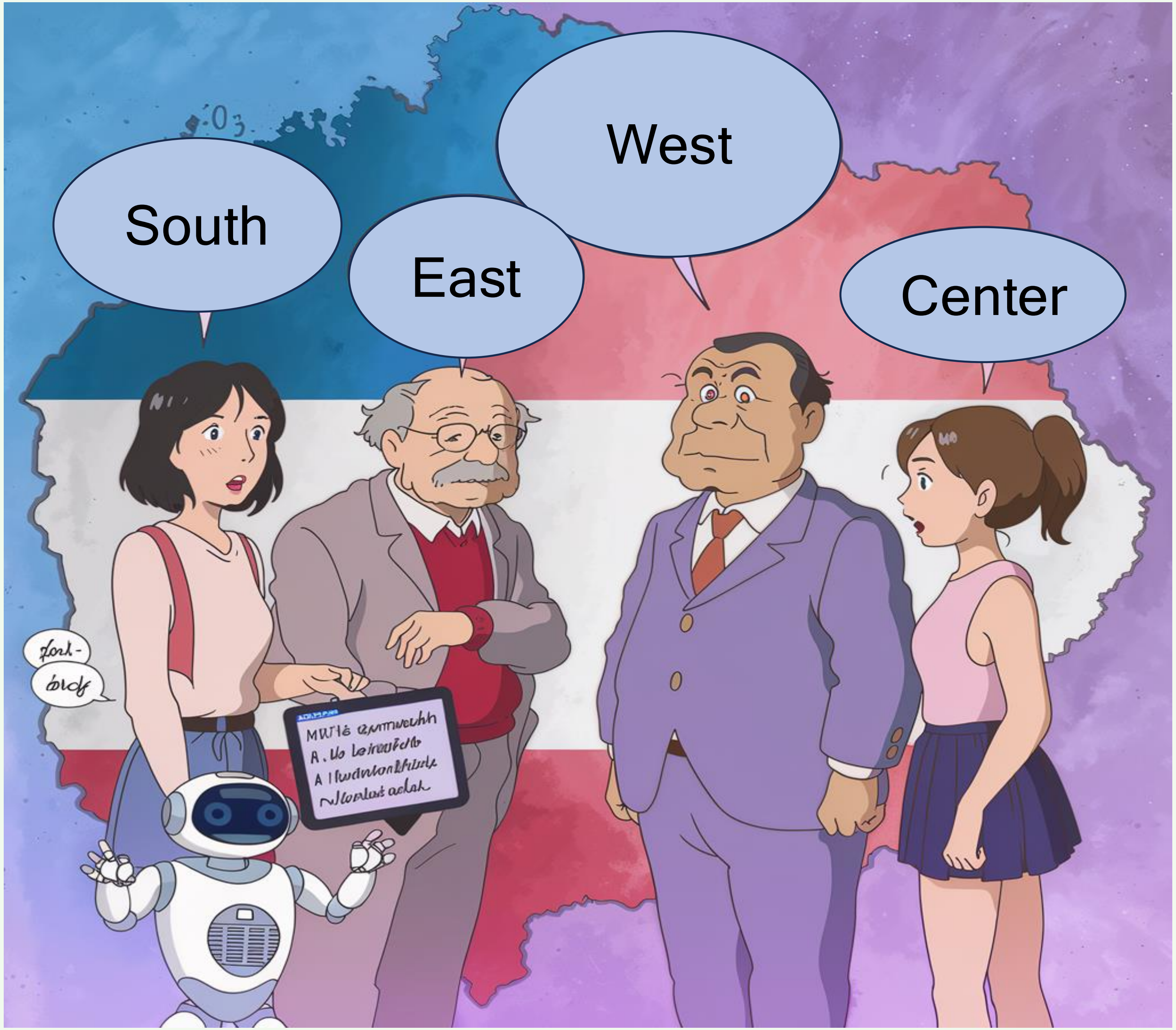
Voices of Luxembourg: Tackling Dialect Diversity in a Low-Resource Setting

