International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Lightweight Contextual Encoding and Ensemble Classification for Multilingual Text Disambiguation

C. Bagath Basha[1*], M. Ramana Kumar[2*], Edem Gunotham[3], Syed Khasim[3], Chirra Ashritha[3]
[1]Professor, [2]Associate Professor, [3]UG Student,[1,2,3]Department of Computer Science and Engineering
[1,2,3]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.
*Correspondence: C. Bagath Basha (basha@kpritech.ac.in), M. Ramana Kumar (ramana.reah@gmail.com)

Abstract
The rapid expansion of global connectivity has resulted in the widespread use of thousands of languages across digital platforms, with many users frequently communicating in multiple languages. Despite this linguistic diversity, a significant portion of multilingual content remains inaccurately classified due to the limitations of existing language identification techniques. Traditional manual approaches are time-consuming and error-prone, particularly when handling short, informal, or code-mixed text. Moreover, conventional algorithms often struggle to capture deeper semantic and contextual relationships inherent in multilingual data. To address these challenges, this study proposes a transformer-based multilingual language identification framework leveraging advanced Natural Language Processing (NLP) techniques. The process begins with a multilingual dataset subjected to preprocessing steps such as tokenization, stopword removal, and lemmatization. Exploratory Data Analysis (EDA) is then conducted to identify patterns and data distributions. Semantic features are extracted using Miniature Language Model (MiniLM), a lightweight transformer model capable of generating meaningful contextual embeddings. These embeddings are utilized by multiple machine learning classifiers, including Decision Tree (DT), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Random Forest (RF), to perform classification. Random Forest is employed as the primary model due to its robustness in handling high-dimensional data and its superior predictive performance. By integrating transformer-based embeddings with classical machine learning techniques, the proposed framework effectively handles short texts, informal language, and multilingual variations. The system is implemented as a Flask-based web application, enabling real-time classification and interactive user engagement.

Keywords: Multilingual Language Identification, Transformer-based Models, Natural Language Processing (NLP), MiniLM, Text Classification, Machine Learning, Code-Mixed Text

1. Introduction

Human societies are characterized by an extraordinary range of languages, encompassing thousands of spoken forms and diverse writing systems. These languages reflect the cultural heritage and identity of communities across the globe. Among the widely used languages, Urdu and English hold significant international relevance, being commonly spoken in regions such as South Asia, the Middle East, Africa, Europe, and parts of South America. In nations like Pakistan, both languages function as official modes of communication and are deeply integrated into major sectors such as governance, education, commerce, finance, and legal systems. Their widespread adoption emphasizes their importance in facilitating interaction across varied social and professional contexts. In addition to linguistic variation, visual communication through text plays a crucial role in daily life, particularly in public environments where information is presented through signboards, directional indicators, banners, and advertisements. People rely heavily on such visual text to navigate spaces and make decisions effectively. Research indicates that individuals are more inclined to focus on textual content within images compared to other visual elements, highlighting its significance in interpreting real-world scenes, as shown in figure 1.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

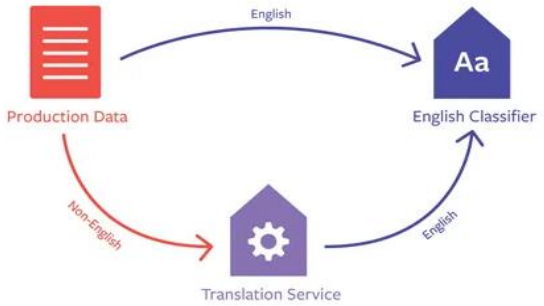

Figure. 1: Multilingual text classification.

This underscores the growing need for reliable text detection and recognition technologies. Extracting textual information from images enables numerous practical applications, including instant language translation, analysis of public information, advertisement assessment, and multimedia indexing. Furthermore, it contributes to enhanced scene interpretation in advanced fields like robotics and industrial systems. For instance, autonomous vehicles depend on precise text recognition to understand traffic signs and environmental instructions, ensuring safe and efficient navigation. As these technologies advance, the demand for accurate and robust text detection in complex real-world settings continues to grow.

