A Comparative Evaluation of Deep Learning Architectures for Amharic Riddle Meaning Detection

Mihret Desta[1*], Negash Woldemeskel[2], Dagim Firide[1], Biruk Gebrehiwot[3], Sharew Aknaw[3], Erimias Abebe[4]

[1]Department of Software Engineering, College of Computing, Debre Berhan University, Ethiopia

[2]Department of Information Technology, College of Computing, Debre Berhan University, Ethiopia

[3]Department of Computer Science, College of Computing, Debre Berhan University, Ethiopia

[4]Department of Mathematics, College of Natural and Computational Sciences, Debre Berhan University, Ethiopia

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1307000327

Received: 03 August 2026; Accepted: 08 August 2026; Published: 18 August 2026

ABSTRACT

Language is a rule-governed system of symbols that enables communication, expresses thought, and preserves culture across diverse contexts. Amharic, a morphologically rich Semitic language and the official working language of Ethiopia, serves as a vital medium for cultural heritage, particularly through the tradition of riddles (እንቆቅልሽ). Riddles are sophisticated cultural tools characterized by metaphorical ambiguity and deep semantic complexity. Despite their educational value, automated interpretation remains a significant challenge due to Amharic's morphological complexity and the scarcity of specialized Natural Language Processing (NLP) frameworks. This study aims to develop a model for Amharic riddle meaning detection. Using an experimental research design with a dataset of 8,457 samples, we implemented a robust pre-processing pipeline involving normalization, tokenization, encoding, and sequence padding. We evaluated four deep learning architectures: Convolutional Neural Network (CNN), Deep Neural Network (DNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and Bidirectional Gated Recurrent Unit (Bi-GRU). Experimental results demonstrate that the Bi-LSTM model achieved superior performance, reaching a peak accuracy of 98.64%. The findings establish the viability of automated semantic detection in Ethiopian folk literature using deep learning techniques. However, the relatively compact base dataset compared with the 763 semantic classes presents a risk of overfitting and may limit model generalization, while the study is restricted to conventional deep learning architectures without integrating modern Transformer-based models. Future research should therefore expand and diversify the dataset across regional and cultural sources, investigate Transformer-based models for low-resource languages, and extend the approach to other Ethiopian languages to support robust, real-time, and multilingual cultural preservation.

Keywords: Ethiopian Folk Literature, Amharic language, Riddle Meaning Detection, Deep Learning, Semantic Interpretation.

INTRODUCTION

Language is a fundamental means of communication, cultural transmission, and the representation of thought. It is a structured system of symbols governed by rules of phonology, morphology, syntax, and semantics, allowing speakers to convey meaning across diverse contexts [1]. Beyond its structural aspects, language functions as a social institution, shaping identity, reinforcing cultural values, and facilitating interaction within and across communities [2]. The scientific study of language spans multiple disciplines, including linguistics,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

psychology, anthropology, and computational sciences, each contributing unique perspectives to understand its nature and use [3].

Among the world’s languages, Amharic occupies a distinctive position as both a national and cultural symbol of Ethiopia. Amharic (አማርኛ) is a member of the Ethio-Semitic branch of the Semitic language family, closely related to Tigrinya and Ge’ez, yet distinct in its phonological and morph syntactic development [4]. As the official working language of the Federal Government of Ethiopia, it is widely spoken as both a primary and secondary language across the country. The Amharic writing system, known as Fidel (ፊደል), is an abugida derived from Ge’ez, and it has been central to Ethiopia’s literary, religious, and administrative traditions for centuries [5].

Riddles (እንቆቅልሽ) stand as a cornerstone of Ethiopian oral tradition and cultural heritage. Riddles serve not only as entertainment but also as pedagogical tools, transmitting cultural wisdom and sharpening cognitive skills. They function as a medium of intergenerational knowledge transfer, fostering creativity through problem-solving [6]. Within Ethiopian society, riddles are performed in communal settings, reinforcing social bonds and transmitting moral lessons embedded in local traditions and proverbs [7].

Amharic riddles are characterized by figurative density and cultural subtlety, making them resistant to literal interpretation and requiring contextual knowledge for comprehension [8]. As such, riddles exemplify the interplay between language, culture, and cognition. The automated classification and interpretation of Amharic folk literature represents a significant frontier in Natural Language Processing (NLP) due to the intricate intersection of figurative complexity and severe resource scarcity [9].