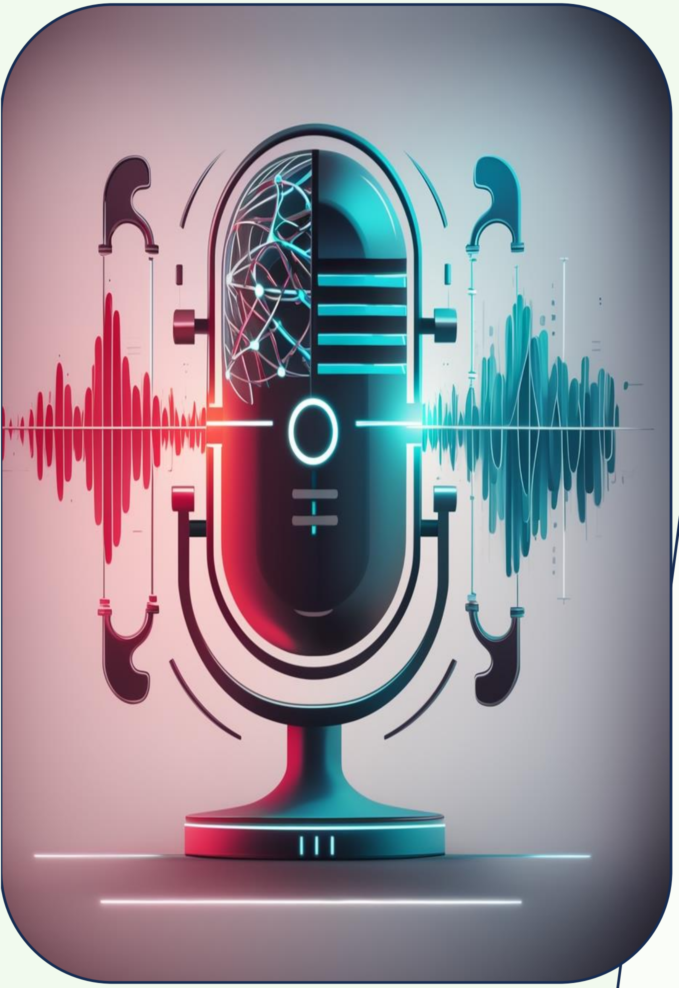
Presented by: **Nina Hosseini-Kivanani**
Authors: Nina Hosseini-Kivanani, Christoph Schommer, Peter Gilles
University of Luxembourg, Department of Computer Science
RESOURCEFUL-Tallinn 2025


Ech hätt gär e croissant.
Ech wëllt e croissant.
Luxembourgish is hard... Send help!

Same language, different vibes– welcome to Luxembourgish dialects!



Why Dialect Classification Matters?



Importance of Dialect Research

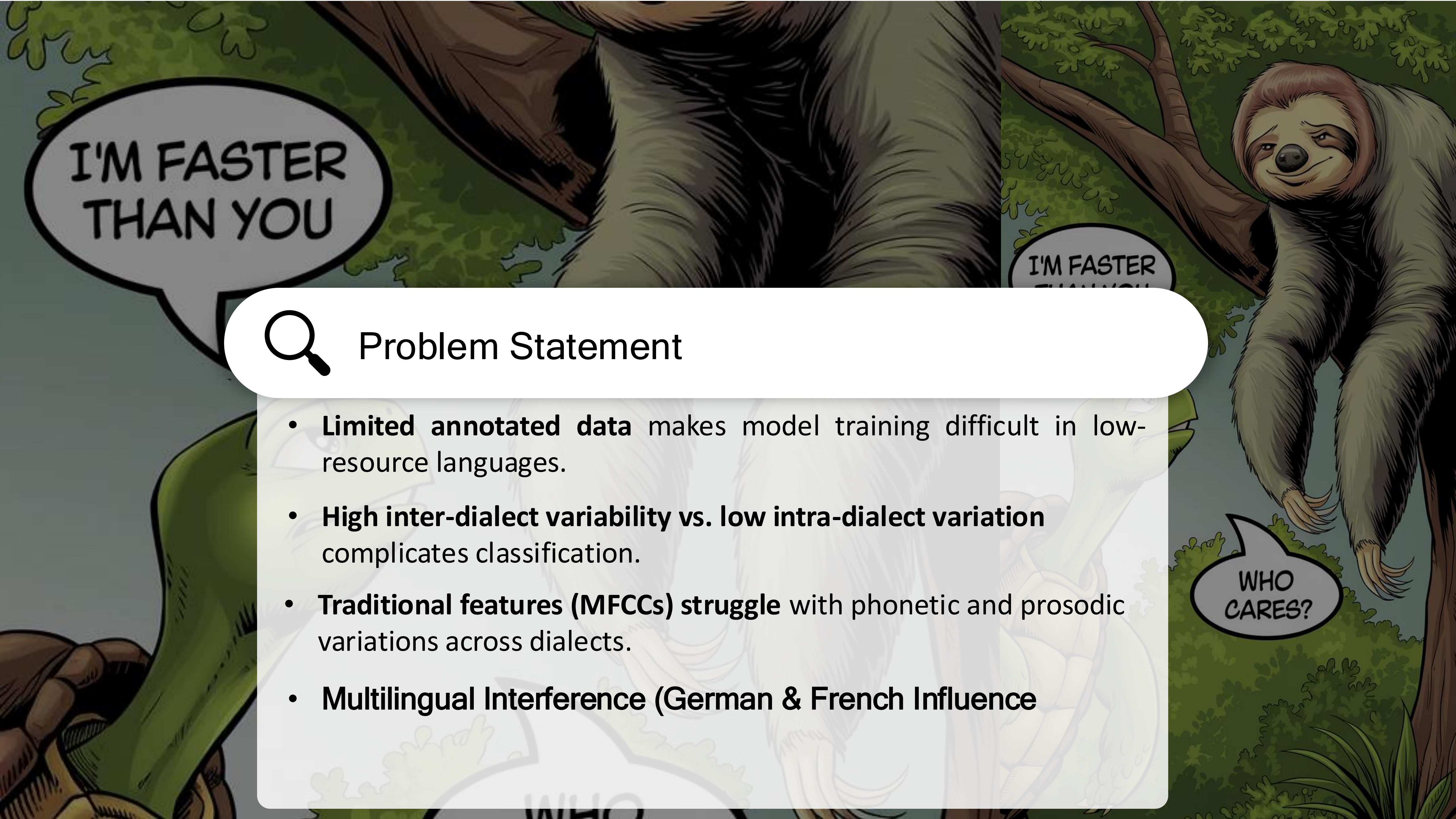
- Luxembourgish dialects preserve linguistic diversity and cultural identity.
- Automatic dialect classification aids ASR, NLP, and digital archiving.

