemble approaches [12].

A. AMHARIC ALPHABET AND WRITING SYSTEM

Amharic follows the right direction in writing systems like Latin scripts and others. It has 34 base characters [21], [29], and the remaining characters are derived from those characters by adding small modifications to the base characters to create the vowel sounds [21], [22], [25], as shown in the sample in **TABLE 1**.

In addition to the characters found in **TABLE 1**, the Amharic alphabet includes special characters that generate unique "ua" sounds. Many of these unique characters are modified versions of the base alphabets, typically altered at the bottom. These special characters are displayed in **TABLE 2**.

TABLE 1. Part of Amharic alphabets [20].

	Base sound	Orders						
		1st (ä)	2nd (u)	3rd (ï)	4th (a)	5th (e)	6th (ə)	7th (o)
1	H	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
2	L	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
3	H	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
32	p'	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
33	F	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
34	P	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

TABLE 2. Amharic special characters.

ሏ	ሟ	ሯ	ሷ	ሿ	ቧ
ቷ	ቿ	ኗ	ኟ	ዟ	ዧ
ዷ	ጇ	ጧ	ጯ	ጿ	ፏ

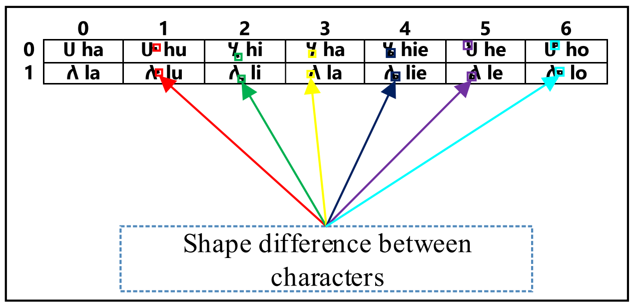

FIGURE 2. Amharic alphabets shape the difference between the base and derived characters [21].

As shown in **FIGURE 2**, the modifications occur in the middle, bottom, and top. Some modifications also happened inside and outside of the base alphabet, which creates a complex structure. Character symbols in handwriting appear to be nearly identical to the base alphabet. This becomes difficult to identify due to the writer's handwritten variation.

B. HANDWRITTEN AMHARIC SCRIPT RECOGNITION

Recognizing handwritten Amharic words is a difficult task because of the intricate nature of the Amharic characters. Additionally, annotated datasets are not widely available, which makes it even more challenging. However, in recent times, some studies have been conducted to address the gaps that exist in handwritten Amharic script recognition.

Alemu and Fuchs [9] published a research paper outlining their approach to recognizing handwritten Amharic bank checks. The main technique they apply in their study is the

Hidden Markova Random Field (HMRF). In their study, they took 7,240 handwritten character images to train the model written by different individuals to consider handwritten variation. They used 20 checks with 713 characters to test the model's performance. By applying contextual information, they achieved an accuracy rate of 99.44%, whereas without contextual information, they achieved an accuracy rate of 89.06%. They recommend using parallel processing to speed up HMRF computations.

Assabie and Begun [8] conducted a study focusing on the recognition of offline handwritten Amharic words. In their study, they used a direction tensor field rather than segmentation to extract features. 177 writers collected the data for the study from various documents written in an unrestricted environment. For training purposes, they extracted 10,932 distinct words from 307 pages with 300 dpi-resolution images.

In their study, "Amharic Handwritten Character Recognition Using Combined Features and Support Vector Machine," Reta et al. [30] integrated various feature extraction techniques to address the challenges and complexities associated with Amharic handwritten character recognition. The study utilized linear discriminant analysis (LDA) to reduce feature dimensions. A multiclass support sector machine (SVM) with error-correcting output codes (ECOC) served as the classifier. The algorithm was trained and tested on both the Amharic dataset and the Chars74K benchmark numeric dataset. Model validation was carried out using a 10-fold cross-validation technique. Because of this, the multiclass SVM classification method showed a strong algorithm that combined features well enough to correctly recognize handwritten characters.

In his study, Tamir [6] investigated the application of CNN for the recognition of Amharic characters in handwriting. The architecture consists of two convolutional layers specifically designed for the classification of handwritten Amharic characters. Tamir compiled a dataset comprising 30,446 characters sourced from approximately 130 different writers. Among these, 27,413 and 3,033 characters were utilized for training and testing purposes, respectively. He implemented and assessed the suggested model design using the Keras and TensorFlow frameworks. Following the training phase, Tamir achieved approximately 99.52% training accuracy after the first epoch, representing the lowest accuracy among the epochs. Meanwhile, the testing accuracy reached approximately 99.71%.

In a study by Agegnehu et al. [31], deep learning was employed to identify Amharic punctuation marks and handwritten digits using a CNN architecture. In their study, they mainly focused on 20 handwritten Ge'ez digits and nine punctuation marks found in Amharic script. They collected 5800 handwritten images from 100 individuals. After preprocessing, they used 32-by-32-pixel binarized images for model development and evaluation. They developed four different models with different CNN architectures. Their models have two, three, four, and five convolutional layers. At the end of each model, there is one fully connected layer with 29 classes in addition to the input layer. The architecture is simple and precise according to the number of layers, which has the advantage of reducing model complexity. They utilized CNN's automatic feature extraction, eliminating the need for handcrafted methods. During the model configuration, they applied trial-and-error approaches. They selected the model with three convolutional layers that generated the best result out of the four model configurations.

The lack of openly available digit datasets limits research on Ge'ez numeral recognition. Before Agegnehu et al. [31], Nur et al. [26] proposed Ge'ez digit recognition by creating a recognition model with better accuracy and an experimental dataset consisting of 51,952-digit images written by 524 people. The researchers suggested exploring various deep-learning techniques for recognizing multiple digits. Their CNN model resulted in 96.21% accuracy.

The majority of researchers focused on the recognition of printed text in the case of the Amharic language. The availability of a public dataset for Latin text motivated researchers to focus on the area. However, there is no enough available dataset publicly for studying the Amharic words in this area. This discourages researchers from making efforts to recognize handwritten Amharic words. In this study, developing a recognition model with better accuracy and preparing the experimental dataset advance the state of the art in handwritten Amharic word recognition.

This study aims to identify handwritten Amharic words in images using an attention mechanism. To accomplish this, we incorporate additional DL techniques such as CNN for automatic feature extraction from the input image, RNN for temporal context modeling from the image, and CTC for efficient sequential classification.

III. METHODS

In this section, we explore handwritten Amharic word recognition by featuring an attention mechanism. We dedicate our study to developing a recognition model that can accurately identify handwritten words within scanned images. Here, we provide a comprehensive overview of the proposed model along with insights into the dataset used for experimentation.

