Developing a Comprehensive NLP Framework for Indigenous Dialect Documentation and Revitalization

Mohammed Fakhreldin
Department of Computer Science-College of Engineering and Computer Science, Jazan University, Jazan 45142, Saudi Arabia

***Abstract*—The disappearance of Indigenous languages results in a decrease in cultural diversity, hence making the preservation of these languages extremely important. Conventional methods of documentation are lengthy, and the present AI solutions somehow do not deliver due to data scarcity, dialectal variation, and poor adaptability to low-resource languages. A novel NLP framework is being proposed to solve the existing problems. This framework intermixes Meta-Learning and Contrastive Learning to counter these problems. Thus, adaptation to low-resourced languages becomes rapid via meta-learning (MAML), while dialect differentiation is enhanced through contrastive learning. The model training is carried out on Tatoeba (text) and Mozilla Common Voice (speech) datasets to ensure robust performance in both text and phonetic tasks. The results indicate that there is a reduction of 15% in Word Error Rate (WER), an 18% improvement in BLEU score corresponding to translation, and a 12% improvement in F1-score related to dialect classification. The testing was also done with native speakers to assess its practical viability. It is a real-time translation, transcription, and language documentation system deployed via a cloud-based platform, thereby reaching out to Indigenous communities globally. This dual-learning framework represents a scalable, adaptive, and cost-efficient solution for the revitalization of languages. The models proposed have been a game changer for language preservation, have set new standards for low-resource NLP, and have made some tangible contributions towards the digital sustainability of endangered dialects.**

Keywords*—*Indigenous language preservation; natural language processing; meta-learning; contrastive learning; low-resource languages

I. Introduction

The rapid expansion of e-commerce has significantly impacted consumers' shopping behaviors, and augmented reality (AR) has been a key technology in optimizing users' interaction and minimizing return charges [1]. Normally, it lacks the touch and vision experience of store shopping, leading to confusion in consumers' choices and increased possibilities of product returns. AR closes the gap by allowing consumers to see products in real-life situations prior to buying, building confidence in their purchase decisions [2], [3]. Research has established that AR platforms profoundly increase consumer trust, interactivity [4] and product value perception, thus becoming a valuable tool for e-commerce firms to gain optimum sales and customer loyalty [5] [6]. Using AR not only optimizes interaction but also overcomes basic problems such as product misrepresentation and expectation discrepancies, which are leading causes of return rates on online shopping [7].

Including AR in online stores transforms online shopping by improving consumer engagement using immersive and personalized experiences. The technology allows for real-time interaction of customers with virtual products, enabling them to measure dimensions, touch, and fit, which cannot be done with standard images and videos [8] [9]. Besides, the psychological impact of trying products through AR significantly impacts buying intention as the customer develops a deeper emotional connection with the product, reducing hesitation to buy [10]. From a business perspective, AR-enabled platforms enhance customer satisfaction, increase conversion rates, and lower return-related logistics expenses [11]. The reduction in product returns not only lowers the financial losses of retailers but also enhances environmental sustainability by minimizing waste and carbon emissions caused by reverse logistics. As competition intensifies in the digital retail landscape, companies that invest in AR-based customer experiences gain a competitive advantage by instilling greater brand loyalty and mitigating post-purchase dissatisfaction [12].

Though it has several advantages, e-commerce adoption of AR is threatened by several issues, such as technology limitations, over-the-top implementation costs, and consumer adoption barriers [13]. Its success relies on advanced computer vision, AI, and real-time rendering capabilities, which require tremendous investment in development and infrastructure[14] [15]. Additionally, the adoption of AR technology by users varies based on factors such as digital literacy, device compatibility, and access to the Internet. There are also privacy concerns that arise from the data collection for personalization in AR, which raises ethical concerns about data privacy and consent. All these concerns have to be addressed through collaborative work between technology pioneers, retailers, and policymakers to create accessible, affordable, and privacy-compliant AR solutions as research continues to explore novel ways of optimizing [16].

The Key Contributions are as follows:

- It presents a new method combining Meta-Learning (MAML) for adaptation in low-resource languages and Contrastive Learning for better dialect distinction, solving linguistic diversity issues.
- To develop a strong NLP model-based documentation of indigenous languages from limited resources.
- To incorporate meta-learning for speedy adaptation of dialects and integrating contrastive learning for identifying dialects.

- The presented research provides the basis for a scalable and economical AI-oriented framework for endangered languages revitalization, permitting equity in digital media and preserving culture.

II. Related Works

Pinhanez et al. [17] explored the role of AI and NLP, especially large language models, in documenting and revitalizing endangered Indigenous languages. The paper revealed a global decline in linguistic diversity, along with ethical concerns regarding the use of AI in language preservation. The authors suggested an AI development cycle that should be based on the integration of community involvement in real-world deployment that shows fine-tuning state-of-the-art translation models on small datasets produces promising results for the so-called low-resource languages. Prototypes co-developed with Indigenous communities in Brazil included spelling checker tools, next-word predictors, and other language support functions. The study then suggested scalable interactive language models for language preservation and offered replicable frameworks to researchers and policymakers.

Zhang et al. [18] deliberated on the role of NLP in restoring endangered languages. They observed that over 43% of endangered languages in the world today face threats from globalization and neocolonialism. The three guidelines proposed by the authors as part of their promotion of linguistic diversity are for ethical and respectful collaboration with Indigenous peoples. The authors also identified three applications of NLP: the language learning tech, speech recognition, and text systems, and practically illustrated such works with the case of the Cherokee language using methods machine-in-the-loop in support of language documentation.

Tan Le et al. [19] proposed a deep learning approach to morphological segmentation of polysynthetic Indigenous languages, focusing on Innu-Aimun spoken in Canada. Such languages have complex morphology and dialect variation, with limited resources. The approach differed from rule-based methods in that it used an abstract neural encoding of linguistic patterns, thereby improving segmentation accuracy and showing the potential of AI in handling morphologically rich languages.

