Classifying Pakistan's Diverse Languages Through Speech Using Deep Neural Networks

Presented By :

Muhammad Ibad, Eshal Khalid, Aina Shakeel

Habib University | Fall 2024

INTRODUCTION

- Pakistan’s lingua franca and regional languages.
- Lack of effective communication & services between regional languages
- Limited research on regional language classification in Pakistan

Rank	Language	2023 Census
1	Punjabi	36.98%
2	Pashto	18.15%
3	Sindhi	14.31%
4	Saraiki	12.00%
5	Urdu	9.25%
6	Balochi	3.38%
7	Hindko	2.32%
8	Brahui	1.16%
9	Mewati	0.46%
10	Kohistani	0.43%
11	Kashmiri	0.11%
12	Shina	0.05%
13	Balti	0.02%
14	Kalasha	0.003%
15	Others	1.38%

PROBLEM STATEMENT

Given an audio clip containing speech,

detect and identify the regional language being spoken

DATASET

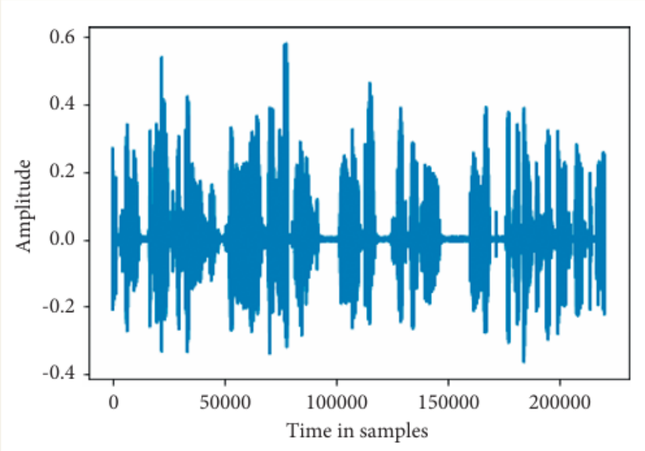
Sources

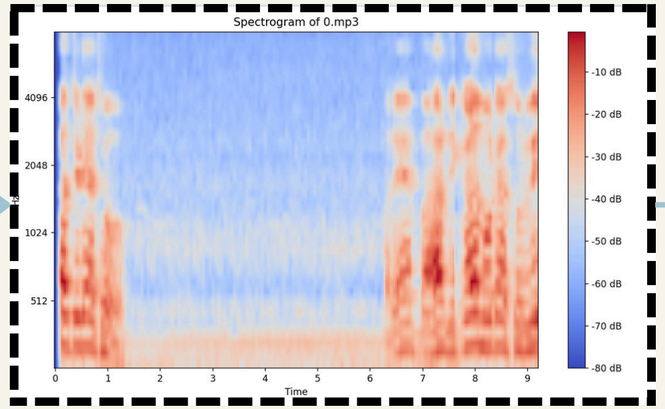
- Audio data was collected from multiple sources including:
  - Mozilla’s Common Voice dataset
  - Open-source platforms such as YouTube

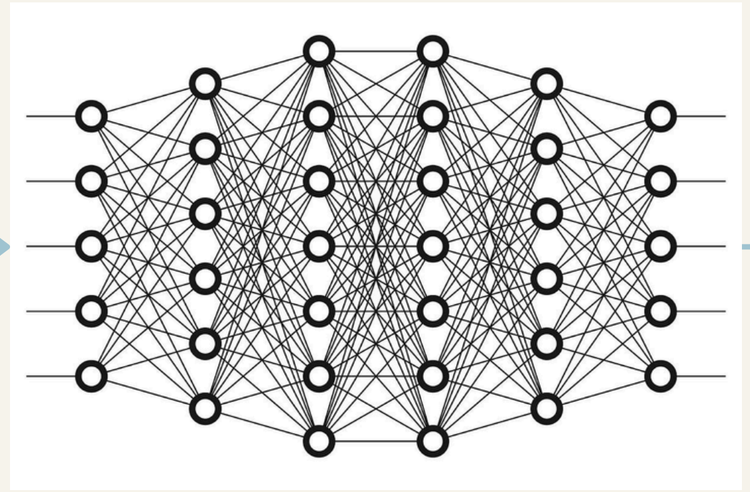
Languages

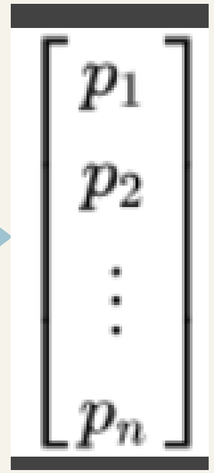
- Urdu, Punjabi, Sindhi, Pushto, Seraiki
- 5000 sample per language (~5 seconds per sample)