A. THE DATASET PREPARATION

During this research, we realized that having a large dataset is crucial for developing and testing deep learning models [12]. We conducted a thorough search for a dataset that aligned with the specific needs of our deep learning task to overcome the challenge of insufficient data. We used the "HARD-I" dataset, which is accessible on the Google Drive repository (https://sites.google.com/view/hawdb-v1) in this study. The dataset comprises handwritten Amharic word images, prepared by Abdurahman et al. [7], and collected from 60 individuals of varying ages and educational backgrounds. This diversity helps the model effectively handle variations in handwriting.

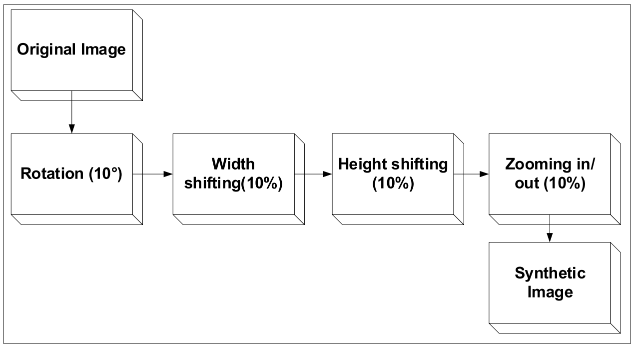

FIGURE 3. Synthetic dataset preparation steps.

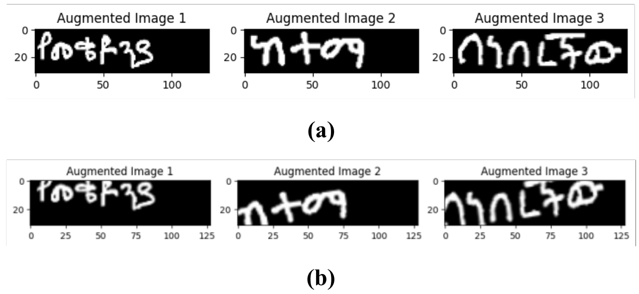

FIGURE 4. Describes the presentation of sample images from the HARD-I database: (a) depicts the original image, while (b) features the augmented image.

This dataset contained 12,047 handwritten Amharic word images that featured 300 unique Amharic alphabets. We classified the dataset into three subsets: training, validation, and testing. To facilitate precise evaluation, a validation dataset was crafted, encompassing 10% of the training dataset.

When faced with the challenge of a limited dataset for DL tasks, data augmentation techniques were implemented to increase its size. Despite resource constraints, we successfully incorporated approximately 22,000 augmented images.

Utilizing image augmentation is instrumental in enabling DL models to effectively learn from limited labeled data, ultimately leading to enhanced recognition performance. This technique has evolved into a crucial tool for overcoming data scarcity, maximizing DL algorithm potential, and finding applications across diverse domains [12]. The following subsections describe the steps and criteria for preparing synthetic datasets.

1) IMAGE AUGMENTATION STEPS

Generating the synthetic data follows the following algorithmic steps and the flowcharts of the steps are depicted in **FIGURE 3**. Image augmentation replicates real-world scenarios by applying various transformations to the original images, allowing the model to better recognize and handle different image variations. These transformations increase the variability of the dataset, helping the model generalize better to new, unseen data. The augmentation steps applied in this study are rotation, width shifting, height shifting and zooming.

During augmentation, images randomly rotate up to 10 degrees. Additionally, we randomly adjust the width and height up to 10%. Also, randomly zooming in or out up to 10%.

2) SAMPLE AUGMENTED IMAGES

The synthetic generation of the dataset provides a more diverse and challenging set of images, enabling a comprehensive evaluation of the OCR model's capabilities in handling Amharic words. Sample handwritten Amharic word images are generated to reflect this diversity. **FIGURE 4** provides visual insights comparing the original dataset with the augmented dataset, highlighting improvements in our proposed model's performance.

B. THE PROPOSED MODEL

In this paper, we leverage the strengths of some DL algorithms: CNN, Attention Mechanism, RNN, and CTC. Each of these algorithms plays a crucial role in different stages of the recognition process [25]. CNN layers excel at capturing spatial hierarchies in data, making them ideal for extracting meaningful features from the raw pixel data of handwritten Amharic characters. This feature extraction transforms the input into an abstract representation that highlights relevant patterns. RNN layers, especially GRU cells, are adept at capturing sequential dependencies and temporal context. For handwritten Amharic word recognition, understanding the order of characters and their dependencies is vital. RNN layers model these dependencies, contextualizing characters within the entire word. By combining CNNs for spatial feature extraction and RNNs for temporal context modeling, our model leverages the strengths of both architectures. The CNN extracts spatial features, which are then processed by the RNN to understand the sequence. This approach ensures effective learning of both local and global features, crucial for accurate handwritten Amharic word recognition.

By combining their capabilities, we've crafted an inclusive and operative method for recognizing handwritten Amharic words. The selection and combination of these DL algorithms are key contributors to the accomplishment of this research, ensuring heightened accuracy and reliability in recognition.

The following sections provide a detailed breakdown of the specific roles each algorithm plays in our holistic approach.

1) INPUT LAYER

In this recognition model, the input layer is the initial part of the model that accepts the input image. It takes 32 by 128-pixel images from scanned images with handwritten Amharic words and makes them ready for the feature extraction part. We reshaped the input image to form 128 rows and 32 columns.

2) CNN LAYERS

Extracting deeper or lower features is important for the recognition of handwritten words. CNN has the capability to extract deeper features [32], [33]. So, this study aimed to extract deeper features from handwritten word images using CNN, which is known for its exceptional ability to identify patterns and significant information within images [11]. The scalability of CNN makes it a selective technique for different image processing applications. Previous studies have successfully applied CNN to a variety of tasks [34], and CNN has demonstrated its power, particularly in document analysis [31]. Therefore, we employed these techniques to precisely identify the unique features of handwritten words.