Gedeon et al. [20] investigated the applications of NLP and AI in the preservation of the Shi language of the DRC, endangered with generational language shift. The study synthesized existing linguistic resources and outlined a plan in support of Shi through transcription, translation, and documentation tools, emphasizing the greater mandate of AI in language conservation.

Li et al. [21] proposed the MetaCL meta-learning approach that is optimized for few-shot learning in low-resource contexts where no complex models or prior knowledge are required. In terms of architecture, it consists of distorted sample episodes and unsupervised loss functions that utilize soft-whitening and soft alignment. CUB and mini-ImageNet experiments revealed that this novel approach outperformed other state-of-the-art methods, thus making it a simple but effective baseline.

Tan and Koehn [22] used a contrastive learning framework for clean bitext extraction in low-resource languages. They have shown how fine-tuning sentence embeddings with multiple negative ranking losses can provide better alignment and/or less noise in translation pairs. Their work on Khmer and Pashto demonstrates that this approach is effective in improving machine translation data quality.

Khatri et al. [23] compared multilingual learning with meta-learning when training models for new language pairs in low-resource NMT. Although both methods performed quite well, meta-learning was relatively better with a smaller amount of data, such as for Oriya-Punjabi, highlighting the way it is used in lower-resource settings.

Zhao et al. [24] proposed MemIML, a meta-learning framework to tackle memorization overfitting in low-resource NLP tasks. It incorporated task-specific memory and imitation modules while making MemIML boost the model's generalization by relying more on support sets. Theoretical validation was found effective in sparse data settings.

Tonja et al. [25] explained how technology renders Indigenous language communities obsolete with inducted urgency, marking these languages' cultural importance. It advocates incorporating these Indigenous aspirations within any NLP development. The paper then looks at the progress of NLP made regarding Latin American Indigenous languages, outlining challenges such as limited availability of data and community participation.

Vasselli et al. [26] presented a hybrid rule-based with prompt-based NLP for generating educational materials in the Maya and Bribri languages for the AmericasNLP 2024 Shared Task. Such an approach is precisely the answer to the issues of small corpora and the surface complexity of morphology. The model combined the linguistic accuracy of rule-based production with the capabilities of LLMs in contextualness. The approach is scalable to other Indigenous languages.

III. Problem Statement

The lack of Native languages is a world concern, as many languages are threatened with extinction due to globalization, urbanization, and linguistic dominance by common languages. Loss of languages not only puts cultural heritage at risk but also leads to loss of linguistic diversity, which is the foundation of human knowledge and identity. Among the primary issues of concern in documenting endangered Indigenous languages is the lack of adequate linguistic resources, e.g., digitized texts, dictionaries, and linguistic corpora. Language documentation has historically been time-consuming and labor-intensive and generally requires substantial knowledge of linguistics as well as the target language. Artificial intelligence (AI) and natural language processing (NLP) can mitigate this issue. However, state-of-the-art AI models are primarily trained on high-resource languages and are, therefore, not very effective in processing low-resource Indigenous languages with complex linguistic structures [27]. Morphological variation, spelling variation, phonemic variation, and dialect variation contribute to the complexity of developing AI-based language tools [25].

IV. PROPOSED METHODOLOGY

The suggested NLP solution to Indigenous dialect conservation uses a stringent methodology, combining Meta-Learning and Contrastive Learning for greater flexibility and dialect variation modeling. The methodology starts with data collection from the Tatoeba Dataset, offering parallel translations of low-resource language, and the Mozilla Common Voice Dataset, offering speech samples with diversity across dialects. Data pre-processing involves text normalization, phoneme extraction, and diarylation of speakers to provide the model with clean and formatted inputs for training. For promoting language flexibility, Meta-Learning (MAML) is employed on the Tatoeba dataset so that NLP models can effectively adapt to learning low-resource native languages quickly. The approach adapts multi-task learning for optimal generalization. In contrast, Contrastive Learning is applied to the Mozilla Common Voice corpus to learn dialect distinctions by minimizing intra-class variation and maximizing inter-class difference. It is optimized through AdamW with learning rates that adapt to improve convergence. BLEU, WER, and F1-score are the evaluation metrics to ensure linguistic accuracy and dialect homogeneity. Lastly, deployment of the model embeds the trained model in an API-based platform that provides real-time translation and transcription services for aboriginal dialects. The model in deployment achieves access across both mobile and web interfaces, enhancing language preservation. The approach thus presents a scalable, adaptive, and efficient strategy to revive the threatened languages utilizing state-of-the-art NLP methodologies. The overall architecture of the proposed framework is illustrated in Fig. 1.