OVERALL APPROACH









Audio .wav file

Pre-processing

Deep Learning Model

Classifying Languages

EVALUATION METRICS

- **Accuracy:** The percentage of correct predictions out of all predictions.
- **Precision**: The proportion of true positives out of all predicted positives.
- **Recall**: The proportion of true positives out of all actual positives.
- **F1-Score**: Measures the balance between precision and recall.

MODELS TRAINED

- **Baseline CNN**
  - **Layers** = 21
  - **Input Size** = (375, 256, 3)
- **AlexNet**
  - **Layers** = 17
  - **Input Size** = (224, 224, 3)
- **ResNet50**
  - **Layers** = 59
  - **Input Size** = (224, 224, 3)

TRAINING

- **Epochs =** 40 (Early Stopping Enabled)
- **Learning Rate =** 0.001
- Learning Rate Scheduler
- **Batch Size =** 16
- **Optimizer =** Adam
- **Loss Function =** Sparse Categorical Cross Entropy Loss

CNN (LR Sched, BSize=16)

Languages	Precision	Recall	F1-Score
Punjabi	0.84	0.74	0.79
Pushto	0.74	0.8	0.77
Saraiki	0.83	0.82	0.82
Sindhi	0.98	0.98	0.98
Urdu	0.74	0.78	0.76

Final Model Accuracy

82%

AlexNet (Def LR, BSize=16)

Languages	Precision	Recall	F1-Score
Punjabi	0.75	0.6	0.67
Pushto	0.63	0.6	0.62
Saraiki	0.81	0.65	0.72
Sindhi	0.83	0.99	0.9
Urdu	0.6	0.75	0.67

Final Model Accuracy

72%

ResNet50 (LR Sched, BSize=16)

Languages	Precision	Recall	F1-Score
Punjabi	0.87	0.57	0.69
Pushto	0.51	0.9	0.65
Saraiki	0.86	0.71	0.78
Sindhi	1	0.88	0.93
Urdu	0.74	0.65	0.69

Final Model Accuracy

74%

FURTHER TESTING

- **Epochs =** 40 (Early Stopping Enabled)
- **Learning Rate =** 0.001
- Learning Rate Scheduler
- **Batch Size =** 32
- **Optimizer =** Adam
- **Loss Function =** Sparse Categorical Cross Entropy Loss

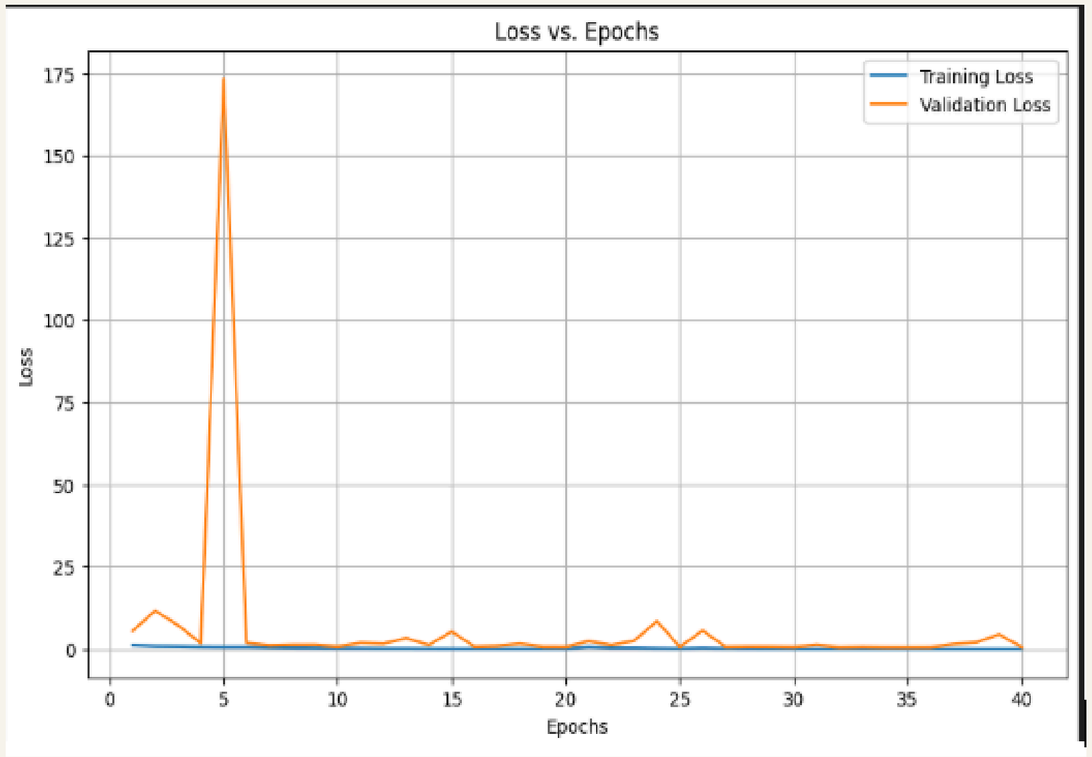
CNN (Sched LR, BSize=32)