To illustrate the effectiveness of CNN in feature extraction, let's explore the mathematical representation. Consider an input image of a handwritten word with pixel values denoted by $I_{(i,j)}$, where i and j are the spatial coordinates of the image. Let $o_{k,l}$ represent the output feature map at the spatial location (k, l), $W_{(m,n)}$ denote the learnable weights of the kernel at position (m, n). Each convolution applies a set of kernels to the input image, allowing for the computation of the output in "(1)". In equation "(2)", we compute the activation function element-wise after convolution to introduce non-linearity.

$$O_{k,l} = \sigma\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(I_{i+m,j+n}W_{m,n}\right)\right) \tag{1}$$

$$O_{k,l} = \mathrm{ReLU}(O_{k,l}) \tag{2}$$

Pooling, often called max-pooling, reduces the spatial dimensions of the feature maps, aiding in computational efficiency and mitigating overfitting [12]. We compute the pooled output in "(3)", which signifies the maximum value within the designated pooling region.

$$P = \max(I_{i:i+F,\,j:j+G}) \tag{3}$$

where P is the pooled output, and $\max(I_{i:i+F,\,j:j+G})$ represents the maximum value within the specified pooling region. The dimensions of the pooling filter are denoted by symbols F and G.

3) ADDITIVE ATTENTION LAYER

The attention mechanism, renowned in machine translation for its proficiency in unfixed order prediction [35], plays a central role in our recognition model. It is strategically designed to refine the model's focus on pertinent regions within handwritten words. Specifically, we leverage the additive attention mechanism, a subtype of soft attention mechanism, which dynamically assigns weights to different segments of the sequence during recognition.

In additive attention, the computation of attention weights involves a learned function that amalgamates information from both the source (input) and target (output) sequences. The term "additive" is aptly used because it encompasses a summation of weighted values.

This mechanism operates dynamically by adjusting the weights assigned to each element in the input sequence based on its relevance to the output. Notably, salient regions in the static representation can dynamically emerge to the forefront [36]. This dynamic adjustment supplements the model's attention to critical features and characters, proving particularly advantageous when addressing challenges like inter-class similarity and structural complexity [12].

Handwritten Amharic characters often exhibit subtle variations that can lead to ambiguity, especially among visually similar characters. The additive attention mechanism's ability to dynamically adjust attention weights helps mitigate confusion between characters with similar appearances. By emphasizing critical features during recognition, it improves the model's ability to distinguish between these visually similar characters.

The structural complexity of handwritten Amharic, characterized by diverse character forms and ligatures, poses significant challenges to recognition models. Additive attention enhances the model's ability to interpret and adapt to this complexity by allowing it to focus selectively on different parts of the sequence. This selective attention mechanism helps capture intricate details and variations within characters, improving overall recognition performance.

Our model describes the mathematical representation of the attention operation in "(4)".

$$c_i = \sum_{j=1}^{T} \alpha_{ij} h_j \qquad \text{where, } i = 1, 2, \ldots, T \tag{4}$$

Here, h_i is the hidden state at the time step i, T is the length of the sequence, α_{ij} represents the attention weights, and c_i corresponds to the context vectors.

The attention weights α_{ij} are computed using a scoring function that measures the relevance of the input at time step j to the output at time step i. In our approach, we adopt the dot product as the scoring function, as shown in "(5)".

$$\mathrm{score}_{ij} = h_i h_j \tag{5}$$

Subsequently, the scores are normalized using the SoftMax function to obtain attention weights, as outlined in "(6)".

$$\alpha_{ij} = \frac{e^{\mathrm{score}_{ij}}}{\sum_{k=1}^{T} e^{\mathrm{score}_{ik}}} \tag{6}$$

In summary, additive attention, as a soft attention mechanism, significantly enhances the interpretability and adaptability of the attention process in sequence-to-sequence tasks, particularly in the context of OCR.

4) RNN LAYERS

Recurrent neural networks which is segmentation-free networks generate impressive result [37]. In this study, we leverage BGRU networks which is less complex and more efficient computationally [38], to proficiently model intricate sequential dependencies inherent in handwritten words. The incorporation of BGRU layers substantially enhances the network's prowess in capturing contextual information efficiently [34], leading to a notable enhancement

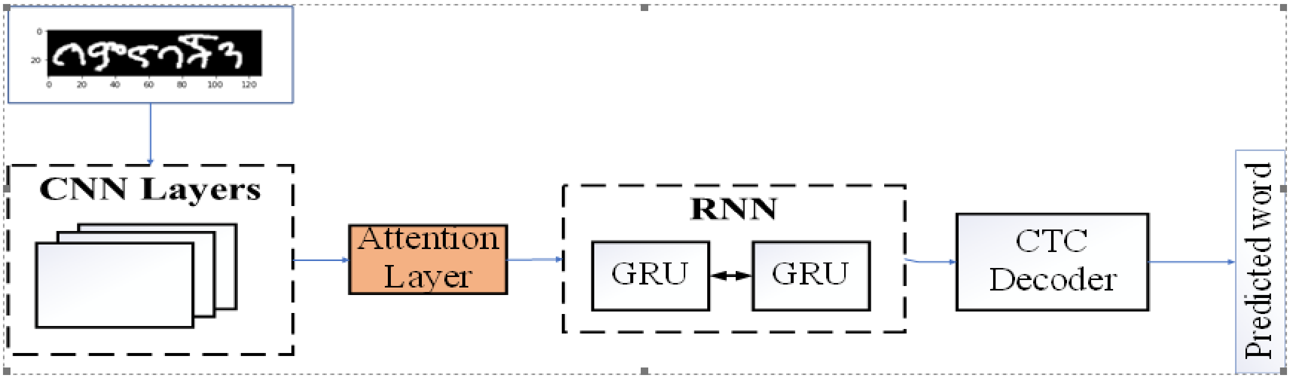

FIGURE 5. The proposed handwritten Amharic word recognition model architecture.

in recognition accuracy [15]. Integrating BGRU layers into the recognition architecture empowers the model with a holistic understanding of the inherent structure and context of handwritten words. This strategic utilization contributes significantly to the overall effectiveness of our recognition model.

5) CTC LAYER

The utilization of the CTC technique proves to be indispensable for addressing sequential recognition tasks, a characteristic inherent in handwritten word recognition. This choice is grounded in the inherent sequential nature of handwritten word recognition tasks, where understanding the temporal order of characters is crucial.

In our specific case, employing CTC for recognizing handwritten Amharic words emerges as a sound technique. CTC provides the model with the ability to learn from sequences of varying lengths, eliminating the need for explicit alignment between input images and their corresponding labels. This inherent flexibility makes CTC a powerful tool for facilitating end-to-end training and decoding processes [12].