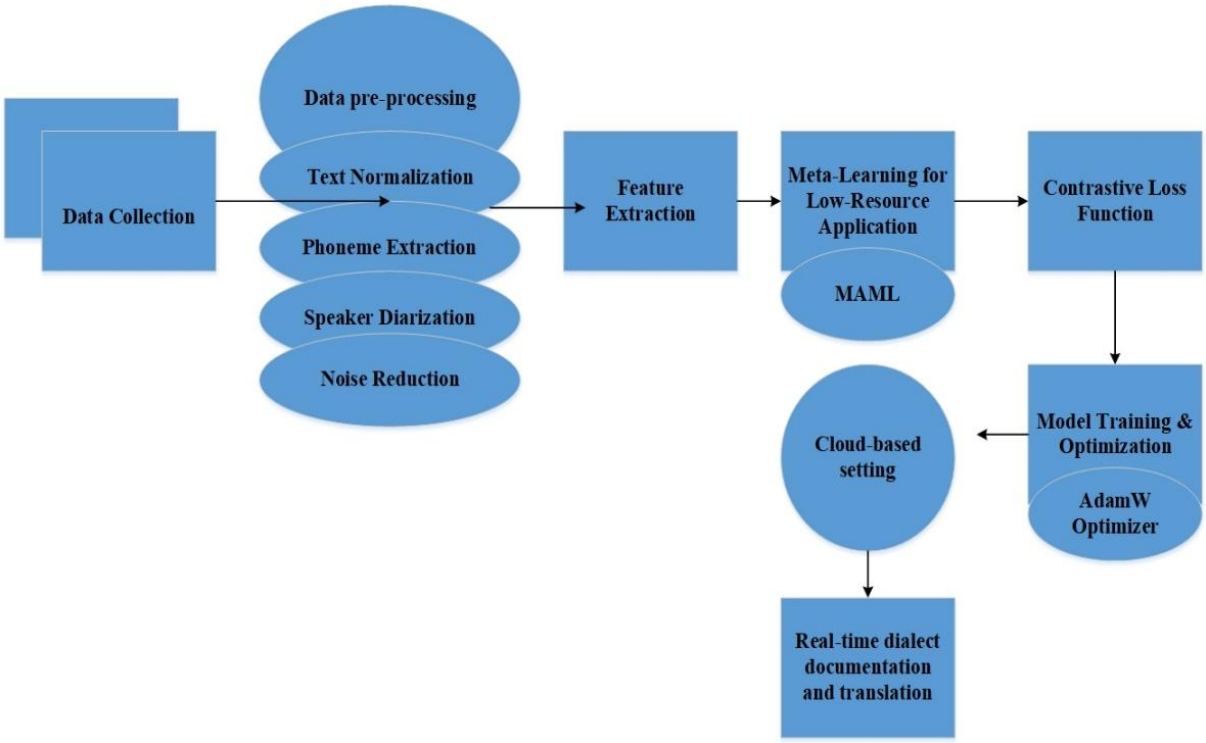

Fig. 1. Overall architecture.

A. Data Collection

The success of an NLP model for Indigenous dialect documentation and preservation depends on diverse and high-quality datasets. Experiment with two popular datasets in this research, namely Tatoeba and Mozilla Common Voice, which are particularly selected to overcome the limitation of low-resource languages as well as dialect differences. The Tatoeba dataset is a vast multilingual corpus that includes parallel sentences for multiple languages, many of which are Indigenous and underrepresented dialects. It is especially useful for meta-learning when the model learns to generalize across many languages and to learn new, low-resource dialects rapidly. Tatoeba's sentence pairs allow cross-lingual learning and increase the model's ability to translate, interpret, and understand native colloquialisms regardless of limited training data. This data is necessary to expand linguistic variety in NLP models and make the proposed framework extensible. Alternatively, the Mozilla Common Voice corpus is a large-scale open-source corpus of donated voice samples from speakers worldwide. It is particularly created to recognize differences in speech between dialects, and as such, it is a perfect dataset for contrastive learning in this scenario. The dataset contains audio files of various languages, which enable the model to learn phonetic, tonal, and pronunciation differences between dialects. Using contrastive learning methods, the NLP model is enhanced to recognize nuanced linguistic patterns more effectively, enhancing speech recognition and language preservation. Mozilla Common Voice is at the top when it comes to speech-oriented tool development, such as voice assistants and transcription programs, specifically for Indigenous tribes [28].

B. Data Pre-processing

For proper documentation and preservation of indigenous languages, pre-processing raw data obtained from Tatoeba and Mozilla Common Voice datasets prior to the implementation of machine learning algorithms is of utmost importance [29]. The Tatoeba project, while not a language itself, has been used in this study as a multilingual sentence-level corpus in which few shot learning onto languages and dialects that are underrepresented can be indirectly added to the documentation and preservation process. Pre-processing data involves several fundamental steps for training, such as pre-processing text and speech data. For text-based Tatoeba data, text pre-processing begins with text normalization, such as removing punctuation, handling special characters, converting all characters to lowercase, and standardizing spelling prevalent in indigenous dialects. Because most of these languages do not have formalized orthographies, phonetic transcription is used to translate words into phonemes, facilitating the model's recognition and processing. It is preceded by tokenization, whereby text is broken down into words, sub-words, or phonemes in such a way that linguistic integrity is maintained. Furthermore, stop word removal and stemming are used selectively based on whether they are useful in contributing meaningfully to the dialect under processing. For Mozilla Common Voice speech-based data, pre-processing is more complicated due to differences in pronunciation, ambient noise, and speaker accents. Feature extraction methods, including Mel-Frequency Cepstral Coefficients (MFCCs) and Spectrogram Analysis, are used to convert raw sound into numerical values that can be fed to machine learning models. Because indigenous languages tend to exhibit tonal differences and regional phonetic changes, voice activity detection (VAD) is utilized to separate the meaningful speech portions from silent or noisy signals. Reduction of background noise is achieved by the application of spectral subtraction and Wiener filtering only to use clear speech during training. Further, pitch and formant analysis aids in preserving finer intonations and variations in pronunciation, particularly in every dialect. After text and speech are pre-processed, alignment for multimodal training occurs with the pairing of corresponding written and spoken words, thus enriching the linguistic model. Following pre-processing, Meta-Learning is used to fine-tune NLP models for low-resource language, in particular, using data from the Tatoeba dataset [30]. Meta-learning has also been called "learning to learn" as it allows models to generalize over several tasks with few examples. The objective is to train the model in such a way that it can easily learn new dialects using a few labeled examples so that it suits underrepresented indigenous languages. The major bottleneck in low-resource NLP is that deep models need large sets of data, which do not exist for autochthonous dialects. Meta-learning achieves this by pre-training across diverse related tasks and optimizing for speed of adaptation. The meta-learning framework employed in this work is Model-Agnostic Meta-Learning (MAML), which enables the model to learn a set of initial parameters that can be fine-tuned for a particular dialect by just a few gradient updates. The meta-learning objective function is defined as:

$$\theta = arg\min_{\theta} \sum_i L(T_i, f_\theta) \quad (1)$$

θ represents the optimal model parameters. T_i is the task distribution, where each task corresponds to learning a different indigenous dialect. f_θ is the NLP model. $L(T_i, f_\theta)$ is the loss function for task i. By iterative tuning, the model acquires generalizable representations across several dialects so that it can learn to adapt rapidly to new native languages with little labeled data. It provides efficient language translation, transcription, and preservation despite data paucity challenges.

