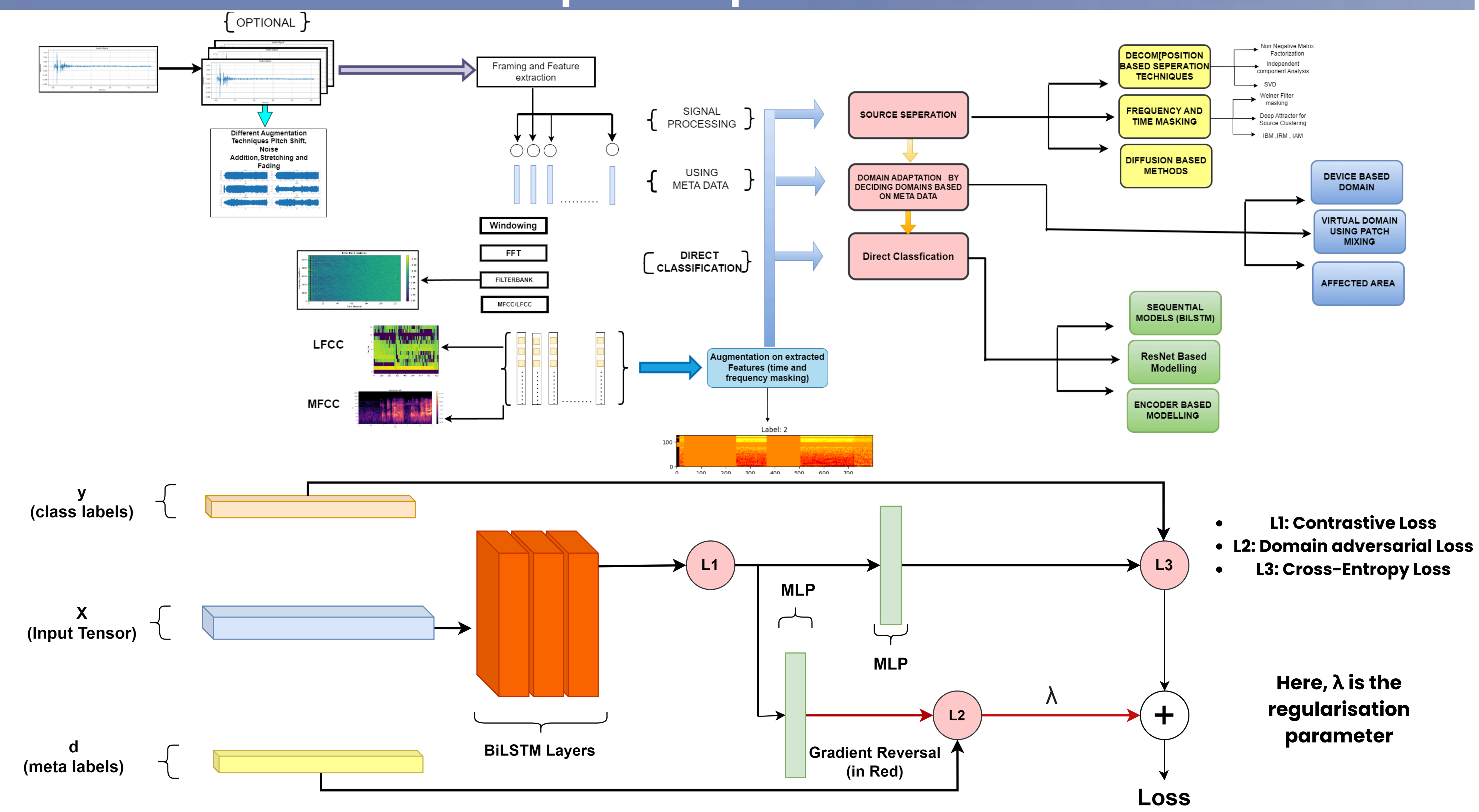
Detecting NasoLaryngeal Anomalies in Pediatric Population

