A comparative study of natural language inference in Swahili using monolingual and multilingual models

Hajra Faki Ali, Adila Alfa Krisnadhi
Department of Computer Science, Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Article Info

Article history:

Received Feb 15, 2024
Revised Oct 29, 2024
Accepted Nov 14, 2024

Keywords:

Monolingual model
Multilingual model
Natural language inference
Swahili
XNLI dataset

ABSTRACT

Recent advancements in large language models (LLMs) have led to opportunities for improving applications across various domains. However, existing LLMs fine-tuned for Swahili or other African languages often rely on pre-trained multilingual models, resulting in a relatively small portion of training data dedicated to Swahili. In this study, we compare the performance of monolingual and multilingual models in Swahili natural language inference tasks using the cross-lingual natural language inference (XNLI) dataset. Our research demonstrates the superior effectiveness of dedicated Swahili monolingual models, achieving an accuracy rate of 69%. These monolingual models exhibit significantly enhanced precision, recall, and F1 scores, particularly in predicting contradiction and neutrality. Overall, the findings in this article emphasize the critical importance of using monolingual models in low-resource language processing contexts, providing valuable insights for developing more efficient and tailored natural language processing systems that benefit languages facing similar resource constraints.

Corresponding Author:

Adila Alfa Krisnadhi
Department of Computer Science, Faculty of Computer Science, Universitas Indonesia
UI Campus Depok, West Java 16424, Indonesia
Email: adila@cs.ui.ac.id

1. INTRODUCTION

Swahili is a rich language widely used by over 100 million people in countries across East and Central Africa, such as Tanzania, Kenya, and Uganda [1], [2]. Its prominence leads to its presence in international media, such as the British Broadcasting Corporation (BBC) [3] and Europe Media Monitor [4]. The significant presence of Swahili motivates the recent adoption of natural language processing (NLP) technologies to provide more assistance to Swahili-speaking users. Some examples of such technologies include the Swahili chatbots of vodacare and ada health [5], [6]. However, these chatbots are pretty limited in their ability to respond to users' queries and, thus, cannot interact much with users.

The recent advancements in large language models (LLMs) provide an opportunity to improve such applications in various domains. Nonetheless, existing LLMs fine-tuned for Swahili or a collection of African languages, such as robustly optimized bidirectional encoder representations from transformers (RoBERTa)-base-wechsel-Swahili [7] and African bidirectional encoder representations from transformers (AfriBERTa) [8], often rely on pre-trained multilingual models like multilingual bidirectional encoder representations from transformers (mBERT) [9] or cross-lingual language model with recurrent neural networks (XLM-R) [10], where the pre-training data includes many languages around the world. Consequently, the Swahili portion of the training data is often tiny compared to other languages in these multilingual models [11]. For example, the portion of Swahili pre-training data for mBERT makes up less than 1% of its vocabulary [12].

On the other hand, Swahili bidirectional encoder representations from transformers (SwahBERT) [13] is a monolingual model specifically trained for Swahili. It was trained for the following NLP tasks: named entity recognition (NER), achieving 88.50% accuracy; news classification with 90.90% accuracy; emotion classification with 64.46% accuracy; and sentiment analysis with 70.94% accuracy. One important NLP task missing from these results is natural language inference (NLI).

NLI (also known as textual entailment) is the task of determining the entailment relationship between a "premise" and a "hypothesis"-whether the hypothesis is true (entailment), false (contradiction), or undetermined (neutral) given the premise [14], [15]. Entailment in NLI differs from standard logical entailment in that it does not require strict logical semantics of the sentences to obtain and derive the entailment. Instead, in NLI, the premise entails the hypothesis of whether typical human reading would justify that the entailment holds. From an application perspective, a solution to the NLI task can often be used in other, more complex NLP tasks, such as question-answering [16], document summarization [17], and information extraction [18]. In the context of this paper, several Swahili chatbot applications could be much improved if they are equipped with NLI models because currently, those chatbots are pretty simplistic in their approach, namely using string-based pattern matching and handcrafted rules to enable the applications to respond to user's queries.

In addition to motivation from application areas, the gap remains regarding whether any Swahili monolingual model could provide a better solution to the NLI task than multilingual models. It has been shown that for low-resource languages, monolingual models often outperform their multilingual counterparts [19]–[23]. This applies even to high-resource languages when specific NLP tasks are considered [24], [25]. Thus, answering this question can encourage further research and development of Swahili monolingual models.