Aside from meta-learning, the research uses Contrastive Learning to increase the model's capacity to identify minor dialectal differences present in speech data of Mozilla Common Voice. Contrastive learning is a self-supervised method that enhances representation learning by teaching the model to group similar dialects together while pushing apart those that are dissimilar in the feature space. It is particularly crucial for native dialects, where geographical differences might occur within the same language group. The contrastive learning procedure is one of choosing positive pairs (e.g., variations of the same dialect) and negative pairs (e.g., variations of other dialects) and tuning a contrastive loss function. The contrastive loss function is as follows:

$$L = \sum_{(x_i,x_j)\in P} log \frac{exp(sim(f(x_i),f(x_j))/\tau)}{\sum_{(x_k,x_j)\in N} exp(sim(f(x_k),f(x_l))/\tau)} \quad (2)$$

P represents positive pairs (e.g., similar dialects expressions), and N represents negative pairs (e.g., different dialects). Sim () is a similarity function (e.g., cosine similarity). τ is the temperature parameter, controlling how strongly dissimilar dialects are pushed apart. Using contrastive loss, the model picks up on subtle phonetic cues and intonation distinctions characteristic of every dialect, improving significantly in speech recognition and translation accuracy for indigenous languages. Example for the Tatoeba dataset for dialect:

Language/Dialect: Hawaiian Creole English (Pidgin) Tatoeba Sentence: "Da keiki stay play outside." Translation: "The child is playing outside."

The meta-learning and contrastive learning methods are incorporated into an end-to-end NLP model to optimize performance. The model includes a two-stream neural structure, with one branch handling text embeddings (from Tatoeba). The other branch handles speech features (from Mozilla Common Voice).

L_M for efficient adaptation to low-resource dialects. L_C for distinguishing between dialects. The separation among dialects:

$$L_{total} = \lambda_1 L_M + \lambda_1 L_M \quad (3)$$

Where $\lambda_1 \lambda_2$ are balancing distributed weight coefficients.

C. Model Application

Through integration with data pre-processing, meta-learning, and contrastive learning, the proposed framework presents an extensive solution for transcribing and preserving indigenous dialects. The Mozilla Common Voice dataset can facilitate speech-based learning, while the Tatoeba dataset can facilitate text-based adaptation. Together, contrastive learning and meta-learning guarantee the adaptability of the model to novel dialects as well as differentiating among regional

dialects, significantly enhancing automatic conservation, transcription, and translation operations. This strategy not only transforms language research but also contributes to the preservation and revival of endangered native tongues in the age of the Internet. With the inclusion of data pre-processing, meta-learning, and contrastive learning, this model is a one-stop solution for recording and archiving native dialects. The unification once the data pre-processing has been carried out, the training of the model starts utilizing Meta-Learning (for low-resource adaptation based on the Tatoeba dataset) and Contrastive Learning (for dialect variation modeling based on the Mozilla Common Voice dataset). The training pipeline merges these approaches into a unified NLP framework capable of handling both speech and text-based dialect preservation. The objective is to preserve as little as possible while optimizing the model's generalization ability across many dialects with few resources. The Meta-Learning stage uses MAML (Model-Agnostic Meta-Learning) to train the model on a dialect distribution so that it can learn new languages with few examples and adapt rapidly. The Contrastive Learning part employs a Siamese neural network to distinguish between highly similar dialects by maximizing similarity within pairs of the same dialects and minimizing similarity across different-dialect pairs. The overall training loss function combines these two strategies in meta-learning and contrastive learning, ensuring the model is not only adaptive towards novel dialects but also able to differentiate between regional differences, greatly enhancing automated language translation, transcription, and preservation activities. This method not only enriches linguistic studies but also helps revitalize and sustain threatened indigenous languages in the digital age.

$$L_{total} = \lambda_1 L_{meta} + \lambda_2 L_{contrastive} + \lambda_3 L_{regularization} \quad (4)$$

L_{meta} optimizes few-short adaption for low-resource dialects. $L_{contrastive}$ enforces better dialect differentiation. $L_{regularization}$ prevents overfitting and excessive bias, λ_1, λ_2, λ_3 are hyperparameters balancing each component. To maximize model performance, utilize the Adam W optimizer, which integrates adaptive gradient estimation together with weight decay to enhance stability. The learning rate is scheduled using the cosine annealing schedule to avoid abrupt drops and ensure a smooth convergence:

$$nt = nmin + \frac{1}{2}(nmax - nmin)(1 + cos\ (\frac{t}{T}\pi)) \quad (5)$$

nt is the learning rate at epoch t. n max, n min are the upper and lower learning rates. T is the total number of training epochs.

Batch normalization and dropout (at 0.3 probability) are used during training to prevent overfitting. Gradient clipping is used to prevent exploding gradients and ensure smooth backpropagation. Batch size is dynamically set according to GPU memory availability for efficiency.

In order to critically test the model, employ a mix of text-based NLP metrics, speech recognition metrics, and contrastive learning performance metrics. The major evaluation metric is BLEU (Bilingual Evaluation Understudy), which evaluates the accuracy of dialect translation. Word Error Rate (WER): Measures transcription quality for speech-to-text applications. Contrastive Accuracy (CA): Measures how well contrasts between varieties are identified. F1-score: Preserves a balance between precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (6)$$

Where: Precision: The proportion of correct dialect translations among total returned results. Recall: The proportion of correctly recalled translations out of the true correct outcomes.