2. Literature Survey

Mihailo Skoric, et al. [1] investigated the application of composite language models for processing texts in morphologically rich and highly inflected languages, with a specific focus on Serbian. Such languages present significant challenges due to their complex grammatical structures and word variations. To address this, the authors developed a perplexity-based dataset generated using transformer-based models trained on multiple representations of a Serbian language corpus. The dataset included diverse sentence categories such as expert translations, machine-generated translations, and deliberately corrupted text samples. By analysing perplexity scores across these categories, the study evaluated the capability of language models to distinguish between high-quality and degraded linguistic content. Their findings demonstrated that composite language models can effectively improve classification accuracy in binary tasks, particularly in linguistically complex environments.

Al-onazi, et al. [2] proposed a transformer-based framework for speech emotion recognition (SER), addressing the limited availability of research on Arabic emotional speech processing. The study emphasized the importance of robust feature representation and introduced a preprocessing pipeline that included data augmentation to enhance dataset diversity. A total of 273 acoustic features were extracted and provided as input to the transformer model. The system was evaluated across four benchmark datasets BAVED, EMO-DB, SAVEE, and EMOVO achieving high accuracy levels, with a peak performance of 95.2% on the BAVED dataset. The results highlighted the effectiveness of transformer architectures in capturing emotional nuances in speech, particularly for underrepresented languages like Arabic. Kwon, et al. [3] introduced an advanced multi-learning framework based on a one-dimensional enhanced Convolutional Neural Network (1D CNN) designed to extract both local and global emotional features from speech signals. Their approach incorporated a dynamic feature fusion mechanism to improve the discriminative power of extracted features. The model was tested on widely used datasets such as IEMOCAP and EMO-DB, achieving accuracy rates of 73% and 90%, respectively. Additionally, the framework demonstrated the ability to model both short-term and long-term dependencies within speech signals. However, the study also identified a limitation in terms of

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

increased computational cost, as the model required longer training and inference times compared to simpler approaches.

Tang, et al. [4] developed an end-to-end deep neural network model for speech emotion recognition, aiming to improve performance in regression-based emotional prediction tasks. Their proposed DiCCOSER-CS model incorporated techniques such as root mean square (RMS) aggregation and context stacking to better capture temporal and contextual information within speech signals. Experimental results showed significant improvements over baseline CNN-LSTM models, with increases of 9.5% in arousal concordance correlation coefficient (CCC) and 12.7% in valence CCC. This study demonstrated the potential of end-to-end architectures in enhancing continuous emotion prediction accuracy. In another study [5], researchers focused on detecting anger in Arabic speech by constructing a dataset derived from real-world conversational dialogues. The study emphasized the importance of acoustic features such as fundamental frequency (pitch), signal energy, and formants in accurately identifying emotional states. Using support vector machine (SVM) classifiers, the system achieved an accuracy exceeding 77% in real-time anger detection. The findings underscored the relevance of carefully selected acoustic features for emotion recognition tasks, particularly in conversational and real-world scenarios.

Masethe, et al. [6] addressed the challenge of lexical ambiguity in Sesotho sa Leboa, a language characterized by homonyms and polysemous words that complicate semantic interpretation. The study highlighted the difficulty of accurately determining word meanings and grammatical roles in such contexts. To overcome these challenges, the authors proposed a hybrid word sense disambiguation (WSD) framework that integrates transformer-based contextual embeddings with deep learning techniques. This approach improved semantic understanding and demonstrated the potential of hybrid architectures for low-resource and morphologically complex languages. Shafi, et al. [7] conducted research on semantic tagging for Urdu, utilizing a manually annotated dataset consisting of 8000 tokens across multiple domains. Their approach employed supervised multi-target classification techniques to assign semantic labels to words. The study achieved an impressive accuracy of 94% in coarse-grained semantic classification. Additionally, the authors discussed the effectiveness of various supervised learning algorithms including neural networks, K-nearest neighbors (KNN), support vector machines (SVM), decision trees, and Naive Bayes in performing word sense disambiguation tasks. Their work highlights the importance of annotated datasets and supervised methods in improving semantic analysis for Urdu.