This paper aims to perform a comparative analysis between SwahBERT [13] and mBERT [9]. The former is the only known Swahili monolingual model, while the latter is a widely used multilingual model pre-trained in 104 languages, including Swahili [26]. Our work looks at the performance of both models on the Swahili NLI task. Specifically, we fine-tune both models for the downstream NLI task on a Swahili subset of the cross-lingual natural language inference (XNLI) dataset [27]. The results fill the Swahili NLI research gap as neither model was trained for the Swahili NLI task.

The rest of the paper is organized, making up less than 1%. We follow this with an explanation of our method in section 2, where we explain the models, the dataset, and the evaluation scenarios. After that, section 3 details the evaluation results, revealing exciting insights into the performance of our models. Finally, section 4 concludes the paper, summarizing our findings and suggesting avenues for future research.

2. METHOD

In our research, we prepared our dataset and selected SwahBERT for the monolingual model and mBERT for the multilingual model. We fine-tuned both SwahBERT and mBERT for the NLI task and evaluated both models. Finally, we conducted a comparative analysis between the two models. Our general approach follows a standard workflow, as depicted in Figure 1.

2.1. Data collection and pre-processing

This study utilizes the Swahili segment of the XNLI dataset, which contains 392,702 training sets, 2,490 validations sets, and 5,010 test sets of the pairs of sentences [27]. Each sentence pair consists of a premise, hypothesis, and label. The labels indicate whether the hypothesis is entailed by the premise (0), contradicted by the premise (1), or neutral concerning the premise (2). Table 1 shows some samples from the XNLI dataset. Table 1 contains English translations in italics so the reader can understand them. While the use of the monolingual model has shown promising results in some downstream tasks in Swahili, further research is needed to fine-tune the existing monolingual model in Swahili, such as SwahBERT for NLI, to determine it is strength in NLI tasks to address the unique linguistic and cultural challenges of Swahili NLI [28]. To ensure the quality of the data and compatibility with the models, the dataset undergoes thorough pre-processing. This involves several steps, including text cleaning to remove special characters and punctuation. Additionally, BertTokenizer is applied to tokenize the input text into subwords suitable for input into a BERT-based model.

2.2. The models: SwahBERT and mBERT

The methodology employed in this study utilizes two transformer-based neural network architectures, namely SwahBERT and mBERT. SwahBERT is a model specifically fine-tuned for the Swahili NLI task, leveraging the capabilities of the BERT model. On the other hand, mBERT is a pre-trained model designed to handle multiple languages.

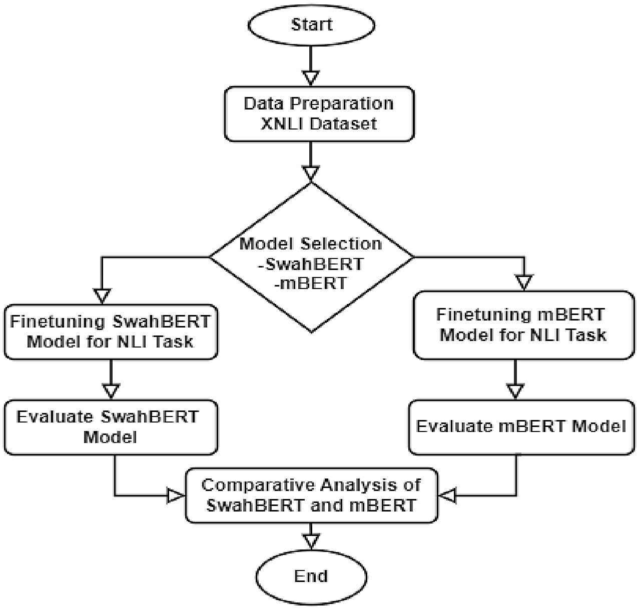


Figure 1. Experimental flow chart

Table 1. Sample data from the used dataset

Premise	Hypothesis	Label
Nilipitia seti ya milango ya kabati, na nikaanguka chini. I went through a set of cupboard doors, and I fell.	*Nilipasua milango na kuanguka chini.* I broke the doors and fell.	0
Katikati ya muunganiko huu wa ajabu wa tamaduni ni shauku ya kuendelea. A passion for progress is at the heart of this remarkable fusion of cultures.	*Shauku ya kuendelea sio muhimu zaidi ya tamaduni hizi.* This passion for progress is not the most important thing about these cultures	1
Burudani kwa watu wazima na Watoto Entertainment for adults and children.	*Furaha kwa watoto.* Fun for kids only.	2
Masuala katika usanifu wa data Issues in data architecture.	*Matatizo katika usanisi wa data* Problems in data synthesis.	0
Ona hiyo sio mbaya sana kwa masaa kadhaa See, that's not too bad for a few hours	*Sipendi kusubiri, hata kwa dakika 10.* I do not like waiting, not even for 10 minutes	1