Amharic riddles exhibit high levels of metaphorical density and semantic ambiguity, which demand a deep understanding of contextual and cultural nuances that traditional architectures fail to capture [10]. Unlike literal text, these structures rely on deliberate semantic shifts and localized proverbs that current deep learning (DL) models struggle to decode, as they are typically optimized for high-resource languages with more accessible, literal semantic mappings [11].

Furthermore, the morphological richness and unique syllabic structure of the Ge'ez script add layers of structural variability that complicate automated feature extraction and pattern recognition [12]. While DL has revolutionized NLP for dominant languages, Amharic remains an under-resourced language lacking dedicated computational tools and standardized, high-quality annotated datasets specific to specialized genres like folk literature [13]. Existing models often fail to translate effectively to these unique syntactic demands, resulting in a persistent gap between labor-intensive manual classification and the need for robust automated systems. Consequently, there is a critical research need to evaluate which neural architectures specifically Convolutional Neural Network (CNN), Deep Neural Network (DNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and Bidirectional Gated Recurrent Unit (Bi-GRU) can optimally capture the local patterns and bidirectional context required for accurate riddle meaning detection in a low-resource setting [14] [15].

To address this challenge, the present study frames Amharic riddle meaning detection as a supervised multi-class text classification problem. Automatically detecting the intended meaning of a riddle is a complex task. Given a riddle text as input (e.g., “ቀን ቀን አይጠጋም ማታ ማታ አብሮ ይጋደም”), the system must predict its correct semantic interpretation from a predefined set of classes. This task is closely related to idiom interpretation, metaphor detection, and sarcasm recognition, but is uniquely complicated by Amharic’s morphological richness and the cultural grounding of its figurative language.

While DL models have demonstrated strong performance in figurative language tasks for high-resource languages, investigations for Amharic remain limited. Current literature has primarily focused on metaphor detection in foreign languages [16] [17] and idiomatic expression recognition in Ethiopian languages [18] [13]. However, these studies do not adequately address the unique semantic challenges of Amharic riddles. The core difficulty lies in the Wax and Gold (ሰምና ወርቅ) system, which creates a deliberate gap between the literal surface meaning and the hidden metaphorical truth. This is facilitated by Amharic's morphological complexity and homonymy, where words with identical sounds but different roots create deceptive double meanings. Furthermore, these riddles rely heavily on cultural allusions and traditional imagery that require deep, context-specific knowledge. To fill this gap, the current study facilitates the automated detection of Amharic riddle meanings.

The remaining parts of the paper are organized as follows: Section 2 reviews related works, Section 3 describes the methodology, Section 4 presents the experimental results, Section 5 discusses the results, and Section 6 provides concluding remarks.

Related Works

This section reviews existing literature on riddle meaning detection and figurative language processing within international and local contexts. While direct computational research on riddle meaning detection is currently limited, methodologies from related domains such as idiom and metaphor analysis provide a foundational framework.

The survey conducted by Giadikiaroglou et al. [16] noted the absence of a systematic framework for evaluating Large Language Models (LLMs) within complex reasoning domains. The study introduces a novel taxonomy that bifurcates puzzles into rule-based categories (requiring deterministic deduction) and rule-less categories (necessitating lateral thinking and figurative interpretation). By analyzing methodologies including Chain-of-Thought (CoT), Tree-of-Thought (ToT), and neuro-symbolic translation across benchmarks like BrainTeaser and RiddleSense, the researchers identify a persistent "logic gap" where models struggle to replicate human-level creative problem-solving. While structured prompting techniques achieved success rates up to 93.2% in deterministic environments, models remain limited by metaphorical ambiguity and counterfactual reasoning.

Panagiotopoulos et al. [17] addressed the cognitive limitations of LLMs in tasks requiring lateral thinking and abstract reasoning. They introduced RISCORE, an automated framework generating contextually reconstructed puzzles to isolate underlying reasoning patterns rather than surface-level features. Evaluating architectures such as Llama3, Mistral-7B, and Qwen2-7B across BrainTeaser and RiddleSense benchmarks, the authors demonstrated that RISCORE outperforms traditional few-shot selection strategies, establishing targeted contextual framing as a critical requirement for advanced riddle resolution.