Demlew, et al. [8] explored word sense disambiguation in Amharic, a low-resource and morphologically rich language facing challenges similar to Sesotho sa Leboa. Their approach combined both supervised and unsupervised techniques, leveraging neural word embeddings along with Bidirectional Gated Recurrent Unit (BiGRU) models and transformer-based contextual embeddings. This hybrid strategy enabled more accurate handling of polysemy and homonymy, demonstrating the effectiveness of combining traditional and modern deep learning approaches for semantic disambiguation in under-resourced languages. Researchers in [9] developed a comprehensive dataset consisting of one hundred polysemous Arabic expressions, each associated with multiple meanings ranging from three to eight interpretations. Each expression was accompanied by ten example sentences to illustrate contextual usage. Statistical analysis of this dataset provided insights into linguistic variability and complexity. Building on this, BERT-based approaches [10] were introduced for word sense disambiguation, utilizing contextual embeddings and similarity measures to map words to their appropriate meanings. These methods demonstrated strong performance in capturing contextual nuances in Arabic language processing.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Rahali, et al. [11] conducted an extensive survey of transformer-based models, focusing on their architectural design and the role of self-attention mechanisms in capturing long-range dependencies within textual data. The study provided a comparative analysis of different transformer variants, evaluating their strengths and limitations across various NLP tasks. Additionally, the authors identified key research challenges and proposed potential future directions for improving transformer-based systems. The seminal work by Vaswani, et al. [12] introduced the Transformer architecture, which marked a significant breakthrough in deep learning and natural language processing. By replacing recurrent and convolutional structures with self-attention mechanisms, the Transformer enabled more efficient parallel processing and improved performance across a wide range of NLP tasks. Subsequent research [13] demonstrated that attention-based models consistently outperform traditional CNN and RNN architectures, particularly in tasks involving long-range dependencies.

Finally, Lakew, et al. [14] highlighted the effectiveness of transformer models in multilingual applications, showing that they outperform both bilingual models and RNN-based approaches, especially in zero-shot translation scenarios. In general, natural language processing tasks involving sequence-to-sequence (S2S) modelling rely on encoder-decoder architectures [15], where transformers have become the dominant framework due to their scalability, efficiency, and superior performance.

3. Proposed Methodology

The proposed methodology presents a systematic framework for the analysis and classification of multilingual textual data by integrating NLP and machine learning techniques. The overall pipeline begins with data acquisition and organization, followed by preprocessing and semantic feature extraction. Initially, raw text data is cleaned and standardized to eliminate noise, inconsistencies, and irrelevant information. A transformer-based embedding model, Mini LM, is then employed to generate contextual representations of the text, effectively capturing semantic relationships and linguistic structures.

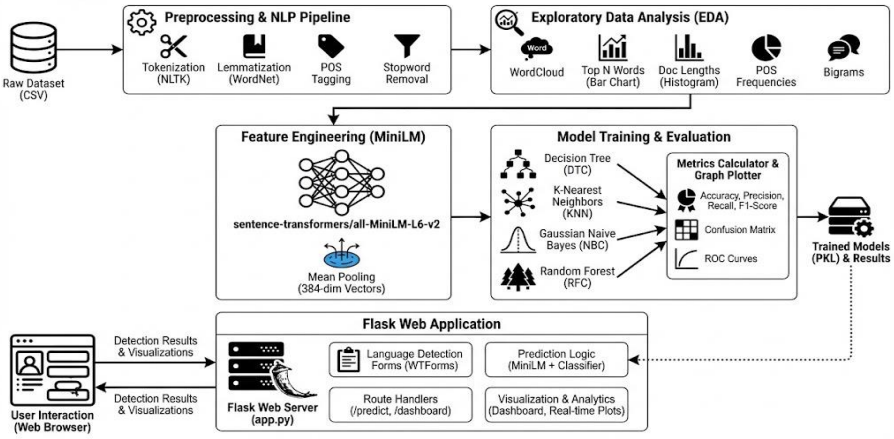

Figure. 2: System architecture.