SwahBERT is a variant of the BERT model specifically trained on a Swahili dataset comprising 105 MB with 16 M words sourced from news websites, forums, and Wikipedia. It is designed for four key tasks: news classification, named entity recognition, emotion classification, and sentiment classification. The architecture of SwahBERT includes 12 encoder blocks with 768 hidden units, and it uses self-attention mechanisms and feedforward neural networks to capture complex token relationships. During pre-training, SwahBERT utilizes masked language modeling (MLM) and next sentence prediction (NSP) to predict masked tokens and assess sentence coherence, respectively. For task-specific NLI adaptation, further training on labelled data is essential, involving parameter updates of the final classification layer to align with the target task's requirements. This fine-tuning process, commonly minimizing cross-entropy loss, optimizes parameters iteratively, including weights and biases associated with each class label, typically instantiated within the BertForSequenceClassification model.

mBERT, a versatile pre-trained model, encompasses 104 languages, employing MLM and NSP techniques. It is built upon the BERT architecture with multiple transformer encoder layers. mBERT extracts contextual information from both left and right token contexts. The mBERT-base cased model, comprising 12-layer Transformer blocks with 110 M parameters and a hidden size 768, serves as the foundation. Although inherently multilingual, it undergoes further fine-tuning for Swahili, enhancing its suitability for Swahili NLI tasks. Fine-tuning involves training the model specifically for Swahili NLI, enabling it to understand Swahili nuances better and improve task performance.

2.3. Training setup

The training process involves two transformer-based neural network architectures: SwahBERT and mBERT. These models are fine-tuned for the Swahili NLI task using the pre-processed data. The models are trained for 100 epochs with early stopping using a batch size of 32 and the Adam optimizer with a learning rate 1e-5. We use the DGX-A100 server equipped with Tesla A100 SXM4 40 GB. GPUs, each with 40 GB. of GPU memory. This setup is ideal for executing our models, which require approximately 15 hours to run.

2.4. Model evaluation

The performance of the models is evaluated using separate test datasets, employing metrics such as accuracy (computed in 1), precision (computed in 2), recall (computed in 3), and F1-score (computed in 4) across various classes in Swahili NLI. These metrics, which utilize true positive (TP), true negative (TN), false positive (FP), and false negative (FN), provide a comprehensive overview of the model's strengths and weaknesses across different NLI categories. Furthermore, macro and weighted averages are derived to summarize the overall performance, offering valuable insights into how effectively the models interpret the relationships between sentences in Swahili NLI tasks. Referring to state-of-the-art learning algorithms such as [29], [30], using a confusion matrix enhances the evaluation process by providing deeper insights into the models' performance across different classes, thereby enriching the development of NLI models for low-resource languages.

$$Accuracy = \frac{TP(entailment)+TP(contradiction)+TP(neutral)}{(Total\ samples)} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{4}$$

3. RESULTS AND DISCUSSION

The study investigated the effectiveness of a developed monolingual model for NLI tasks in Swahili. While previous research primarily focused on NLI models within multilingual frameworks, they did not explicitly address the performance of monolingual models in Swahili NLI tasks. The gap is addressed by evaluating the impact and effectiveness of monolingual models in Swahili NLI tasks compared to multilingual models. The study results will shed light on monolingual models' potential advantages and limitations compared to their multilingual counterparts for Swahili NLI tasks.