Central to the effectiveness of the CTC approach is the CTC loss function, which plays a pivotal role in training the model. This loss function, as represented in "(7)", quantifies the dissimilarity between the predicted sequence and the ground truth labels. Incorporating this loss function into our training regimen facilitates the model's learning of the intricate relationships within the sequential data, contributing to its ability to accurately recognize and decode handwritten Amharic words.

$$L_{CTC} = -\sum\nolimits_{i,j} I(y_{i,j} \neq \text{blank}) \log(\hat{y}_{i,j}) \tag{7}$$

Here, L_{CTC} denotes the CTC loss, $\hat{y}_{i,j}$ represents the predicted probability distribution for each time step, and $y_{i,j}$ corresponds to the ground truth label for the given time step.

The integration of the CTC layer, coupled with the use of its associated loss function, significantly enhances the model's ability to comprehend and recognize handwritten Amharic words in a sequential manner. This strategic combination aligns with the inherent characteristics of the task and contributes to the overall robustness and efficacy of our recognition model.

The probability $P(y \mid x)$ in "(8)" is calculated by summing over all valid alignments between the input sequence and the label sequence, denoted by A. This set of valid alignments considers the presence of a blank symbol as well as allowed repetitions and removals of labels during the alignment process.

$$\mathrm{P}(y \mid x) = \sum\nolimits_{a \in A} \prod\nolimits_{t=1}^{T} \hat{y}_{t,a_t} \tag{8}$$

This consideration ensures that the model accommodates potential variations in the alignment process, providing a robust mechanism for capturing diverse patterns within the handwritten Amharic words.

In essence, the integration of the CTC layer, coupled with a comprehensive understanding of the associated loss function and alignment considerations, contributes significantly to the adaptability and precision of our recognition model. This combination aligns seamlessly with the nuances of handwritten Amharic word recognition, ensuring a thorough and context-aware approach.

The overarching architecture for Amharic word recognition is depicted in **FIGURE 5**. The figure is customized from our previous study [12]. The figure illustrates the steps of the recognition process, commencing with the image data sourced from the HARD-I database. The image, measuring 32 by 128 pixels in our case, undergoes a sequence of transformations through seven convolutional layers, each equipped with a ReLU function. Additionally, two batch normalization layers with a momentum value of 0.8 each and four max-pooling layers contribute to the model's architecture, as outlined in **TABLE 3**. Following the CNN layers, the extracted features are reshaped into a 2D format and then directed to the attention layers. Here, the reshape input is permuted with a permutation size of (2,1). Then we applied a fully connected layer (dense) with the Softmax activation function for each time step. Then the dimension is reduced by selecting the mean value.

TABLE 3. The proposed model CNN layers and hyper-parameter values configuration.

S. No	Layers	Configuration	Number of Parameters
1.	Input	Rows:128, Columns: 32, Channels:1, datatype: float32	0
2.	Conv_1	Feature:64, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	640
3.	MaxPooling_1	Pooling size: (2,2), strides:2	0
4.	Conv_2	Feature:128, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	73856
5.	MaxPooling_2	Pooling size: (2,2), strides:2	0
6.	Conv_3	Feature:256, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	295168
7.	Conv_4	Feature:256, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	590080
8.	MaxPooling_3	Pooling size: (1,2)	0
9.	Conv_5	Feature:512, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	1180160
10.	BN_1	Momentum: 0.8	248
11.	Conv_6	Feature:512, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	2359808
12.	BN_2	Momentum: 0.8	248
13.	MaxPooling_4	Pooling size: (1,2)	0
14.	Conv_7	Feature:512, kernel:(2,2), activation: ReLU	1049088

TABLE 4. The attention layer hyper-parameter values configuration.

S. No	Layers	Configuration	Number of Parameters
1.	Reshape	Rows: 31, Columns: 512	0
2.	permute	Size:(2,1), Rows:512, Columns:31	0
3.	dense	Units: timestep, activation: Softmax	992
4.	dim_reduction	Lambda:31	0
5.	repeat_vector	Rows:512, Columns:31	0
6.	attention_vector	Size:(2,1), Rows:31, Columns:512	0
7.	attention_multiply	Rows:31, Columns:512	0

TABLE 5. The proposed model RNN-CTC layers and hyper-parameter values configuration.

S. No	Layers	Configuration	
1.	Bidirectional_1	BGRU: 128, return sequence: True, droout:0.25	656384
2.	Bidirectional_2	BGRU: 128, return sequence: True, droout:0.25	394240
3.	last_dense	Units: no_class+1, activation: SoftMax	77357
4.	Output + CTC	max_len:11, input length: varies, Label length: varies	0

Then apply the repeat vector to the reduced dimension using the input dimensions. To get the probability of the region permutation, it is applied again with size (2,1). Finally, weighted average multiplication has been applied between the input and probability results. The attention layer hyper-parameter configuration and the number of parameters are depicted in **TABLE 4**. Subsequently, the sequence is processed by two BGRU layers, each comprising 128 hidden units, as specified in **TABLE 5**. A fully connected (dense) layer with a Softmax activation function further refines the features for classification. The number of units in the dense layer corresponds to the classes or categories in the recognition task, with a total of 301 classes, encompassing 300 alphabet symbols and a blank CTC space [12]. In this study, the model is compiled using the Adam optimizer and CTC loss.

C. PERFORMANCE EVALUATION METRICS

In the field of deep learning, it's important to assess the performance of models before launching them in the market. This involves using different metrics to ensure that the model is functioning as expected. When it comes to recognition models, the WER is a commonly used metric for evaluating performance. The WER is calculated mathematically in "(9)", where the variables I_w, D_w, S_w, and GT represent the number of word insertions, deletions, substitutions, and the total number of words in the ground truth text, respectively.

$$\mathrm{WER} = \left(\frac{(I_w + D_w + S_w)}{\mathrm{GT}}\right) * 100 \tag{9}$$

IV. EXPERIMENT SETUP, RESULTS, AND DISCUSSION

In this section, we elucidate the setup and execution of the experiments, present the outcomes, and deliberate on our findings. We did four tests to show how well our Amharic word recognition model works. We also discuss how we picked the best settings for our model and show the results we got with those settings. Finally, we take a close look at the error analysis in the results.

A. EXPERIMENT SETUP

In this study, we harnessed the computational power of a robust GPU to accelerate image processing tasks, leveraging the capabilities of Google Collaboratory—an efficient, cloud-based service enabling remote code execution. Python, a versatile scripting language, served as our programming foundation, while the Keras framework, backed by TensorFlow, facilitated the seamless development and training of our recognition model.