These embeddings are subsequently utilized as input features for multiple machine learning classifiers, including DT, KNN, GNB, and RF, to perform language classification. Additionally, a validation module is incorporated to assess and refine textual inputs prior to classification. The system is supported by a graphical interface that facilitates user interaction for dataset management, preprocessing, feature extraction, model training, evaluation, and prediction, as illustrated in Figure 2. A lightweight storage mechanism is used to manage trained models and processed data, while a server-based architecture enables real-time prediction and analysis. Continuous evaluation and retraining ensure improved accuracy and adaptability.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

User Interface (Client Application): The system provides a user-friendly graphical interface developed using either a desktop-based or web-based platform. This interface allows users to perform operations such as login, dataset upload, preprocessing, feature extraction, model training, performance comparison, and prediction. Users can either manually input text or upload text files for analysis. All inputs received through the interface are forwarded to the backend processing modules for further analysis.

Flask Application Server: The Flask server functions as the core communication layer of the system. It handles incoming requests from client applications and directs them through the processing pipeline. The server is responsible for executing preprocessing, feature extraction, classification, and response generation tasks. It also supports remote access, enabling users to submit text data and receive classification results efficiently.

Lightweight Database (Authentication & Storage): A lightweight database is utilized to manage system data and user authentication. It stores user credentials, processed datasets, and trained machine learning models. The database interacts with the application layer to facilitate secure login verification and efficient data retrieval. This ensures reliable storage and quick access to essential resources.

Text Dataset (Multilingual Data Collection): The dataset serves as the primary input for the system and consists of multilingual textual data collected from various sources. It includes sentences, phrases, and documents representing diverse linguistic patterns. This dataset is used for both training and evaluating the performance of classification models.

Text Preprocessing and Feature Extraction: The collected raw text undergoes several preprocessing steps, including lowercasing, tokenization, stopword removal, and lemmatization. Unnecessary elements such as special characters and noise are removed to enhance data quality. Following preprocessing, the Mini LM transformer model is applied to extract deep semantic embeddings, converting textual data into numerical feature vectors suitable for machine learning algorithms.

ML Classification Models: The generated feature vectors are analysed using multiple machine learning classifiers to perform language identification:

- **DT:** Utilizes rule-based hierarchical structures for classification.
- **KNN:** Determines class labels based on similarity between feature vectors.
- **GNB:** Applies probabilistic modelling based on feature distributions.
- **RF:** Employs an ensemble of decision trees to improve classification robustness.

Each model independently predicts the language class, allowing comparative evaluation of performance.

Text Analysis and Validation Module: A validation component is integrated into the system to ensure the quality and reliability of input text. It checks whether the input meets predefined criteria such as minimum length and content validity. This module filters out unsuitable inputs and enhances the overall reliability of the classification process.

Prediction Results and Output Generation: After processing, the system generates predictions indicating the language category of the input text. The results are displayed through the graphical interface along with confidence scores. Additionally, model-wise predictions and performance metrics are presented, enabling users to interpret and compare results effectively.

Remote Text Prediction Workflow: The architecture supports remote text classification through a client–server communication model. Users can submit text inputs to the Flask server from external systems. The server processes these inputs using trained models and returns structured predictions, enabling scalable and real-time classification.

Model Evaluation and Retraining: The performance of classification models is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Visualization tools, including

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

confusion matrices and performance graphs, are used to analyse results. The system also supports retraining with new data, ensuring continuous learning and improved adaptability over time.

3.1 Mini LM Feature Extraction

The MiniLM performs feature extraction through a knowledge distillation framework involving two networks: a teacher model and a student model. The Teacher model is a large pre-trained transformer that captures deep contextual relationships between words using multiple transformers blocks as shown in Figure 3. The student model, being smaller and faster, learns to replicate the internal attention behaviours of the Teacher through two key processes Attention Transfer and Value-Relation Transfer. This allows the student model to preserve the semantic richness of the Teacher while maintaining computational efficiency, producing powerful language embeddings suitable for multilingual language identification.