3.1. Key findings

Our study revealed that the Swahili NLI monolingual model achieved an impressive accuracy of 69%, outperforming the multilingual model's accuracy of 66%, as displayed in Table 2. This indicates a consistent alignment between the model's predictions and the actual labels in most test samples. Notably, the monolingual model exhibited enhanced precision, recall, and F1 scores, especially in predicting contradiction and neutrality, highlighting its proficiency in accurate optimistic predictions while minimizing errors. The model's consistent performance across micro and weighted averages suggests unbiased predictions unaffected by imbalanced data distribution. In contrast, the multilingual model showed unexpected strength in predicting entailment, underscoring the importance of leveraging the unique capabilities of each model. Additionally, the confusion matrix (Figure 2) shows a higher number of correctly predicted instances in all classes for the monolingual model compared to the multilingual model, as illustrated in Figure 2(a) for the monolingual (SwahBERT) model and Figure 2(b) for the multilingual (mBERT) model. These underscore the importance of tailored approaches for enhancing low-resource language processing, offering valuable insights for developing effective NLP systems for underrepresented languages like Swahili.

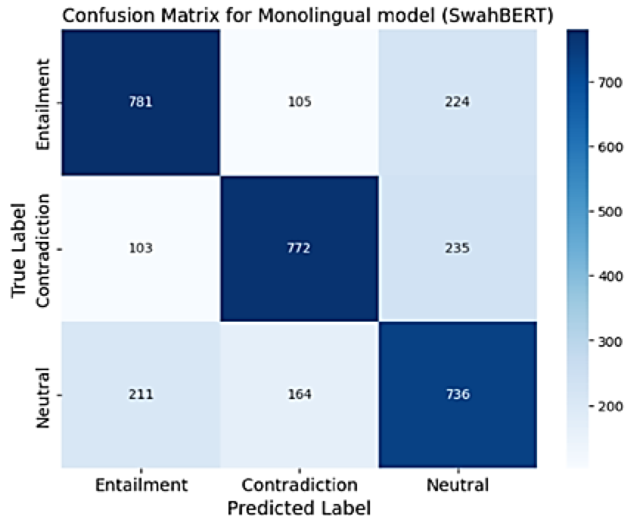
Table 2. Comparison of monolingual and multilingual models

	Monolingual model (SwahBERT)			Multilingual model (mBERT)		
Accuracy (%)	69			66		
	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-score (%)
Entailment	71	70	71	81	67	74
Contradiction	74	70	72	63	71	67
Neutral	62	66	64	57	60	58
Macro average	69	69	69	67	66	66
Weighted average	69	69	69	67	66	66

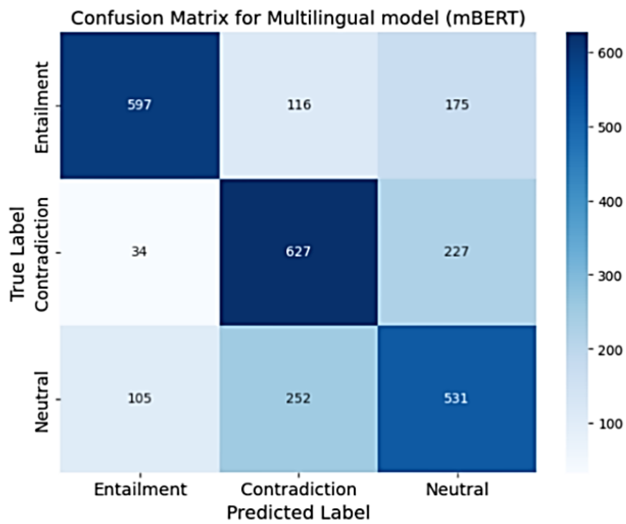
3.2. Comparative analysis and interpretation of results

Our study conducted a detailed comparative analysis between monolingual and multilingual models in Swahili NLI tasks, revealing their unique strengths and weaknesses. The monolingual model, utilizing the entire Swahili vocabulary, demonstrated a comprehensive understanding of the language, contrasting with multilingual models that allocate a smaller portion of their vocabulary to Swahili [12]. This approach potentially boosts performance in Swahili NLI tasks, as evidenced by the higher number of accurate labels identified by the monolingual model compared to the multilingual model. The performance of both models is visually depicted in the confusion matrices presented in Figure 2(a) for the monolingual (SwahBERT) model and Figure 2(b) for the multilingual (mBERT) model.

Furthermore, as outlined in Table 2, our detailed comparison underscored the monolingual model's superior performance, achieving an accuracy score, macro average, and weighted average of 69%, marginally outperforming the multilingual model's 66%. Class-wise comparisons revealed the monolingual model's proficiency in predicting contradiction and neutral labels, achieving the highest precision, recall, and F1-scores in these categories. Despite its excellence, the monolingual model outperformed the multilingual model in predicting entailment, showcasing its ability to discern logical relationships across diverse languages. These findings emphasize the importance of a balanced approach that leverages the strengths of both models to optimize overall performance without compromising accuracy in specific task domains.



