I.J. Information Technology and Computer Science, **2025, 1, 67-80**
Published Online on February 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2025.01.05

Modelling Misinformation in Swahili-English Code-switched Texts

Cynthia Amol*
Maseno University, Department of Computer Science, Maseno, Kenya
E-mail: cynthia@maseno.ac.ke
ORCID iD: https://orcid.org/0009000208707126
*Corresponding Author

Lilian Wanzare
Maseno University, Department of Computer Science, Maseno, Kenya
E-mail: ldwanzare@maseno.ac.ke
ORCID iD: https://orcid.org/000000026718900X

James Obuhuma
Maseno University, Department of Computer Science, Maseno, Kenya
E-mail: jobuhuma@gmail.com
ORCID iD: https://orcid.org/0000000213604562

Received: 22 July 2024; Revised: 08 October 2024; Accepted: 21 December 2024; Published: 08 February 2025

Abstract: Code-switching, which is the mixing of words or phrases from multiple, grammatically distinct languages, introduces semantic and syntactic complexities to sentences which complicate automated text classification. Despite code-switching being a common occurrence in informal text-based communication among most bilingual or multilingual users of digital spaces, its use to spread misinformation is relatively less explored. In Kenya, for instance, the use of code-switched Swahili-English is prevalent on social media. Our main objective in this paper was to systematically re- view code-switching, particularly the use of Swahili-English code-switching to spread misinformation on social media in the Kenyan context. Additionally, we aimed at pre-processing a Swahili-English code-switched dataset and developing a misinformation classification model trained on this dataset. We discuss the process we took to develop the code- switched Swahili-English misinformation classification model. The model was trained and tested using the PolitiKweli dataset which is the first Swahili-English code-switched dataset curated for misinformation classification. The dataset was collected from Twitter (now X) social media platform, focusing on text posted during the electioneering period of the 2022 general elections in Kenya. The study experimented with two types of word embeddings - GloVe and FastText. FastText uses character n-gram representations that help generate meaningful vectors for rare and unseen words in the code-switched dataset. We experimented with both the classical machine learning algorithms and deep learning algo- rithms. Bidirectional Long Short-Term Memory Networks (BiLSTM) algorithm showed the best performance with an f-score of 0.89. The model was able to classify code-switched Swahili-English political misinformation text as fake, fact or neutral. This study contributes to recent research efforts in developing language models for low-resource languages.

Index Terms: Code-switching, Swahili, Low-resource Languages, Misinformation, Text Classification.

1. Introduction

Social media platforms such as Twitter (now X) offer uncensored channels for information dissemination which em- boldens users to share unverified news or misinformation. In the review of misinformation, the study by [1] concluded that the two most popular definitions of misinformation are ‘false and misleading information’ and ‘false and misleading information spread unintentionally’. Despite social media platforms’ strict policies against misinformation amplification and penalisation of misinformation spreaders, cases of false news and misinformation are still being flagged on the plat- forms. In the instance of the 2022 general elections in Kenya, misinformation in the form of fake polls, unverified electoral results, and unsubstantiated statements from political parties [2] were common on Twitter (now X). The major challenge has always been how cases of misinformation can be automatically flagged, especially on

these platforms. There have been several works on misinformation detection and fake news detection in the recent past as discussed in section 2.

However, there are still open questions about not only the unavailability or quality of available training data but also the efficiency of existing models [3], especially for languages that are considered low-resource.

The main issue with this classification task is availability of high-quality, labelled datasets. Misinformation classification depends on labelled datasets for the particular language. However, low-resource languages do not have adequate labelled data [4-6]. Furthermore, the multilingual nature of communities leads to the use of code-switching while communicating on social media, which complicates automated text classification. Code-switching is the shift from one language to another within a single utterance [7] that introduces complexity to a phrase or sentence's semantics. This grammatical complexity not only necessitates that the human annotators be conversant with both languages used during data annotation but also complicates the word embedding process which generates vector representation of words. The lack of data problems hinders efforts to build robust code-switching models since the available monolingual data do not follow the same patterns as code-switching [8].

Since the accuracy of misinformation detection classifiers is dependent on the training dataset used, there is a need to use a high-quality, task-specific dataset. For this study, we used the PolitiKweli[1] dataset created by [9], which is the first code-switched Swahili-English dataset curated from posts from Twitter on the electioneering period in Kenya. The dataset contains 6,345 code-switched Swahili-English texts manually annotated in three categories: fake, fact and neutral. The dataset was processed by data cleaning to remove missing values, detailed anonymization, removal of stopwords, tokenization and feature extraction.

Handling code-switching is challenging as the nuances of the two or more languages in the text should be captured should be captured in order to flag a text as either fake or factual accurately. Previous researchers working on code-switching have experimented with different word embeddings to capture the code-switching [10-12], as shown in section 2.2. In this paper, we explore the use of an ensemble of word embedding models for a code-switched dataset to be able to handle the semantic complexities of code-switching due to semantic variations of words. We used a combination of two embeddings, Global Vectors (GloVe) [13] which provides accurate contextual embedding by emphasising the co-occurrence probabilities of words within a corpus of texts [14] and FastText [15] which represents words by a sum of its character n-grams [16] considering sub-word information (morphology) thus is able to accurately represent out-of- vocabulary words which exist in a code-switched dataset such as the one used in this study.

In order to build baseline models, we experimented with both traditional machine learning models and transformer-based models. For traditional machine learning models, we utilized Logistic Regression (LR) and the Multilayer Perceptron (MLP). For transformer-based model, we utilized the Bidirectional Encoder Representations from Transformers (BERT) [17] and Text-to-Text Transfer Transformer (T5) [18]. We got best results using Bidirectional Long Short-Term Memory Networks (BiLSTM) that showed improved precision in classifying the text.

Our contributions in this study are therefore:

- A systematic review of existing datasets and models for classification of code-switched data in low-resource languages.
- Generation of word embeddings for a Swahili-English code-switched dataset which capture semantic and syntactic nuances of code-switching.
- A BiLSTM-based code-switched Swahili-English misinformation classification model trained using a high-quality code-switched dataset.
- Validation of the code-switched dataset using both traditional and deep learning models.

The rest of the paper is structured as follows: section 2 discusses related work while section 3 outlines the dataset used. Section 4 describes the methodology; section 5 illustrates the model design while section 6 discusses results from experiments. We offer ethical considerations, conclusions and recommendations for future work in sections 7 and 8 respectively.

2. Related Works

A number of previous studies [10-12, 16, 19-23] have developed machine learning datasets and models for classification tasks such as misinformation, fake news, sentiment and hate speech classification. Despite most of them being for high-resource languages, mostly English, there has been increased effort in development of language resources for low-resource languages such as Swahili and other indigenous African languages [4, 24-30]. This section reviews different approaches to the task of misinformation classification, the challenge of bilingual or multilingual communities including the art of code-switching and its influence on informal text-based communication. It also highlights some of the models developed for classification tasks involving code-switching and existing datasets, especially code-switched datasets in low-resource languages.