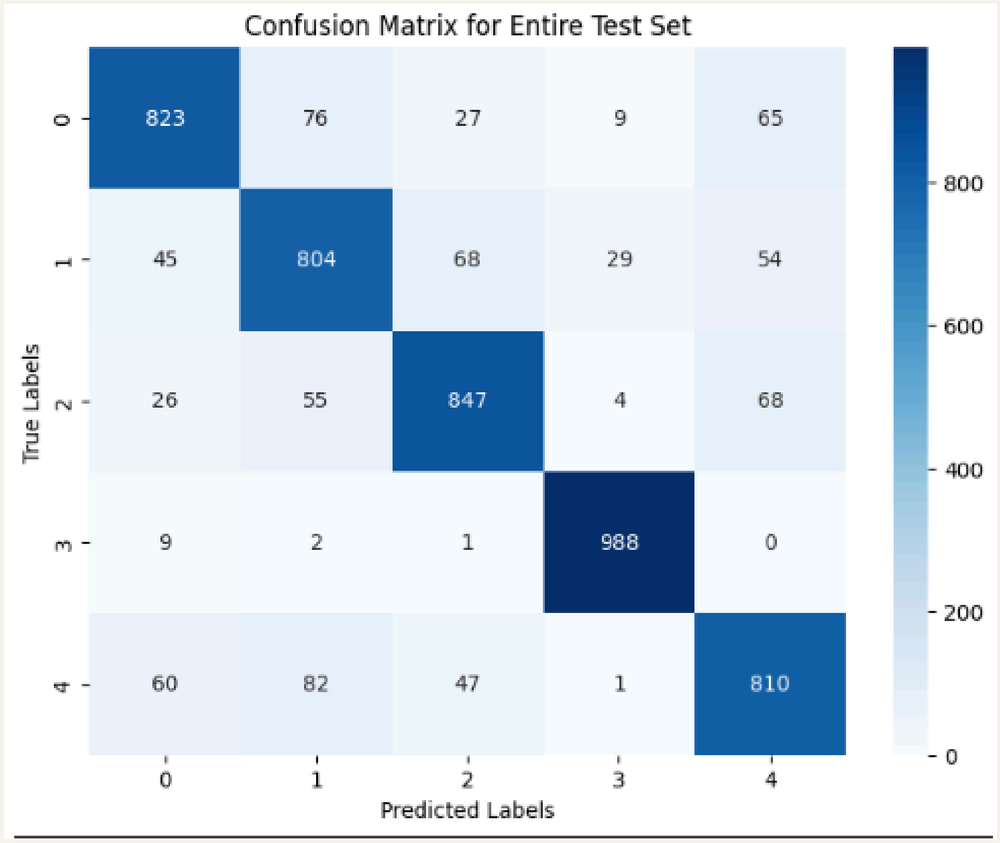
Languages	Precision	Recall	F1-Score
Punjabi	0.85	0.82	0.84
Pushto	0.79	0.8	0.8
Saraiki	0.86	0.85	0.85
Sindhi	0.96	0.99	0.97
Urdu	0.81	0.81	0.81

