International Journal of Scientific Research in
Computer Science and Engineering
Vol.12, Issue.6, pp.41-51, December 2024
E-ISSN: 2320-7639
Available online at: www.isroset.org

Research Article

A CNN-Based Digraph Extraction Model for Enhanced Swahili Natural Language Processing

Tirus Muya Maina[1*], Aaron Mogeni Oirere[2], Stephen Kahara[3]

[1,2,3]Computer Science Department, Murang'a University of Technology, Murang'a, Kenya

Corresponding Author: tmuya@mut.ac.ke

Received: 20/Oct/2024; **Accepted**: 21/Nov/2024; **Published**: 31/Dec/2024

Abstract— Swahili, a prominent language in East Africa, is integral to the region's communication, commerce, and cultural exchange. Enhancing the accuracy of Swahili speech recognition systems is critical for improving accessibility, Transcription, linguistic translation, text analysis, speech recognition, sentiment analysis and aiding individuals with disabilities. However, the unique linguistic features of Swahili, particularly its digraphs, pose significant challenges to existing speech recognition technologies. This study addressed these challenges by introducing a novel approach for the extraction of Swahili digraphs from speech data. The research involved the extraction of a specialized Swahili digraph dataset and the design of an advanced Digraph Extraction Model. This model leverages Dense layer and Convolutional Neural Networks architecture to improve the precision and efficiency of Natural Language Processing tasks related to Swahili. Employing Design Science Research Methodology, the study systematically designs, implements, and evaluates the digraph extraction model. Results from the study demonstrate the model's robust performance across several metrics. The low Mean Absolute Error and Root Mean Squared Error values indicate that the model is highly accurate, with predictions closely aligning with the actual values. Furthermore, the R-squared value of 0.89 demonstrates that the model effectively captures and this accounts for a substantial part of the variation in the dataset. The low-test loss suggests effective generalization to new, unseen data, affirming the model's reliability for practical applications. This research significantly advances the field of Swahili speech recognition by enhancing accessibility and usability for Swahili speakers while supporting the preservation of the language's oral traditions. The innovations introduced, including the annotated Swahili digraph corpus and the advanced Swahili Digraph Extraction Model, provide a substantial foundation for future research and development in Swahili digraph recognition technology. The model's potential for successful deployment in real-world scenarios offers promising implications for improving Swahili language processing across various applications.

Keywords— Swahili Digraph Recognition, Digraph Extraction Model, Dense Neural Networks, Convolutional Neural Networks (CNN), Natural Language Processing (NLP), Swahili Corpus

1 Introduction

1.1 Background and Motivation

Swahili is a widely spoken language in East Africa, serving as a lingua franca in countries such as Kenya, Tanzania, Uganda, and the Democratic Republic of Congo. It plays a critical role in regional communication, trade, and cultural exchange. According to [1] technology is crucial in enabling and improving interactions among humans and between humans and computers. This entails grouping spoken words into distinct digraph, consonant, and vowel sound categories [2]. Speech-enabled applications that depend on correctly identifying phonetic units of speech need ASR [3].

The development of accurate Swahili speech recognition systems is essential for several reasons. First, in the realm of communication technologies, the increasing prevalence of voice-activated assistants and smart devices necessitates support for Swahili. This inclusion is vital for ensuring that Swahili speakers can interact with technology in their native language, thereby enhancing user experience and accessibility. Second, speech recognition technology can significantly impact language education [4]. Interactive tools that facilitate learning pronunciation and conversational skills are particularly important in regions where Swahili is taught as a second language or in multilingual societies where language proficiency is crucial for social and economic mobility.

Moreover, for people with disabilities, including those with visual impairments or restricted mobility, speech recognition technology offers a different method for engaging with digital devices. Incorporating support for Swahili in this technology can significantly boost these users' independence and improve their overall quality of life [5]. Swahili's being rich oral tradition underscores the importance of developing speech recognition systems to aid in the documentation and preservation of cultural narratives and folklore, ensuring their accessibility to future generations.

Swahili presents unique challenges for speech recognition technology due to its distinct linguistic characteristics. One major challenge is the inadequate of digital resources and state of art recognition and detection of digraphs, which are combinations of two letters representing a single sound, such as "ch," "dh," "gh," "kh," "ng'," "ny," "sh," "th," and "ng" [6]. These digraphs do not always have direct equivalents in other languages, and their accurate recognition is crucial for correct transcription. Additionally, Swahili's phonetic complexity poses significant difficulties. The language includes a mix of consonants, vowels, and digraphs, creating unique sound patterns. ASR systems, especially those trained on high-resource languages with different phonetic rules, often struggle to handle this complexity accurately [7], [8], [9].

While there are existing corpora for Swahili speech, they often lack detailed annotations or focus solely on phonemes without considering digraphs. This scarcity of annotated data makes it challenging to train accurate and robust speech recognition models for Swahili digraphs. Furthermore, existing corpora often focus solely on phonemes, neglecting the specific challenges posed by digraphs [10] [11], Digraphs, which are combinations of two letters representing a single sound, are integral components of Swahili speech and play a significant role in its phonetic structure. However, due to the lack of emphasis on digraphs in existing corpora, speech recognition models trained on these corpora may struggle to accurately recognize and transcribe digraphs in real-world scenarios [12], [13].

Moreover, there is a notable absence of pre-existing models or algorithms designed specifically for extracting Swahili digraphs from audio corpora. Existing speech recognition models may struggle to accurately identify and differentiate Swahili digraphs due to their complex phonetic structure and the lack of specific training data [14]. Without a dedicated model for Swahili digraph extraction, speech recognition systems may produce inaccurate transcriptions, leading to reduced performance and usability in real-world applications [15], [16], [17].

The lack of available resources and tools for Swahili digraph recognition hampers research efforts in this area and limits the development of effective speech recognition solutions for Swahili. To address these challenges, comprehensive datasets with annotated Swahili digraphs must be created, along with specialized models and algorithms that can accurately extract and recognize these digraphs from audio data [18]. By successfully navigating these challenges, researchers have the potential to advance the domain of Swahili speech recognition, thereby fostering greater accessibility and usability of speech technology for individuals who speak Swahili.

Section 1 of the study introduces the significance of processing the Swahili language and outlines the challenges posed by its distinct linguistic characteristics. Section 2 reviews existing literature on digraph recognition methods and discusses their shortcomings when applied to Swahili. Section 3 explains the methodology utilized, encompassing data gathering and the development of an annotated digraph corpus for Swahili. Section 4 elaborates on the model architecture, specifically the CNN-based Digraph Extraction Model, and outlines the key steps in its design. Section 5 presents the findings and a discussion, evaluating the model's effectiveness through various performance metrics. Finally, Section 6 summarizes the study's contributions and recommends potential avenues for future research to enhance Swahili language processing.