Our approach involved the creation of four distinct recognition models, each addressing specific considerations. In Case I, we examined the original image without incorporating attention mechanisms, aiming to establish a baseline performance. In Case II, we introduced attention mechanisms to the original images, enhancing the model's ability to focus on crucial features. Case III involved the exploration of augmented images without attention mechanisms, diversifying the training dataset for improved generalization. Finally, in Case IV, we combined the benefits of augmented images with attention mechanisms, seeking to achieve a comprehensive and high-performing recognition model. For training consistency across all cases, we opted for 25 epochs and a batch size of 128. This deliberate choice ensures a balance between model convergence and computational efficiency during the training process. The numerical selection of these parameters was driven by a careful consideration of the trade-off between training time and model accuracy.

B. RESULTS

In this section, we present the accomplishments of the four models developed for recognizing handwritten Amharic words. Each model underwent rigorous training, with both training and validation loss metrics being reported throughout the training phase. Additionally, the models were evaluated using the same testing dataset to ensure a fair and consistent comparison of their effectiveness. **FIGURE 6** shows the training and validation loss graphs for Case I, while **FIGURE 7** displays the same for Case II. **FIGURE 8** and **FIGURE 9** illustrate the training and validation losses for Cases III and IV, respectively. These figures provide valuable insights into each model's learning progress and performance throughout the training process.

By examining these graphs, we can assess how well each model configuration converges and generalizes to unseen data. The comparison of training and validation losses across different cases helps identify the most effective model for recognizing handwritten Amharic words.

We have also provided the percentage of correctly predicted characters and words. **FIGURE 10** displays a comparison of the correct character prediction results for all four cases, while **FIGURE 11** shows the comparison of correct word prediction results for all four cases.

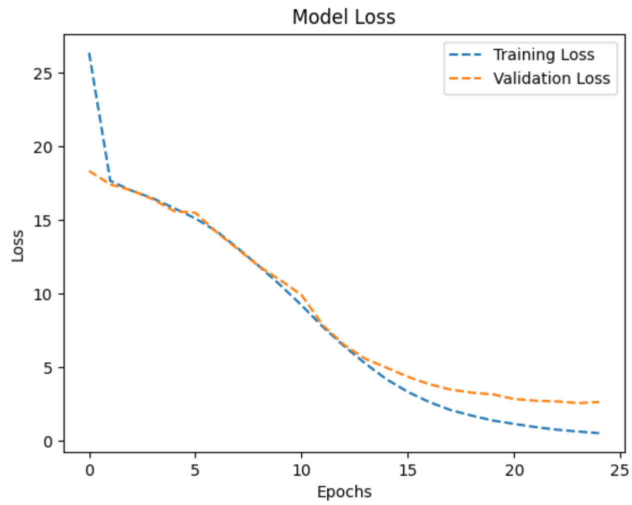

FIGURE 6. Loss of training and validation graph for case I.

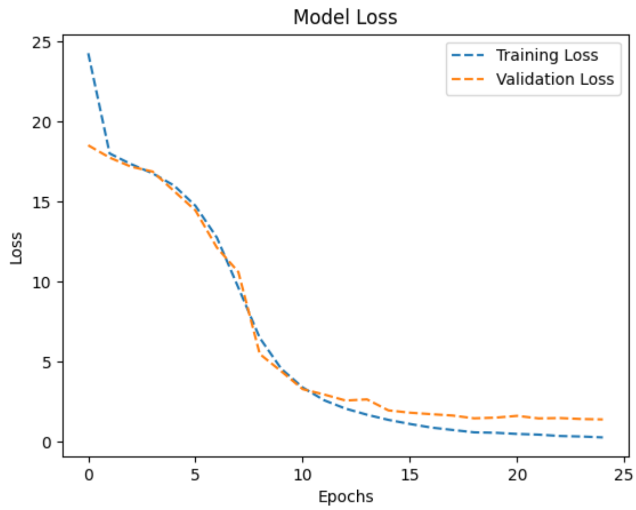

FIGURE 7. Loss of training and validation graph for case II.

C. DISCUSSION

Since our experimental dataset is small, we applied the augmented technique to increase the number of images from 12,047 to 34,047. The paper conducts a comparative analysis of four recognition models, taking into account the attention mechanism used for handwritten Amharic word recognition, through the implementation of four scenarios outlined in the experiment setup subsection.

As shown in **TABLE 6**, the results indicate that the models trained with augmented images perform better in terms of average CER and average WER on the testing dataset of 1200 images. The first model, Case I, achieved an average CER of 11.44% and an average WER of 33.75% for the

TABLE 6. Testing results of the developed four models.

Cases	Augmentation	Attention	Average CER	Average WER	Correct characters predicted	Correct words predicted
Case I	no	no	11.44%	33.75%	80.51%	36.72%
Case II	no	yes	7.00%	21.42%	90.31%	43.56%
Case III	yes	no	4.75%	15.33%	94.29%	46.93%
Case IV	yes	yes	2.84%	9.75%	96.09%	50.02%

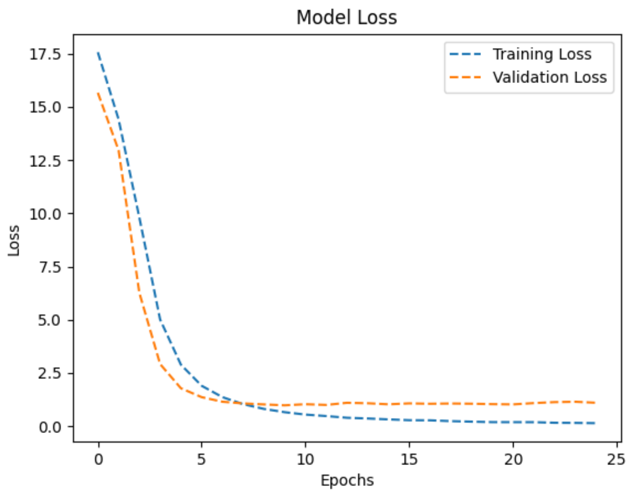

FIGURE 8. Loss of training and validation graph for case III.

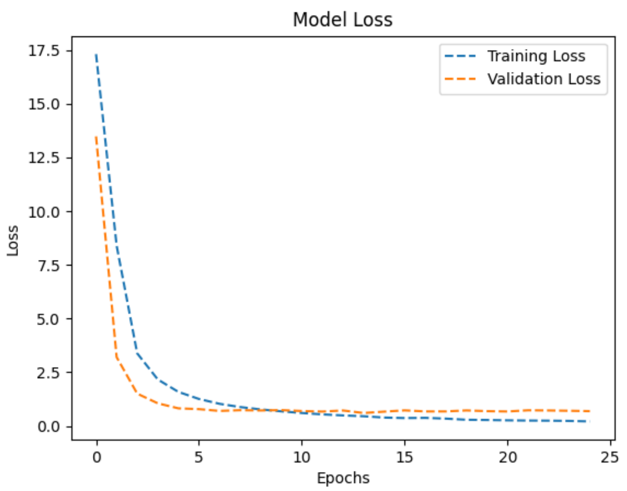

FIGURE 9. Loss of training and validation graph for case IV.