Challenges in Dialect Classification

- Phonetic, prosodic, and lexical variations across four main dialects.
- Limited annotated data.
- Influence of German and French

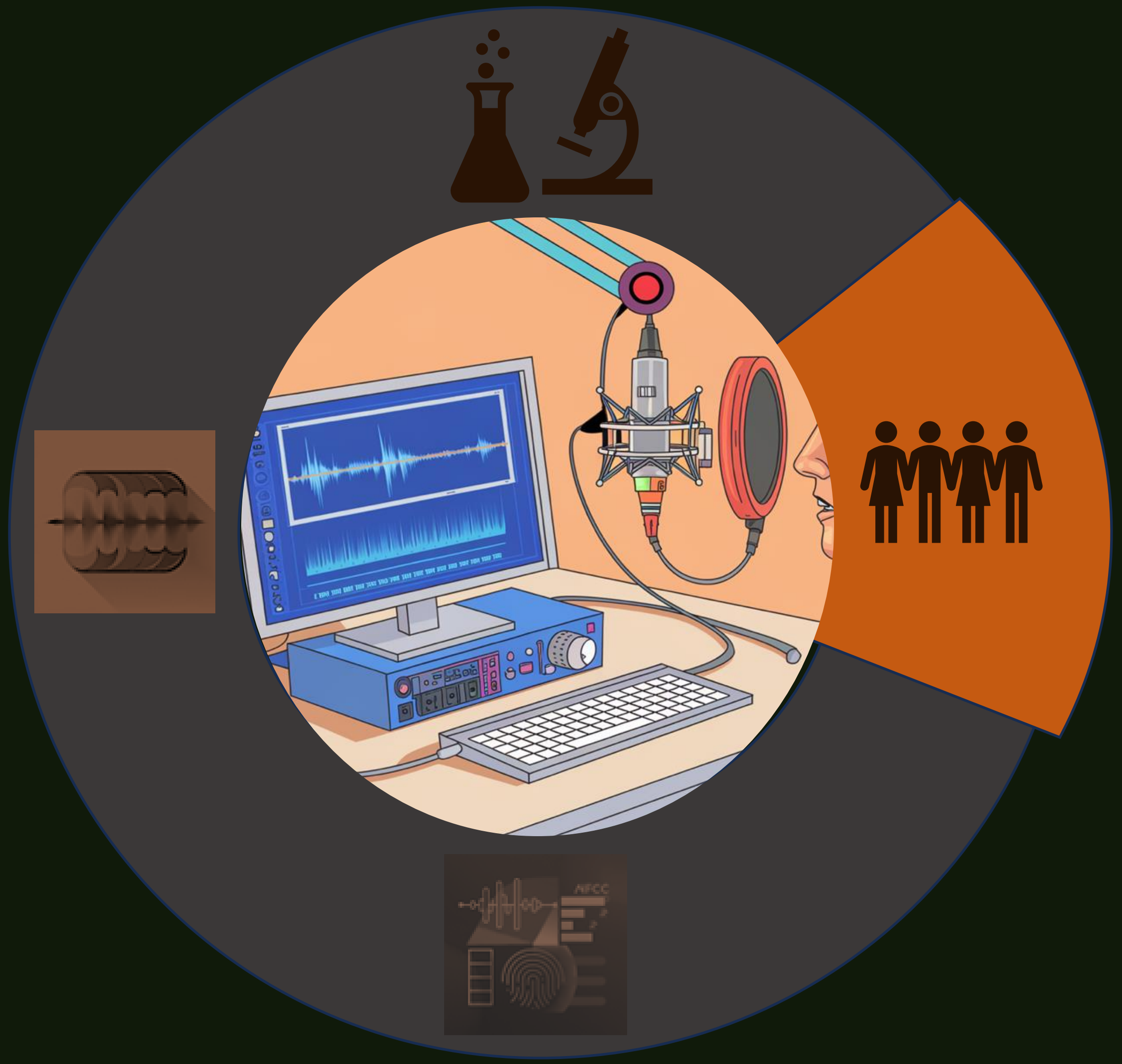
Contributions

- First systematic approach to Luxembourgish dialect classification.
- Benchmarking multiple models (Wav2Vec2, Whisper, CNN).
- Data augmentation techniques