2 Related work

There is no widely established, specialized model exclusively designed for digraph recognition analogous to those developed for speech recognition. Instead, digraph recognition is generally integrated within broader frameworks and methodologies in natural language processing and pattern recognition.

The concept of digraph recognition can be integrated within broader frameworks and models in natural language processing (NLP) and pattern recognition, rather than being the focus of standalone, widely recognized models. It can encompass various methodologies primarily developed within the realms of speech recognition and NLP in general. These models employ diverse techniques that include rule-based systems, machine learning algorithms, and statistical methods, and, to identify and process digraphs within textual or speech data [19].

Rule-based systems rely on established linguistic rules and patterns, which have played a crucial role in the early development of NLP and speech recognition technologies. These systems established a foundational framework for subsequent advancements in the field. They demonstrate particular efficacy in languages characterized by well-defined and consistent digraph structures. For example, they excel in languages like English, where digraph patterns like 'sh,' 'ch,' and 'th' are predictable and consistent. However, the milestone of applying rule-based systems to languages like Swahili reveals their limitations. Swahili's digraphs are highly context-dependent and diverse, making it challenging for rule-based approaches to maintain accuracy and effectiveness. This context dependency means that digraphs in Swahili may change meaning based on the surrounding text or speech, something rule-based systems struggle to handle without extensive and often complex rule sets [20].

Statistical methods, including N-gram models and Hidden Markov Models (HMMs), have made substantial contributions to digraph recognition by capturing statistical dependencies between adjacent characters or phonemes. N-gram models serve to estimate the probabilities of various sequences occurring, while HMMs present a probabilistic framework for representing sequential data., enabling effective pattern recognition in both text and speech. These methods represent a significant improvement over rule-based systems, providing contextually relevant predictions, particularly for languages with complex linguistic structures. [21].

However, these models exhibit notable limitations, particularly in handling intricate linguistic phenomena and requiring large datasets for accurate performance. For example, they struggle with context-dependent digraphs in languages like Swahili, where the same sequence may have multiple meanings. The demand for extensive training data further restricts their applicability. These challenges have led to advanced methods like neural networks and deep learning, which provide greater flexibility and efficiency in handling the complexities of natural language. Thus, while statistical models have advanced natural language processing, their limitations highlight the need for more adaptive, data-efficient approaches to achieve better performance. [21].

Machine learning-based approaches, particularly deep learning architectures such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have demonstrated significant potential in digraph recognition. These models excel at automatically learning hierarchical representations from data, capturing intricate patterns and dependencies. The emergence of end-to-end models that directly map input speech signals to digraph representations has further streamlined the recognition pipeline, resulting in competitive performance [22] [23] [24]

Despite these advancements, existing models face notable limitations. They often struggle with morphologically complex languages like Swahili, where digraph patterns are highly context-dependent and influenced by word structure. Additionally, the scarcity of annotated data and linguistic resources for under-resourced languages poses further challenges to effective model development [18].Addressing these limitations requires innovative approaches and increased investment in linguistic resources to improve the adaptability and accuracy of machine learning-based digraph recognition models.

Research on digraph recognition in other languages, such as English, Mandarin, and Arabic, provides valuable insights. Techniques explored in these languages include machine learning, algorithms, statistical models and rule-based systems [14]. However, the applicability of these findings to Swahili may be limited due to its unique phonetic characteristics and digraph structures.

Adapting and extending existing techniques to the Swahili context requires further research, considering its specific linguistic features and digraph patterns. Developing annotated datasets targeting Swahili digraphs is essential for advancing research and addressing the challenges effectively.

2.1 Problem Statement

Despite Swahili's prominence as a key language in East Africa, facilitating regional communication, commerce, and cultural exchange, existing Natural Language Processing (NLP) technologies face significant challenges in accurately processing the language’s unique linguistic characteristics, particularly its digraphs. These limitations undermine the effectiveness of Swahili speech recognition systems, resulting in reduced accessibility, inaccurate transcription, and inadequate support for language education and individuals with disabilities. Therefore, enhancing the accuracy of Swahili speech recognition is essential.

A major obstacle to achieving this enhancement lies in the absence of a specialized corpus for Swahili digraphs and the lack of an advanced extraction model designed to accommodate the language’s distinct acoustic and phonetic features. This gap presents a critical need for an innovative approach to improve the precision and efficiency of Swahili language processing. In response, this research proposes the development of a Convolutional Neural Network (CNN)-based digraph extraction model, incorporating Dense layers, to enhance the accuracy of Swahili NLP tasks and to advance the field of Swahili digraph recognition technology.

2.2 Objectives of the Research

The following objectives guided the research study

1. To identify and categorize the unique acoustic and phonetic features of Swahili digraphs that significantly impact their recognition in speech data.
2. To develop a CNN-based model for digraph extraction that includes the identified features specific to Swahili.
3. To analyse the effectiveness of the CNN-based digraph extraction model with the annotated corpus, focusing on its accuracy and efficiency.
4. To develop a Swahili digraph corpus that aids in training and evaluating the CNN-based extraction model.

3 Methodology

The Design Science Research Methodology (DSRM) was used as the guiding framework for the development of the Swahili Digraphs Model. DSRM provides a structured approach that facilitates the systematic design, implementation, and evaluation of innovative solutions to complex problems [25]. In this study, DSRM provides a strong methodology for tackling the challenges of Swahili speech recognition, particularly emphasizing the extraction of digraphs.

The principles of DSRM were applied in developing the Swahili Digraphs Model in each stage of the research process, from problem identification to model refinement [26]. Initially, the researchers utilized DSRM to identify the specific challenges posed by Swahili digraphs, considering factors such as phonetic complexity and corpus limitations. Subsequently, the design phase involved conceptualizing and developing algorithms for digraph extraction, drawing upon insights from linguistic principles and computational techniques. The implementation phase entailed the actual coding and implementation of the digraph extraction model, ensuring that it aligns with the identified requirements and objectives [26]. Finally, the evaluation phase employed DSRM principles to rigorously assess the performance of the model through empirical validation, iteratively refining its design based on feedback and insights gained from testing.

The adoption of DSRM is justified by its suitability for addressing the complex challenges of Swahili speech recognition, particularly in relation to digraph extraction [26]. DSRM emphasizes a systematic and problem-solving-oriented approach, which is essential for navigating the complexities of developing an effective digraph extraction model [27]. By following the principles of DSRM, researchers ensured that the Swahili Digraphs Model is not only technically sound but also practically relevant and effective in accurately transcribing Swahili speech. This approach not only contributes to advancements in Swahili speech recognition technology but also enhances the accessibility and usability of speech recognition systems for Swahili speakers.