Input Encoding: The process begins with an input text sequence represented as tokens x1, x2, x3, …, xnx_1, x_2, x_3, \dots, x_nx1, x2, x3, …, xn, which are converted into embedding vectors. These embeddings are fed into multiple stacked Transformer Blocks in both the Teacher and Student models. Each block applies self-attention mechanisms that capture dependencies among tokens, helping the model understand how words relate contextually within a sentence.

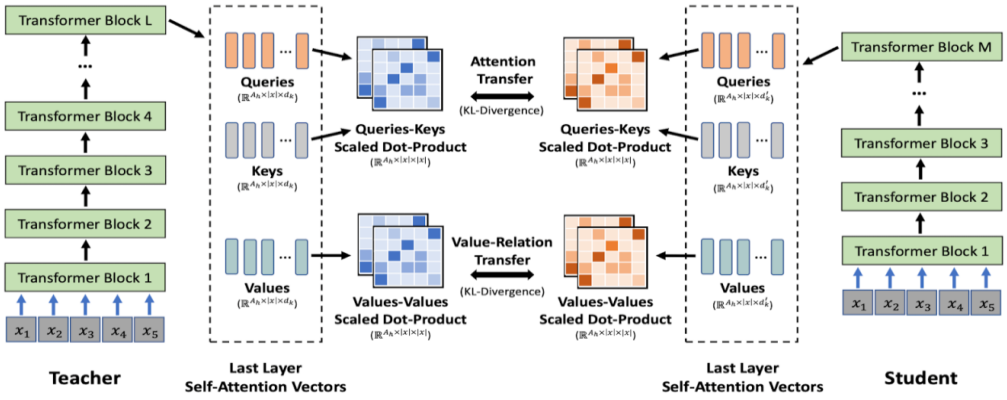

Figure. 3: Architecture of Mini LM feature extraction.

Teacher Model Representation Learning: The Teacher network consists of several transformer layers (Transformer Block 1 to L). Each layer computes Queries (Q), Keys (K), and Values (V) vectors through self-attention, modelling how each token attends to others in the sequence. The Teacher's final layer generates the last-layer self-attention vectors, which encapsulate rich syntactic and semantic relationships learned from large-scale language data. These representations serve as the foundation for knowledge transfer to the student model.

Knowledge Distillation via Attention Transfer: Mini LM employs Attention Transfer to distil contextual understanding from the Teacher to the Student. The student learns the Queries–Keys scaled dot-product relationships that define the Teacher's attention maps. Using Kullback–Liebler (KL) Divergence), the student minimizes the difference between its own attention distribution and that of the Teacher. This ensures the student effectively inherits how the Teacher model allocates attention among words, preserving contextual focus and linguistic nuances.

Value-Relation Transfer: In addition to attention-based learning, Mini LM performs Value-Relation Transfer, aligning the Values–Values scaled dot-product between the Teacher and Student. This process captures how semantic meaning flows and interacts across tokens. Again, KL-Divergence is used to align the relational structure of the Value vectors, ensuring that even with fewer parameters, the student model retains the Teacher's semantic depth and relational consistency.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Student Model Learning and Embedding Generation: The Student model, comprising fewer transformer blocks (Transformer Block 1 to M), receives the same input token sequence. Through the Attention and Value-Relation transfer processes, it learns to replicate the Teacher's representational behaviour. After training, the student generates compressed but semantically rich embeddings that carry contextual and linguistic information equivalent to the Teacher's representations but at a fraction of the computational cost.

Feature Extraction for Language Identification: The final embeddings produced by Mini LM serve as high-quality feature vectors that encode cross-lingual, syntactic, and semantic information. These embeddings are extracted and used as inputs for downstream classifiers such as DT, KNN, GNB, and RF in the language identification pipeline. Their strong contextual representation enables the system to accurately distinguish between multiple languages, even in short, noisy, or code-mixed text samples.