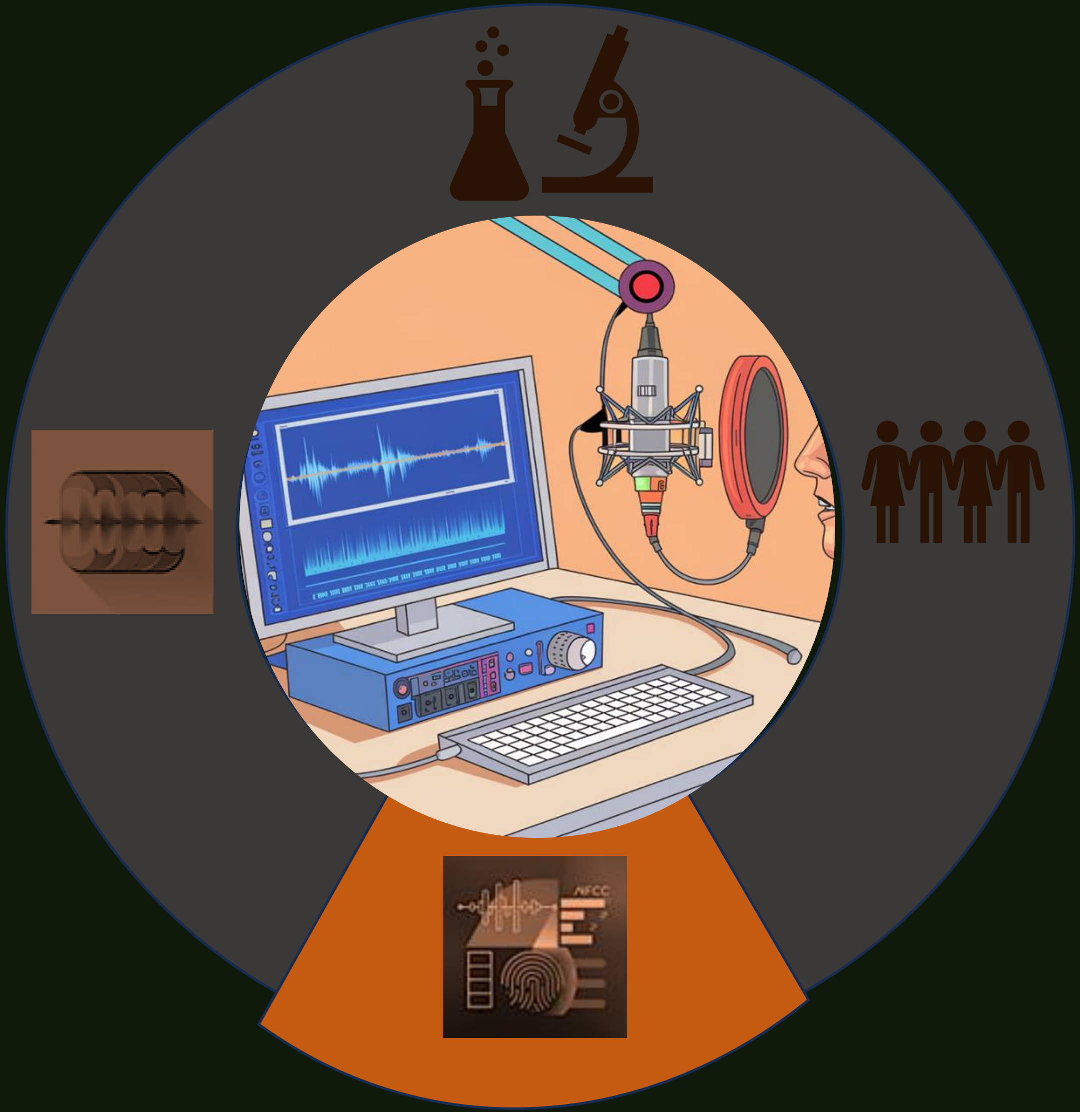
Problem Statement

- **Limited annotated data** makes model training difficult in low-resource languages.
- **High inter-dialect variability vs. low intra-dialect variation** complicates classification.
- **Traditional features (MFCCs) struggle** with phonetic and prosodic variations across dialects.
- Multilingual Interference (German & French Influence