[1] https://github.com/jayneamol/kweli

Table 1. Example Datasets on misinformation detection and corresponding labels

Study	Task	Label
[9]	Misinformation classification	'fake', 'fact', 'neutral'
[11]	Misinformation classification	'misinformation', 'no misinformation'
[19]	Misinformation classification	'misinformation spreaders', 'real news spreaders'
[20]	Fake news detection	'fake news', 'not fake news'
[31]	Fake news detection	'fake', 'real'
[32]	Fake news detection	'fake', 'real'
[33]	Fake news detection	'fake news', 'real news'

2.1. Misinformation Classification

Dissemination of unverified news is often described broadly as fake news detection or misinformation classification. Fake news is defined by [34] as news reports that are untrue or exaggerated. The study proposes the use of machine learning, deep learning and pre-trained models to identify linguistic and user features in data for fake news detection. Misinformation, on the other hand, is defined by [35] as inaccurate information that is deliberately created, intentionally or unintentionally spread. [36] points out that misinformation is any kind of false or misleading information. Despite this subtle difference in definition, the key characteristics that define both fake news and misinformation are authenticity of the information and the intent of the spreader [37].

Studies such as [11, 19, 38] examine misinformation classification. The study by [11] explores the use of deep learning models in binary classification by classifying posts on Facebook and Twitter as either 'misinformation' or 'no misinformation'. In the study [19], users are classified as misinformation or real-news spreaders based on posted web links and posting history on the social media platform Reddit. In their modelling of a framework for fact verification, the study [39] uses a T5 model to create the 'LisT5' framework.

In [20, 31-33, 40-42], fake news detection is discussed. Using both deep learning and transformer-based models such as Hindi-BERT, Hindi-TPU-Electra, mBERT and DistilBERT, [41] models a propaganda detection model in the low-resource language, Hindi. The study [14] uses deep learning models to detect fake news detection in mainstream news channels. On the other hand, [31] uses a combination of both traditional and deep learning methods to detect fake news on Twitter. The study by [32] experiments with BERT and LSTM models to classify news as either real or fake. [20], using the dataset 'FakeNewsNet', built a diffusion graph of news spreaders and used Graph Attention Network to do binary classification of the resulting diffusion tree while [33] used domain reputations and content characteristics to classify fake and real news. These recent studies achieved the best performances using either deep learning or transformer- based classification models demonstrating their ability to classify text better compared to traditional models

The output classes of classification models depend on the labels assigned to texts during annotation. Most misinformation and fake news classification models use data labelled in binary, fake or false. As shown in Table 1, the most used labels are fake or real/fact.

A. Use of Code-Switching in Misinformation Dissemination on Social Media

Social media platforms such as Twitter are channels for social commentary and political engagement [43]. Most of the discussions can impact the agenda of mainstream media and offline political life [44]. During such discussions, speakers, especially those from multilingual communities, often mix languages as they converse. Mixing languages is a common occurrence among speakers of more than one language, especially among peers who are fluent in each language [45]. Similar to the use of Spanglish (Spanish-English) in the Americas and Hinglish (Hindi-English) in India, Swahili-English is used among some speakers of Swahili in East and Central Africa due to colonial influence [46].

The alteration between two or more languages in the same sentence is considered code-switching (CS) [47] and is common in speech and casual text as found in social media. CS is a common consequence of multilingualism [48]. Among multilinguals, a degree of CS is expected to occur in almost every scenario [45]. [49] argue that 3.5% of tweets are code-switched. This is a significant percentage considering the massive number of tweets shared on a daily basis. CS has variations - from intra-sentential to inter-sentential to emblematic [50, 51] make social media text challenging to process [47]. Table 2 shows some of the common code-switching variations in textual data.

Code-switching is often employed to express ethnic identity and convey various emotions [52]. In mostly bilingual countries such as Kenya, users of social media use code-switching to express their views. Despite the increased use of code-switching, this field is relatively unexplored, especially in its use to spread misinformation. Table 3 shows some of the existing code-switched datasets.

Misinformation can be spread in social media and the use of code-switching makes it even harder to detect. In the case of the 2022 general elections in Kenya, the social media platform Twitter (now X) flagged and labelled multiple posts relating to misleading news like fake polls, unverified electoral results and unsubstantiated statements from individuals or organisations. These cases can potentially fuel more chaos and distrust online and spiral into violence offline [2, 9]. Despite cybercrime mitigation efforts such as the signing of the Computer Misuse and Cybercrimes Act 2018 that exposes internet users who publish false or misleading information to stiff penalties [43], the anonymity

offered by social media emboldens users [53] to share unverified news.

With the increased use of code-switching among bilingual and multilingual communities in casual text-based communication such as social media [45], the development of more accurate language models trained using code-switched data will contribute immensely to research into more inclusive language technologies.

Table 2. Code-switching variations

Language	Text	Type
|en |swa |en |swa	|voting |ni[is] |jukumu[responsibility] |of |wananchi[citizens]	Intra-sentential
|swa |eng	|tunapenda nchi yetu [we love our country] |maintain peace	Inter-sentential
|eng |swa	|your candidate won |ama [right]	Emblematic

Table 3. Example Code-switched datasets and feature extraction for various classification tasks

Study	Feature extraction	Classification model
Hindi-English hate speech classification [10]	FastText, word2vec, GloVe, BERT	LSTM, CNN
Luganda-English misinformation classifi-cation [11]	TweetToSparseFeatureVector filter	DMNB, SVM
Swahili-English hate speech detection [12]	TD-IDF and GloVe	NB, SVM, KNN, DT, RF, CNN
English-Yoruba, and Nigerian Pidgin abu-sive language and hate speech detection [54]	Twitter-RoBERTa	Twitter-RoBERTa-base
Urdu-English propaganda detection [55]	BERT, RoBERTa, BERT, DeBERTa	mBERT,

B. Existing Datasets on Misinformation Detection

Previous studies [19, 56] have developed datasets for misinformation classification in languages that are considered high-resource. In this study, we focus on languages that are considered to be low-resource such as indigenous languages in parts of Africa.

Increased communication in digital and social media in a continent like Africa with over 2,000 languages [27] has sparked interest in computational linguistics, steadily increasing research efforts into low-resource language resources and by extension code-switching. Despite the small-scale size of these datasets and their task-specific nature [57], curation of these datasets is a crucial step towards language inclusion. Some the existing code-switched datasets for misinformation classification were curated by [9] and [11].

PolitiKweli by [9] is a dataset for misinformation classification. The dataset, sourced from the social media platform Twitter, contains 29,510 texts: 6,345 in code-switched Swahili-English, 22,954 in Kenyan English and 211 in Swahili. The data was labelled as fake, fact or neutral by a group of human annotators. The dataset built by [11] is a Luganda- English code-mixed COVID-19 misinformation classification dataset consisting of 12,049 tweets and 430,000 Facebook posts collected from Twitter and Facebook respectively using APIs. The text was annotated as either “misinformation” or “no-misinformation”.

C. Existing Code-switched Datasets for other Classification Tasks

Despite the increasing research into misinformation classification, there are few datasets curated for this task. In this section, we review some of the code-switched datasets for other classification tasks including propaganda and hate speech detection which follow a similar methodology as misinformation classification.

Studies such as [10, 12, 54, 55] focus on classification tasks for code-switched English-Yoruba, English-Urdu, Hindi- English, and Swahili-English language pairs respectively. EkoHate by [54] is a code-switched English, Yoruba, and Nigerian Pidgin (Naija) abusive language and hate speech dataset consisting of 3,398 annotated tweets posted during the 2023 general elections in Nigeria. The text was annotated by human annotators in two categories: normal or offensive. The study by [55] curated a propaganda detection dataset comprising of 1,030 texts code-switching between English and Roman Urdu. The text is annotated using 20 propaganda techniques.