Table 1: DSRM Principle and Stages in Developing Swahili Digraph Model

DSRM Stage	DSRM Principle	Application in Developing the Swahili Digraphs Model
1.Problem Identification	Identify and justify the problem.	Identify the challenges in recognizing and processing Swahili digraphs, and justify the need for a specialized model to address these challenges.
2.Objectives Definition	Define the objectives and requirements of the solution.	Set objectives for the model, such as improved accuracy in digraph recognition, computational efficiency, and scalability for various NLP tasks.
3. Design and Development	Develop the artifact (model, system, method) that addresses the problem.	Design and implement the Swahili Digraphs Model using Dense Layers and CNN architectures for effective feature extraction and digraph recognition.
4. Demonstration	Demonstrate how the artifact solves the identified problem.	Test the model on Swahili text datasets to demonstrate its effectiveness in recognizing digraphs, and compare its performance with baseline models.
5. Evaluation	Measure and observe the artifact's effectiveness in solving the problem.	Evaluate model performance using metrics like accuracy, precision, recall, and F1 score; conduct error analysis and gather user feedback.
6.Model Refinement	Refine the artifact iteratively based on evaluation and feedback.	Refine the model by adjusting architectures, retraining with diverse datasets, and reassessing performance to ensure continuous improvement.

3.1 Feature Extraction for Swahili Digraph Corpus

Feature extraction aimed to identify and isolating meaningful attributes from the dataset for further analysis or modelling. The primary focus for the Swahili digraph corpus is on the digraphs and their corresponding vowels. The process involves several key steps:

3.1.1 Steps in Feature Extraction

Step 1: Data Understanding and Preparation

The preliminary phase involved a comprehensive understanding and cleansing of the dataset to ensure its accuracy and usability. The dataset encompasses:

Total Number of Words per Digraph: The aggregate count of words in which each digraph appears.

Vowel Classification: The number of words containing each of the vowels (a, e, i, o, u) within each digraph.

Step 2: Identification of Key Features

The extraction process focused on several crucial features

Digraph: The specific two-letter combination analysed.

Total Number of Words: The frequency with which each digraph appears across the dataset.

Vowel Distribution: The count of words containing each vowel within the scope of each digraph.

Step 3: Data Structuring

The data was systematically organized into a tabular format to facilitate analysis

Step 4: Application of Feature Extraction Techniques

Various techniques are used to extract meaningful features from the data.

One-Hot Encoding: Each digraph is encoded as a binary vector to indicate its presence in the corpus. This approach allows for the representation of categorical data in a format that is compatible with machine learning algorithms.

Frequency Distribution: This technique normalizes the frequency of words containing each vowel within each digraph, enabling comparison across digraphs and vowels.

Proportion of Vowels: The proportion of words containing each vowel is calculated relative to the total number of words for each digraph. This provides insight into the distribution of vowels within each digraph.

This feature extraction process, including both the distribution and proportion tables, provides a comprehensive overview of vowel usage patterns associated with different digraphs in Swahili.

3.2 Data Collection

For the study, secondary corpora were utilized by exploring various sources to ensure diversity and representativeness in the collected corpus. These corpora were sourced from multiple platforms, each offering a different collection of Swahili audio recordings. The inclusion of secondary datasets aimed to enrich the corpus with a wide range of speech samples, encompassing diverse speakers, accents, dialects, and recording conditions.

Some of the sources adopted for secondary corpora include:

Table 2: Swahili Corpora

Source	Size	Reference
Mendeley Data	1,570 audio files containing 23,487	Rono, Kiptoo (2021), "A Kiswahili dataset", Mendeley Data, V1, doi: 10.17632/rbn6nmygcn.1
Harvard Dataverse		Wanzare, Lilian D.A; Indede, Florence; McOnyango, Owen; Ombui, Edward; Wanjawa, Barack; Muchemi, Lawrence, 2022, "KenTrans: A Parallel Corpora for Swahili and local Kenyan Languages", https://doi.org/10.7910/DVN/NOAT0W, Harvard Dataverse, V2
Zenodo		Shivachi Casper Shikali, & Mokhosi Refuoe. (2019). Language modelling data for Swahili (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3553423
Kaggle		https://www.kaggle.com/datasets/victorolufemi/swahili-audio-dataset

The utilization of secondary corpora contributed to the effectiveness of the study in addressing the challenges of Swahili speech recognition and advancing the development of speech recognition models tailored to Swahili. The annotation process involved identifying and labelling Swahili digraphs within the collected audio corpus. Each audio recording was transcribed into text. The study developed a model that extracts, segments, and Classifies Swahili digraphs with accuracy and consistency. Additionally, quality checks and validation procedures were implemented to verify the correctness of the annotations.

3.3 Sample and Sampling Technique

The study population comprises a total of 31,197 extracted Swahili digraphs, which have been categorized based on their features and vowel sequences. The aim of the study was to classify these digraphs according to their characteristics and vowel sequences to improve NLP tasks.

3.3.1 Stratified random sampling

Stratification is the process of grouping before sampling [28]. In machine learning, groups are labelling such as digraph and digraph with respective vowels. In classification problems, stratified sampling is important, especially if the dataset is imbalanced. This ensures that the training and testing datasets maintain the same proportion of classes as the population.

For digraph classification, stratified sampling ensured that each class (digraph) is adequately represented and classified based on vowel sequences. Classification is essential in NLP, as it systematically categorizes text into predefined classes or labels [29]. Machine learning algorithms effectively categorize text based on extracted features, enabling efficient management and analysis of large datasets. Such effective classification enhances the organization and interpretation of textual data, thereby supporting informed decision-making and actionable insights across diverse NLP tasks [30]. The corpus contains distinct digraphs *"ch," "dh," "gh," "kh," "ng'," "ny," "sh," "th," and "ng"* [6], stratifying based on these digraphs ensures that each digraph is well represented in the sample. This is particularly important if some digraphs appear less frequently than others.

Further refinement involved classifying words based on the vowels they contained, leading to more homogeneous subsets. This refined corpus provides a valuable resource for linguistic studies, facilitating in-depth analysis of Swahili digraphs and their phonological characteristics. The corpus was divided into groups based on the Distribution of Words by Digraph. Then the distribution by words is further stratified into digraphs with respective vowels the *a, e, i, o, u.* they contained, leading to more homogeneous subsets as shown on the table 3 below [31].