Hemansh Shridhar, Department of Electronics and Communications, Indian Institute of Information Technology, Allahabad Mentor: Dr.Tushar Sandhan , Department of Electrical Engineering, Indian Institute of Technology Kanpur

Abstract

- In this Project a framework for detecting Breathing related anomalies in pediatric population has been presented.
- The main aim of this project was to design a complete workflow for the detection of NasoLaryngeal diseases in infants using breathing data.
- Various State of the Art feature extraction and Augmentation techniques were explored during the course of the project.
- Experiments were carried out on primarily two datasets one containing infant sounds and one containing adult respiratory sounds.
- The adult respiration dataset was created by combining various other publicly available datasets. This was done inorder to obtain a dataset that contains samples recorded using different instruments and examination methods
- In the end various loss functions like contrastive loss and domain adversarial methods were applied on a BiLSTM model to obtain a methodology that is able to generalize well across a data which contains inter class similarity.

Proposed Pipeline



- **L1: Contrastive Loss**
- **L2: Domain adversarial Loss**
- **L3: Cross-Entropy Loss**

Here, λ is the regularisation parameter

Proposed Methodology

- ***Input***: The model takes extracted MFCCs and LFCCs as extracted features .In this project the number of extracted MFCCS and LFCCs has been set at 13. For dataset containing respiratory or breathing data, usually some metadata containing the patient or device information is provided. The meta data serves as our meta labels.
- ***Embedding:*** The BiLSTM model is used to embed each sample in input tensor (containing batch of input data samples) to embedding *h*. This embedding ***h*** is normalized to normalized embedding ***z***.
- ***Contrastive Loss:*** Contrastive loss is calculated among the samples in the batch containing the embeddings as shown in the optimisation section. For any anchor ***z*** inside the batch, it forces the encoder to maximize the cosine similarity between ***z*** and its positive ***z+***, while simultaneously minimize the ones between z and all its negatives ***zk*** (k = 1..N). Thus, the embeddings which are compact in intra-class and well separated in inter-class are obtained **.**
- ***Domain classifier Loss:*** Since respiratory data recorded using different recording instrument is prone to device -based domain based similarity, we used domain adversarial loss to introduce domain invariance within the data. The gradients of the domain classifier loss are inverted before they reach the feature extractor. This is done by multiplying the gradients by a negative constant. The gradient reversal feature extractor forces the model to reduce the distribution shift between different stethoscope classes.
- Thus the final loss is calculated as the sum of CE loss with weighted sum of CL and DA loss with regularisation parameters controlling their contribution.

Proposed optimisation

The Various Loss functions and parameters the model leveraged are mentioned below:

Contrastive Loss [1] : To force the BiLSTM encoder to produce embeddings that are close for samples of the same class (positive pairs) and distant for samples of different classes (negative pairs).

$$\mathcal{L} = -\log \frac{\exp\left(z^\top z^+\right)}{\exp\left(z^\top z^+\right) + \sum_{k=1}^{N} \exp\left(z^\top z_k\right)}$$

Here, $z^\top z^+$ computes the cosine similarity between the normalized vectors z (anchor) and z^+ (positive sample), while $z^\top z_k$ computes the cosine similarity between z and each negative sample z_k.

Domain Adversarial Loss [2]: used in domain adversarial training (DAT) to promote domain invariance in the learned feature representations. The main objective is to ensure that the features extracted by the model are not specific to any particular domain.

$$\mathcal{L}_{\mathrm{DA}} = -\sum_{i=1}^{n} d_i \log(\hat{d}_i).$$

Cross-Entropy Loss: Measures the discrepancy between the true class labels y and the predicted $\hat{y}$ probabilities for the primary classification task.