The study by [12] resulted in a code-switched Swahili-English hate speech dataset of 260,000 tweets collected from Twitter using Twitter API. The data is then pre-processed by removal of unusable tweets. Annotation was done by a team of human annotators. Indian Political Memes (IPM) by [10] is a code-switched dataset Hinglish (Hindi-English) for hate speech classification sourced from social media. The memes (satirical images with text) were collected by identifying keywords and downloading images based on them, resulting in a corpus of images containing 5,000 images. The data was processed by removing blurry images and images with no text. The clean data was then annotated as hate-inducing or not by human annotators.

2.2. Feature Extraction

The accuracy of language models in Natural Language Processing (NLP) is anchored on the quality of the dense word representations of text in the training data. Several word embeddings, including Word2Vec [58], FastText [15], GloVe [13], BERT (Bidirectional Encoder Representations from Transformers) [17] and ELMo (Embeddings from Language Models) [59] have been developed over the years.

Feature extraction, in this case, embedding, is the process of generating numerical data from textual data. Word embedding helps understand semantics and find contextual clues [34] which is essential in NLP. During embedding, words and phrases are mapped into n-dimensional dense vectors of real numbers that effectively capture the semantic and syntactic relationship with neighbouring words in a geometric way [60].

We reviewed studies on classification tasks involving code-switching and the choice of various embedding to represent the features in the text or images in the datasets used. The study [12] that used GloVe embeddings and Term Frequency- Inverse Document Frequency (TF-IDF) vectorizers noted that the use of a single embedding, especially in a code-switched Swahili-English set-up, impacts model accuracy negatively.

The study [10] experimenting with FastText, Word2Vec, GloVe and BERT embeddings with a code-switched Hindi- English dataset concluded that a combination of word embeddings (FastText and GloVe) yields better results and improves precision [16].

GloVe, pre-trained embeddings trained on large sets Common Crawl, Wikipedia and Twitter data, emphasises the co-occurrence probabilities of words within a corpus of texts in order to embed them in meaningful vectors making it better at contextual embedding [14]. FastText, on the other hand, represents words by a sum of its character n-grams considering sub-word information (morphology), allowing for sharing the representations across words, thus the ability to learn reliable representations for rare words. A combination of the two is able to capture unique datasets such as code-switched datasets.

Table 3 shows the feature extraction techniques used by some of the code-switched datasets for misinformation classification tasks. Studies that use transformer-based models [54, 55] utilise the self-attention mechanisms [61] of these transformer models for feature extraction. However, models implemented using deep learning algorithms [10-12] have an embedding layer that uses mostly GloVe and FastText as the models are pre-trained on social media data.

2.3. Classification Models

Classification tasks involving code-switched datasets have experimented with both traditional machine learning mod- els such as Support Vector Machine (SVM), Random Forest (RF), Decision Trees (DT), deep learning models such as Recurrent Neural Networks (RNNs) and pre-trained language models such as transformer-based BERT. Due to the semantic complexity of text in code-switched texts, RNNs ability to learn from training input sequentially enables better learning of semantic features of text thus more accurate classification

Traditional machine learning algorithms have limitations in text classification and representation of semantic features. Deep learning methods overcome these limitations [62]. RNNs such as Bidirectional Long Short-Term Memory Network (BiLSTM) offers better predictions for sequential data because of additional training by traversing the input data twice (i.e., left-to-right, and right-to-left) [63].

The study [14] experimented with SVM, RF, DT, Long Short-Term Memory Network (LSTM) and Gated Recurrent Unit (GRU) sequential deep learning models using the dataset 'ISOT'. The results of the study show that LSTM and GRU models performed better than machine learning models SVM, RF and Decision Trees for performance evaluation done *for* accuracy, recall, precision, and f-score. An accuracy of 98.9% was achieved by the model compared to 77.5% by Random Forest, 76.3% by Decision Trees and 78.1% by SVM. This showed that deep learning methods outperform traditional machine learning models.

The studies by [12, 64] discuss hate speech detection. [64] utilise Bidirectional Auto-Regressive Transformers (BART) large model and few-shot prompting using Generative Pre-trained Transformer- 3 (ChatGPT-3) to classify misogyny in code-mixed Hindi-English (Hinglish). This study noted better results with the zero-shot learning approach. In the study, [10] used the code-switched dataset 'IPM' with LSTM and CNN models for multi-modal data classification. The performance metrics used are precision, recall and f-1 score. This study notes that deep learning techniques, with an accuracy of 82% compared to Random Forest's 72%, perform better. Additionally, the use of a combination of word embeddings is noted to have improved precision to up to 84%.

3. Dataset

As explained earlier, the development of models for misinformation detection relies on the availability of annotated datasets to train such models. For this study, we used the PolitiKweli dataset[2] [9]. The PolitiKweli dataset is the first Swahili-English code-switched dataset for misinformation classification curated from texts on the electioneering process posted during the 2022 general elections in Kenya on the social media platform X (formerly Twitter)[3].

The dataset curation process involved the data collection and pre-processing steps:

- Collection of textual data using Twitter API (now X API v2).[4]
- Manual language identification to eliminate any non-Swahili-English code-switched sentences.
- Data cleaning to remove non-alphanumeric characters, obscenities and non-textual data.

[2] This dataset is publicly available at https://zenodo.org/records/10228602
[3] https://twitter.com/
[4] https://developer.x.com/en/docs/x-api

- Data anonymisation to remove all personal identifiers of individuals to protect their identities.
- Data annotation into three categories: fake, fact or neutral by teams of human annotators.
- Data normalisation by removal of punctuation marks and stop-words, lower-casing and removal of missing values.

3.1. Dataset Analysis

The dataset has a total of 29,510 texts in code-switched Swahili-English (swa-en), Kenyan English (en-ke) and Swahili (swa) labelled as fake, fact or neutral. The code-switched Swahili-English (swa-en) dataset, which was the main focus of the study, has 6,345 texts whose label distribution is as shown in Table 4. The M-index [65] of the dataset is 0.48 indicating moderate code-switching (an almost equal number of Swahili and English words in each sentence). This points to a highly diverse dataset.

Table 4. Distribution of texts

Label	Number of texts
Fake	1,030
Neutral	4,064
Fact	1,221

3.2. Data Sampling

There was a substantial class imbalance in the distribution of text in the fake, neutral and fact classes which could skew classification [66]. The majority class (neutral) had more data points compared to the minority classes (fake and neutral), requiring the use of resampling techniques to avoid the model classifying instances of the minority class in the majority class [67]. According to the review of sampling techniques for imbalanced datasets by [68], standard classification algorithms' performance is negatively affected by imbalanced datasets necessitating the use of sampling procedures. These algorithms assume that the number of samples is equal, with equal cost of misclassification during training [69]. The use of hybrid techniques, unlike single techniques such as Condensed Nearest Neighbours (CNN) and Neighborhood Cleaning Rule (NC) [68] boosts the quality of the resulting dataset, increasing model performance [66]. This is because hybrid techniques achieve optimal data balance by minimizing the loss of data that often represent important features while ensuring proper generalization of the model.

We used a hybrid of Synthetic Minority Oversampling Technique (SMOTE) algorithm to oversample the minority class and Random Under-sampling (RUS) techniques to undersample the majority class [69]. RUS combined with SMOTE ensures valuable information from the minority class is retained. This hybrid technique effectively captures nuances in the data without over-generating synthetic data points.

4. Methodology

We used an experimental research design in this study. This section explains the methodology we used, starting from dataset preparation, feature extraction and model development as depicted in Figure 1.