Final Model Accuracy

85%

Observations for CNN





0: Punjabi 1: Pushto 2: Saraiki 3: Sindhi 4:Urdu

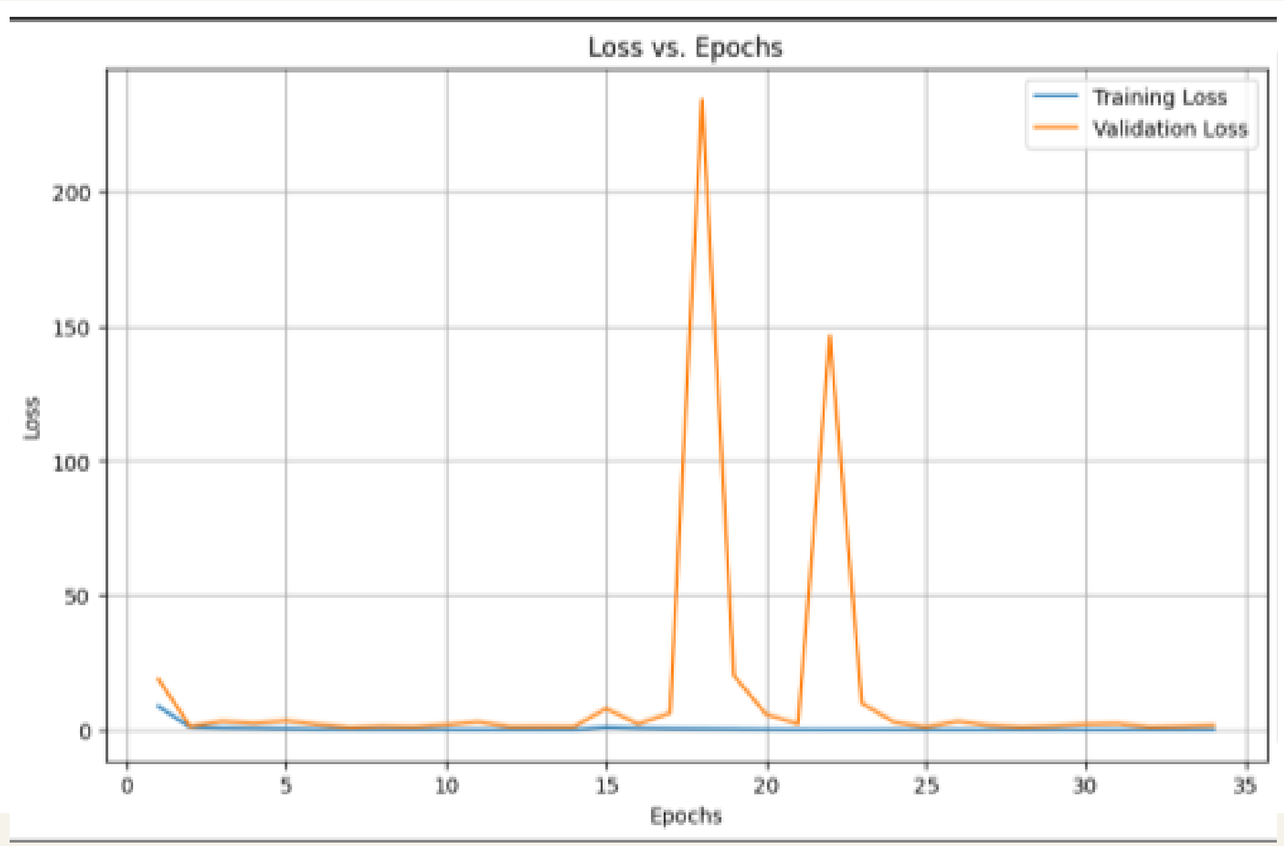
AlexNet (Sched LR, BSize = 32)