Table 3: Distribution of Words by Digraph

Digraph	Words with 'a'	Words with 'e'	Words with 'i'	Words with 'o'	Words with 'u'	Total
ch	2,022	1,594	1,716	2,232	1,919	9,483
dh	1,155	130	1,159	92	323	2,859
gh	126	293	340	30	52	841
kh	24	11	9	0	0	44
ng'	62	43	26	45	23	199
ny	1,000	472	495	255	63	2,285
sh	1,003	576	741	466	703	3,489
th	68	134	142	12	37	393
ng	2,439	2,000	2,369	2,395	2,401	11,604
Total	7,899	5,253	6,997	5,527	5,521	31,197

3.3.2 Sample Size Determination

Given the size of the population of 31,197 [31], corpus is relatively large, the study adopted the use of a reasonable sample for training, validation, and testing using the power analysis sampling technique. The power analysis identified the sample size needed for achieving statistical confidence in the classification results [32] [33]. Power analysis is a statistical method used to ascertain the minimum sample size needed to reliably detect an effect, taking into account a predetermined significance level and the desired statistical power. This is useful for determining whether you have enough data to achieve a certain level of predictive accuracy [32].

The general formula for sample size using power analysis is:

$$N = \frac{Z^2 . P(1-P)}{E^2} \qquad (1)$$

Where:
N = sampled size
Z = Z-score corresponding to the desired confidence level (1.96 for 95% confidence)
P = estimated proportion of the population (since each class might be different, a conservative estimate is (P = 0.5)
E = margin of error 0.01 (1% margin of error)

Given that P=0.5, Z=1.96(corresponding to a 95% confidence level), and E represents the margin of error of 0.01, The sample size is:

$$N = \frac{1.96^2 .0.5(1-0.5)}{0.01^2} = 9{,}604\ samples \tag{2}$$

This means you would need a sample size of 9,604 to achieve a 95% confidence level with a 1% margin of error, which is a suitable number for training, validation, and testing the corpus.

3.3.3 Sample Distribution of Words by Digraph

To break down the minimum sample size of 9,604 samples across the individual digraphs for training, validation, and testing, the samples can be allocated proportionally to each digraph based on the total word count for each digraph. This approach ensures that the distribution of samples reflects the relative frequency of each digraph, thereby maintaining the integrity of the corpus for accurate model training and evaluation.

Step 1: Calculate the proportion of each digraph in the total corpus.

The total word count across all digraphs is **31,197**.

For each digraph, the proportion is calculated as:

$$Proportion = \frac{Total\,Word\,Count\,for\,Digraph}{Total\,Word\,Count\,across\,All\,Digraphs} = \frac{Word\,Count\,for\,Digraph}{31{,}197} \tag{3}$$

Step 2: Allocate the minimum sample size (9,604) proportionally to each digraph.

For each digraph, the sample size is calculated as:

Sample Size for Digraph(n)=Proportion × 9,604

Step 3: Divide the sample size into three distinct segments: allocate 70% for training purposes, 15% for validation, and the remaining 15% for testing.

Table 4: Proportional Sample Sizes for Each Digraph

Digraph	Total Word Count	Proportion	Sample Size n (9,604 × Proportion)	Training (70%)	Validation (15%)	Testing (15%)
ch	9,483	0.3038	2,917	2,042	438	438
dh	2,859	0.0916	879	615	132	132
gh	841	0.027	259	181	39	39
kh	44	0.0014	18	10	4	4
ng'	199	0.0064	61	43	9	9
ny	2,285	0.0732	703	492	106	106
sh	3,489	0.1118	1,073	751	161	161
th	393	0.0126	121	85	18	18
ng	11,604	0.372	3,574	2,502	536	536
Totals	31,197		9,605	6,721	1,443	1,443

In this study, the corpus was systematically divided into three distinct sets: 70% was allocated for training purposes, 15% for validation, and the remaining 15% for testing. This allocation provides a balanced approach to model development, with the majority of the data (70%) dedicated to training, while 15% is set aside for validation and another 15% for testing. Based on the calculated minimum sample size of 9,604, the distribution of samples is as follows: 6,721 samples are assigned to the training set, 1,443 to the validation set, and 1,443 to the testing set.

The training subset, which comprises 70% of the total data, was used to help the model learn patterns, relationships, and key features within the dataset. By utilizing a significant portion of the data for training, the model can develop a strong understanding of Swahili digraphs, thereby improving its predictive accuracy. The validation subset, which constitutes 15% of the entire corpus, is used to fine-tune the model's parameters. This subset is essential for optimizing the model's performance, as it aids in avoiding overfitting and guarantees that the model performs effectively on new data. The testing subset, which also comprises 15% of the overall dataset, is utilized to evaluate how well the model performs on on unseen data. This provides an unbiased evaluation of its accuracy and ability to generalize.

This 70-15-15 split guarantees that the model undergoes thorough training, validation, and testing, promoting high-quality performance across all phases of the model development process.

4 Model Architecture for Utilizing Extracted Features from the Swahili Digraph corpus

To effectively leverage the extracted features from the Swahili digraph corpus for natural language processing (NLP) tasks, a well-designed neural network architecture is essential [34]. This architecture integrates both dense and convolutional layers to extract and learn from the rich feature set. This study provides an in-depth explanation of how each layer functions and contributes to the overall model performance.

4.1 Data Preprocessing

Before training the model, preprocessing was done to ensure that the features were in a suitable format;

Normalization: Scales the vowel proportion features to a standard range, improving numerical stability and convergence during training.

One-Hot Encoding: Converts categorical digraph features into binary vectors, which enables the model to process discrete features effectively.

4.2 Model Architecture

The proposed neural network architecture effectively integrates convolutional and dense layers to utilize the rich feature set derived from the Swahili digraph corpus. Conv1D layers extract local patterns and features, while dense layers perform high-level reasoning and prediction. Dropout layers ensure robust generalization by mitigating overfitting. This architecture is well-suited for various NLP tasks, including classification, language modelling, and phonological analysis, providing a comprehensive approach to leveraging the extracted features.

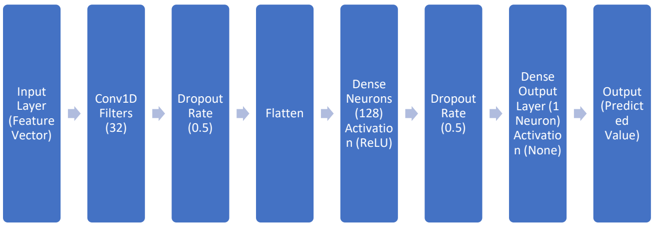
Input Layer (Feature Vector)
Conv1D Filters (32)
Dropout Rate (0.5)
Flatten
Dense Neurons (128) Activation (ReLU)
Dropout Rate (0.5)
Dense Output Layer (1 Neuron) Activation (None)
Output (Predicted Value)

Figure 1: Model Architecture

Architecture Input Layer

The purpose of the input layer is to receive the pre-processed features, including normalized vowel proportions and one-hot encoded digraphs. The input shape is defined by the number of features after preprocessing to ensure that the model can appropriately ingest the data.