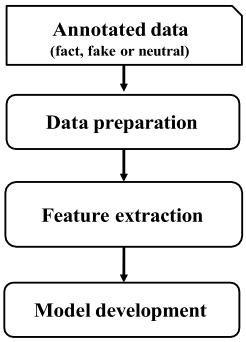

Fig.1. Experimental research design

4.1. Dataset Preparation

The textual data in the dataset was already pre-processed by [9] where any unusable texts or characters were removed and anonymised to protect the identity of users mentioned. For the data preparation step, we removed stopwords, split the dataset and tokenized the text.

A. Removal of Stopwords

We removed English stopwords using the Python Gensim library[5] version 4.3.2 and Swahili stopwords using the Python NoelNLP[6] package. Stopwords are frequently occurring words in a corpus that might only be included for syntactic purposes [70]. We removed English words like 'the', 'in', 'they', 'why' and Swahili words like 'hii' (this), 'na' (and), 'ni' (it is), among others. For Swahili, we relied on the Swahili stopword list by [71] and the Swahili stopwords for NLP generated using the Python NoelNLP package.

B. Data Tokenization

The code-switched Swahili-English tweets were already in phrases and sentences. We used the Python Gensim library to split the sentences and phrases into sub-words. This word-level tokenization process generated sequences of sub- words tokens from the sentences. A sentence such as 'lazima citizens kuwa responsible piga kura' (citizens need to be responsible vote) was broken down to ['lazima', 'citizens', 'kuwa', 'responsible', 'piga', 'kura']. The tokenization process resulted in a total of 89,290 tokens from the phrases or sentences. This tokenization process is necessary as the sub-word tokens (morphemes or groups of morphemes) allow language models to build stronger representations with less data. The representations of complex words can be computed from the representations of the sub-words [72].

4.2. Feature Extraction

In order to convert the tokens to meaningful vectors, we used a concatenation of two embeddings: GloVe [13] and FastText [15]. GloVe contains pre-trained embeddings trained on large sets of Common Crawl, Wikipedia and Twitter data. This study used the 100-dimensional GloVe vectors trained on 2 billion tweets. FastText contains word vectors pre-trained on Webcrawl and Wikipedia data. We experimented with two versions of FastText: continuous bag of words (CBoW) and skip-gram based. The CBoW version generated more accurate vector representations of words. For a morphologically rich language such as Swahili, FastText's ability to break down individual words to respective n-grams resulted in better representation of the words in the vector space.

The combined data frame had a vocabulary size (number of unique tokens) of 18,083. This large vocabulary size included the number of out-of-vocabulary words such as misspellings, abbreviations, slang which is commonly used in social media messaging and hybrid words resulting from Swahili-English conjugation. Some of the words include 'msee' (person)- slang, tenje (phone) - slang and uda (abbreviation of a popular political party's name).

4.3. Model Development

We implemented the model using Keras framework with a Tensorflow backend on Google Colaboratory (Colab) with T4 GPU. For model evaluation, we used the confusion matrix for assessing multi-class classification tasks [73] such as this study. We used the confusion matrix's standard metrics of accuracy, precision, recall and f-score for model evaluation. In addition, we set up experiments using classical machine learning algorithm Logistic Regression (LR), Artificial Neural Network (ANN) algorithm Multilayer Perceptron (MLP) and the transformer-based Bidirectional Encoder Representations from Transformers (BERT) as baselines. The model design is explained in Section 5 while the experimental results are discussed in Section 6.

5. Model Design

The model design had four components: the input (embedding layer), regularisation, BiLSTM and output (dense) layers.

Embedding layer: The embedding layer contained dense vectors generated by using a concatenation of two pre-trained embeddings generated from the feature extraction step (see Section 4.2). These dense vectors captured the semantic and syntactic meanings of the code-switched texts in the dataset for accurate text classification.

Regularisation layers: We implemented a dropout layer of size 0.4 to control weights during feature selection and experimented with both L1 (Lasso) and L2 (Ridge) regularisation techniques during training. Due to the imbalanced nature of the dataset, regularisation of weights was essential in preventing overfitting. Regularisation layers also increased the generalisability of the model, ensuring that the model performed well with unseen data.

BiLSTM layers: A BiLSTM model is a Bidirectional Recurrent Neural Network (RNN) [74] with three gates: forget, input and output. The forget gate determines whether to forget or keep information in the previous state by looking at the values of input xt and hidden state ht−1and output value while the input gate decides how much information in the input text and hidden state should pass through to update the cell state [75].

BiLSTM consists of two LSTM layers, forward and backward therefore representing the full context of the input data.

The forward LSTM can be represented as:

$$h_t^{forward} = LSTM^{forward}(x_t, h_{t-1}^{forward}) \quad (1)$$

[5] https://pypi.org/project/gensim/
[6] https://pypi.org/project/NoelNLP/

The backwards LSTM can be represented as:

$$h_t^{backward} = LSTM^{backward}(x_t, h_{t+1}^{backward}) \quad (2)$$

A combination of the two reads input reviews in both directions, forward and backwards and can represented as:

$$h_t = (h_t^{forward}, h_t^{backword}) \quad (3)$$

The BiLSTM layer has two aspects that determine a model's complexity: the number of units (hidden states) and the number of layers. We used three BiLSTM layers, each sized 64 to balance the model's ability to learn dependencies from the dataset while avoiding overfitting and computational implications of complex models. Given the complex nature of the Swahili-English code-switched dataset that was used, we experimented with one layer initially and then gradually increased the layers to three where the highest value of accuracy was achieved This was to ensure that the model learnt representations of the phrases optimally.

Output layer: The output layer consisted of two dense layers for both feature extraction and classification. The first dense layer of size 64 aggregated features from the BiLSTM layers while the second dense layer of size 3 and Softmax activation function served as the misinformation classification (output) layer for the three classes: fake, fact and neutral.

The model architecture is represented in Figure 2. In Figure 2, the labelled data in the dataset is embedded to convert the words to dense vectors. The three BiLSTM layers capture dependencies and contextual information from the word representations, while the dense layers classified text in the three classes defined in the labels (fake, fact or neutral).

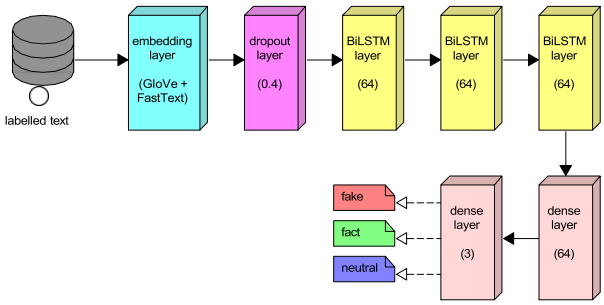

Fig.2. Model architecture

6. Experimentation and Results

We set up a number of experiments: to test the efficiency of the word embeddings, to demonstrate initial and final performance of the model before and after fine-tuning and to set a baseline for model evaluation with traditional Machine Learning models. This section first discusses the different experiments, then the model performance across the different experiments, and later discusses qualitative error analysis.

6.1. Experiments

A. Efficiency of Word Embeddings

In order to establish a baseline with different embeddings, we set up two experiments with individual embedding models (GloVe only and FastText only) to test the effectiveness of using an ensemble of word embeddings compared to using only one embedding. The results are shown in Table 5.