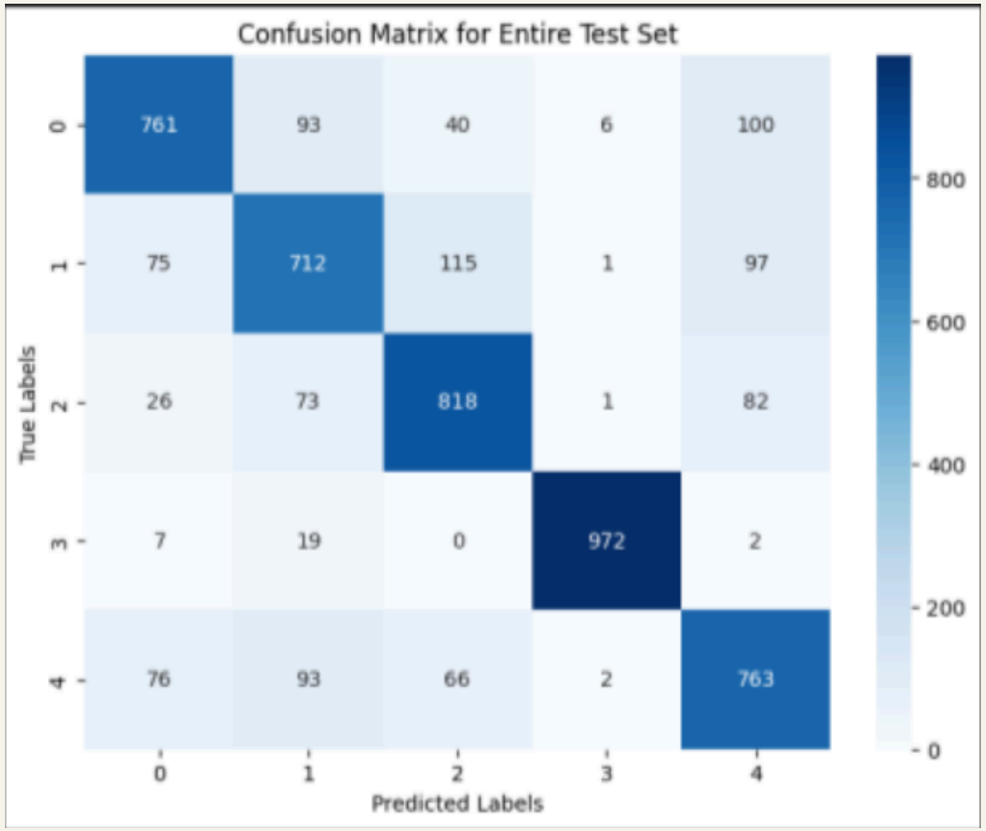
Languages	Precision	Recall	F1-Score
Punjabi	0.81	0.76	0.78
Pushto	0.72	0.71	0.72
Saraiki	0.79	0.82	0.8
Sindhi	0.99	0.97	0.98
Urdu	0.73	0.76	0.75

Final Model Accuracy

81%

Observations for AlexNet





0: Punjabi 1: Pushto 2: Saraiki 3: Sindhi 4:Urdu

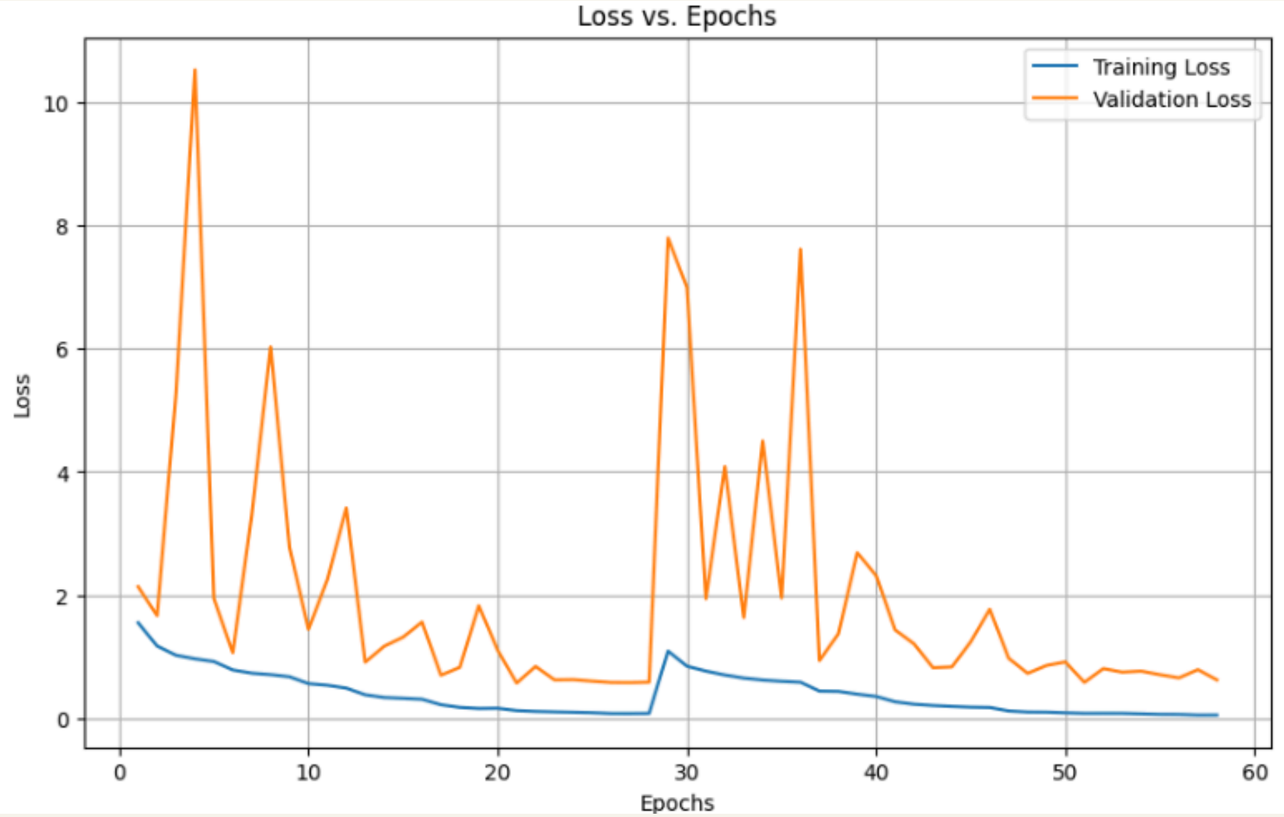
ResNet50 (Sched LR, BSize = 32)

