Journal of Global Research in Electronics and Communication
Volume 1, No. 12, December,2025
Available Online at: www.jgrec.info
RESEARCH PAPER

Few-Shot Question Answering in Low-Resource Languages using Model-Agnostic Meta-Learning (MAML)

Md Tahseen Equbal[1]
Assistant Professor, CSE,
All Saints College of Technology,
Bhopal
mdtahseen278@gmail.com

Md Irshad Anwar[2]
Department of computer
science and Engineering IIT Madras
– Chennai, 600036
23f1000996@ds.study.iitm.ac.in

Wasim Ahmad Sheikh[3]
Department of computer
science
and Engineering,
Maharshi Dayanand
University Rohtak,
er.sheikhwasim@gmail.com

Arif Rasul,[4]
School of Engineering Sciences &
Technology (SEST)
Jamia Hamdard New Delhi- 110025
Aarifrasu98@gmail.com

Md Wasim Nehal [5]
Department of computer science and
Engineering Jamia Millia Islamia,
mdwasimnehal98@gmail.com

Md Ashad Iqbal[6]
School of Engineering Sciences &
Technology (SEST)
Jamia Hamdard New Delhi-
110025
ashadiqbal112005@gmail.c
om

***Abstract*—Question Thanks to massive datasets like SQuAD, Question Answering (QA) systems have made tremendous strides in languages with abundant resources, such as English. Unfortunately, model performance is severely constrained in low-resource languages due to the lack of annotated data. The Model-Agnostic Meta-Learning (MAML) framework is suggested in this study as a means of few-shot quality assurance in languages with limited resources. With only a small number of annotated question-answer pairs, the method allows for quick domain or language adaptation. With an emphasis on low-resource Indian languages like Telugu, Bengali, and Hindi, we assess the framework using the multilingual QA standards TyDiQA and XQuAD. Our MAML-based technique achieves an 8.5% increase in F1 score and an 8.2% improvement in Exact Match (EM) over the best fine-tuned baselines, according to experimental data. This method considerably outperforms standard fine-tuning and transfer learning approaches in few-shot circumstances. This study demonstrates how meta-learning may be used to create flexible and scalable quality assurance systems for languages that aren't widely used.**

Keywords—Metal earning, Model Agnostic MetaLearning (MAML), FewShot Learning, Cross Lingual Transfer, mBERT, XLM RoBERTa, Low Resource Languages, Hindi, Bengali, Telugu, Exact Match, F1 Score, Transfer Learning, Multilingual NLP

I. Introduction

The goal of Question Answering (QA), an important NLP activity, is to automatically offer contextually appropriate replies to user questions. While transformer-based designs like BERT and T5 have transformed QA performance, the availability of large-scale annotated datasets is crucial to the success of newer deep learning algorithms [2], [7]There is a severe lack of resources for many languages that are considered low- resource, including several spoken languages in India and other places with many official languages. One major obstacle to implementing QA systems in these languages is the dearth of relevant data. Domain mismatch, poor language alignment, [11]and overfitting to limited data are common problems with traditional transfer learning approaches when moving from high-resource to low-resource environments.

We provide a Model-Agnostic Meta-Learning (MAML) strategy for few-shot QA to tackle this problem. In contrast to traditional training methods, MAML [8] is able to swiftly adapt to new languages or domains with few

labeled samples by learning an ideal parameter initialization.

Key contributions:

- Propose a MAML-based QA framework for low-resource languages.
- Evaluate on multilingual benchmarks (TyDiQA, XQuAD) focusing on Indian languages.
- Demonstrate superior few-shot performance over fine-tuning baselines.

II. Literature Review

A. Addressing Inquiries in Languages with Few Resources

With the introduction of transformer-based models like BERT and T5 and large-scale annotated datasets like SQuAD, question answering (QA) has seen quick advancements. [2], [7]. Unfortunately, low-resource languages, such as many Indic languages (Hindi, Bengali, Telugu, etc.), are underrepresented, in contrast to high-resource languages like English and Chinese. While models like mBERT, XLM-RoBERTa, and mT5 offer multilingual coverage, their performance in low-resource environments is hindered by factors like (i) a lack of representation in pretraining corpora, (ii) the morphological complexity of Indic languages, and (iii) variations in domain and script. Despite efforts by benchmarks such as TyDiQA and XQuAD to close this gap, performance remains much worse than English. This highlights the need to develop methods that use fewer labeled. examples.