Table 5. Model Performance: GloVe (G), BiLSTM (B) and FastText (F)

Model	Accuracy	Precision	Recall	F-score
G + B	0.63	0.81	0.97	0.88
F + B	0.64	0.80	0.97	0.88
G + F + B	0.66	0.82	0.99	0.89

The experiments done using GloVe and FastText embeddings independently draw a comparison to their effectiveness in the generation of meaningful vectors to low-resource languages. GloVe vectors and FastText vectors both used with a BiLSTM model resulted in an F-Score of 0.88. When the vectors are combined by concatenation, the F-score slightly improved to 0.89 showing the usefulness of using an ensemble of word embeddings compared to a single one.

B. Results before and after Fine-tuning

Afterwards, we used the combined embeddings and a BiLSTM model with default configurations of a fixed embedding layer, one BiLSTM layer (size 32), one dense layer (size 3) as the output layer, learning rate of 0.0001, no regularisation layer, Adam's optimizer and sparse categorical cross-entropy. This basic configuration helps the model learn slowly and steadily, making it less likely to overlook important patterns. The initial configuration resulted in an accuracy of 0.65.

We then fine-tuned the model to achieve optimal performance. The final dimensions were based on gradual experimentation. The final model had two BiLSTM layers (size 64), two dense layers (size 64 and 3), learning rate of 0.0004, L2 regularisation layers on every layer to prevent overfitting and Adam's optimizer with a Softmax activation function for the multi-class classification task. The additional BiLSTM layer helped the model recognize more detailed patterns in the data thus classify the data accurately while the added layer size boosted the model's feature selection capability. The first dense layer (with 64 units) processed the features better before the final classification, helping the model differentiate between classes while the second dense layer acted as the output layer for the three classes. By increasing the learning rate gradually from 0.0001 to 0.0004, the model was able improve performance. The L2 regulariser helps prevent overfitting, enabling the model to generalise unseen data better.

The model was trained over 20 epochs with early stopping. The results of model performance after fine-tuning is as shown in Table 5. The fine-tuning process marginally improved accuracy to 0.66. Despite the simplicity of the model and the fewer data points due to the size of the dataset available, the model was able to learn the data features and classify the data in the three classes: fake, fact or neutral.

C. Baseline Models

We evaluated the model in comparison to other traditional machine learning models such as Logistic Regression (LR), Multilayer Perceptron (MLP), transformer-based monolingual Bidirectional Encoder Representations from Transformers (BERT) and Text-to-Text Transfer Transformer (T5). The results of the model performance against the baseline models are shown in Table 6.

Table 6. Comparison of model performance

Model	Accuracy	Precision	Recall	F-score
MLP	0.62	0.60	0.62	0.61
LR	0.66	0.62	0.67	0.58
BERT	0.53	0.52	0.76	0.62
T5	0.80	0.64	0.79	0.71
BiLSTM	0.66	0.82	0.99	0.89

- Logistic Regression (LR): LR is a statistical model that can be modified for multi-class classification tasks using the 'multinomial' strategy. It yields less discriminative and generalisable models in label classification [76] and is simpler to implement. As a baseline model, LR provides a reference base for more complex deep learning models. The LR model used in this experiment had a layer size of 100, L2 regulariser, multinomial class and was trained in 100 epochs until convergence.
- Multilayer Perceptron (MLP): MLP is a class of feed-forward artificial neural networks that consists of fully connected hundreds of thousands of neurons that can be used for NLP [77]. For the baseline model in this study, MLP offers a middle-ground comparison, performance-wise, between a simple LR to a more complex RNN such as the BiLSTM. The MLP model's embedding layer generates meaningful vectors to represent the textual data. Additionally, the multiple hidden layers capture complex patterns in the data for misinformation classification. The configuration of the model used by this study was an input layer consisting of TF-IDF vectors, hidden layer size 100, L2 regulariser to prevent over-fitting and rectified linear unit 'relu' activation function. The model was trained over 300 epochs until convergence.
- Bidirectional Encoder Representations from Transformers (BERT): BERT model [17] is a state-of-the-art transformer- based model that is pre-trained on Wikipedia data. It has a self-attention mechanism that enables it to observe relationships between words making it appropriate for a semantically complex dataset such as the one in this study. The BERT model used by this study had one dense layer, 0.1 dropout, Adam's optimizer, categorical cross-entropy, trained in 5 epochs initially, then 10 epochs with early stopping [9].
- Text-to-Text Transfer Transformer (T5): T5 [18] is a full encoder-decoder transformer-based model pre-trained on The Colossal Clean Crawled Corpus, which is a cleaned version of Common Crawl's text. As this was a text-based task, the T5 model's 'text-to-text' approach in input/output generation, generalisability and pre-training on natural language corpora made it a viable baseline for model evaluation. For experimentation, we used the 't5-small' version of the T5 model, and trained batch sizes of 8 in 10 epochs at a learning rate of 0.00002.

6.2. Model Evaluation

The model performance comparison is plotted in the histogram in Figure 3. As shown in Figure 3, the BiLSTM-based model shows improved performance compared to the baseline models (LR, MLP, BERT and T5). This can be attributed to the feature extraction process, which utilised an ensemble of word embeddings pre-trained on social media data. The BiLSTM model was also fine-tuned to achieve optimal values of accuracy, precision, recall and f-score. From the scores, it is evident that for a code-switched dataset with a morphologically rich language such as Swahili, accurate representation of out-of-vocabulary words through the use of pre-trained embeddings improves the model classification accuracy.

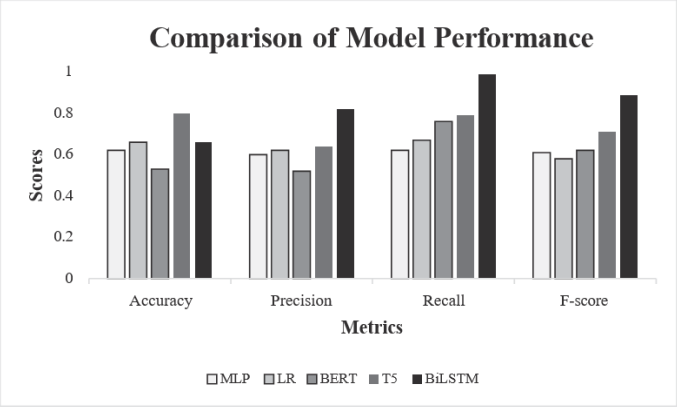

Fig.3. Comparison of model performance

This work can be compared to previous studies by [10] and [16] which experimented with an ensemble of word embeddings and a code-switched dataset with deep learning models, respectively. [16] achieved an f-sore of 0.81 using a combination of GloVe and FastText embeddings with an English dataset while [10] achieved an f-score of 0.76 using a code-switched Hindi-English hate speech classification dataset and ensemble of word embeddings with deep learning models. The model developed by this study performed at per with notable improvements in performance. Given the complexity of the model architecture and the diversity of the dictionary for a code-switched dataset, the model presented in this study performs comparatively better.

6.3. Error Analysis

We conducted an error analysis by examining the false positives (FP) (incorrectly predicted as misinformation) and false negatives (FN) (incorrectly predicted as not misinformation). The number of FP and FN cases determines the classification model's ability to make predictions. We went further to interrogate possible reasons for incorrect predictions and identify patterns in the errors with the aim of offering insights into improving the model's performance. Table 7 shows sample text with its correct label against the model's prediction.