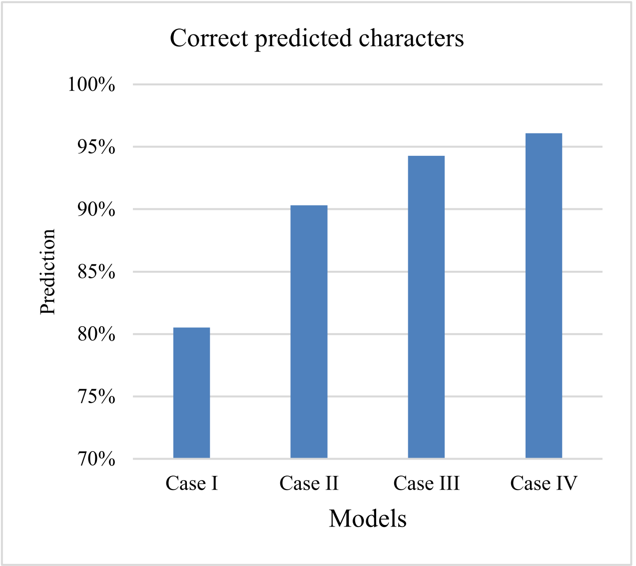

FIGURE 10. A graph of the correct characters' predictions for all cases.

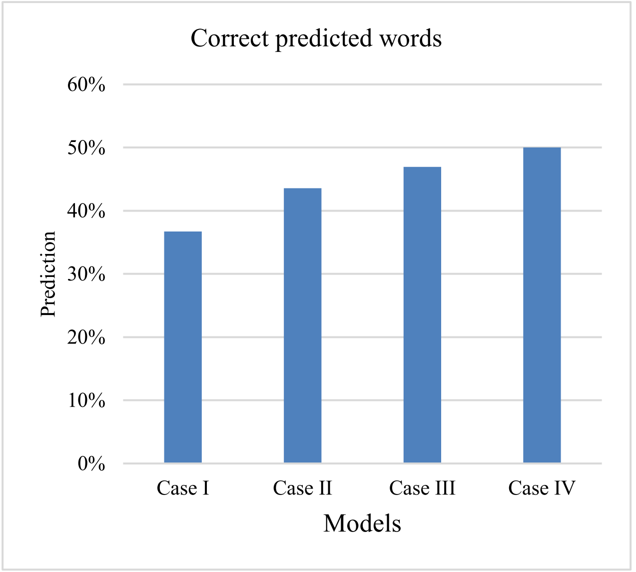

FIGURE 11. A graph of the correct word predictions for all cases.

testing dataset. The second model, Case II, achieved an average CER of 7.00% and an average WER of 21.42% for the testing dataset. The third model, Case III, achieved an average CER of 4.75% and an average WER of 15.33% for the testing dataset. The fourth model, Case IV, achieved an average CER of 2.84% and an average WER of 9.75% for the testing dataset.

The inclusion of an attention mechanism in the image training enabled the model to perform better. For further reference, it is possible to cross-check prediction results from FIGURE 10 and FIGURE 11.

1) ERROR ANALYSIS

In this study, we tried to analyze errors that exist in the Amharic word recognition model. Two sample errors are

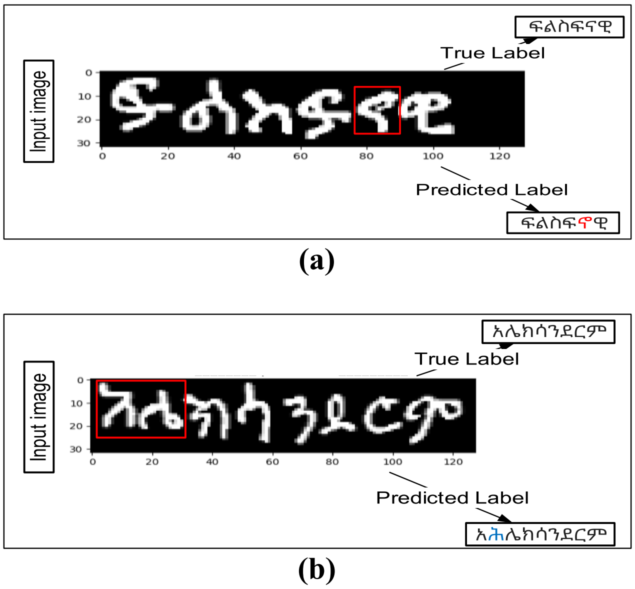

FIGURE 12. Sample error results generated for case IV.

TABLE 7. Effectiveness of the proposed model.

S. No	Study	Technique	Test Data Samples	Results (CER) %
1.	[12]	CNN, RNN, CTC	6,150	17.95
2.	[39]	Gated CNN, Self-Attention encoder	58,000	8.22
3.	Proposed	CNN, Attention, RNN, CTC	1,200	2.84

depicted in **FIGURE 12**, considering case IV. The first sample handwritten image, which is labeled as **(a)**, has the word "ፍልስፍናዊ" and the second sample handwritten image, which is labeled as **(b)**, has the word "አሌክስአንደርንም". In this figure, characters incorrectly recognized are marked in a red rectangle.

From the first handwritten word image, the character "ና" is predicted as ኖ. In this case, the confusion happened due to the shape similarity of the characters when written by hand.

In the second handwritten word image, ሕ is predicted due to two አ and ሌ letters being connected to each other during handwriting. Each character is predicted correctly, but the character ሕ inserted a character during the prediction. The model is confused about character connectivity. When the space between characters is not sufficient, then the possibility of an error happening is high.

2) METHOD COMPARISON

While comparing handwritten recognition systems across various scripts and experimental conditions can be challenging, we highlight two studies related to this paper to indicate progress in Amharic handwriting recognition.

Malhotra and Addis [12] employed CNN, RNN, and CTC techniques, achieving a Character Error Rate (CER) of 17.95% on 6,150 testing samples. Tadesse et al. [39] proposed a "Gated Convolution and Stacked Self-Attention Encoder–Decoder-Based Model for Offline Handwritten Ethiopic Text Recognition," resulting in a CER of 8.22%.

Our proposed method, utilizing CNN, Attention, RNN, and CTC, achieved a CER of 2.84%, outperforming the previous two studies. As shown in **TABLE 7**, our model demonstrates effectiveness in recognizing Amharic characters.