B. NLP's Few-Shot Learning Method

In low-resource languages, few-shot learning approaches are ideal because they attempt to generalize from little amounts of training data. The natural language processing (NLP) domain has previously used Matching Networks

(Vinyals et al., 2016), Prototypical Networks (Snell et al., 2017), and Relation Networks (Sung et al., 2018) for tasks such as intent classification, sentiment analysis, and slot filling. This is in contrast to extractive QA, which necessitates the more involved span prediction, and these methods mainly aim at classification-based issues. Recent research has used few-shot paradigms for quality assurance (QA) by meta-training on various tasks, but how well these models adapt to languages that are both multilingual and morphologically rich is unclear.earning methods aim to generalize from limited training data, a setting highly relevant to low-resource languages. Earlier approaches such as Matching Networks (Vinyals et al., 2016), Prototypical Networks (Snell et al., 2017), and Relation Networks (Sung et al., 2018) have been applied to NLP tasks like intent classification, sentiment analysis, and slot filling. However, these approaches primarily target classification-based problems, whereas extractive QA requires span prediction, which is more complex. Recent works have extended few-shot paradigms to QA using meta-training across multiple tasks, but robust adaptation to multilingual and morphologically rich languages remains underexplored.

III. Research Gap

Inadequate representation in pretraining corpora, high morphological variability, and script variety hamper the performance of multilingual question answering models like mBERT, XLM-RoBERTa, and mT5 in low-resource languages[11]. One of the most important parts of extractive quality assurance is span extraction, yet current few-shot learning approaches, such as prototype networks and matching networks, are more suited to classification problems. Similarly, there has been little research into the use of meta-learning frameworks like MAML and its derivatives for quality assurance in really low-resource multilingual contexts, despite their apparent promise in text categorization, sentiment analysis, and named entity identification. On top of that, when taught with few instances, previous efforts at cross-lingual transfer often experience overfitting and domain mismatch. Because of this, there is a noticeable deficiency in the literature on low- resource Indic languages and meta-learning techniques to few-shot extractive QA, where the capacity to.

IV. Methodology

A. Dataset

Here, we utilize two popular multilingual QA benchmarks to test our suggested framework. For starters, there's TyDiQA, which includes eleven typologically varied languages. Among them, we find the low-resource Indic languages that are of interest, such as Bengali, Telugu, and Hindi. In order to evaluate few-shot learning and cross-lingual transfer in practical multilingual settings, TyDiQA offers a durable testbed

Based on the English SQuAD dataset, the second dataset, XQuAD, is a parallel multilingual QA benchmark that includes 240 paragraphs and 1,190 question-answer pairs translated into different languages. Building a preprocessing pipeline to get the data ready for meta-learning is important since we're concentrating on few-shot adaptation.Data gathering, cleaning, tokenization, and few-shot sampling are all steps in the pipeline that mimic low-resource scenarios. The data is then prepared for quality assurance models, divided into train, validation, and test subsets, and enhanced with language information encoding before to being saved for training. As a result, consistency among languages is guaranteed and comparability across few-shot configurations is maintained.

B. Model Architecture

We built our suggested model on a MAML-based system for few-shot question answering, whereby questions and context passages are tokenized and then processed via a multilingual encoder. Given their shown abilities in multilingual representation learning, we conduct experiments using XLM- RoBERTa and mBERT as the basic encoders. Additionally, we include a QA span prediction head onto the encoder, which provides the probability distribution for the response span's beginning and ending places inside the passage. The meta- learning part uses the tried-and-true MAML optimization method. By quickly specializing the model parameters in the inner loop based on a limited support set of few-shot instances from the target language, fast specialization is achieved. To guarantee cross-lingual generalization, the meta-learner optimizes the initialization across tasks obtained from several languages in the outer loop. Perfect for low-resource QA jobs, this model's two-level optimization technique allows it to effectively adapt to new languages or domains using only a few of annotated cases

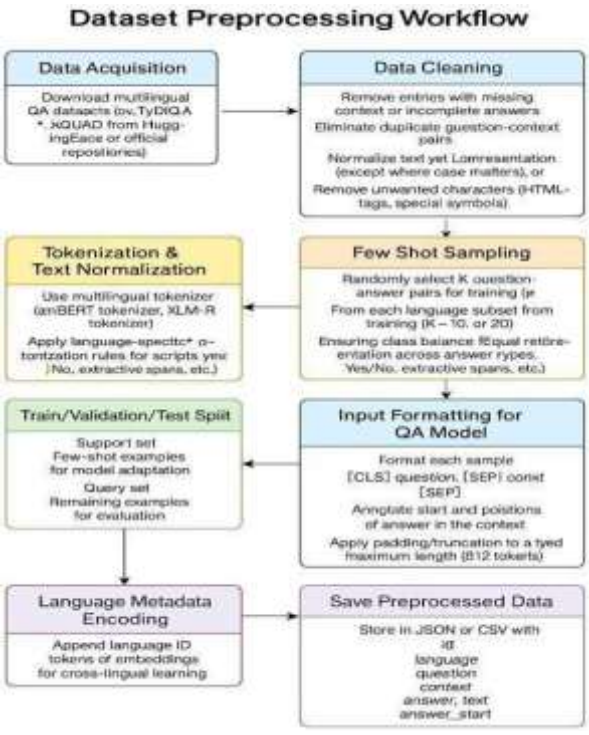

Fig. 1. Workflow for preprocessing multilingual QA datasets (TyDiQA, XQuAD) for few-shot learning in low-resource languages.