Limenih et al. [18] explored processing non-literal idiomatic expressions in Amharic. Evaluating CNN, LSTM, and GRU architectures against traditional machine learning algorithms on a dataset of 3,556 sentences, the results indicated that deep learning models significantly outperform conventional methods, with the CNN architecture achieving a peak accuracy of 93%. The study concluded that deep learning effectively captures the complex semantic dependencies of Amharic idioms.

In a related study, Endalie et al. [13] developed an automated recognition system using CNN integrated with a pre-trained FastText embedding model to detect idioms within Amharic text. Sourced from 3,300 isolated phrases, the proposed CNN architecture achieved a testing accuracy of 80%, outperforming Support Vector Machine (76%), Random Forest (72%), and K-Nearest Neighbor (68%).

A critical gap remains in the literature. International frameworks rely on high-resource language data that cannot be directly translated to Amharic's morphological structures, while local studies focus primarily on binary classification of idioms. Consequently, a critical gap exists in developing a comparative framework that moves beyond simple recognition toward full meaning detection. To address this, our research separately evaluates CNN, DNN, Bi-LSTM, and Bi-GRU models to measure each architecture's ability to capture the complex semantic mappings of riddles (እንቆቅልሽ).

METHODOLOGY

The proposed study is designed to detect the meaning of Amharic riddles using CNN, DNN, Bi-LSTM, and Bi-GRU. Each algorithm is selected based on its unique strengths in processing textual data: CNNs extract

INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

local n-gram features, DNNs provide a baseline for high-level representations, Bi-GRUs capture long-term dependencies efficiently, and Bi-LSTMs enhance sequence modeling by incorporating context from both directions. To compare model performance systematically, this study follows an experimental research design.

Architecture of the Proposed Model

Figure 1 below shows the overall architecture of the Amharic riddles meaning detection model encompasses data collection, preprocessing, dataset splitting, training, validation, and testing. The architecture follows a systematic pipeline that begins with an Amharic Riddles Dataset subjected to cleaning, normalization, tokenization, sequence padding, and target encoding. The processed data is divided into training, validation, and test sets. Training sets are used to train the models, validation sets are utilized for hyperparameter tuning, and the unseen test set is reserved for final evaluation.

