LOW-RESOURCE SPEECH RECOGNITION WITH SELF-SUPERVISED LEARNING

Niket Agrawal
M23CSA520

Ritesh Lamba
M23CSA544

PROBLEM WITH EXISTING WORK

- Automatic Speech Recognition (ASR) for low-resource languages is challenged by the scarcity of labeled data. Traditional supervised approaches require large annotated datasets, which are often unavailable for these languages.

PROPOSED METHODOLOGY

- Common Voice: pre-trained Wav2Vec2.0 and XLS-R models
- FLEURS: multilingual dataset for ASR
- We have chosen Hindi as Low Resource Language
- Apply data augmentation techniques to improve generalization

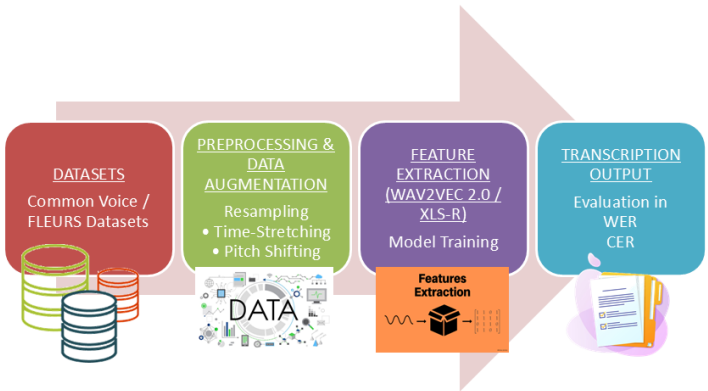
RESULTS AND ANALYSIS

- Fine-tuning pretrained Wav2Vec2.0 and XLS-R models on low-resource datasets
- Utilize LoRA for efficient adaptation

Model	WER (%)	CER (%)
Pre-trained	1.00	1.06
Fine-tuned (CTC)	1.00	1.00
Fine-tuned (LoRA)	1.00	1.00

CONCLUSION

- Fine-tuning, self-supervised models enhance ASR performance in low-resource languages, contributing to digital inclusivity and linguistic diversity.



References

- A. Baevski et al: Wav2Vec 2.0: A Framework for Self-Supervised Learning of Speech, 2020
- R. Ardila et al: Common Voice: A Massively Multilingual Speech Corpus, 2020