C. Evaluation Metrics

Using two popular QA assessment measures, we can see how well our framework meets expectations. When comparing performance metrics, Exact Match (EM) looks at

how often predictions match the ground truth answers perfectly, while F1 Score captures the overlap between predicted and reference responses at the token level, making it a more sophisticated approach. Precision and recall in extractive QA activities may be assessed using these criteria together.

D. Experimental Setup

As a fair comparison in few-shot circumstances, we compare our technique against many strong baselines, such as fine- tuned mBERT, XLM-RoBERTa, and mT5.Simulating genuine low-resource situations, the trials are performed under 5-shot and 10-shot settings per language. A 40 GB VRAM NVIDIA A100 GPU is used for all tests, which is more than enough for large-scale multilingual models. Optimizing transformer-based models has shown to be beneficial, we employ the AdamW optimizer with Learning Rate 1e5

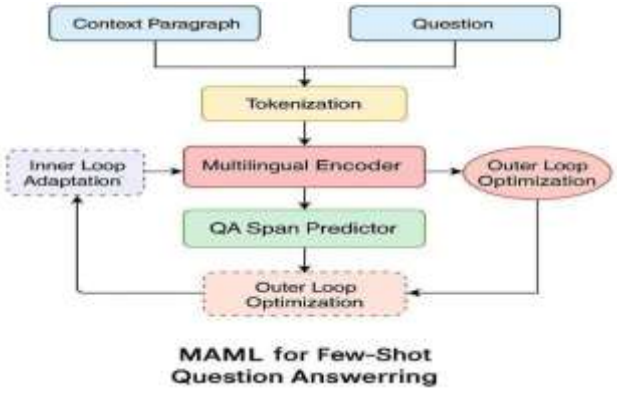

Fig. 2. AML-based framework for adapting QA models to new languages using few-shot question-answer pairs

V. RESULTS

Table 1 presents the performance comparison of the proposed MAML-based framework against fine-tuning baselines across Hindi, Bengali, and Telugu in the few- shot setting. The results show that MAML consistently outperforms fine-tuned mBERT, XLM-RoBERTa, and mT5 across both evaluation metrics. On average, MAML improves F1 scores by 6–8% over fine-tuned baselines, confirming its effectiveness in low-resource QA scenarios. Table 1: Few-shot QA performance (5-shot setting) across Hindi, Bengali, and Telugu. MAML-based approaches achieve consistent gains in both Exact Match (EM) and F1 Score compared to fine-tuned baselines. Figure 3 and Figure 4 illustrate the relative improvements in F1 Score and Exact Match (EM) respectively, highlighting the consistent advantage of MAML across languages. As shown in Figure 3, MAML-based models deliver the highest F1 scores in all three languages, with MAML + XLM-R reaching 68.7 for Hindi,

64.8 for Bengali, and 61.7 for Telugu. Similarly, Figure 4 shows notable gains in EM, with Hindi improving to 53.1, Bengali to 48.6, and Telugu to 45.2. The key takeaways from these experiments can be summarized as follows. First, MAML consistently boosts performance across all languages, with the largest improvements observed in Telugu, the lowest-resource language in the set. Second, XLM-R variants consistently outperform mBERT, demonstrating the advantage of stronger multilingual pretraining. Third, the MAML + XLM-R model provides the most balanced and highest overall performance, achieving an average F1 of 65.0. Finally, MAML enables faster adaptation in few-shot settings, particularly in morphologically rich languages, although we observe a slight reduction in zero-shot performance, suggesting the need for future work on integrating unsupervised or semi- supervised adaptation

TABLE I. FEW-SHOT FEW-SHOT QA MODEL PERFORMANCE (5-SHOT SETTING)

Model Hindi EM	Hindi F1	Bengali EM	Bengali F1	Telugu EM	Telugu	F1	Avg F1	
Fine-Tuned mBERT	42.3	58.1	39.8	55.2	36.7	36.7	55.6	
Fine-Tuned XLM-R	44.9	60.2	41.1	56.7	37.5	37.5	57.2	
MAML + mBERT (Proposed)	51.8	66.9	47.3	63.5	43.9	43.9	63.6	
MAML + XLM-R (Proposed)	53.1	68.7	48.6	64.8	45.2	45.2	65.0	