In our study, we acknowledge the significant work of Abdurahman et al. [7] in the area of offline handwritten Amharic word recognition. While their approach using a CRNN demonstrates strong performance, our research introduces an alternative method that integrates an additive attention mechanism. Although our results, with a CER of 2.84%, do not surpass those reported by Abdurahman et al., our approach offers distinct advantages, including greater adaptability and improved interpretability across different datasets. Also, the differences in how our work and that of Abdurahman et al. was done show how important it is to look into different ways to make Amharic text recognition systems more reliable and useful in a wider range of situations.

3) FUTURE WORKS

One direction for future research involves incorporating a multilingual word recognition model into our study. Expanding the model to support multiple languages will make it more versatile and applicable in diverse linguistic contexts. This enhancement would require the collection and processing of multilingual handwriting datasets and the development of advanced techniques to handle the intricacies of different scripts and writing styles.

Currently, our study relies on synthetic datasets for training and evaluation. Future work will focus on utilizing real-life data sources to better capture the variability and complexity of handwritten documents encountered in everyday situations. This transition to authentic datasets will improve the model's generalizability and reliability when applied to practical applications.

Handwritten documents in real-world scenarios often contain various forms of noise, degradation, and variability, such as smudges, blurs, ink bleed, and inconsistent writing styles. Enhancing the model's robustness to these factors is crucial for reliable performance. Future research will aim to develop and integrate techniques that make the model more resilient to these common issues. Potential approaches include advanced preprocessing algorithms, noise reduction techniques, and augmentation strategies to simulate realistic conditions during training.

By pursuing these future directions, we aim to enhance the capabilities and applicability of our handwriting recognition model, making it more robust, versatile, and user-friendly in real-world scenarios.

V. CONCLUSION

In conclusion, this research on handwritten Amharic word recognition with an additive attention mechanism has successfully addressed the challenges associated with recognizing handwritten Amharic words. Leveraging a deep learning approach, specifically, the integration of additive attention with RNN, our proposed recognition model, consisting of seven CNNs and two RNNs using the CTC strategy, has demonstrated promising results.

The study not only identified and navigated the complexities of variations in individual handwriting, connectivity of words without spaces, similarities in alphabet characters, and noise from scanned images, but also tackled the issue of data insufficiency through effective augmentation techniques. The utilization of an original dataset of 10,784 handwritten Amharic words, augmented with 24,000 images, contributed to a more robust and diverse training set.

The positive outcomes achieved in the recognition of handwritten Amharic words demonstrate the additive attention mechanism's effectiveness in improving precision. This research represents a significant stride in bridging the existing gap in OCR technology for the Amharic script, emphasizing the transformative capabilities of deep learning in pattern recognition.

Moving forward, the insights gained from this study can pave the way for further advancements in the field of handwritten language recognition, not only for Amharic but also for other languages facing similar challenges. The success of our approach highlights the potential for broader applications of additive attention mechanisms to improve the accuracy and efficiency of handwritten word recognition systems.

ACKNOWLEDGMENT

The authors extend their heartfelt gratitude to the esteemed faculty members, dedicated staff, and fellow students at Delhi Technological University. Their unwavering guidance and support have been invaluable throughout the completion of this manuscript, playing a pivotal role in their research project's success. Their continuous assistance and expertise have not only enriched their academic journey, but they have also significantly contributed to their growth as researchers.

They also acknowledge the indispensable role that "ChatGPT (GPT-3.5)" and "Wordtune" played in the development of this study. Their assistance has proven to be extremely beneficial, enhancing the overall quality and depth of their research. They used these tools for language refinement, grammar checking, and enhancing the clarity of their writing.

REFERENCES

[1] M. Meshesha and C. V. Jawahar, "Recognition of printed Amharic documents," in *Proc. 8th Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 2, Seoul, Korea (South), 2005, pp. 784–788.

[2] B. H. Belay, T. A. Habtegebrial, and D. Stricker, "Amharic character image recognition," in *Proc. IEEE 18th Int. Conf. Commun. Technol. (ICCT)*, Chongqing, China, Oct. 2018, pp. 1179–1182.

[3] M. Meshesha, "Recognition and retrieval from document image collections," Ph.D. dissertation, Comput. Sci. Eng., Int. Inst. Inf. Technol. Hyderabad, Hyderabad, India, 2008.

[4] Y. Assabie and J. Bigun, "Offline handwritten Amharic word recognition," *Pattern Recognit. Lett.*, vol. 32, no. 8, pp. 1089–1099, Jun. 2011.

[5] E. Yohannes Obsie, H. Qu, and Q. Huang, "Amharic character recognition based on features extracted by CNN and auto-encoder models," in *Proc. 13th Int. Conf. Comput. Model. Simul.*, Melbourne, VIC, Australia, Jun. 2021, pp. 58–66.

[6] K. Tamir, "Handwritten Amharic characters recognition using CNN," in *Proc. IEEE AFRICON*, Accra, Ghana, Sep. 2019, pp. 1–4.

[7] F. Abdurahman, E. Sisay, and K. A. Fante, "AHWR-net: Offline handwritten Amharic word recognition using convolutional recurrent neural network," *Social Netw. Appl. Sci.*, vol. 3, no. 8, pp. 1–11, Aug. 2021.

[8] Y. Assabie and J. Bigun, "HMM-based handwritten Amharic word recognition with feature concatenation," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, 2009, pp. 961–965.

[9] W. Alemu and S. Fuchs, "Handwritten Amharic bank check recognition using hidden Markov random field," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun. 2003, p. 28.

[10] D. M. Kassa and H. Hagras, "An adaptive segmentation technique for the ancient Ethiopian Ge'ez language digital manuscripts," in *Proc. 10th Comput. Sci. Electron. Eng. (CEEC)*, Sep. 2018, pp. 83–88.

[11] H. T. Weldegebriel, H. Liu, A. U. Haq, E. Bugingo, and D. Zhang, "A new hybrid convolutional neural network and eXtreme gradient boosting classifier for recognizing handwritten Ethiopian characters," *IEEE Access*, vol. 8, pp. 17804–17818, 2020.

[12] R. Malhotra and M. T. Addis, "End-to-end historical handwritten Ethiopic text recognition using deep learning," *IEEE Access*, vol. 11, pp. 99535–99545, 2023.

[13] M. Meshesha and C. V. Jawahar, "Optical character recognition of Amharic documents," *Afr. J. Inf. Commun. Technol.*, vol. 3, no. 2, pp. 53–66, Aug. 2007.