As shown in Table 7, we randomly selected 10 sentences to ensure a non-biased sample of the model's predictions for analysis. The model was able to accurately classify 6 out of the 10 classes correctly indicating moderate accuracy. While the model demonstrates good performance overall, especially in classifying neutral texts, there are specific areas where the model struggles. The model accurately classifies data as neutral due to the larger number of training data labelled as neutral compared to the rest of the classes. Despite the resampling techniques employed to counter model bias towards the majority class, neutral, the model still showed elements of bias towards predicting neutral labels, making it less effective at distinguishing between fact and misinformation.

Table 7. Error analysis

S.No.	Sample text	True label	Predicted label
1	winner including nyinyi wote ni wetu na tunawapenda	neutral	neutral
2	independent mps joined kenya kwanza government coalition makofi	fact	neutral
3	user remember thou promises kazi ni kazi	neutral	neutral
4	vote peacefully usisahau yeye ni jirani yako	fact	fact
5	ukiwa unregistered allowed vote any polling station uchaguzi	fake	fake
6	primary data iko iebc portal media tallying determine winner iebc	fact	neutral
7	sasa mbona ulikubali kufuata losing team	neutral	neutral
8	siasa ni kujipanga friend user	neutral	fact
9	iebc summons user mp wednesday vote rigging claims	fact	fact
10	hakuna mtu kirinyaga who voted baba	fake	neutral

For instance, examples 2 and 6 are labelled as fact, yet the model classifies them as neutral. This is likely because the content of these texts contains neutral language, even though the underlying information is factual. This suggests that the model struggles to detect more subtle factual cues and relies heavily on surface-level patterns, such as tone or phrasing. Moreover, in example 8, the model incorrectly predicts a neutral text as fact. This highlights the difficulty the model has in distinguishing between fact and neutral, suggesting that there is a fine line between these two categories in the current dataset. The close similarity between neutral and factual statements can be attributed to the overlap in the linguistic features of both classes, making it harder for the model to make accurate distinctions.

Such issues arising from an originally imbalanced dataset such as the one used in this study can be resolved by developing large-scale, properly balanced datasets to be able to expose the model to varied classes of data and prevent over- fitting or under-fitting. This can help the model learn more robust patterns. The use of more advanced model architectures with more accurate vector representation of words in embeddings may also reduce error rates of this misinformation classification model.

7. Ethical Considerations

The curation process of the dataset used in our study followed the strict data privacy policy by the collection platform X (formerly Twitter)[7]. The data was stored securely, access restricted until authorisation and was only shared according to the terms of the policy. This ensured the safety of the data against unethical access and manipulation. Moreover, the data was thoroughly anonymised to protect the identities of both users whose posts were collected and usernames mentioned by posts.

8. Conclusions and Future Work

This study set out to achieve two major tasks. First, systematically review code-switching, highlighting existing datasets and models developed for classification in low-resource languages. Second, develop a Swahili-English code-switched misinformation classification model. In the literature review, we discuss some of the advancements in the creation of language technologies (datasets and models) for low-resource languages for classification tasks including misinformation classification, hate speech detection and sentiment analysis. This systematic review provides the background to the unique problem of inadequate datasets and accurate models to classify text, in this case, misinformation in code-switched Swahili-English, which is considered low-resource. Code-switching is identified as a major setback in classification tasks as it introduces both syntactic and semantic complexities to text, limiting the accuracy of models trained on monolingual data to classify code-switched text.

Through experimentation, we demonstrated that using an ensemble of word embeddings, in this case, GloVe and FastText, generates a more accurate vector representation of text resulting in better accuracies compared to a single embedding. With an f-score of 0.89, it is evident that the dataset is valuable in this misinformation classification task. The misinformation classification model resulting from this study can be applied to label text on social media of political nature as fake, fact or neutral. Automated flagging of fake texts on social media helps with reporting and removal of these texts to mitigate the spread of misinformation. We, however, recognise the error rate of the classification model, which might lead to the mislabelling of text towards the neutral class. Continuous training with larger, high-quality code-switched datasets will improve the model performance substantially. Moreover, the creation and use of embeddings models pre-trained on large datasets in code-switched Swahili-English will enable the model capture nuances in the data more accurately.

Further work involving the use of both foundation models and fine-tuned Large Language Models (LLMs) is expected to show improved model performance. LLMs are trained on big data and may represent a broader view of the world. We also will look at zero-shot learning (ZSL) in training of future models. Additionally, we will look into experimentation with larger, high quality datasets to increase generalisability of the model.

Acknowledgements

The dataset used in this study was a product of our previous research work supported by the Department of Computer Science of Maseno University, School of Computing and Informatics. We thank the faculty members and students for their contribution in improving the quality of this study and advancing research in under-resourced languages.

References

[1] Sacha Altay, Manon Berriche, Hendrik Heuer, Johan Farkas, and Steven Rathje. A survey of expert views on mis- information: Definitions, determinants, solutions, and future of the field. Harvard Kennedy School Misinformation Review, 4(4):1–34, 2023.