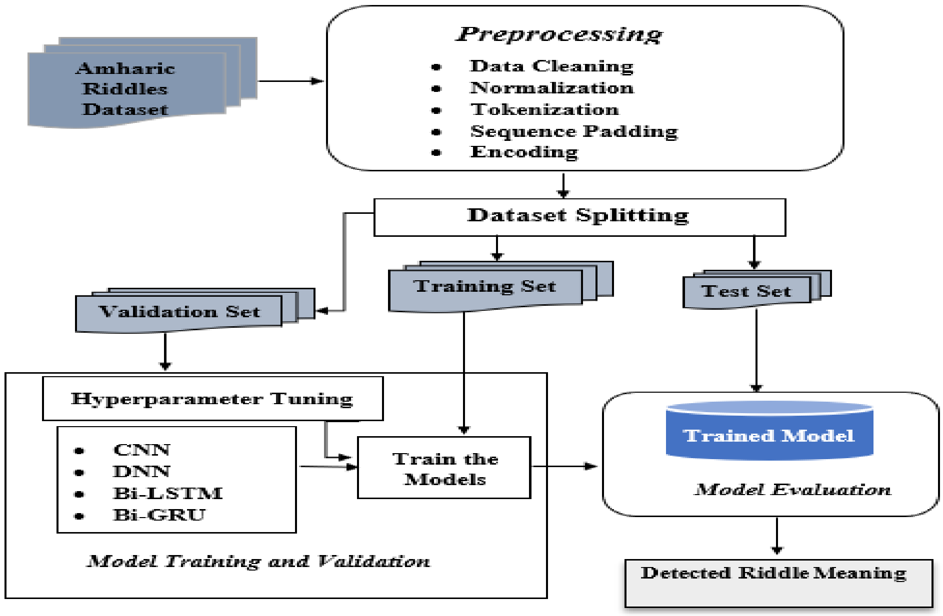

Figure 1: The overall architecture of the proposed Model

Data Collection and Preparation

Amharic riddles (እንቆቅልሽ) are traditional oral literature categorized into natural phenomena, household objects, and animals [19]. Riddles were collected from traditional folklore literature Shiferaw [20], Mamo [21], Demena [22], online platforms MetaAppz [23], and field annotations from community elders. Preprocessing involved cleaning punctuation, special characters, and extra whitespaces. Character-level normalization was performed to map redundant Amharic characters: variations of the "h" sound (ሃ, ኀ, ኃ, ሐ, ሓ, ኸ) were unified into ሀ (and across vowel orders: ሁ, ሂ, ሄ, ህ, ሆ). Similarly, "s" (ሠ -> ሰ), "a" (ዓ, ኣ, ዐ -> አ), "ts" (ጸ -> ፀ), and "wu" (ዉ -> ው) were merged. Tokenization mapped normalized words to unique integer IDs, Sequence Padding standardized input lengths to 500 tokens, and Target Encoding (Label and One-Hot) transformed unique riddle meanings into binary vectors. The compiled dataset contained 1,282 riddle instances paired with their corresponding meanings in CSV format. An 80:10:10 ratio was employed for training, validation, and testing. Reserving a dedicated test set ensured an unbiased assessment of generalization accuracy on unseen data [25].

Model Selection and Training

CNN, DNN, Bi-GRU, and Bi-LSTM architectures were selected due to their proven capabilities in text classification. CNNs extract spatial patterns, DNNs capture complex non-linear relationships, and Bi-GRU/Bi-LSTM capture sequential dependencies and bidirectional context essential for resolving Wax and Gold semantic ambiguities [26] [27].

INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

EXPERIMENTAL RESULTS

Experimental Setup

The dataset comprised 1,282 Amharic riddles across 763 classes. Stratified splitting allocated 80% to training, 10% to validation, and 10% to testing. To address data scarcity, random oversampling by a factor of 8 was applied strictly to the training dataset, increasing training instances from 1,025 to 8,200 while keeping validation and test sets original.

Hyperparameters were optimized using the Adam optimizer with a learning rate of 0.001 and batch sizes of 32 and 64. Models were trained up to 100 epochs with a dropout ratio of 0.4 and early stopping (patience factor of 5 epochs) based on validation loss. ReLU activation was used in hidden layers, Softmax in the output layer, and Categorical Cross-Entropy as the loss function.

Table 1: Summary of Hyperparameter Configurations for the Proposed Models

Hyperparameter	**Value**
Optimizer	Adam
Learning Rate	10^{-3} (0.001)
Batch Size	32, 64
Epochs	100
Dropout Ratio	0.4
Regularization Strategy	Early Stopping (Callbacks)
Monitoring Metric	Validation Loss (val_loss)
Patience Factor	5 Epochs
Hidden Layer Activation	ReLU
Output Layer Activation	Softmax
Loss Function	Categorical Cross-Entropy
Total Classes	763

Performance Results

The performance of CNN, DNN, Bi-LSTM, and Bi-GRU models was evaluated across Accuracy, Macro-Precision, Macro-Recall, and Macro-F1 Score. The comparative performance metrics are summarized in Table 2 and Figure 2.

Table 2: Experimental Results of the Proposed Models

Models	**Accuracy**	**Precision**	**Recall**	**F1-Score**
CNN	98.34%	1.00	0.99	1.00
DNN	98.44%	1.00	0.99	1.00

INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Bi-LSTM	98.64%	1.00	0.99	0.99
Bi-GRU	98.54%	1.00	0.99	0.99

All four deep learning models achieved accuracy scores exceeding 98%. The Bi-LSTM model achieved the highest accuracy at 98.64%, followed by Bi-GRU (98.54%), DNN (98.44%), and CNN (98.34%). Bidirectional recurrent models proved superior in capturing sequential dependencies in Amharic figurative language.