(a)



(b)

Figure 2. Confusion matrix for (a) monolingual (SwahBERT) and (b) multilingual (mBERT) models

3.3. Addressing limitations

Comparing the monolingual and multilingual model predictions provides valuable insights into their misclassifications. Both models encounter challenges due to lexical variation, semantic equivalence, subtle differences, and question formation, as illustrated in Table 3. In instances where the premise and hypothesis discuss techniques utilized by organizations, both models mistakenly predict label 1 (entailment) instead of label 0 (contradiction). This misclassification underscores their failure to discern the nuanced difference between "*kwa kawaida*" (usually) and "*zinazotumiwa sana*" (heavily used), indicating a limitation in comprehending contextual nuances. Similarly, the monolingual model adeptly identifies semantic equivalences, highlighting its proficiency in capturing subtle nuances, while the multilingual model struggles due to linguistic variation across languages. Moreover, the multilingual model encounters challenges with sentence pairs involving questions, where misclassification occurs due to difficulties in navigating linguistic differences. Despite these limitations, the monolingual model generally outperforms, emphasizing the significance of language-specific comprehension in accurate predictions amidst linguistic complexities.

Table 3. Illustrative examples for monolingual and multilingual model predictions

SN	Premise	Hypothesis	True label	Predicted monolingual	Predicted multilingual
1	*Tulijaribu kutambua mbinu ambazo zilitekelezwa kwa kawaida na mashirika katika muda wa miaka 5 iliyopita*. We tried to identify the methods that were commonly implemented by organizations during the last 5 years.	*Tunataka kutambua mbinu zinazotumiwa sana na mashirika katika miaka 5 iliyopita.* We want to identify the methods most used by organizations in the last 5 years.	0	1	1
2	*Ni lazima mataifa yaonyeshe maendeleo ya kuridhisha katika mipango yao ya utekelezaji ya majimbo kuelekea lengo lililoamrishwa na bunge la kurejea hali ya asili katika mbuga za kitaifa na maeneo ya nyika.* States must demonstrate satisfactory progress in their state action plans toward the congressionally mandated goal of restoring natural conditions in national parks and wilderness areas.	*Sio lazima kwa uboreshaji wowote.* It is not necessary for any improvement.	2	2	1
3	*Je, walikuwa humo ndani?* Were they inside?	*Je, walipaswa kuwa humo ndani?* Were they supposed to be inside?	1	1	0

3.4. Implications for future research and model refinement

The findings of this study suggest that language-specific models are crucial for tasks like NLI, where language nuances greatly affect performance. Multilingual models may not effectively capture these nuances, particularly for low-resource languages. Future research could focus on enhancing multilingual model performance by incorporating language-specific pre-processing techniques or exploring transfer learning from language-specific models. Additionally, investigating hybrid models that combine the strengths of monolingual and multilingual approaches could yield promising results in improving overall model performance on NLI tasks.

Furthermore, there is a need to enhance entailment prediction within monolingual models, particularly for languages like Swahili. This could involve integrating external knowledge sources or exploring alternative neural network architectures tailored for entailment tasks. Further validation using meticulously curated datasets is essential to address limitations and gain a comprehensive understanding of the model's capabilities, particularly its impact on Swahili NLI tasks. Recent observations suggest that monolingual models often outperform multilingual ones, highlighting the importance of this line of research. Our study contributes to this discourse by emphasizing the superior prediction capability of monolingual models, supported by higher actual prediction counts and competitive average scores across different classes.

4. CONCLUSION

This study highlights the effectiveness of SwahBERT in capturing the nuances of the Swahili language, outperforming mBERT in the NLI task. This underscores the potential of language-specific models like SwahBERT to enhance NLP applications tailored to low-resource languages. However, addressing existing limitations and refining model performance, particularly in tasks related to entailment interpretation, requires further research and development. Moving forward, this research will focus on refining

SwahBERT's performance by exploring specialized datasets and methodologies to better address the linguistic complexities inherent in low-resource languages like Swahili. Additionally, promoting inclusivity in language processing technologies for low-resource languages remains a priority. Utilizing both monolingual and multilingual approaches, this research aims to develop more robust NLP tools for Swahili-speaking communities and advance low-resource language processing globally, including exploring advanced techniques like fuzzy logic.