[7] https://developer.x.com/en/developer-terms/agreement-and-policy

[2] Mozilla. New research: In kenya, disinformation campaigns seek to discredit pandora papers. https://foundation.mozilla.org/en/blog/new-research-in-kenya-disinformation-campaigns-seek-to-discredit- pandora-papers/, 2021.
[3] Jawaher Alghamdi, Yuqing Lin, and Suhuai Luo. Fake news detection in low-resource languages: A novel hybrid summarization approach. Knowledge-Based Systems, page 111884, 2024.
[4] Barack Wanjawa, Lilian Wanzare, Florence Indede, Owen McOnyango, Edward Ombui, and Lawrence Muchemi. Kencorpus: A kenyan language corpus of swahili, dholuo and luhya for natural language processing tasks. arXiv preprint arXiv:2208.12081, 2022.
[5] Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbo- hungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, et al. Participatory research for low-resourced machine translation: A case study in african languages. arXiv preprint arXiv:2010.02353, 2020.
[6] Michael A Hedderich, Lukas Lange, Heike Adel, Jannik Stro¨tgen, and Dietrich Klakow. A survey on recent ap- proaches for natural language processing in low-resource scenarios. arXiv preprint arXiv:2010.12309, 2020.
[7] Lusiana Kartika Candra and Laila Ulsi Qodriani. An analysis of code switching in leila s. chudori's for nadira. Teknosastik, 16(1):9–14, 2019.
[8] Genta Indra Winata. Multilingual transfer learning for code-switched language and speech neural modeling. Hong Kong University of Science and Technology (Hong Kong), 2021.
[9] Cynthia Amol, Lilian Wanzare, and James Obuhuma. Politikweli: A swahili-english code-switched twitter political misinformation classification dataset. In Speech and Language Technologies for Low-Resource Languages, pages 3–17, Cham, 2024. Springer Nature Switzerland.
[10] Kshitij Rajput, Raghav Kapoor, Kaushal Rai, and Preeti Kaur. Hate me not: detecting hate inducing memes in code switched languages. arXiv preprint arXiv:2204.11356, 2022.
[11] Peter Nabende, David Kabiito, Claire Babirye, Hewitt Tusiime, and Joyce Nakatumba-Nabende. Misinformation detection in luganda-english code-mixed social media text. arXiv preprint arXiv:2104.00124, 2021.
[12] Edward Ombui, Lawrence Muchemi, and Peter Wagacha. Hate speech detection in code-switched text messages. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pages 1–6. IEEE, 2019.
[13] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532– 1543, 2014.
[14] Eslam Amer, Kyung-Sup Kwak, and Shaker El-Sappagh. Context-based fake news detection model relying on deep learning models. Electronics, 11(8):1255, 2022.
[15] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. Transactions of the association for computational linguistics, 5:135–146, 2017.
[16] Nabil Badri, Ferihane Kboubi, and Anja Habacha Chaibi. Combining fasttext and glove word embedding for offen- sive and hate speech text detection. Procedia Computer Science, 207:769–778, 2022.
[17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
[18] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140):1–67, 2020.
[19] Flora Sakketou, Joan Plepi, Riccardo Cervero, Henri-Jacques Geiss, Paolo Rosso, and Lucie Flek. Factoid: A new dataset for identifying misinformation spreaders and political bias. arXiv preprint arXiv:2205.06181, 2022.
[20] Dimitrios Michail, Nikos Kanakaris, and Iraklis Varlamis. Detection of fake news campaigns using graph convolu- tional networks. International Journal of Information Management Data Insights, 2(2):100104, 2022.
[21] Marcos Zampieri, Preslav Nakov, Sara Rosenthal, Pepa Atanasova, Georgi Karadzhov, Hamdy Mubarak, Leon Derczynski, Zeses Pitenis, and C¸ ag˘rı C¸ o¨ltekin. Semeval-2020 task 12: Multilingual offensive language identification in social media (offenseval 2020). arXiv preprint arXiv:2006.07235, 2020.
[22] Deepti Jain, Sandhya Arora, CK Jha, and Garima Malik. Transformer-based models for hate speech classification. In AIP Conference Proceedings, volume 3072. AIP Publishing, 2024.
[23] Hind Saleh, Areej Alhothali, and Kawthar Moria. Detection of hate speech using bert and hate speech word embed- ding with deep model. Applied Artificial Intelligence, 37(1):2166719, 2023.
[24] David Ifeoluwa Adelani, Hannah Liu, Xiaoyu Shen, Nikita Vassilyev, Jesujoba O Alabi, Yanke Mao, Haonan Gao, and Annie En-Shiun Lee. Sib-200: A simple, inclusive, and big evaluation dataset for topic classification in 200+ languages and dialects. arXiv preprint arXiv:2309.07445, 2023.
[25] Odunayo Ogundepo, Tajuddeen R Gwadabe, Clara E Rivera, Jonathan H Clark, Sebastian Ruder, David Ifeoluwa Adelani, Bonaventure FP Dossou, Abdou Aziz DIOP, Claytone Sikasote, Gilles Hacheme, et al. Afriqa: Cross- lingual open-retrieval question answering for african languages. arXiv preprint arXiv:2305.06897, 2023.
[26] David Ifeoluwa Adelani, Marek Masiak, Israel Abebe Azime, Jesujoba Oluwadara Alabi, Atnafu Lambebo Tonja, Christine Mwase, Odunayo Ogundepo, Bonaventure FP Dossou, Akintunde Oladipo, Doreen Nixdorf, et al. Masakhanews: News topic classification for african languages. arXiv preprint arXiv:2304.09972, 2023.
[27] Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif M Mohammad, Sebastian Ruder, et al. Afrisenti: A twitter sentiment analysis benchmark for african languages. arXiv preprint arXiv:2302.08956, 2023.
[28] Colin Leong, Joshua Nemecek, Jacob Mansdorfer, Anna Filighera, Abraham Owodunni, and Daniel White- nack. Bloom library: Multimodal datasets in 300+ languages for a variety of downstream tasks. arXiv preprint arXiv:2210.14712, 2022.
[29] Marta R Costa-jussa`, James Cross, Onur C¸ elebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. No language left behind: Scaling human-centered machine transla- tion. arXiv preprint arXiv:2207.04672, 2022.

[30] Odunayo Ogundepo, Xinyu Zhang, Shuo Sun, Kevin Duh, and Jimmy Lin. Africlirmatrix: Enabling cross-lingual information retrieval for african languages. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 8721–8728, 2022.
[31] Alexandros Zervopoulos, Aikaterini Georgia Alvanou, Konstantinos Bezas, Asterios Papamichail, Manolis Maragoudakis, and Katia Kermanidis. Deep learning for fake news detection on twitter regarding the 2019 hong kong protests. Neural Computing and Applications, 34(2):969–982, 2022.
[32] Nishant Rai, Deepika Kumar, Naman Kaushik, Chandan Raj, and Ahad Ali. Fake news classification using trans- former based enhanced lstm and bert. International Journal of Cognitive Computing in Engineering, 3:98–105, 2022.
[33] Kuai Xu, Feng Wang, Haiyan Wang, and Bo Yang. Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Science and Technology, 25(1):20–27, 2019.
[34] Lu Yuan, Hangshun Jiang, Hao Shen, Lei Shi, and Nanchang Cheng. Sustainable development of information dissemination: A review of current fake news detection research and practice. Systems, 11(9):458, 2023.
[35] Sakshini Hangloo and Bhavna Arora. Fake news detection tools and methods–a review. arXiv preprint arXiv:2112.11185, 2021.
[36] Jon Roozenbeek, Eileen Culloty, and Jane Suiter. Countering misinformation. European Psychologist, 2023.
[37] Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5):1–40, 2020.
[38] Jennifer Jerit and Yangzi Zhao. Political misinformation. Annual Review of Political Science, 23:77–94, 2020.
[39] Kelvin Jiang, Ronak Pradeep, and Jimmy Lin. Exploring listwise evidence reasoning with t5 for fact verification. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 402–410, 2021.
[40] Kai Shu, Guoqing Zheng, Yichuan Li, Subhabrata Mukherjee, Ahmed Hassan Awadallah, Scott Ruston, and Huan Liu. Early detection of fake news with multi-source weak social supervision. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III, pages 650–666. Springer, 2021.
[41] Deptii Chaudhari and Ambika Vishal Pawar. Empowering propaganda detection in resource-restraint languages: a transformer-based framework for classifying hindi news articles. Big Data and Cognitive Computing, 7(4):175, 2023.
[42] Kelly Stahl. Fake news detection in social media. California State University Stanislaus, 6:4–15, 2018.
[43] Lynete Lusike Mukhongo. Participatory media cultures: virality, humour, and online political contestations in kenya. Africa Spectrum, 55(2):148–169, 2020.
[44] Andreu Casero-Ripolle´s. Influencers in the political conversation on twitter: Identifying digital authority with big data. Sustainability, 13(5):2851, 2021.
[45] Sunayana Sitaram, Khyathi Raghavi Chandu, Sai Krishna Rallabandi, and Alan W Black. A survey of code-switched speech and language processing. arXiv preprint arXiv:1904.00784, 2019.
[46] Wilkinson Daniel Wong Gonzales and Yuen Man Tsang. The sociolinguistics of code-switching in hong kong's digital landscape: A mixed-methods exploration of cantonese-english alternation patterns on whatsapp. Journal of English and Applied Linguistics, 2(1):2, 2023.
[47] Navya Jose, Bharathi Raja Chakravarthi, Shardul Suryawanshi, Elizabeth Sherly, and John P McCrae. A survey of current datasets for code-switching research. In 2020 6th international conference on advanced computing and communication systems (ICACCS), pages 136–141. IEEE, 2020.
[48] Xinyi Zhong, Lay Hoon Ang, and Sharon Sharmini. Systematic literature review of conversational code-switching in multilingual society from a sociolinguistic perspective. Theory and Practice in Language Studies, 13(2):318–330, 2023.
[49] Shruti Rijhwani, Royal Sequiera, Monojit Choudhury, Kalika Bali, and Chandra Shekhar Maddila. Estimating code- switching on twitter with a novel generalized word-level language detection technique. In Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: long papers), pages 1971–1982, 2017.
[50] Charlotte Hoffmann. Introduction to bilingualism. Routledge, 2014.
[51] Maya Sari, Adip Arifin, and Ratri Harida. Code-switching and code-mixing used by guest star in hotman paris show. Journal of English Language Learning, 5(2), 2021.
[52] Janet Holmes and Nick Wilson. An introduction to sociolinguistics. Routledge, 2022.
[53] Hans Ko¨chler. Idea and politics of communication in the global age. Digital Transformation in Journalism and News Media: Media Management, Media Convergence and Globalization, pages 7–15, 2017.
[54] Comfort Eseohen Ilevbare, Jesujoba O Alabi, David Ifeoluwa Adelani, Firdous Damilola Bakare, Oluwatoyin Bunmi Abiola, and Oluwaseyi Adesina Adeyemo. Ekohate: Abusive language and hate speech detection for code-switched political discussions on nigerian twitter. arXiv preprint arXiv:2404.18180, 2024.
[55] Muhammad Umar Salman, Asif Hanif, Shady Shehata, and Preslav Nakov. Detecting propaganda techniques in code-switched social media text. arXiv preprint arXiv:2305.14534, 2023.
[56] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data, 8(3):171–188, 2020.
[57] Vivek Srivastava and Mayank Singh. Challenges and considerations with code-mixed nlp for multilingual societies. arXiv preprint arXiv:2106.07823, 2021.
[58] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
[59] Matthew E Peters, Mark Neumann, Luke Zettlemoyer, and Wen-tau Yih. Dissecting contextual word embeddings: Architecture and representation. arXiv preprint arXiv:1808.08949, 2018.
[60] Wazir Ali, Jay Kumar, Junyu Lu, and Zenglin Xu. Word embedding based new corpus for low-resourced language: Sindhi. arXiv preprint arXiv:1911.12579, 2019.
[61] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. ACM computing surveys (CSUR), 54(10s):1–41, 2022.