4. Result Description

The result analysis phase presents the outcomes obtained after applying preprocessing, feature extraction, and machine learning techniques on the textual dataset. It provides a clear understanding of how effectively the system performs in classifying and analysing the input data. The results are evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. These metrics help in assessing the reliability and consistency of different models used in the study. Visualization techniques such as graphs and confusion matrices further support the interpretation of results. By comparing multiple models, the analysis highlights variations in performance and identifies the most suitable approach. This phase plays a vital role in validating the effectiveness of the analytical framework.

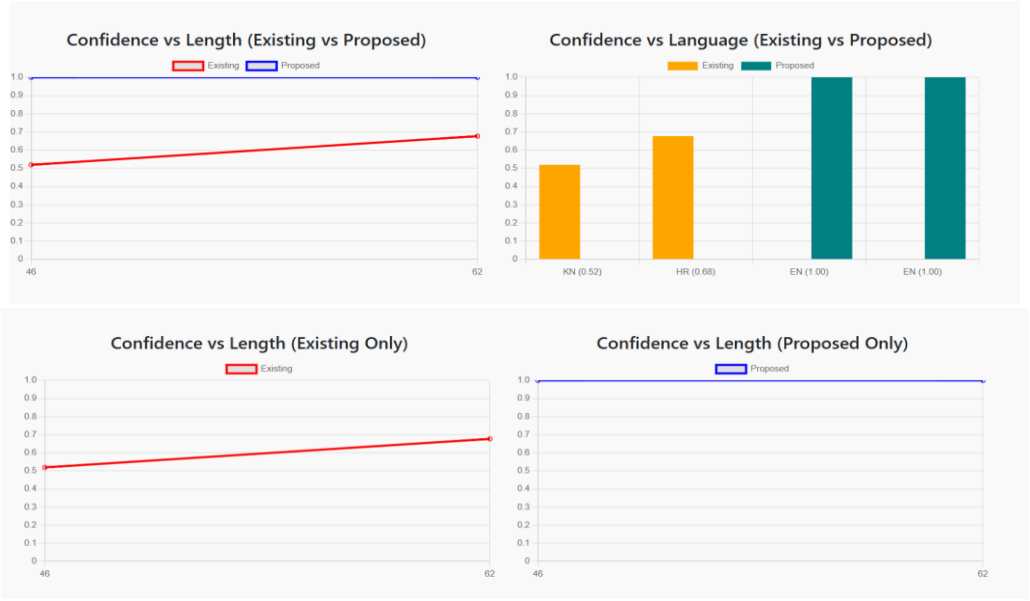

Figure. 4: Visualization (Existing vs Proposed Detection)

Figure 4 offers a direct comparison between the outputs of the RF and Mini LM Transformer for the same set of test inputs. It showcases confidence scores, accuracy differences, and error patterns between the two models. Bar charts and line plots illustrate the Transformer's ability to deliver higher confidence and accuracy across multiple languages. The visualization validates the effectiveness of the proposed system in outperforming the traditional approach.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

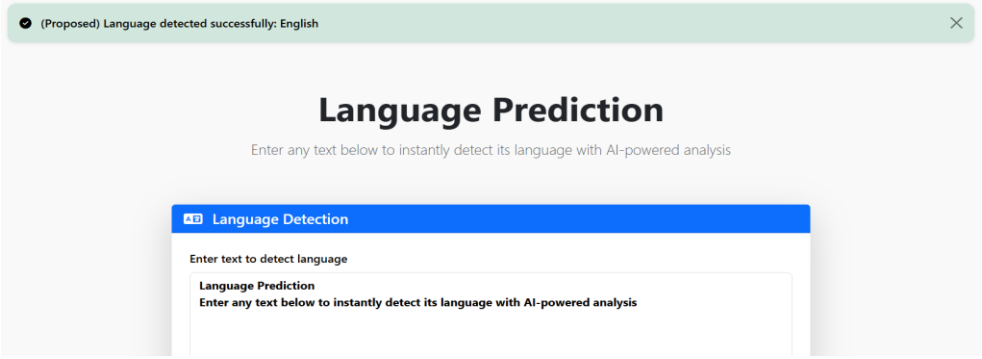

Figure. 5: Language prediction by using the proposed model.