Convolutional Layers (Conv1D):

The Convolutional layers serve to capture local patterns and spatial hierarchies in the input data. Specifically for sequence data like digraph features, Conv1D layers are effective in identifying patterns along the sequence.

Conv1D layers use a series of filters (or kernels) to analyse input data, sliding over the sequence to detect features like edges or local correlations. These filters extract spatial features by convolving with the input data.

a) Filters: The number of filters (e.g., 32) determines the creation of different feature maps, each detecting distinct aspects of the input.

b) Kernel Size: The size of the convolutional kernel (e.g., 2) determines the length of the subsequence analysed at each step. A kernel size of 2 is suitable for identifying patterns in pairs of features.

c) ReLU (Rectified Linear Unit) is employed as the activation function, introducing non-linearity and enabling the model to learn complex patterns. ReLU is defined as ReLU(x) = max (0, x).

Flatten Layer

The Flatten layer is important for converting the multi-dimensional output of the convolutional layers into a one-dimensional vector. This transformation is necessary to connect the convolutional part of the model with the dense layers. By reshaping the output from the convolutional layers into a vector, the Flatten layer ensures that the data is ready for processing by the dense layers.

Dense Layers

Dense layers, also known as fully connected layers, play a crucial role in high-level reasoning by utilizing the features extracted by the convolutional layers [35]. These layers interpret the extracted features and make final predictions. In a dense layer, each neuron is connected to every neuron in the previous layer, which allows the model to integrate information from all the features effectively. There are two main aspects to consider for a dense layer:

i. Units: The quantity of neurons in the dense layer (for example, 128) influences the model's capacity to comprehend intricate patterns and relationships.

ii. Activation Function: The ReLU (Rectified Linear Unit) activation function is often used to introduce non-linearity, which allows the model to learn more intricate relationships between features.

Dropout Layer

Dropout is a strategy used to regularize neural networks by randomly turning off a fraction of neurons during training. This approach reduces the likelihood of the model becoming overly dependent on certain neurons, ultimately improving its performance on new, unseen data. A common dropout rate is 0.5, which indicates that 50% of the neurons are randomly disabled in each training round.

Output Layer

The purpose of this process is to generate the final prediction from the processed features. In regression tasks, the output layer consists of a single neuron that uses a linear activation function. This setup enables the model to predict continuous values. The linear activation function is suitable for regression because it allows for unbounded output values.

4.3 Training and Evaluation

Compilation

Optimizer: The Adam optimizer is employed for its adaptive learning rate capabilities, enabling efficient convergence to the optimal solution.

Loss Function: The Mean Squared Error is used as the loss function for regression tasks. It calculates the average of the squared differences between the predicted values and the actual values.

Training

The corpus is structured into three segments: 70% allocated for training purposes, 15% reserved for validation, and 15% set aside for testing. This distribution provides a well-rounded strategy for model development, with the largest portion (70%) utilized for training, 15% designated for validation to optimize model parameters, and the final 15% reserved for testing to assess the model's effectiveness on new data. Given a minimum sample size of 9,604, the allocation consists of 6,723 samples for training, 1,441 for validation, and 1,441 for testing.

Evaluation

Metrics: Mean Absolute Error evaluates model performance on test data, showing average prediction error.

5 Results and Discussion

The research yielded four key outcomes based on the study-specific objectives: the Swahili digraph features, the Annotated Swahili Digraph Corpus, the Swahili Digraph Extraction Model, and the model performance evaluation.

5.1 Objective 1: Identification and Categorization of Unique Acoustic and Phonetic Features of Swahili Digraphs

This objective focused on the systematic identification and categorization of the unique acoustic and phonetic features of Swahili digraphs that significantly impact their recognition in speech data. The study analysed specific digraphs, including "ch," "dh," "gh," "kh," "ng'," "ny," "sh," "th," and "ng," to understand their distinctive sound patterns. The extraction process emphasized several key features: the specific two-letter combinations (digraphs), the total number of words (frequency of each digraph across the dataset), and vowel distribution (the count of words containing each vowel within the scope of each digraph). This detailed analysis aimed to capture the intricate patterns and dependencies of these digraphs in Swahili speech. By categorizing these features in detail, the study offers significant insights into Swahili's phonetic structure, which aids in developing more refined and effective NLP tools tailored to the unique characteristics of the language.

5.2 Objective 2: The Swahili Digraph Extraction Model

The study designed a neural network architecture to effectively use the features extracted from the Swahili digraph corpus for Natural Language Processing (NLP) tasks. This architecture integrated both dense and convolutional layers to utilize the rich feature set of Swahili digraphs, including their specific two-letter combinations, frequency across the corpus, and vowel distribution.

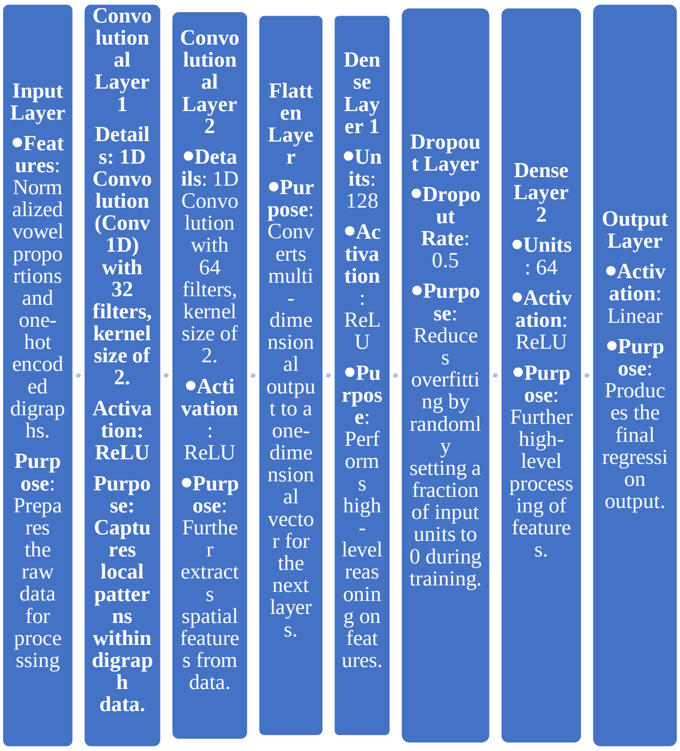
Input Layer
•Features: Normalized vowel proportions and one-hot encoded digraphs.
Purpose: Prepares the raw data for processing
Convolutional Layer 1
Details: 1D Convolution (Conv1D) with 32 filters, kernel size of 2.
Activation: ReLU
Purpose: Captures local patterns within digraph data.
Convolutional Layer 2
•Details: 1D Convolution with 64 filters, kernel size of 2.
•Activation: ReLU
•Purpose: Further extracts spatial features from data.
Flatten Layer
•Purpose: Converts multi-dimensional output to a one-dimensional vector for the next layers.
Dense Layer 1
•Units: 128
•Activation: ReLU
•Purpose: Performs high-level reasoning on features.
Dropout Layer
•Dropout Rate: 0.5
•Purpose: Reduces overfitting by randomly setting a fraction of input units to 0 during training.
Dense Layer 2
•Units: 64
•Activation: ReLU
•Purpose: Further high-level processing of features.
Output Layer
•Activation: Linear
•Purpose: Produces the final regression output.