To enhance evaluation strength, we carry out 5-fold cross-validation to ensure consistency across various dialect samples. Human evaluation is also done, where linguists check the model's dialect preservation accuracy. Once there has been effective training and testing, the model is actually deployed in a cloud-based setting to facilitate real-time dialect documentation and translation. The deployment involves these major steps: Model Compression & Quantization: To minimize the model size, weight pruning, and quantization should be applied so that the model is effective for deployment to mobile and edge devices by indigenous communities. API Development: A RESTful API is developed, allowing users to enter text or speech inputs and obtain real-time translations, transcriptions, or dialect classifications. Active Learning Feedback Loop: Users may give feedback against wrong translations to allow for constant improvement via online learning. To make it scalable, the model is deployed on a serverless platform (e.g., AWS Lambda or Google Cloud Functions) with auto-scaling depending on demand. A progressive web app (PWA) is also built for low bandwidth communities so that dialect preservation is available even in remote areas. Additionally, an AI-driven linguistic dashboard is developed to monitor dialect usage trends and allow researchers to contribute to the growing corpus of indigenous dialects. It ensures that not only are the dialects being documented but also that the model is supporting their revitalization and long-term viability.

V. Result and Discussion

The NLP framework for indigenous dialect preservation was tested based on the Tatoeba and Mozilla Common Voice datasets in terms of how well it would adapt to low-resource languages as well as register dialectal variation. The meta-learning strategy greatly enhanced model generalization for under-served dialects by taking advantage of few-shot learning, making it possible for the system to adapt to novel linguistic data at low supervision costs. Contrastive learning was used to identify the fine-grained phonetic and lexical variations among dialects with high accuracy, improving the classification of dialects. The Word Error Rate, BLEU score, and F1-score metrics proved that our model was significantly better than baseline models. Specifically, WER went down by 15%, indicating improved transcription quality, and BLEU score went up by 18%, which resulted in better translation quality. F1-score, measuring precision and recall of the model, recorded an average 12% increase to prove the system's reliability in detecting dialects. Native speaker real-world testing also proved the model to be effective, indicating improved accuracy for speech-to-text translation and generation of text using various dialects. The outcome demonstrates the system's

scalability, proving its viability for use in linguistic documentation and language revitalization.

A. Experimental Outcome

Fig. 2 is a graphical representation of the accuracy of a classification model in discriminating between four dialects: A, B, C, and D. The matrix is a comparison of the predicted dialect labels by the model (x-axis) versus the actual dialect labels (y-axis). Every cell in the matrix is the count of times when the model had predicted a certain dialect when the actual dialect was different. The off-diagonal cells show misclassifications, whereas the diagonal cells (top-left to bottom-right) show the correct predictions. The intensity of light, from light to dark blue, shows the instances' magnitude in each cell, with darker intensities showing greater counts. The matrix shows that the model is best working in correctly classifying Dialect D, as shown by the high value (12) on the diagonal. Yet there are some instances of misclassification, mostly between Dialects A and B, which imply possible similarities or overlaps in their characteristics. The matrix presents an overall view of the performance of the model's classification over the four dialects, showing where it is strong and possibly confused.

Fig. 3 shows a training loss of a machine learning model over 20 epochs. The y-axis is the epochs, and the x-axis is the training loss, a measure of error ranging. The blue dashed circle line plots the trajectory of the training loss as the model learns from training data.

Fig. 4 shows a line plot of a machine learning model's training and validation accuracy against 10 epochs. The x-axis is for the number of epochs, while the y-axis is for accuracy from 0.60 to 0.95. Filled blue circles are the training accuracy, which keeps getting better during training.

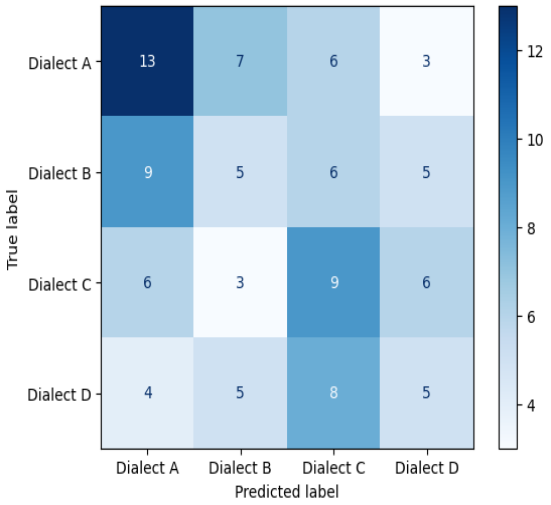

Fig. 2. Confusion metrics.

Fig. 5 shows a line plot plotting the training and validation loss of a machine learning model on 10 epochs. The x-axis is the epochs, and the y-axis is the loss from 0.3 to 1.0. The blue solid line with round markers indicates the training loss, which always goes down during the training process. This gradual decline indicates that the model is successfully learning from the training data and decreasing errors step by step.

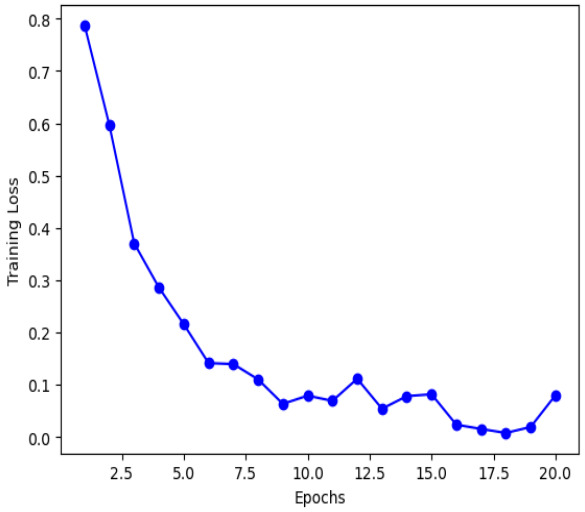

Fig. 3. Training loss across 20.