ACKNOWLEDGEMENTS

This research is supported by internal funding from the Faculty of Computer Science, Universitas Indonesia, and the high-performance computing facilities at the Tokopedia-UI AI Center, Universitas Indonesia. The authors would like to thank all of them for their support.

REFERENCES

[1] C. S. Shikali and R. Mokhosi, "Enhancing African low-resource languages: Swahili data for language modelling," *Data in Brief*, vol. 31, 2020, doi: 10.1016/j.dib.2020.105951.
[2] M. J. Robinson, *A language for the world: the standardization of Swahili*. Athens, Ohio: Ohio University Press, 2022.
[3] A. N. Nwammuo and A. Salawu, "Are radio programmes via indigenous languages the solution? A study of Igbo scholars' assessment of the effectiveness of the British Broadcasting Corporation (BBC) in promoting African languages," *African Renaissance*, vol. 16, no. 1, pp. 83–99, 2019, doi: 10.31920/2516-5305/2019/V16n1a5.
[4] P. Mpofu, I. A. Fadipe, and T. Tshabangu, *Indigenous African Language Media*. Singapore: Springer Nature Singapore, 2023, doi: 10.1007/978-981-99-0305-4.
[5] A. Owoyemi, J. Owoyemi, A. Osiyemi, and A. Boyd, "Artificial intelligence for healthcare in Africa," *Frontiers in Digital Health*, vol. 2, 2020, doi: 10.3389/fdgth.2020.00006.
[6] C. Holst *et al.*, "Development of digital health messages for rural populations in Tanzania: multi- and interdisciplinary approach," *JMIR mHealth and uHealth*, vol. 9, no. 9, 2021, doi: 10.2196/25558.
[7] B. Minixhofer, F. Paischer, and N. Rekabsaz, "WECHSEL: Effective initialization of subword embeddings for cross-lingual transfer of monolingual language models," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 3992–4006, doi: 10.18653/v1/2022.naacl-main.293.
[8] K. Ogueji, "AfriBERTa: Towards viable multilingual language models for low-resource languages," *M.Sc. Thesis*, Department of Mathematics, University of Waterloo, Waterloo, Canada, 2022.
[9] G. L. Martin, M. E. Mswahili, and Y.-S. Jeong, "Sentiment classification in Swahili language using multilingual BERT," *arXiv-Computer Science*, pp. 1–5, 2021.
[10] Y. Xiao *et al.*, "Are BERT family good instruction followers? A study on their potential and limitations," in *ICLR 2024*, pp. 1–20, 2024.
[11] K. Alnajjar and M. Hämäläinen, "Harnessing multilingual resources to question answering in Arabic," *arXiv-Computer Science*, pp. 1–6, 2022.
[12] S. Wu and M. Dredze, "Are all languages created equal in multilingual BERT?," in *Proceedings of the 5th Workshop on Representation Learning for NLP*, 2020, pp. 120–130, doi: 10.18653/v1/2020.repl4nlp-1.16.
[13] G. Martin, M. E. Mswahili, Y.-S. Jeong, and J. Young-Seob, "SwahBERT: language model of Swahili," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 303–313, doi: 10.18653/v1/2022.naacl-main.23.
[14] Z. Wang, L. Li, and D. Zeng, "Knowledge-enhanced natural language inference based on knowledge graphs," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 6498–6508, doi: 10.18653/v1/2020.coling-main.571.
[15] A. De, M. S. Desarkar, and A. Ekbal, "Towards improvement of grounded cross-lingual natural language inference with VisioTextual attention," *Natural Language Processing Journal*, vol. 4, 2023, doi: 10.1016/j.nlp.2023.100023.
[16] Q. Wu, P. Wang, X. Wang, X. He, and W. Zhu, "Question answering (QA) basics," in *Advances in Computer Vision and Pattern Recognition*, Springer, Singapore, 2022, pp. 27–31, doi: 10.1007/978-981-19-0964-1_3.
[17] V. Agate, S. Mirajkar, and G. Toradmal, "Book summarization using NLP," *International Journal of Innovative Research in Engineering*, vol. 11, no. 4, pp. 476–480, 2023, doi: 10.59256/ijire.2023040218.
[18] W. Zhou, "Research on information extraction technology applied for knowledge graphs," *Applied and Computational Engineering*, vol. 4, no. 1, pp. 26–31, 2023, doi: 10.54254/2755-2721/4/20230340.
[19] H. Tanvir, C. Kittask, S. Eiche, and K. Sirts, "EstBERT: A pretrained language-specific BERT for Estonian," in *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, 2021, pp. 11–19.
[20] A. Velankar, H. Patil, and R. Joshi, "Mono vs multilingual BERT for hate speech detection and text classification: a case study in Marathi," in *Artificial Neural Networks in Pattern Recognition (ANNPR 2022)*, 2023, pp. 121–128, doi: 10.1007/978-3-031-20650-4_10.
[21] M. Straka, J. Náplava, J. Straková, and D. Samuel, "RobeCzech: Czech RoBERTa, a monolingual contextualized language representation model," in *Text, Speech, and Dialogue (TSD 2021)*, 2021, pp. 197–209, doi: 10.1007/978-3-030-83527-9_17.
[22] D. Q. Nguyen and A. T. Nguyen, "PhoBERT: Pre-trained language models for Vietnamese," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 1037–1042, doi: 10.18653/v1/2020.findings-emnlp.92.
[23] K. Jain, A. Deshpande, K. Shridhar, F. Laumann, and A. Dash, "Indic-transformers: an analysis of transformer language models for Indian languages," *arXiv-Computer Science*, pp. 1–14, Nov. 2020.
[24] R. Scheible, F. Thomczyk, P. Tippmann, V. Jaravine, and M. Boeker, "GottBERT: a pure German language model," *arXiv-Computer Science*, pp. 1–6, 2020.
[25] A. F. M. D. Paula and I. B. Schlicht, "AI-UPV at IberLEF-2021 DETOXIS task: toxicity detection in immigration-related web news comments using transformers and statistical models," *arXiv-Computer Science*, pp. 1–20, Nov. 2021.
[26] S. Wu and M. Dredze, "Beto, Bentz, Becas: The surprising cross-lingual effectiveness of BERT," *arXiv-Computer Science*, pp. 1–12, 2019.