PARTICIPANTS & TASKS

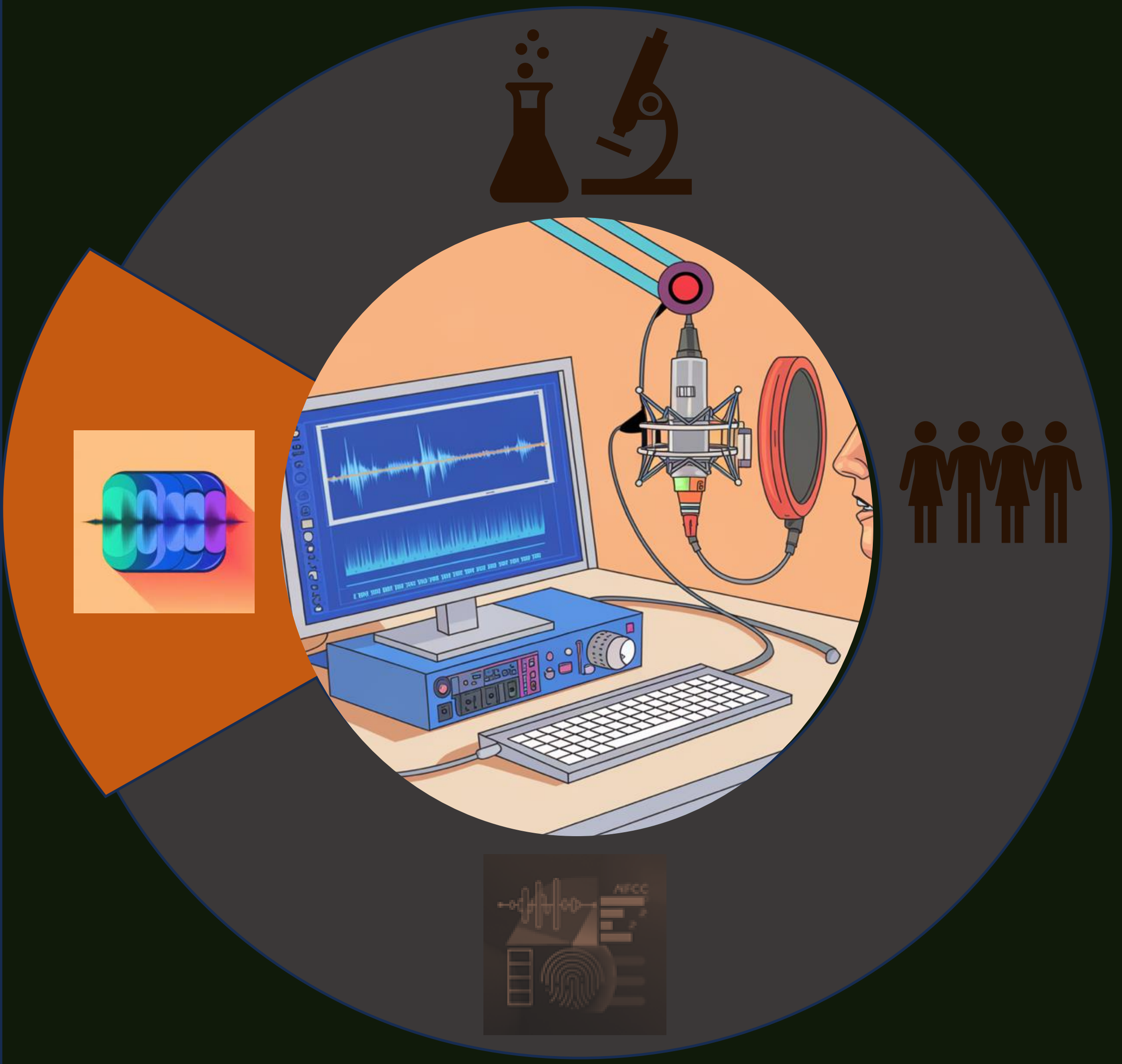
Attribute	Category	Count
Total Audio Files		1720
	Unique Entries	1720
Gender	Female	1210
	Male	510
Age Group	25–34	567
	35–44	377
	45–54	352
	55–64	277
	65+	132
Dialect Region	Center	762
	South	482
	East	293
	North	168



FEATURE EXTRACTION

Mel-Frequency Cepstral Coefficients (MFCCs)