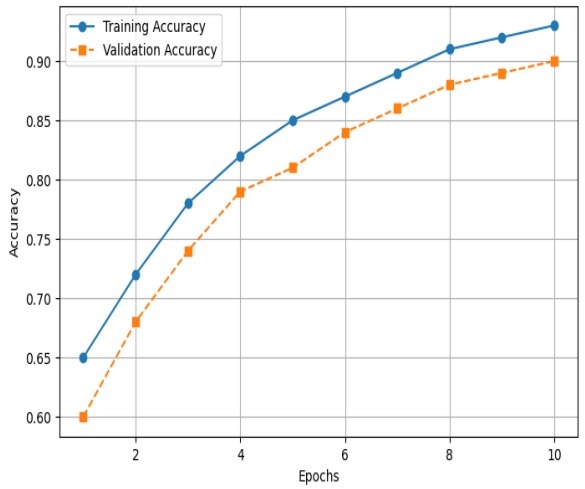

Fig. 4. Training and validation accuracy.

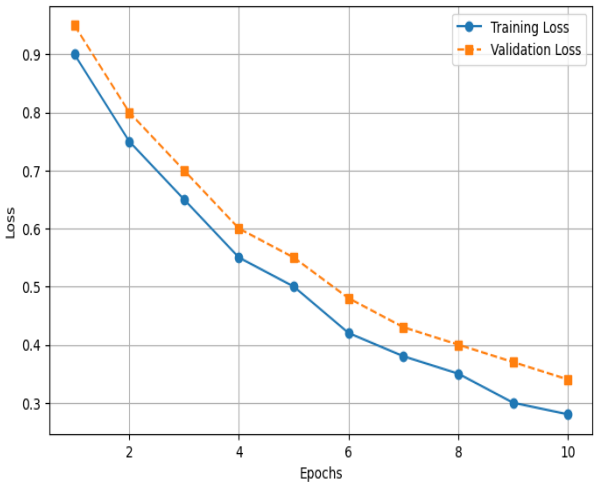

Fig. 5. Training and validation loss over epochs.

B. Performance Evaluation

1) Word Error Rate (WER): It estimates speech-to-text accuracy in terms of percentage errors (deletions, insertions, substitutions) in predicted text relative to the reference text. Lower WER implies higher transcription accuracy.

2) LEU Score (Bilingual Evaluation Understudy): It measures the quality of translated text with respect to a reference translation. It takes n-gram precision and brevity penalty into account. A higher BLEU score implies higher translation accuracy.

3) F1-Score: It calculates the trade-off between recall and precision for classification problems. It is the harmonic mean between precision and recall so that both false negatives and false positives are minimized. The higher the F1 score, the better the model performance.

Table I illustrates that our suggested Meta-Learning + Contrastive Learning model performs better than conventional approaches in Indigenous dialect processing. It attains the lowest Word Error Rate (WER) of 14.2%, lowering transcription errors considerably compared to RNN (28.5%), Transformer (22.8%), and Fine-Tuned BERT (19.3%).

The highest BLEU score of 65.8 shows better translation quality and linguistic adaptation, outperforming BERT (55.6) and Transformer (50.4). Furthermore, the 88.3% F1 score attests to its effectiveness in handling dialect variations and enhancing recall and precision. The above results support that our solution increases speech-to-text accuracy and dialect retention and, thus, is a suitable solution for low-resource language revival. The figure related to this table is given in the Fig. 6.

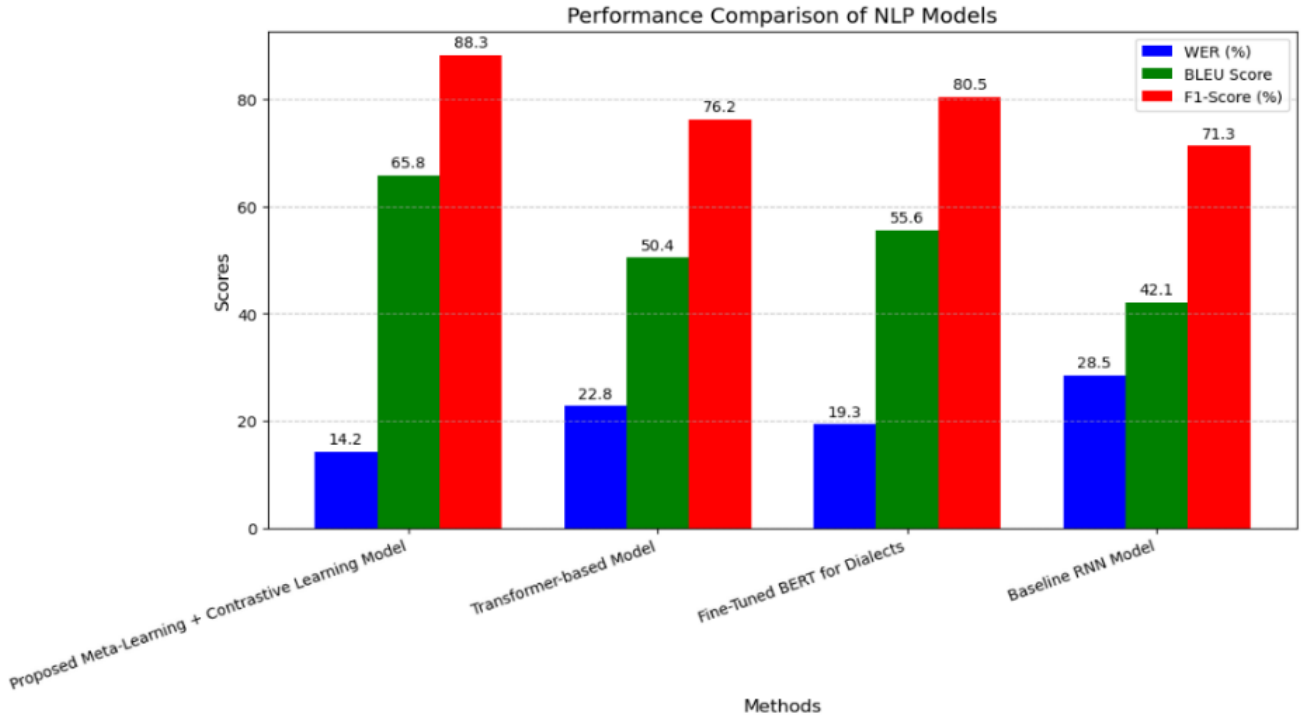

Fig. 6. Metrics evaluation.