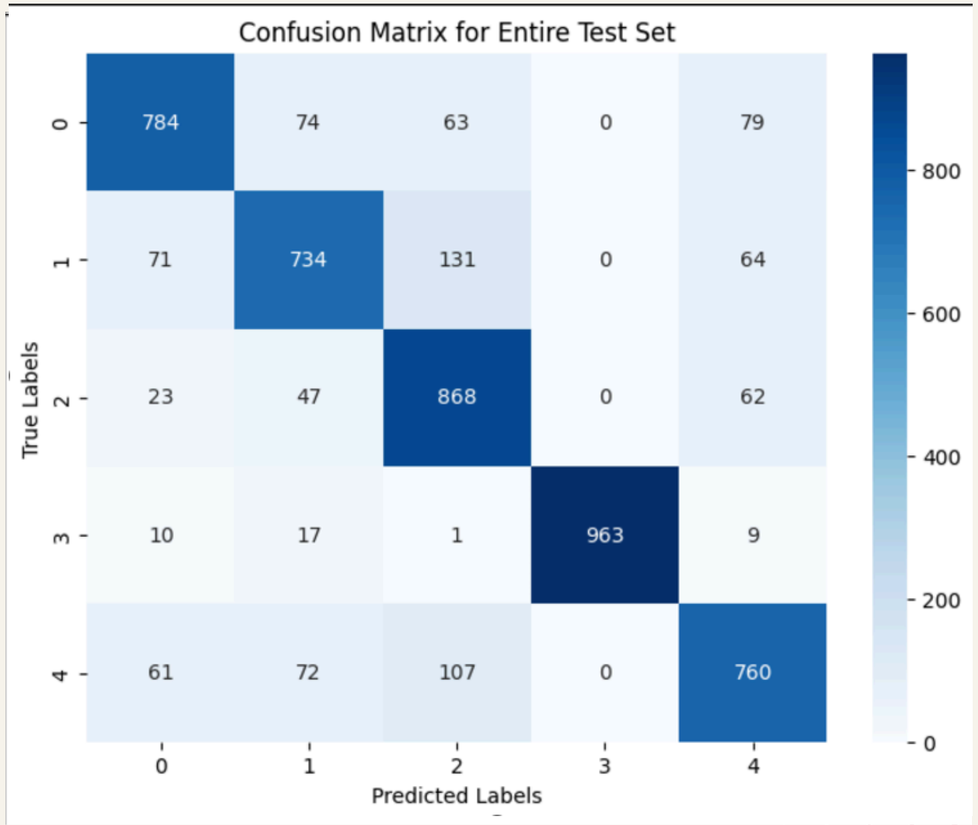
Languages	Precision	Recall	F1-Score
Punjabi	0.83	0.78	0.8
Pushto	0.78	0.73	0.76
Saraiki	0.74	0.87	0.8
Sindhi	1	0.96	0.98
Urdu	0.78	0.76	0.77