[14] E. G. Beyene, "Handwritten and machine printed OCR for geez numbers using artificial neural network," 2019, *arXiv:1911.06845*.

[15] Y. S. Chernyshova, A. V. Sheshkus, and V. V. Arlazarov, "Two-step CNN framework for text line recognition in camera-captured images," *IEEE Access*, vol. 8, pp. 32587–32600, 2020.

[16] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR)," *IEEE Access*, vol. 8, pp. 142642–142668, 2020.

[17] L. Xu, Y. Wang, X. Li, and M. Pan, "Recognition of handwritten Chinese characters based on concept learning," *IEEE Access*, vol. 7, pp. 102039–102053, 2019.

[18] B. H. Belay, "Deep learning for Amharic text-image recognition: Algorithm, dataset and application," Ph.D. dissertation, Dept. Comput. Sci., Technische Univ. Kaiserslautern, Germany, 2021, p. 386.

[19] B. Belay, T. Habtegebrial, M. Meshesha, M. Liwicki, G. Belay, and D. Stricker, "Amharic OCR: An end-to-end learning," *Appl. Sci.*, vol. 10, no. 3, p. 1117, Feb. 2020.

[20] Y. Assabie and J. Bigun, "A hybrid system for robust recognition of Ethiopic script," in *Proc. 9th Int. Conf. Document Anal. Recognit.*, Curitiba, Brazil, 2007, pp. 1–5.

[21] B. Belay, T. Habtegebrial, M. Liwicki, G. Belay, and D. Stricker, "Factored convolutional neural network for Amharic character image recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2906–2910.

[22] Y. Assabie and J. Bigun, "Lexicon-based offline recognition of Amharic words in unconstrained handwritten text," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[23] H. T. Weldegebriel, J. Chen, and D. Zhang, "Deep learning for Ethiopian Ge'ez script optical character recognision," in *Proc. 10th Int. Conf. Adv. Comput. Intell. (ICACI)*, Xiamen, China, Mar. 2018, pp. 540–545.

[24] R. Malhotra and M. T. Addis, "Ethiopic base characters image recognition using LSTM," in *Proc. 2nd Int. Conf. Comput. Methods Sci. Technol. (ICCMST)*, Mohali, India, Dec. 2021, pp. 94–98.

[25] D. Addis, C.-M. Liu, and V.-D. Ta, "Printed Ethiopic script recognition by using LSTM networks," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, New Taipei, Taiwan, Jun. 2018, pp. 1–6.

[26] M. Ali Nur, M. Abebe, and R. S. Rajendran, "Handwritten geez digit recognition using deep learning," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, pp. 1–12, Nov. 2022.

[27] S. Aly and A. Mohamed, "Unknown-length handwritten numeral string recognition using cascade of PCA-SVMNet classifiers," *IEEE Access*, vol. 7, pp. 52024–52034, 2019.

[28] B. Gatos, K. Ntzios, I. Pratikakis, S. Petridis, T. Konidaris, and S. J. Perantonis, "An efficient segmentation-free approach to assist old Greek handwritten manuscript OCR," *Pattern Anal. Appl.*, vol. 8, no. 4, pp. 305–320, Feb. 2006.

[29] Y. Assabie and J. Bigun, "Ethiopic character recognition using direction field tensor," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, 2006, pp. 284–287.
[30] B. Y. Reta, D. Rana, and G. V. Bhalerao, "Amharic handwritten character recognition using combined features and support vector machine," in *Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI)*, May 2018, pp. 265–270.
[31] M. Agegnehu, G. Tigistu, and M. Samuel, "Offline handwritten Amharic digit and punctuation mark script recognition using deep learning," in *Proc. 2nd Deep Learn. Indaba-X Ethiopia Conf.*, Adama, Ethiopia, 2021, pp. 38–47.
[32] S. B. Ahmed, I. A. Hameed, S. Naz, M. I. Razzak, and R. Yusof, "Evaluation of handwritten Urdu text by integration of MNIST dataset learning experience," *IEEE Access*, vol. 7, pp. 153566–153578, 2019.
[33] M. I. Sharif, M. A. Khan, A. Alqahtani, M. Nazir, S. Alsubai, A. Binbusayyis, and R. Damaševičius, "Deep learning and kurtosis-controlled, entropy-based framework for human gait recognition using video sequences," *Electronics*, vol. 11, no. 3, p. 334, Jan. 2022.
[34] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.
[35] Y. Zhu, Z. Xie, L. Jin, X. Chen, Y. Huang, and M. Zhang, "SCUT-EPT: New dataset and benchmark for offline Chinese text recognition in examination paper," *IEEE Access*, vol. 7, pp. 370–382, 2019.
[36] J. Zhang, J. Du, and L. Dai, "Track, attend, and parse (TAP): An end-to-end framework for online handwritten mathematical expression recognition," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 221–233, Jan. 2019.
[37] B. H. Belay, T. Habtegebirial, M. Liwicki, G. Belay, and D. Stricker, "Amharic text image recognition: Database, algorithm, and analysis," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 1268–1273.
[38] Y. Jiang, H. Dong, and A. El Saddik, "Baidu meizu deep learning competition: Arithmetic operation recognition using end-to-end learning OCR technologies," *IEEE Access*, vol. 6, pp. 60128–60136, 2018.
[39] D. A. Tadesse, C.-M. Liu, and V.-D. Ta, "Gated convolution and stacked self-attention encoder–decoder-based model for offline handwritten Ethiopic text recognition," *Information*, vol. 14, p. 654, Dec. 2023.

RUCHIKA MALHOTRA received the master's and Ph.D. degrees in software engineering from the University School of Information, Communication and Technology, Guru Gobind Singh Indraprastha University, Delhi, India. She is currently a Professor with the Department of Software Engineering, Delhi Technological University, Delhi. She has published more than 200 research papers in international journals and conferences. Her research interests include software testing, improving software quality, statistical and adaptive prediction models, software metrics, neural nets modeling, and the definition and validation of software metrics.

MARU TESFAYE ADDIS received the bachelor's degree in computer science and information technology from Arba Minch University, Ethiopia, in 2010, and the master's degree in computer science from Bahir Dar University, Ethiopia, in 2017. He is currently pursuing the Ph.D. degree in computer science and engineering with Delhi Technological University, India. He is also a Lecturer with the Department of Computer Science, Debre Tabor University. His research interests include pattern recognition, deep learning, artificial intelligence, and image processing. He is also dedicated to advancing knowledge and contributing to the development of innovative solutions in these areas.