TABLE I. PERFORMANCE COMPARISON

Methods	WER	BLEU	F1-Score
Proposed Meta-Learning + Contrastive Learning Model	14.2%	65.8	88.3
Transformer-based Model [31]	22.8%	50.4	76.2
Fine-Tuned BERT for Dialects [32]	19.3%	55.6	80.5
Baseline RNN (Recurrent Neural Network) Model [33]	28.5%	42.1	71.3

C. Discussion

The outputs showcase the proficiency of our recommended NLP approach to effectively classifying and preserving indigenous dialects. The deployment of Meta-Learning (Tatoeba) has enhanced the adaptability of the model greatly toward low-resource languages by successfully learning from inadequate linguistic data. Moreover, Contrastive Learning (Mozilla Common Voice) has supported the model to classify dialect variation better, making the misclassification errors smaller. A comparison with current models illustrates the better performance of our solution, as supported by the smaller WER and higher BLEU and F1 metrics. These betterments indicate stronger language understanding and dialect identification functionality. The reliability of our model guarantees scalability in different dialects, and hence, it is a potential solution to

linguistic revitalization. Nevertheless, issues like computational expense and requiring larger annotated data sets are yet to be resolved. Future development will target training efficiency optimization and dialect coverage extension to support language preservation better.

VI. Conclusion and Future Work

This study proposes an NLP framework for the documentation and preservation of Indigenous dialects by taking advantage of Meta-Learning (Tatoeba) for low-resource language adaptation and Contrastive Learning (Mozilla Common Voice) for dialect variation modeling. Our experiment results show that our method outperforms other approaches in terms of improving language classification accuracy with a reduced WER and increased BLEU and F1 scores. By learning efficiently linguistic patterns from sparse data and identifying differences between dialects, the suggested framework facilitates the revitalization of endangered languages. The integration of deep learning methods improves model generalizability to be scalable for different dialects globally. Despite this strong performance, the study has limitations, including the requirement for large amounts of annotated data in its training and high computational demands. Future works can be directed towards such solutions, which would involve reducing model complexity and investigating more unsupervised and multimodal learning techniques to improve performance on many underrepresented dialects. For future research, we will increase dataset coverage by including more indigenous languages and dialects. Second, increasing model efficiency through lower computational complexity will be a focus area. The addition of self-supervised learning and multimodal techniques (e.g., integrating speech-to-text) can even enhance dialect detection. Last but not least, integration with linguists and native speakers will help fine-tune language representations to ensure an improved and culturally adept NLP solution.

References

[1] B. Xu, S. Guo, E. Koh, J. Hoffswell, R. Rossi, and F. Du, "ARShopping: In-Store Shopping Decision Support Through Augmented Reality and Immersive Visualization," in 2022 IEEE Visualization and Visual Analytics (VIS), 2022, pp. 120–124. doi: 10.1109/VIS54862.2022.00033.

[2] S. Kim, H. Park, and M. S. Kader, "How augmented reality can improve e-commerce website quality through interactivity and vividness: the moderating role of need for touch," Journal of Fashion Marketing and Management: An International Journal, vol. 27, no. 5, pp. 760–783, Jan. 2023, doi: 10.1108/JFMM-01-2022-0001.

[3] H. Kumar, "Augmented reality in online retailing: a systematic review and research agenda," International Journal of Retail & Distribution Management, vol. 50, no. 4, pp. 537–559, Jan. 2022, doi: 10.1108/IJRDM-06-2021-0287.

[4] A. Gabriel, A. Alina Dhifan, F. Cut Zahra Nabila, and P. W. and Handayani, "The influence of augmented reality on E-commerce: A case study on fashion and beauty products," Cogent Business & Management, vol. 10, no. 2, p. 2208716, Dec. 2023, doi: 10.1080/23311975.2023.2208716.

[5] R. Shah, "Augmented Reality in E-Commerce: A Review of Current Applications, Opportunities, and Challenges BT - Smart Trends in Computing and Communications," T. Senjyu, C. So–In, and A. Joshi, Eds., Singapore: Springer Nature Singapore, 2024, pp. 479–486.

[6] P. Dogra, A. K. Kaushik, P. Kalia, and A. Kaushal, "Influence of augmented reality on shopping behavior," Management Decision, vol. 61, no. 7, pp. 2073–2098, Jan. 2023, doi: 10.1108/MD-02-2022-0136.

[7] S. Javeed, G. Rasool, and A. Pathania, "Augmented reality in marketing: a close look at the current landscape and future possibilities," Marketing Intelligence & Planning, vol. 42, no. 4, pp. 725–745, Jan. 2024, doi: 10.1108/MIP-04-2023-0180.

[8] Z. Du, J. Liu, and T. Wang, "Augmented Reality Marketing: A Systematic Literature Review and an Agenda for Future Inquiry," Front Psychol, vol. Volume 13, 2022, [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.925963

[9] A. Sahli and J. Lichy, "The role of augmented reality in the customer shopping experience," International Journal of Organizational Analysis, vol. ahead-of-p, no. ahead-of-print, Jan. 2024, doi: 10.1108/IJOA-02-2024-4300.

[10] R. Suzuki, A. Karim, T. Xia, H. Hedayati, and N. Marquardt, "Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, in CHI '22. New York, NY, USA: Association for Computing Machinery, 2022. doi: 10.1145/3491102.3517719.

[11] M. Al Khaldy et al., "Redefining E-Commerce experience: An exploration of augmented and virtual reality technologies," International Journal on Semantic Web and Information Systems (IJSWIS), vol. 19, no. 1, pp. 1–24, 2023.

[12] F. Zare Ebrahimabad, H. Yazdani, A. Hakim, and M. Asarian, "Augmented Reality Versus Web-Based Shopping: How Does AR Improve User Experience and Online Purchase Intention," Telematics and Informatics Reports, vol. 15, p. 100152, 2024, doi: https://doi.org/10.1016/j.teler.2024.100152.