Final Model Accuracy

82%

Observations for ResNet50





0: Punjabi 1: Pushto 2: Saraiki 3: Sindhi 4:Urdu

Overall Results

Model	BS=32, NO LR	BS=32, Def LR	BS=16, NO LR	BS=16, Def LR
CNNs	84.00%	**85.00%**	80.00%	82.00%
AlexNet	64.00%	**81.00%**	72.00%	71.00%
ResNet50	64.50%	**82.00%**	59.60%	74.00%

Discussion

Paper Name	Model	Dataset Size	Languages	Result
Spoken Language Recognition using CNN	CNNs	36,810 samples	German, English & Spanish	0.94 (German), 0.92 (English), 0.91 (Spanish) [F1-Score]
Multiclass Language Identification using Deep Learning on Spectral Images of Audio Signals	ResNet50	7,000 samples	English, Spanish, German, French, Russian	89% [Overall Accuracy]
Spoken Language Identification Using Deep Learning	CNNs	73,620 samples	English, German & Spanish	98.9% [Overall Accuracy]
Our Best Results	**CNNs**	**25,000 samples**	**Punjabi, Pushto, Saraiki, Sindhi, Urdu**	**85% [Overall Accuracy]**
Our Best Results	**AlexNet**	**25,000 samples**	**Punjabi, Pushto, Saraiki, Sindhi, Urdu**	**81% [Overall Accuracy]**
Our Best Results	**ResNet50**	**25,000 samples**	**Punjabi, Pushto, Saraiki, Sindhi, Urdu**	**82% [Overall Accuracy]**

Future Works

- **Test More Models:** Explore different deep learning architectures.
- **Expand Dataset:** Add more samples and try data augmentation.
- **Increase Classes:** Train models for multiple languages.
- **Different Features**: Try MFCCs or wavelet transforms.
- **Real-Time Detection:** Optimize for speed and noise robustness.

REFERENCES

[1] “Spoken Language Recognition Using CNN | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 18, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9031923
[2] “FuzzyGCP: A deep learning architecture for automatic spoken language identification from speech signals | Request PDF.” Accessed: Oct. 18, 2024. [Online]. Available: https://www.researchgate.net/publication/347518701_FuzzyGCP_A_deep_learning_architecture_for_automatic_spoken_language_identification_from_speech_signals
[3] “Spoken Language Identification Using Deep Learning - Singh - 2021 - Computational Intelligence and Neuroscience - Wiley Online Library.” Accessed: Oct. 18, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2021/5123671
[4] “Development of a regional voice dataset and speaker classification based on machine learning | Journal of Big Data | Full Text.” Accessed: Oct. 18, 2024. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00435-9
[5] “Data augmentation and deep neural networks for the classification of Pakistani racial speakers recognition [PeerJ].” Accessed: Oct. 18, 2024. [Online]. Available: https://peerj.com/articles/cs-1053/
[6] “A Language Identification System using Hybrid Features and Back-Propagation Neural Network - ScienceDirect.” Accessed: Oct. 18, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0003682X19313672
[7] S. Revay and M. Teschke, “Multiclass Language Identification using Deep Learning on Spectral Images of Audio Signals,” May 10, 2019, arXiv: arXiv:1905.04348. doi: 10.48550/arXiv.1905.04348.
[8] resourceExport, “Language data for Pakistan,” CLEAR Global, Feb. 10, 2020. https://clearglobal.org/resources/language-data-for-pakistan/
[9] “POPULATION BY MOTHER TONGUE | Pakistan Bureau of Statistics,” www.pbs.gov.pk. https://www.pbs.gov.pk/node/97
[10] “GitHub - hubertmaka/Spoken-language-detection: The aim of the project is to design and build a model that recognizes language from a given sound sample. The assumption is a given number of different languages that the model will be able to recognize.,” GitHub, 2024. https://github.com/hubertmaka/Spoken-language-detection/tree/main

thank you>

```
[ ]
```

```python
# Convert the model.
converter = tf.lite.TFLiteConverter.from_saved_model('saved_model/1')
tflite_model = converter.convert()
```

```
...
```

The Kernel crashed while executing code in the current cell or a previous cell.
Please review the code in the cell(s) to identify a possible cause of the failure.
Click here for more info.
View Jupyter log for further details.

Q/A