Figure 2:CNN-based digraph extraction model

The Figure 2 illustrates the CNN-based digraph extraction model visually, illustrating the sequence of layers and their respective functionalities in processing Swahili digraph features. This Extraction Model architecture effectively integrates Swahili-specific features identified through systematic analysis, enabling precise modelling of the language's phonetic structures. The integration of convolutional layers for feature extraction and dense layers for advanced reasoning, coupled with dropout for regularization, guarantees robust and efficient processing of Swahili digraphs. This comprehensive approach enhances the accuracy and efficiency of various NLP tasks, advancing the field of Swahili speech recognition technology.

5.3 Objective 3: The Annotated Swahili Digraph Corpus

The study Objective **3**, which focused on creating and annotating a Swahili digraph corpus to from the CNN-based extraction model, resulted to The Annotated Swahili Digraph Corpus [31].

Table 5: The Annotated Swahili Digraph Corpus

Digraph	Words with 'a'	Words with 'e'	Words with 'i'	Words with 'o'	Words with 'u'	Total
ch	2,022	1,594	1,716	2,232	1,919	9,483
dh	1,155	130	1,159	92	323	2,859
gh	126	293	340	30	52	841
kh	24	11	9	0	0	44
ng'	62	43	26	45	23	199
ny	1,000	472	495	255	63	2,285
sh	1,003	576	741	466	703	3,489
th	68	134	142	12	37	393
ng	2,439	2,000	2,369	2,395	2,401	11,604
Total	7,899	5,253	6,997	5,527	5,521	31,197

The findings from Table 5 The Annotated Swahili Digraph Corpus underscore the comprehensive and balanced nature of the dataset, which effectively captures the unique phonetic characteristics crucial for Swahili language processing applications. The corpus includes a diverse array of Swahili digraphs, such as "ch," "dh," "gh," "kh," "ng'," "ny," "sh," "th," and "ng," which are essential for representing the phonetic complexity of Swahili sounds. This diversity supports the corpus's role in providing the CNN model with a foundation to discern subtle linguistic distinctions, thereby enhancing the accuracy and effectiveness of digraph recognition.

A key aspect of the corpus is its detailed annotation of each digraph's frequency across Swahili's five vowels: "a," "e," "i," "o," and "u." Notably, the digraph "ch" appears most frequently, with 9,483 instances across vowel contexts, reflecting its prominence in Swahili phonology. In contrast, the digraph "ng" demonstrates a balanced distribution with 11,604 occurrences across vowels, highlighting its phonetic versatility and prevalence within the language. Such a distribution aids in capturing Swahili's phonological diversity, providing the model with the necessary data to generalize effectively across various vowel combinations, which is critical for achieving high accuracy in digraph recognition.
The corpus, comprising a total of 31,197 words across different digraphs and vowel contexts, offers an extensive dataset for training purposes. This extensive representation of digraphs strengthens the model's ability to learn varied pronunciation patterns, particularly those associated with commonly occurring digraphs like "ch" and "ng," which feature prominently within the language. This balanced distribution enhances the model's potential for generalization, facilitating accurate recognition of Swahili digraphs across diverse phonetic contexts.

The corpus included rarer digraphs, such as “kh” and “ng’,” with lower counts of 44 and 199 instances, respectively, offering insights into these less common sounds. While these rare digraphs may necessitate additional emphasis during the training phase to ensure reliable model recognition, they provide valuable data for evaluating the model’s capacity to identify and accurately process linguistically significant, albeit infrequent, phonetic elements.

The Annotated Swahili Digraph Corpus provides a robust, well-distributed dataset crucial for the training and evaluation of a CNN-based Swahili digraph extraction model. The corpus’s balanced representation of both frequent and rare digraphs, coupled with its extensive coverage across vowel contexts, creates a comprehensive foundation for Swahili language processing tasks. This resource encompasses a wide range of Swahili phonetic patterns, thereby enhancing the model's ability to generalize across different linguistic contexts. This methodology greatly advances the development of effective and context-sensitive applications for processing the Swahili language. By utilizing such diverse phonetic data, the creation of tools and applications within the domain of Swahili language processing can be made increasingly robust and effective.

5.4 Objective 4: The model performance evaluation.

This section presents the results of the evaluation metrics applied to the model, which was trained on the dataset above. The primary goal of the evaluation was to assess the model's performance and its generalization capability when applied to previously unseen data. The following subsections outline the evaluation procedure in detail, along with a comprehensive analysis of the results obtained from these metrics.

5.4.1 Model Evaluation for Swahili Digraph Corpus

Step 1: Evaluate Model on Test Data
The researcher evaluated the trained model with the test data set aside during the data splitting phase to gain an initial understanding of the model's performance.
Step 2: Calculate Performance Metrics
Performance metrics, including Root Mean Squared Error, R-squared, and Mean Squared Error, are computed to offer a comprehensive assessment of the model's performance.
Step 3: Analysis of Results
The following is a summary of the evaluation results.

Table 6: Evaluated Performance Metrics

Test	Results
Test Loss	0.024
Test Mean Absolute Error	0.112
Mean Squared Error	0.031
Root Mean Squared Error	0.176
R-squared	0.89

5.5 Discussion

The study, successfully created and annotated a comprehensive corpus consisting of 31,197 extracted Swahili digraphs, classified based on their features and vowel sequences. The primary aim of this research was to enhance natural language processing (NLP) tasks by effectively categorizing these digraphs. This corpus not only serves as a valuable resource for further research in Swahili NLP but also contributes to the understanding of phonetic and phonological patterns in the Swahili language. The corpus consists of rigorously annotated Swahili Corpus, constituting a critical resource for the training and evaluation of the Digraph Extraction Model. These annotations are essential for accurately reflecting the phonological complexities of Swahili, particularly in the case of digraphs, where the combination of two letters results in unique sounds that cannot be adequately captured by their components.