[27] A. K. Upadhyay and H. K. Upadhya, “XNLI 2.0: Improving XNLI dataset and performance on cross lingual understanding (XLU),” in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 2023, pp. 1–6, doi: 10.1109/I2CT57861.2023.10126332.

[28] M. Sadat and C. Caragea, “Learning to infer from unlabeled data: a semi-supervised learning approach for robust natural language inference,” in *Findings of the Association for Computational Linguistics: EMNLP 2022*, 2022, pp. 4763–4776, doi: 10.18653/v1/2022.findings-emnlp.351.

[29] J. D. L. C. Ntivuguruzwa and T. Ahmad, “A convolutional neural network to detect possible hidden data in spatial domain images,” *Cybersecurity*, vol. 6, no. 1, 2023, doi: 10.1186/s42400-023-00156-x.

[30] N. J. D. L. Croix and T. Ahmad, “Toward secret data location via fuzzy logic and convolutional neural network,” *Egyptian Informatics Journal*, vol. 24, no. 3, 2023, doi: 10.1016/j.eij.2023.05.010.

BIOGRAPHIES OF AUTHORS

Hajra Faki Ali is advancing her computer science knowledge through her master's degree at the Universitas Indonesia. Her academic journey began at the University of Dar es Salaam, Tanzania, where she obtained her bachelor of science in computer science in 2017. Her scientific interests are deeply rooted in machine learning, with a particular focus on natural language processing. Her research work includes an intriguing study on natural language inference in low-resource languages such as Swahili. She can be contacted at email: hajra.ali@judiciaryzanzibar.go.tz.

Adila Alfa Krisnadhi received a B.S. in computer science from Universitas Indonesia in 2002, an M.Sc. in computational logic from Technische Universität Dresden, Germany in 2007, and a Ph.D. in computer science from Wright State University, USA in 2015. He joined as a permanent faculty member at the Faculty of Computer Science, Universitas Indonesia in 2007. Since 2019, he also held a position as a co-director of the Tokopedia-UI AI Center of Excellence. He has authored/co-authored more than 80 publications in the area of semantic technologies, knowledge representation, and machine learning. His research interests include knowledge graph construction, reasoning and question answering over knowledge graphs and ontologies, as well as knowledge modelling and engineering. He can be contacted at email: adila@cs.ui.ac.id.