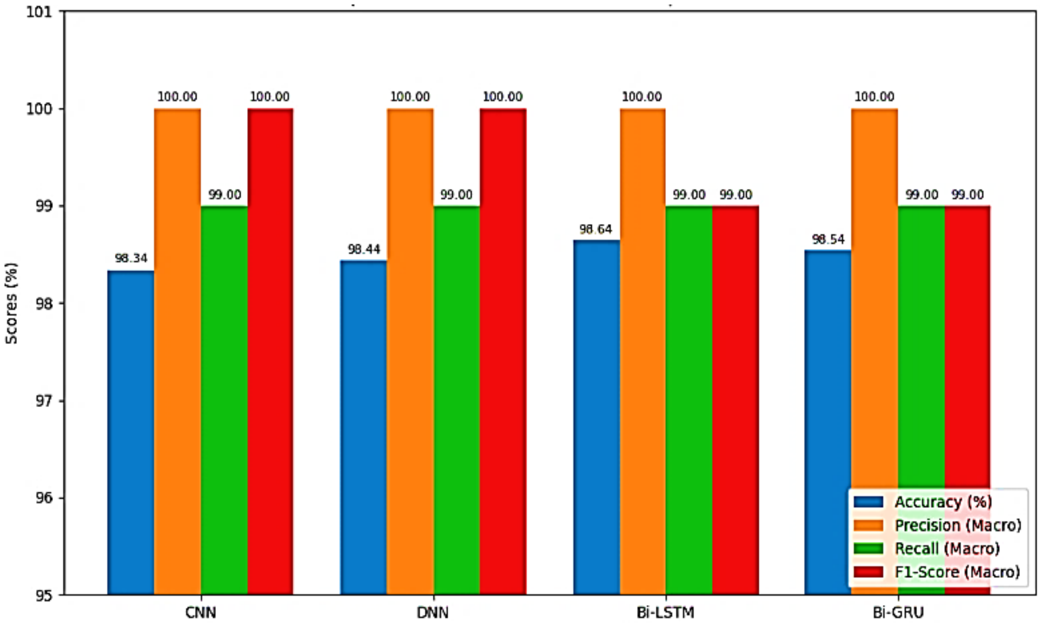

Figure 2: Experimental Results of the Proposed Models

DISCUSSION

Error Analysis based on Confusion Matrix

The error analysis was conducted using the confusion matrix of the Bi-LSTM model, which achieved the highest accuracy of 98.64%. As illustrated in the Bi-LSTM confusion matrix in Figure 3, the model exhibits a dominant diagonal pattern, indicating that for almost all classes, the predicted meanings align perfectly with the actual labels. The sparse off-diagonal elements represent the marginal error rate of 1.36%. A closer inspection of these rare misclassifications suggests they occur primarily due to semantic overlap or orthographic similarity in the Amharic riddle dataset. For instance, riddles with nearly identical sentence structures but different hidden meanings can occasionally lead to slight model confusion. For example, the riddles “የሚሄድ ነገር እግር የሌለው” (ውሃ/water) and “የሚሄድ ነገር ክንፍ የሌለው” (ደመና/cloud) share similar linguistic patterns, although they refer to different concepts. Likewise, “ነጭ ሜዳ ላይ ጥቁር ዘር” (ወረቀት ላይ ፅሁፍ/writing on paper) and “ጥቁር ሜዳ ላይ ነጭ በግ” (ከዋክብት በሰማይ/stars in the sky) exhibit comparable figurative structures while conveying distinct meanings. However, the bidirectional processing of the Bi-LSTM effectively mitigated most of these ambiguities by analyzing the context from both the start and end of the riddle, ensuring that the final hidden answer was identified with high precision

Findings

The findings from this experiment provide several key insights into the automated detection of Amharic riddle meanings. First, the superior performance of Bi-LSTM and Bi-GRU over CNN and DNN confirms that the temporal sequence of words is vital in interpreting figurative Amharic language; the bidirectional approach is particularly effective at capturing the indirect meaning characteristic of traditional riddles. Second, despite the challenge of classifying across 763 distinct meaning categories, the selected model maintained a Macro-

Precision of 1.00, proving that the system is highly stable and does not suffer from class-bias or frequent false positives. Additionally, the high accuracy across all models indicates that the character-level normalization and the systematic expansion of the dataset provided a clean, high-quality signal that allowed the deep learning architectures to converge effectively. Finally, while Bi-LSTM is the most accurate, the Bi-GRU's suggests it is a highly efficient alternative that offers nearly identical performance.