Figure 5 illustrates the prediction results generated by the Mini LM Transformer. The system outputs the detected language and its confidence score with higher reliability and consistency than the RF. The Transformer leverages pretrained multilingual embeddings and contextual understanding, which enables it to handle diverse inputs, including short phrases and longer text passages. This figure emphasizes the effectiveness of the proposed approach in delivering accurate, real-time language predictions for multilingual scenarios.

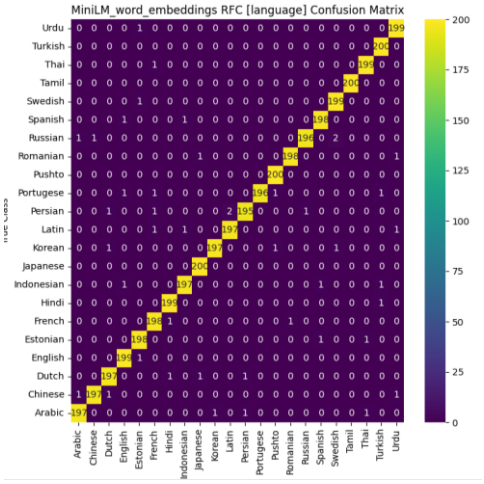

Figure. 6: Confusion matrix obtained using Mini LM word embeddings of RF.

Figure 6 showcases confusion matrices for language classification using MiniLM word embeddings across RF demonstrates near-perfect classification with a sharp, dominant diagonal and minimal off-diagonal values (e.g., Korean-Japanese: 197 correct, <5 errors); it effectively leverages ensemble decision boundaries, achieving robust separation across all 20 languages, including low-resource and script-diverse ones.

Figure 7 presents ROC curves for language classification using MiniLM word embeddings across RF demonstrates ideal ROC curves hugging the top-left corner, achieving AUC = 1.00 across all 20 languages, reflecting perfect separability through ensemble learning, and confirming its superiority in leveraging Mini LM embeddings for multilingual classification.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

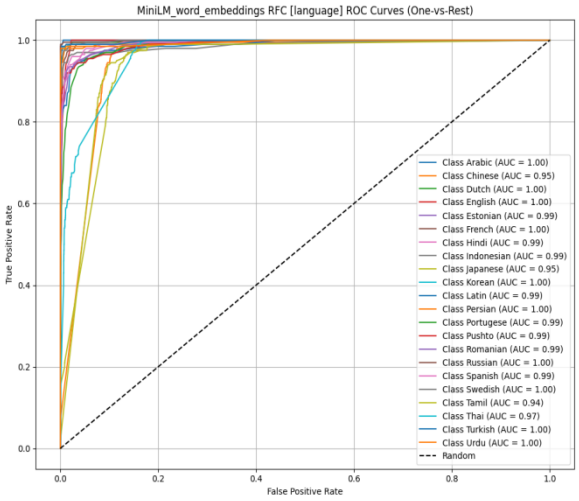

Figure. 7: ROC Curve obtained using Mini LM word embeddings of RF.

Table 1 presents the performance comparison of Mini LM-based feature embeddings across different machine learning algorithms and reveals a clear hierarchy in predictive effectiveness. The DT achieved moderate performance, with an accuracy of 64.16% and an F1-score of 63.37%, indicating limited generalization for language detection. The GNB improved slightly, achieving 72.86% accuracy and 71.43% F1-score, while the KNN classifier demonstrated better performance with 80.68% accuracy and 78.97% F1-score, reflecting its ability to leverage similarity in feature space more effectively. Notably, the RF significantly outperformed all baseline algorithms, achieving near-perfect results across all metrics (99% accuracy, precision, recall, and F1-score), highlighting its robustness and superior capability in capturing complex patterns in Mini LM embeddings for reliable language detection.