The high level of annotation detail enhances the model’s capacity to learn and differentiate among various Swahili digraphs in Corpus, thereby improving its effectiveness in handling real-world audio inputs. Moreover, the precision of these annotations plays a pivotal role in ensuring the model’s robustness and ability to generalize across different linguistic variations, including speaker accents and dialects. Consequently, the corpus not only facilitates the development of a reliable digraph extraction model but also serves as an indispensable benchmark for evaluating the model’s performance in recognizing and extracting Swahili digraphs across diverse speech environments.

The evaluation of the model applied to the Swahili Digraph Corpus produced a set of performance metrics, as outlined in Table 6, which reflects its efficacy. The test loss was measured at 0.024, signifying a minimal error rate and a high level of accuracy in the model's predictions on the test corpus. The Mean Absolute Error was determined to be 0.112, signifying a minimal average absolute deviation between the predicted and actual values. This underscores the model's accuracy and precision in its predictions. The Mean Squared Error stood at 0.031, demonstrating the model's ability to handle larger discrepancies effectively, as evidenced by its low average squared error. The Root Mean Squared Error was calculated to be 0.176, signifying a relatively low level of error in the model's predictions when expressed in the same units as the target variable. Furthermore, the R-squared value of 0.89 signifies that the model accounts for 89% of the variance in the dependent variable. This considerable explanatory power underscores the model's robustness and reliability in effectively capturing the underlying relationships within the corpus.

The evaluation findings underscore the model's impressive performance across various metrics. The low Mean Absolute Error and Root Mean Squared Error values highlight the model's high accuracy, with predictions closely matching actual observations. Additionally, the notable R-squared value of 0.89 demonstrates the model's robust predictive capability, effectively capturing and explaining a significant portion of the variance within the corpus. The low-test loss further indicates the model's proficiency in generalizing to new, unseen data, underscoring its reliability for practical applications and its potential for successful real-world deployment. Collectively, these results confirm the model's effectiveness and readiness for implementation in relevant contexts.

6 Conclusion and Future Scope

The achievement of developing The Annotated Swahili Digraph Corpus addresses the significant gap left by the absence of a specialized corpus for Swahili digraphs. This accomplishment is set to be recognized academically, marking a notable milestone in research efforts to support training and further research in the fields of natural language processing, digraph analysis, and transcription. The lack of a dedicated corpus has posed a major obstacle to enhancing NLP applications for Swahili, making this development crucial for advancing research and practical applications in the area.

The research has effectively demonstrated the strength of feature extraction and advanced modelling techniques in improving Swahili digraph recognition. The notable findings emphasize the model's impressive accuracy and strong predictive capability, as indicated by a test loss of 0.024, a Mean Absolute Error of 0.112, and an R-squared value of 0.89. These metrics underscore the model's precision in recognizing Swahili digraphs and their associated vowel distributions. The feature extraction process, incorporating normalization of vowel counts and one-hot encoding for digraphs, effectively isolated meaningful attributes from the corpus, thereby facilitating accurate modelling.

This research contributes significantly to the field of Swahili language processing by developing a unique corpus that classifies Swahili digraphs and their respective vowels into specific categories. It provides a comprehensive tool for linguistic analysis, enhancing the understanding of Swahili phonology and orthography. The structured corpus and the feature extraction methodology established in this study serve as valuable resources for future research and development, enabling the creation of advanced language models and tools. Moreover, the model's implications for speech recognition applications are substantial, offering improved accuracy and customization for Swahili speakers. It supports the development of multilingual speech recognition systems, promoting inclusivity for diverse linguistic communities. Future research should focus on validating the model across varied language datasets, exploring more sophisticated models, and applying these findings to practical scenarios to further advance language processing technologies.

Ethical Considerations

This study did not involve human subjects. All data used in the research were derived from publicly available datasets, including those from Harvard Dataverse, Mendeley Data, Zenodo, and Kaggle. As such, no informed consent was required. The study adhered to relevant ethical guidelines and institutional policies for research involving non-human data.

Data Availability

The data underpinning the conclusions of this study, including the annotated Swahili digraph corpus, can be obtained from the corresponding author upon reasonable request.

Conflict of Interest

Authors declare that they do not have any conflict of interest with anyone for publication of this work.

Funding Source

None

Authors' Contributions

All authors contributed equally to this study. Tirus Muya Maina, Aaron Mogeni Oirere, and Stephen Kahara collaboratively developed the study's conceptual framework, designed the methodology, and carried out data collection and analysis. Each author played a key role in model development, interpretation of results, and manuscript preparation. All authors reviewed and approved the final manuscript for submission**.**

Acknowledgements

The authors would like to note that no external contributions were made beyond those of the listed authors.

References

[1] K. Kayabol, "Approximate Sparse Multinomial Logistic Regression for Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2020.**

[2] J. H. Hansen and G. Liu, "Unsupervised accent classification for deep data fusion of accent and language information," *Speech Communication*, vol. 78, pp. 19–33.

[3] S. A. M. Yusof, A. F. Atanda, and H. Husni, "Improving the Performance of Multinomial Logistic Regression in Vowel Recognition by Determining Best Regression Coefficients," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*, Sakheer, Bahrain, **2020.**

[4] M. Y.-C. Jiang, M. S.-Y. Jong, W. W.-F. Lau, C.-S. Chai, and N. Wu, "Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom," *Journal of Computer Assisted Learning*, vol. 39, no. 1, pp. **125–140, 2023.**

[5] M. M. Waqar, M. Aslam, and M. Farhan, "An Intelligent and Interactive Interface to Support Symmetrical Collaborative Educational Writing among Visually Impaired and Sighted Users," *Symmetry*, vol. 11, no. 2, p. **238, 2019.**

[6] W. H. Finch, J. E. Bolin, and K. Kelley, *Multilevel Modeling Using R*, United Kingdom: CRC Press/Taylor & Francis Group, **2019.**

[7] M. S. Azmi, "Development of Malay Word Pronunciation Application using Vowel Recognition," *International Journal of u- and e-Service, Science and Technology*, vol. 9, no. 1, pp. **221–234, 2016.**

[8] M. S. Azmi, "Malay Word Pronunciation Test Application for Pre-School Children," *Int Journal of Interactive Digital Media*, vol. 4, no. 2, pp. 2289–4098, 2016.

[9] K. Y. Chan and M. D. Hall, "The importance of vowel formant frequencies and proximity in vowel space to the perception of foreign accent," *Journal of Phonetics*, vol. 77, p. 100919, 2019.