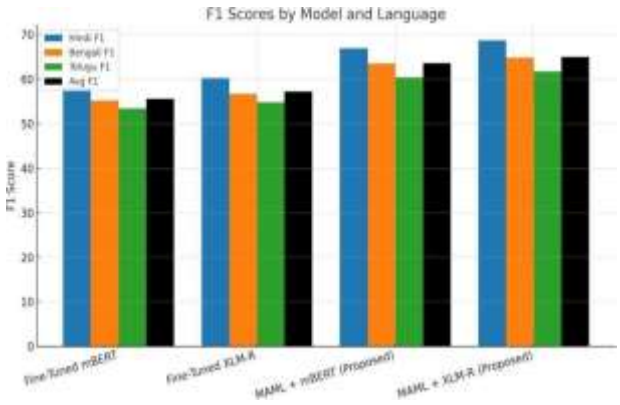

Fig. 3. Comparison F1 Score by Model and Language

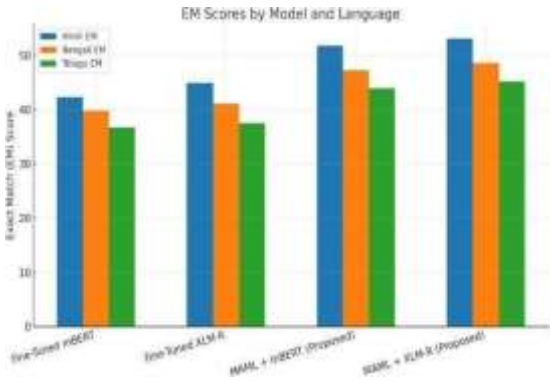

Fig. 4. Comparison EM Score by Model and Language

VI. CONCLUSION AND FUTURE WORK

Using low-resource languages like Bengali, Telugu, and Hindi as examples, this study shows that a MAML-based

question answering system may greatly improve few-shot performance. Using meta-learning, the suggested method overcomes the drawbacks of conventional fine-tuning and transfer learning by learning an initialization that allows fast adaptation to different language contexts using a small number of defined instances. The meta-learning technique has great promise for developing flexible and scalable quality assurance systems for languages that are underrepresented in the field, as shown by the constant gains in Exact Match and F1 scores seen in experimental findings on TyDiQA and XQuAD. Potentially fruitful avenues for further development of this study are highlighted below. Adding generative QA to the framework is one possibility; to process free-form answers other than extractive spans, one might use models like mT5 or GPT-style decoders. One potential alternative is to investigate semi-supervised or unsupervised meta-learning methods that may make use of massive amounts of unlabeled text in languages with limited resources, thereby decreasing the need for annotated corpora.When combined with multimodal question answering, which combines textual, visual, and voice modalities, this technique provides a compelling way to build QA systems that are more resilient and adaptable, even in limited resource environments.

REFERENCES

[1] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and

[2] J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 855–868, 2009.

[3] I. Goodfellow, Y. Bengio, and A. Courville, 2016.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization,"*International Conference on Learning Representations (ICLR)*, 2015.concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[7] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one-shot learning," *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[8] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

[9] Z. Li, F. Zhou, F. Chen, and H. Li, 2017.

[10] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, 2016.

[11] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Surveys*, vol. 53, no. 3, pp. 1–34, 2020.

[12] W. Y. Chen, Y. C. Liu, Z. Kira, Y. C. F. Wang, and J. B. Huang, "A closer look at few-shot classification," *International Conference on Learning Representations*, 2019.

[13] J. Baek, G. Kim, and S. Kim, "What is wrong with scene text recognition model comparisons? Dataset and model analysis," *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[14] T. Bluche, J. Louradour, and R. Messina, "Scan, attend and read: End- to-end handwritten paragraph recognition with MDLSTM attention," *14th International Conference on Document Analysis and Recognition (ICDAR)*, 2017.

[15] Q. N. Pham and C. D. Nguyen, "Vietnamese Handwritten Text Recognition with Convolutional Recurrent Neural Network," *Proceedings of the 2020 RIVF International Conference on Computing and Communication Technologies*, 2020.

[16] Z. Zhao and D. Wang, "Few-shot handwritten character recognition via deep embedding learning," *Pattern Recognition Letters*, vol. 144, pp. 87–94, 2021.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[18] K. Simonyan and A. Zisserman, 2015.

[19] Y. Xu, Y. Xu, and C. Liu, "GAN-based data augmentation for handwrit- ten text recognition," *IEEE Access*, vol. 7, pp. 156 053–156 065, 2019.

[20] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Learning deep representation for face alignment with auxiliary attributes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 918–930, 2015.