Table 1: Performance comparison of DT, KNN, GNB and RF algorithms.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	64.16	64.57	64.16	63.37
KNN	80.68	85.68	80.68	78.97
GNB	72.86	73.57	72.86	71.43
RF	99.00	99.00	99.00	99.00

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

5. Conclusion

The developed language identification framework combines Mini LM-based transformer embeddings with classical machine learning models such as DT, KNN, GNB, and RF to achieve reliable and efficient classification of multilingual text. By utilizing rich semantic representations, the system ensures consistent performance across all stages, including preprocessing, feature extraction, model training, and evaluation. This approach effectively balances the contextual understanding provided by transformer models with the simplicity and interpretability of traditional classifiers, making it suitable for real-world applications such as multilingual communication, content analysis, and information retrieval. Implemented using open-source tools like Hugging Face Transformers and scikit-learn, the framework is scalable, flexible, and cost-efficient. It delivers a dependable and interpretable solution, with potential improvements through real-time deployment, incorporation of advanced architectures, and better adaptation to diverse linguistic variations.

References

[1] Skorić, M.; Utvić, M.; Stanković, R. Transformer-Based Composite Language Models for Text Evaluation and Classification. Mathematics 2023, 11, 4660. https://doi.org/10.3390/math11224660

[2] Al-onazi, B.B.; Nauman, M.A.; Jahangir, R.; Malik, M.M.; Alkhammash, E.H.; Elshewey, A.M. Transformer-Based Multilingual Speech Emotion Recognition Using Data Augmentation and Feature Fusion. Appl. Sci. 2022, 12, 9188. https://doi.org/10.3390/app12189188

[3] Kwon, S. MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach. Expert Syst. Appl. 2021, 167, 114177.

[4] Tang, D.; Kuppens, P.; Geurts, L.; van Waterschoot, T. End-to-end speech emotion recognition using a novel context-stacking dilated convolution neural network. EURASIP J. Audio Speech Music Process. 2021, 2021, 1–16.

[5] Khalil, A.; Al-Khatib, W.; El-Alfy, E.S.; Cheded, L. Anger detection in arabic speech dialogs. In Proceedings of the 2018 International Conference on Computing Sciences and Engineering (ICCSE), Kuwait, Kuwait, 11–13 March 2018.

[6] Masethe, H.D.; Masethe, M.A.; Ojo, S.O.; Owolawi, P.A.; Giunchiglia, F. Hybrid Transformer-Based Large Language Models for Word Sense Disambiguation in the Low-Resource Sesotho sa Leboa Language. Appl. Sci. 2025, 15, 3608. https://doi.org/10.3390/app15073608

[7] Shafi, J.; Nawab, R.M.A.; Rayson, P. Semantic Tagging for the Urdu Language: Annotated Corpus and Multi-Target Classification Methods. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 2023, 22, 1–32.

[8] Demlew, G.; Yohannes, D. Resolving Amharic Lexical Ambiguity using Neural Word Embedding. In Proceedings of the 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), Bahir Dar, Ethiopia, 28–30 November 2022; IEEE: Piscataway, NJ, USA, 2022.

[9] Kaddoura, S.; Nassar, R. EnhancedBERT: A feature-rich ensemble model for Arabic word sense disambiguation with statistical analysis and optimized data collection. J. King Saud Univ. Comput. Inf. Sci. 2024, 36, 101911.

[10] Agbesi, V.K.; Chen, W.; Yussif, S.B.; Hossin, A.; Ukwuoma, C.C.; Kuadey, N.A.; Agbesi, C.C.; Samee, N.A.; Jamjoom, M.M.; Al-Antari, M.A. Pre-Trained Transformer-Based Models for Text Classification Using Low-Resourced Ewe Language. Systems 2025, 12, 1.

[11] Rahali, A.; Akhloufi, M.A. End-to-End Transformer-Based Models in Textual-Based NLP. AI 2023, 4, 54-110. https://doi.org/10.3390/ai4010004

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[12] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.

[13] Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-attention with relative position representations. arXiv 2018, arXiv:1803.02155.

[14] Lakew, S.M.; Cettolo, M.; Federico, M. A comparison of transformer and recurrent neural networks on multilingual neural machine translation. arXiv 2018, arXiv:1806.06957.

[15] Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 3104–3112.