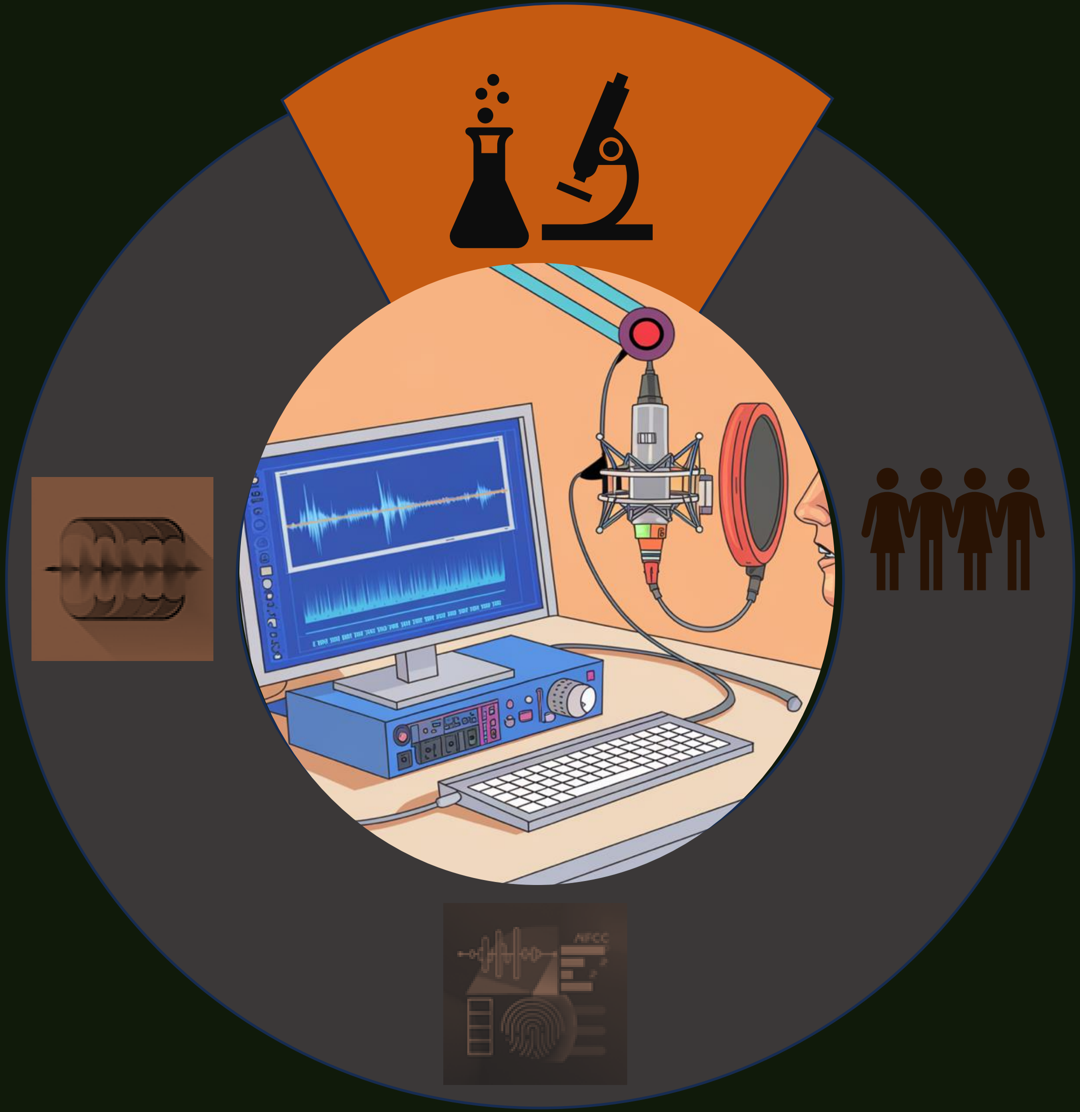
- Captures phonetic and acoustic characteristics.
- Used in Random Forest, CNNs, Wav2Vec2, XLSR-Wav2Vec2, Whisper.



SPEECH DATA AUGMENTATION

- Increase dataset diversity & make the model more robust to real-world speech variations.

- Augmentation Methods Used:
  - Time Stretching
  - Pitch Shifting



TRAINING & EVALUATION

- Training Setup:
  - 5-fold cross-validation (ensures all speakers appear in both train/test sets).
  - Adam optimizer, categorical cross-entropy loss (standard for classification).
  - Early stopping (patience=10 epochs) to prevent overfitting.
- Evaluation Metrics Metrics: Accuracy, Precision, Recall

RESULTS

- **Moderate accuracy (55%-73%)** across models.

- **Random Forest struggled**

- Northern:
  - CNN-Spectrogram (72%),
- Central:
  - CNN-LSTM (73%) showed the highest accuracy.
- Southern:
  - CNN-Spectrogram & CNN-LSTM (72%) led in classification.
- Eastern: Most challenging across all models (~55%-70%).