[10] H. Meutzner, S. Araki, M. Fujimoto, and N. Tomohiro, "A generative-discriminative hybrid approach to multi-channel noise reduction for robust automatic speech recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. **5740–5744, 2016.**

[11] F. Wu, L. P. García-Perera, D. Povey, and S. Khudanpur, "Advances in Automatic Speech Recognition for Child Speech Using Factored Time Delay Neural Network," in *Interspeech*, **2019.**

[12] S. Ghorbani, S. Khorram, and J. Hansen, "Domain Expansion in DNN-Based Acoustic Models for Robust Speech Recognition," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Singapore, **2019.**

[13] C. Shan et al., “Investigating End-to-end Speech Recognition for Mandarin-English Code-switching,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, **2019.**

[14] X. Li, “Low-Resource Speech Recognition for Thousands of Languages,” Carnegie Mellon University, **2023.**

[15] S. Amuda et al., “Engineering Analysis and Recognition of Nigerian English: An Insight into Low Resource Languages,” *Transactions on Engineering and Computing Sciences*, **2014.**

[16] V. Hai, X. Xiao, E. S. Chng, and H. Li, “Cross-lingual phone mapping for large vocabulary speech recognition of under-resourced languages,” *IEICE Transactions on Information and Systems*, vol. 97, no. 2, pp. **285–295, 2014.**

[17] D. A. Gonc¸alves et al., “Facial Expressions Animation in Sign Language based on Spatio-temporal Centroid,” in *22nd International Conference on Enterprise Information Systems*, **2020.**

[18] M. Mehraj et al., “Automatic Speech Recognition Approach for Diverse Voice Commands,” *International Journal of Advanced Research in Computer Science*, vol. 8, no. 9, **2017.**

[19] G. Korvel et al., “Speech Analytics Based on Machine Learning,” in *Machine Learning Paradigms. Intelligent Systems*, Springer, Cham, 2019.

[20] F. Ọ. Asahiah, “Comparison of rule-based and data-driven approaches for syllabification of simple syllable languages and the effect of orthography,” *Computer Speech & Language*, vol. 70, 2021.

[21] R. Zevallos et al., “Automatic speech recognition of Quechua language using HMM toolkit,” in *Annual International Symposium on Information Management and Big Data*, pp. **61–68, 2019.**

[22] H. Tang et al., “End-to-End Neural Segmental Models for Speech Recognition,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, 2017.

[23] M. Alam et al., “Survey on Deep Neural Networks in Speech and Vision Systems,” *Neurocomputing*, vol. 417, pp. **302–321, 2020.**

[24] M. Z. Alom et al., “A State-of-the-Art Survey on Deep Learning Theory and Architectures,” *Electronics*, vol. 8, no. 3, **2019.**

[25] J. O. De Sordi, *Design Science Research Methodology*, Springer International Publishing, 2021.

[26] A. R. Kivaisi, Q. Zhao, and J. T. Mbelwa, “Swahili Speech Dataset Development and Improved Pre-Training Method for Spoken Digit Recognition,” *ACM Transactions on Asian and Low-Resource Language Information Processing*, **2023.**

[27] J. vom Brocke, A. Hevner, and A. Maedche, *Introduction to Design Science Research*, Springer, Cham, 2020.

[28] T. Yamane, *Elementary Sampling Theory*, New Jersey: Prentice-Hall, 1967.

[29] S. K. Daroch and P. Singh, “An Analysis of Various Text Segmentation Approaches,” in *Proceedings of International Conference on Intelligent Cyber-Physical Systems. Algorithms for Intelligent Systems*, Singapore, **2022.**

[30] M. K. Najm et al., “Text Classification Accuracy Enhancement Using Deep Neural Networks,” in *2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*, Al-Muthana, Iraq, 2023.

[31] T. M. Maina, “The Swahili Digraph Corpus,” *Mendeley Data*, vol. 2, 2024.

[32] N. Colegrave and D. R. Graeme, *Power Analysis: An Introduction for the Life Sciences*, Oxford University Press, **2021.**

[33] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Taylor & Francis, **2013.**

[34] S. Sarma and N. Pathak, “Design and Implementation of an Assamese Language Chatbot Using,” *International Journal of Scientific Research in Computer Science and Engineering*, vol. 11, no. 6, pp. **13–18, 2023**.

[35] Deepanshu et al., “Convolutional Neural Network-Based Automated Acute Lymphoblastic Leukemia Detection and Stage Classification from Peripheral Blood Smear Images,” *International Journal of Scientific Research in Computer Science and Engineering*, vol. 12, no. 3, pp. **21–28, 2024.**

AUTHORS PROFILE

Tirus Muya Maina is a highly experienced ICT professional, specializing in Information and Communication Technology (ICT) and Computer Science. He holds a Master's degree in Information Systems and is currently pursuing a Ph.D. in Computer Science. With over ten years of experience, he has held roles such as Senior ICT Technologist II at Murang’a University of Technology. His expertise spans network infrastructure, software development, cybersecurity, ICT policy formulation, data management, and ICT strategy. Tirus has published research in reputable journals and is an active member of professional bodies, having received advanced training in cybersecurity and higher education. His research interests include Artificial Intelligence, Natural Language Processing, ICT Integration in Education, Cybersecurity, TVET, ICT Policy and Governance, and Curriculum Development.

Aaron Mogeni Oirere, received his B.Sc. degree in Computer Science from Periyar University, Salem, Tamilnadu, India in 2007, the M.Sc. degree in Computer Science from Bharathiar University, Coimbatore, Tamilnadu, India in 2010, and the Ph.D. degree in Computer Science from Dr. Babasaheb Ambedkar Marathwada University, Maharashtra, India in 2016. He currently works at the Department of Computer Science, school of computing and Information Technology, Murang’a University of Technology. His research interest includes Automatic Speech Recognition, Human-computer Interaction, Information Retrieval, Database Management Systems (DBMS), Data Analytics and Hardware & Networking.

Stephen Kahara is a Lecturer of Computer Science at Murang'a University of Technology and the Director of Performance Contracting and ISO. He holds a Ph.D. in Computer Science from Murang'a University of Technology, an M.Sc. in Computer Systems from JKUAT, an M.Sc. in Organizational Development from USIU-Africa, and a B.Sc. in Information Sciences from Moi University. Dr. Kahara has 16 years of experience in the ICT industry. He is a certified QMS and ISMS auditor and a DAAD-UNILEAD alumnus. His research interests include machine learning, network security, distributed systems, and computational biology. He has published extensively in academic journals and conferences. Dr. Kahara is a member of the Association of Computing Practitioners - Kenya.