[62] Qian Li, Hao Peng, Jianxin Li, Congying Xia, Renyu Yang, Lichao Sun, Philip S Yu, and Lifang He. A survey on text classification: From traditional to deep learning. ACM Transactions on Intelligent Systems and Technology (TIST), 13(2):1–41, 2022.
[63] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. The performance of lstm and bilstm in forecasting time series. In 2019 IEEE International conference on big data (Big Data), pages 3285–3292. IEEE, 2019.
[64] Haneul Yoo, Yongjin Yang, and Hwaran Lee. Csrt: Evaluation and analysis of llms using code-switching red-teaming dataset. arXiv preprint arXiv:2406.15481, 2024.
[65] Ruthanna Barnett, Eva Codo´, Eva Eppler, Montse Forcadell, Penelope Gardner-Chloros, Roeland Van Hout, Melissa Moyer, Maria Carme Torras, Maria Teresa Turell, Mark Sebba, et al. The lides coding manual: A document for preparing and analyzing language interaction data version 1.1—july, 1999. International Journal of Bilingualism, 4(2):131–132, 2000.
[66] Carla Vairetti, Jose´ Luis Assadi, and Sebastia´n Maldonado. Efficient hybrid oversampling and intelligent undersam- pling for imbalanced big data classification. Expert Systems with Applications, 246:123149, 2024.
[67] Tarid Wongvorachan, Surina He, and Okan Bulut. A comparison of undersampling, oversampling, and smote meth- ods for dealing with imbalanced classification in educational data mining. Information, 14(1):54, 2023.
[68] Asif Newaz and Farhan Shahriyar Haq. A novel hybrid sampling framework for imbalanced learning. arXiv preprint arXiv:2208.09619, 2022.
[69] Zhaozhao Xu, Derong Shen, Tiezheng Nie, and Yue Kou. A hybrid sampling algorithm combining m-smote and enn based on random forest for medical imbalanced data. Journal of Biomedical Informatics, 107:103465, 2020.
[70] J Sangeetha and U Kumaran. A hybrid optimization algorithm using bilstm structure for sentiment analysis. Mea- surement: Sensors, 25:100619, 2023.
[71] Bernard Masua and Noel Masasi. Enhancing text pre-processing for swahili language: Datasets for common swahili stop-words, slangs and typos with equivalent proper words. Data in Brief, 33:106517, 2020.
[72] Edward Gow-Smith, Harish Tayyar Madabushi, Carolina Scarton, and Aline Villavicencio. Improving tokenisation by alternative treatment of spaces. arXiv preprint arXiv:2204.04058, 2022.
[73] Mohammadreza Heydarian, Thomas E Doyle, and Reza Samavi. Mlcm: Multi-label confusion matrix. IEEE Access, 10:19083–19095, 2022.
[74] Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. IEEE Transactions on Signal Pro- cessing, 45(11):2673–2681, 1997.
[75] Zabit Hameed and Begonya Garcia-Zapirain. Sentiment classification using a single-layered bilstm model. Ieee Access, 8:73992–74001, 2020.
[76] Qi Dong, Xiatian Zhu, and Shaogang Gong. Single-label multi-class image classification by deep logistic regression. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 3486–3493, 2019.
[77] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature, 521(7553):436–444, 2015.

Authors' Profiles

Cynthia Jayne Amol is a final year Masters in Computer Science student at Maseno University, Kenya and holds a Bachelor of Science in Computer Science (Maseno University). An AI enthusiast, she is a Natural Language Processing (NLP) researcher working on low resource languages and has contributed to dataset curation and language modelling for African languages. She is also an experienced systems and data analyst, working at Maseno University's ICT Directorate.

Dr. Lilian D. A. Wanzare is the co-founder of KenCorpus and the Chairperson of the Department of Computer Science at Maseno University. She holds a Bachelor of Science in Computer Science (University of Nairobi), Master of Science in Computer Science (Free University of Bozen-Bolzano), Master of Science in Language Technology (Universita¨t des Saarlandes) and Doctor of Philosophy in Computational Linguistics (Universita¨t des Saarlandes). Dr. Wanzare is one of the supervisors for Cynthia Amol and skilled in Natural Language Processing, Text mining, Knowledge Extraction, Knowledge acquisition, Programming (Python, Java, R) and Machine Learning. Her interests and passions lie in Natural Language Processing (NLP), Natural Language Understanding (NLU), Data Science, Machine Learning, building chatbots and personal assistants.

Dr. James Obuhuma is a Computer Science faculty, Department of Computer Science, Maseno University. He holds a PhD in Computer Science from Maseno University, an MSc in Computer Science from the University of Nairobi and a BSc in Computer Science and Technology from Maseno University. His MSc thesis focused on Road Traffic Analysis using GPS Technology which opened his interest in Intelligent Systems. This heavily influenced his PhD research topic which was in the area of Intelligent Transportation Systems (ITS) with a focus on Vehicle Driver Behaviour Modeling. His research interest is in the application of Machine Learning in different domains, including, ITS, Cybersecurity and IoT. Dr. Obuhuma is one of the MSc supervisors for Cynthia Amol, whose MSc research resulted in this publication. Dr. Obuhuma is also a passionate Design Thinking Coach who fosters for Social Innovations across the globe.

How to cite this paper: Cynthia Amol, Lilian Wanzare, James Obuhuma, "Modelling Misinformation in Swahili-English Code-switched Texts", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.1, pp.67-81, 2025. DOI:10.5815/ijitcs.2025.01.05