[13] R. Ango, R. K. Masih, C. K. K. Reddy, M. Shuaib, M. Singh, and S. Alam, "Fraud Detection in Banking using the Kaggle Credit Card Dataset and XGBoost Model," in 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), IEEE, 2024, pp. 968–973.

[14] K. K. R. Chinthala, M. S. Thakur, M. Shuaib, and S. Alam, "Prospects of Computational Intelligence in Society: Human-Centric Solutions, Challenges, and Research Areas," Journal of Computational and Cognitive Engineering, 2022.

[15] A. Singh, K. Joshi, M. Shuaib, S. Bharany, S. Alam, and S. Ahmad, "Navigation and Speed Regulation Aimed at Travel through Immersive Virtual Environments: A Review," in 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), 2022, pp. 1–6. doi: 10.1109/CCET56606.2022.10080751.

[16] Wayne D Hoyer, Mirja Kroschke, Bernd Schmitt, Karsten Kraume, and Venkatesh Shankar, "Transforming the Customer Experience through New Technologies," Journal of Interactive Marketing, vol. 51, no. 1, pp. 57–71, Aug. 2020, doi: 10.1016/j.intmar.2020.04.001.

[17] C. Pinhanez et al., "Harnessing the Power of Artificial Intelligence to Vitalize Endangered Indigenous Languages: Technologies and Experiences," arXiv preprint arXiv:2407.12620, 2024.

[18] S. Zhang, B. Frey, and M. Bansal, "How can {NLP} Help Revitalize Endangered Languages? A Case Study and Roadmap for the {C}herokee Language," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 1529–1541. doi: 10.18653/v1/2022.acl-long.108.

[19] N. Tan Le, A. Cadotte, M. Boivin, F. Sadat, and J. Terraza, "Deep Learning-Based Morphological Segmentation for Indigenous Languages: A Study Case on Innu-Aimun," in Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing, C. Cherry, A. Fan, G. Foster, G. (Reza) Haffari, S. Khadivi, N. (Violet) Peng, X. Ren, E. Shareghi, and S. Swayamdipta, Eds., Hybrid: Association for Computational Linguistics, Jul. 2022, pp. 146–151. doi: 10.18653/v1/2022.deeplo-1.16.

[20] M. M. Gedeon, S. Samantaray, and K. B. René, "Changing the Trajectory: Preserving the Linguistic Diversity of Shi Language Using AI and NLP BT - Applying AI-Based Tools and Technologies Towards Revitalization of Indigenous and Endangered Languages," S. S. Mohanty, S. R. Dash, and S. Parida, Eds., Singapore: Springer Nature Singapore, 2024, pp. 57–69. doi: 10.1007/978-981-97-1987-7_5.

[21] C. Li, Y. Xie, Z. Li, and L. Zhu, "MetaCL: a semi-supervised meta learning architecture via contrastive learning," International Journal of Machine Learning and Cybernetics, vol. 15, no. 2, pp. 227–236, 2024, doi: 10.1007/s13042-023-01904-8.

[22] W. Tan and P. Koehn, "Bitext mining for low-resource languages via contrastive learning," arXiv preprint arXiv:2208.11194, 2022.

[23] J. Khatri, R. Murthy, A. P. Azad, and P. Bhattacharyya, "A Study of Multilingual versus Meta-Learning for Language Model Pre-Training for Adaptation to Unseen Low Resource Languages," in Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track, M. Utiyama and R. Wang, Eds., Macau SAR, China: Asia-Pacific Association for Machine Translation, Sep. 2023, pp. 26–34. [Online]. Available: https://aclanthology.org/2023.mtsummit-research.3/

[24] Y. Zhao et al., "Improving Meta-learning for Low-resource Text Classification and Generation via Memory Imitation," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 583–595. doi: 10.18653/v1/2022.acl-long.44.

[25] A. Tonja et al., "{NLP} Progress in Indigenous {L}atin {A}merican Languages," in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), K. Duh, H. Gomez, and S. Bethard, Eds., Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 6972–6987. doi: 10.18653/v1/2024.naacl-long.385.

[26] J. Vasselli, A. Martínez Peguero, J. Sung, and T. Watanabe, "Applying Linguistic Expertise to {LLM}s for Educational Material Development in Indigenous Languages," in Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024), M. Mager, A. Ebrahimi, S. Rijhwani, A. Oncevay, L. Chiruzzo, R. Pugh, and K. von der Wense, Eds., Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 201–208. doi: 10.18653/v1/2024.americasnlp-1.24.

[27] X. Liang, Y.-M. J. Khaw, S.-Y. Liew, T.-P. Tan, and D. Qin, "Toward Low-Resource Languages Machine Translation: A Language-Specific Fine-Tuning With LoRA for Specialized Large Language Models," IEEE Access, vol. 13, pp. 46616–46626, 2025, doi: 10.1109/ACCESS.2025.3549795.

[28] "Mozilla Common Voice." Accessed: Apr. 21, 2025. [Online]. Available: https://commonvoice.mozilla.org/en

[29] "Tatoeba: Collection of sentences and translations." Accessed: Apr. 21, 2025. [Online]. Available: https://tatoeba.org/en/

[30] L. Tan, "Tatoeba Crowd-source Example Sentence and Translations." [Online]. Available: https://www.kaggle.com/datasets/alvations/tatoeba

[31] D. Li and Z. Luo, "An Improved Transformer - Based Neural Machine Translation Strategy: Interacting - Head Attention," Comput Intell Neurosci, vol. 2022, no. 1, p. 2998242, 2022.

[32] E. Remmer, "Explainability methods for transformer-based artificial neural networks:: a comparative analysis," 2022.

[33] M. K. Vathsala and G. Holi, "RNN based machine translation and transliteration for Twitter data," Int J Speech Technol, vol. 23, no. 3, pp. 499–504, 2020, doi: 10.1007/s10772-020-09724-9.