By comparing various models, the study identifies architectures capable of capturing the sequential dependencies and contextual semantics. The main contributions of this study are outlined as follows. First, an attempt is made to develop a comprehensive dataset of Amharic riddles specifically curated for the training and evaluation of computational models in figurative language. Second, this paper provides a foundational benchmark for future research in Amharic natural language processing (NLP), expanding the field's scope from conventional text classification toward deeper semantic and figurative understanding. Finally, it shows practical applications for the study's findings within educational software and digital platforms for language acquisition and cultural heritage.

CONCLUSION

Language is a rule-governed system of symbols that enables communication, expresses thought, and preserves culture across diverse contexts. Amharic, as a morphologically rich Semitic language and the official working language of Ethiopia, serves as a vital vehicle for cultural heritage, particularly through the tradition of riddles. These riddles are not merely entertainment but sophisticated pedagogical tools characterized by figurative density, metaphorical ambiguity, and deep cultural grounding. However, the automated interpretation of this folk literature is a major NLP frontier due to the combination of semantic complexity and resource scarcity. Most deep learning models are optimized for literal, high-resource languages, making them struggle with the morphological richness of the Ge'ez script and the figurative nature of Ethiopian oral traditions. Furthermore, a persistent research gap exists, as no prior automated systems have been developed specifically for Amharic riddle meaning detection.

To address these challenges, we implemented an experimental research design using a dataset of 8,457 samples. To prepare the data for analysis, we applied preprocessing techniques, including data cleaning, normalization, tokenization, encoding, and sequence padding. We evaluated four deep learning architectures for feature extraction and meaning detection: CNN, DNN, Bi-LSTM, and Bi-GRU. To achieve high performance, hyperparameters were selected using a manual search technique. We conducted four specific experiments by varying the batch size (32 and 64) while maintaining consistent hyperparameter settings and dataset distributions. The experimental results revealed that the Bi-LSTM emerged as the superior architecture, achieving the highest testing accuracy of 98.64%, a Macro-Precision of 1.00, and a Recall of 0.99.

Future Scope

This study successfully demonstrates the automated detection of semantic meanings in Amharic riddles using deep learning techniques. However, despite these promising results, several limitations provide important directions for future research. First, the relatively small size of the original dataset compared with the large number of 763 semantic classes may limit the model's ability to generalize and increase the risk of overfitting, even with the application of random oversampling. Second, the study focuses primarily on conventional deep learning architectures and does not investigate recent pre-trained Transformer-based language models, such as Afro-XLMR and multilingual BERT, which may offer improved representation and classification capabilities for low-resource languages.

Future research should focus on substantially expanding the original dataset by collecting Amharic riddles from a wider range of regional, cultural, and traditional sources. Increasing the diversity and volume of the data could improve model generalization across the 763 semantic classes and support more reliable classification. Future studies should also investigate state-of-the-art Transformer architectures and pre-trained language models designed or adapted for low-resource and African languages, and compare them with the recurrent and convolutional models evaluated in this study. Furthermore, extending the approach to other

Ethiopian languages could support the development of multilingual intelligent systems for cultural preservation, education, and digital heritage applications.