Baseline (Without Augmentation)				
Model	**Northern**	**Central**	**Southern**	**Eastern**
Random Forest	63/61/62	58/60/60	56/57/57	55/55/55
Wav2Vec2	**70/72/72**	69/70/70	70/71/71	69/69/70
Whisper	67/69/68	66/67/66	68/69/69	64/65/65
XLSR-Wav2Vec2	68/70/69	66/68/67	69/70/69	63/64/64
CNN-Spectrogram	72/71/73	71/71/71	**72/74/73**	**70/69/71**
CNN-LSTM	72/70/72	**73/72/71**	69/72/70	68/71/72

RESULTS

- **Overall accuracy improvements (3%-5%)** across all models.
- **CNN-Spectrogram achieved the highest accuracy** in
  - Northern (76%), Southern (79%), and Eastern (78%).
- **CNN-LSTM remained strong**, leading in Central dialect (75%).

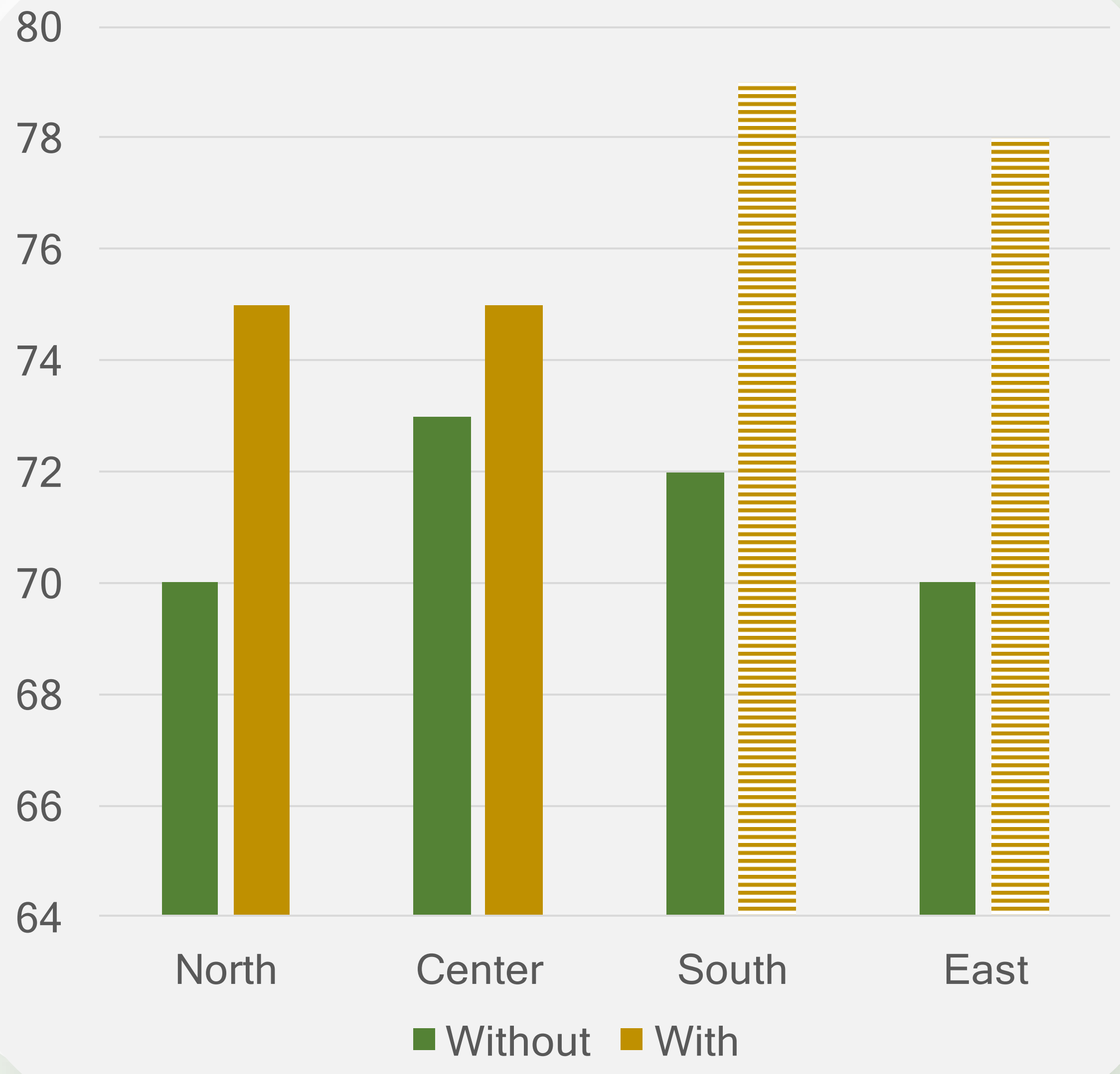
Optimized (With Augmentation)