REFERENCES

1. Fromkin, V., Rodman, R., & Hyams, N. (2018). An Introduction to Language. Cengage Learning.
2. Mesthrie, R., Swann, J., Deumert, A., & Leap, W. (2009). Introducing Sociolinguistics. Edinburgh University Press.
3. Akmajian, A., Farmer, A. K., Bickmore, L., Demers, R. A., & Harnish, R. M. (2017). Linguistics: An Introduction to Language and Communication (7th ed.). MIT Press.
4. Meyer, R., Wakjira, B., & Leyew, Z. (Eds.). (2023). The Oxford Handbook of Ethiopian Languages. Oxford University Press.
5. Meshesha, M., & Jawahar, C. (2008). Indigenous Scripts Of African Languages. Indilinga African Journal of Indigenous Knowledge Systems, 6(2). https://doi.org/10.4314/indilinga.v6i2.26422
6. Amharic Folkloric Oral Traditions: Collections for Insiders and for Outsiders. (2026). Aethiopica, 25. https://doi.org/10.15460/aethiopica.25.0.1563
7. Beyene, Y. (2025). Ethiopic Historiographical Texts. Harrassowitz Verlag.
8. Amberber, M. (2023). Amharic. In R. Meyer, B. Wakjira, & Z. Leyew (Eds.), The Oxford Handbook of Ethiopian Languages (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198728542.013.21
9. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed.). Pearson.
10. Yimam, S. M., Alemayehu, H. M., Ayele, A., & Biemann, C. (2020). Exploring Amharic Sentiment Analysis from Social Media Texts: Building Annotation Tools and Classification Models. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 1048–1060). https://doi.org/10.18653/v1/2020.coling-main.91
11. Tonja, A. L., Kolesnikova, O., Gelbukh, A., & Kalita, J. (2024). EthioMT: Parallel Corpus for Low-resource Ethiopian Languages. In Proceedings of the Fifth Workshop on Resources for African Indigenous Languages @ LREC-COLING 2024 (pp. 107–114).
12. Ejigu, Y. A., & Asfaw, T. T. (2024). Large Scale Speech Recognition for Low Resource Language Amharic, an End-to-End Approach. Preprints, 2024020813. https://doi.org/10.20944/preprints202402.0813.v1
13. Endalie, D., Haile, G., & Taye, W. (2023). Deep learning-based idiomatic expression recognition for the Amharic language. PLOS ONE, 18(12), e0295339. https://doi.org/10.1371/journal.pone.0295339
14. Ayalew, A. A., et al. (2025). Amharic event text classification from social media using hybrid deep learning. International Journal of Electrical and Computer Engineering (IJECE), 15(2), 2264–2270. https://doi.org/10.11591/ijece.v15i2.pp2264-2270
15. Marilign, D. M., & Alemu, E. N. (2025). Amharic News Topic Classification: Dataset and Transformer-Based Model Benchmarks. In Proceedings of the 9th Widening NLP Workshop (pp. 130–135). https://doi.org/10.18653/v1/2025.winlp-main.23
16. Giadikiaroglou, P., Lymperaiou, M., Filandrianos, G., & Stamou, G. (2024). Puzzle Solving using Reasoning of Large Language Models: A Survey. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 11574–11591). https://doi.org/10.18653/v1/2024.emnlp-main.646
17. Panagiotopoulos, I., Filandrianos, G., Lymperaiou, M., & Stamou, G. (2025). RISCORE: Enhancing In-Context Riddle Solving in Language Models through Context-Reconstructed Example Augmentation. In Proceedings of the 31st International Conference on Computational Linguistics.
18. Limenih, H. H., Adege, A. B., Alene, A. Y., Demasu, H. T., & Belachew, H. M. (2025). Automatic Identification of Amharic Text Idiomatic Expressions Using a Deep Learning Approach. IEEE Access, 13, 83095–83106. https://doi.org/10.1109/ACCESS.2025.3567811
19. ʿAzaza, F. (1998). Unheard Voices: Drought, Famine, and God in Ethiopian Oral Poetry. Addis Ababa University Press.
20. Shiferaw, A. (2000). Zemenawi enkokeleshe. Far East Trading P.L.C.
21. Mamo, W. (2010). Alifa Liyu Yelijoch Enkokilish. Kuraz P.L.C.
22. Demena, S. (2018). Enkokilshina Teretoch. Kuraz P.L.C.

23. MetaAppz. (2026). Amharic Riddles - Amharic Enkokelesh. https://www.MetaAppz.com/Games/Enqoqelesh_Riddles
24. HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. PLOS ONE, 15(5), e0232525. https://doi.org/10.1371/journal.pone.0232525
25. Joseph, V. R. (2022). Optimal Ratio for Data Splitting. Statistical Analysis and Data Mining: The ASA Data Science Journal, 15(4), 531–538. https://doi.org/10.1002/sam.11583
26. Amajd, M., Kaimuldenov, Z., & Voronkov, I. (2017). Text Classification with Deep Neural Networks. CEUR Workshop Proceedings, 1989, 362–370.
27. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning Based Text Classification: A Comprehensive Review. arXiv preprint, arXiv:2004.03705. https://doi.org/10.48550/arXiv.2004.03705
28. Shi, Z. (2025). Data Augmentation for Low-resource Neural Machine Translation: A Systematic Analysis. In Proceedings of the 14th International Joint Conference on Natural Language Processing (pp. 510–522). https://doi.org/10.18653/v1/2025.findings-ijcnlp.30