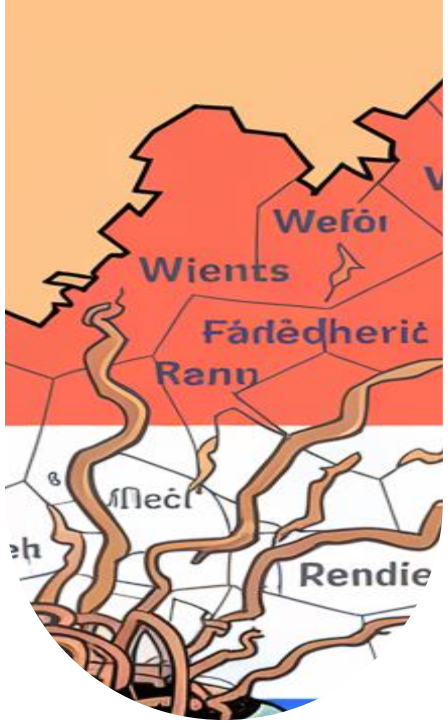
Model	Northern	Central	Southern	Eastern
Random Forest	71/69/71	65/63/65	63/61/63	59/58/59
Wav2Vec2	**75/74/75**	72/71/72	73/72/73	70/71/71
Whisper	72/72/73	70/70/70	72/72/72	67/69/68
XLSR-Wav2Vec2	72/73/72	69/70/70	71/72/71	66/66/66
CNN-Spectrogram	76/74/76	74/73/74	**79/76/78**	**78/75/76**
CNN-LSTM	76/73/74	**75/74/73**	77/75/77	72/70/71

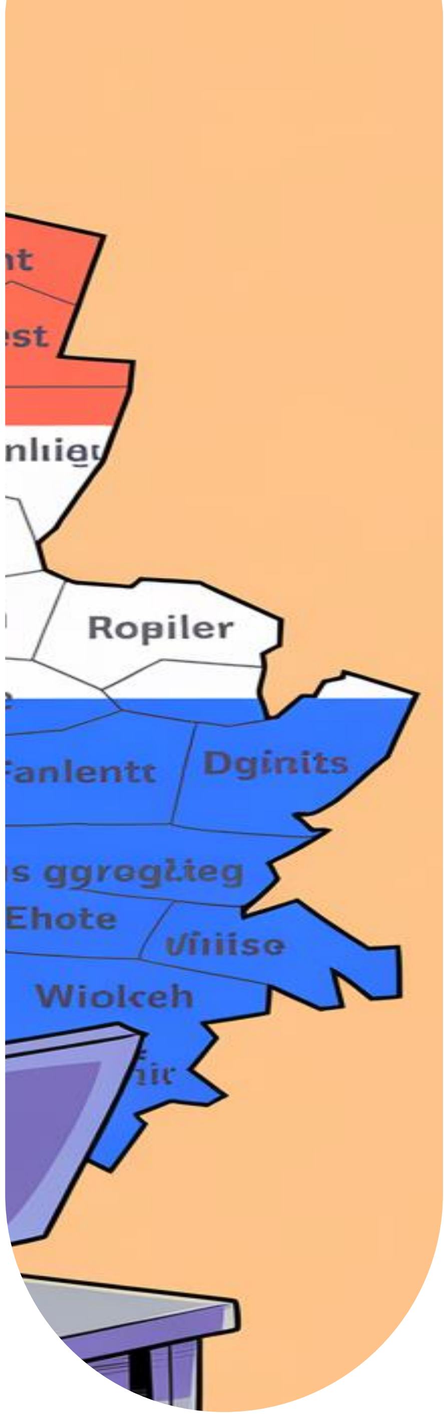
Results

CNN-Spectrogram achieved the highest accuracy in Northern (76%), Southern (79%), and Eastern (78%).

✅ **CNN-LSTM remained strong**, leading in Central dialect (75%).







Deep learning models outperformed traditional methods like Random Forest.

Dialects with more data performed better

CNN-Spectrogram & CNN-LSTM outperformed other models, demonstrating their strength in phonetic and prosodic feature extraction.

CNN-Spectrogram achieved highest accuracy in Northern, Southern, and Eastern dialects.

- six models for dialect classification (ML & DL).
- CNN-Spectrogram and CNN-LSTM performed best.
- Data augmentation improved classification, especially for underrepresented dialects.

- Expand the dataset to include more diverse speakers & spontaneous speech
- Refine fine-tuning for Whisper & XLSR-Wav2Vec2 leveraging multilingual transfer learning

six models for dialect classification (ML & DL).
CNN-Spectrogram and CNN-LSTM performed best
Data augmentation improved classification, especially for underrepresented dialects.

- Expand the dataset to include more diverse speakers & spontaneous speech.

- Refine fine-tuning for Whisper & XLSR-Wav2Vec2, leveraging multilingual transfer learning

Thank You

Ninahkivanani

https://scholar.google.com/citations?user=H6JYohsAAAAJ&hl=en

Nina.hosseinikivanani@uni.lu

Illustrations in this presentation were AI-generated.