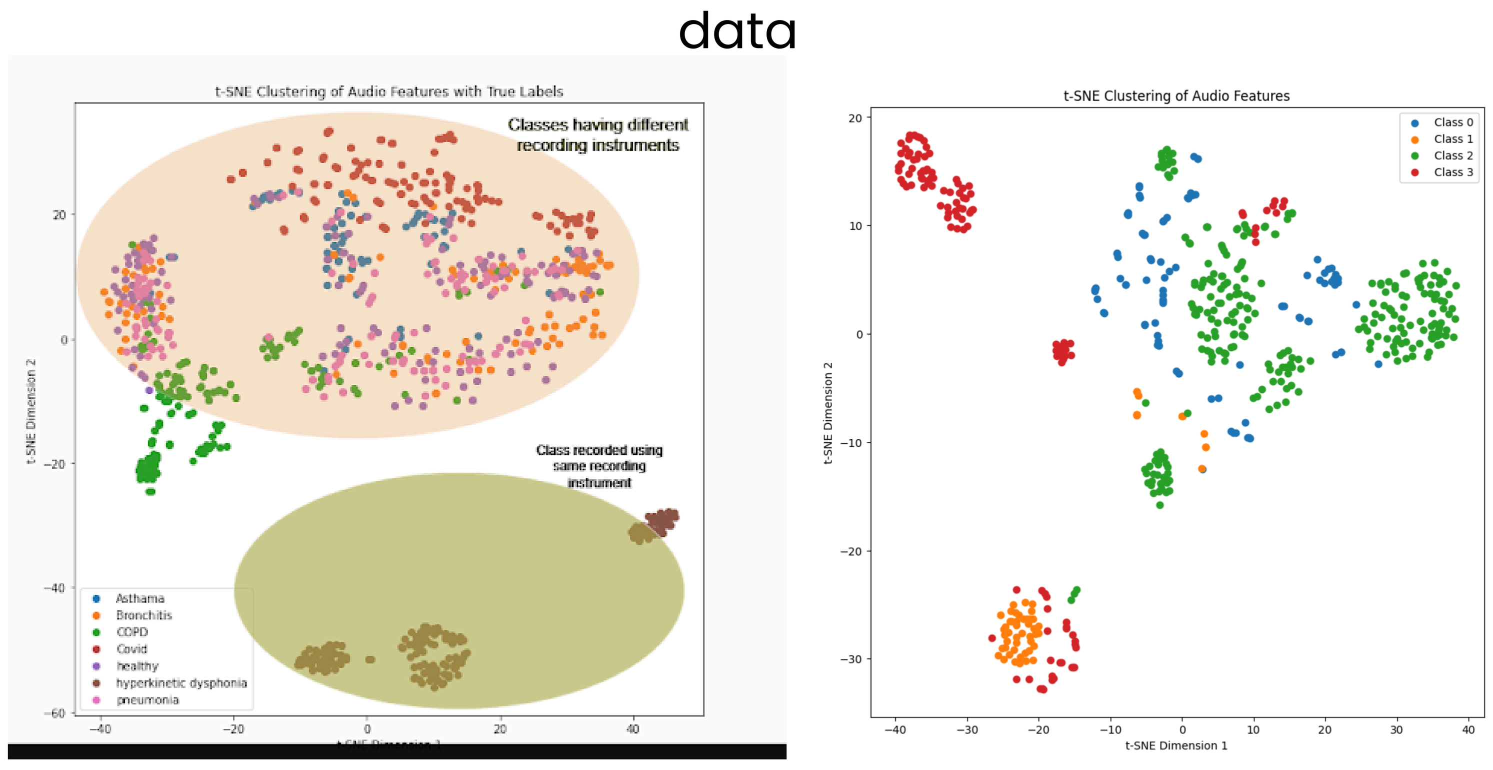
$$\mathcal{L}_{\mathrm{CE}} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i)$$

The total loss is calculated as : $L_{total} = L_{CE} + \lambda_1 L_{DA} + \lambda_2 L_{CL}$

Optimization: Adam optimizer was used to optimize the loss with a learning rate of 0.0001 over 100 epochs.

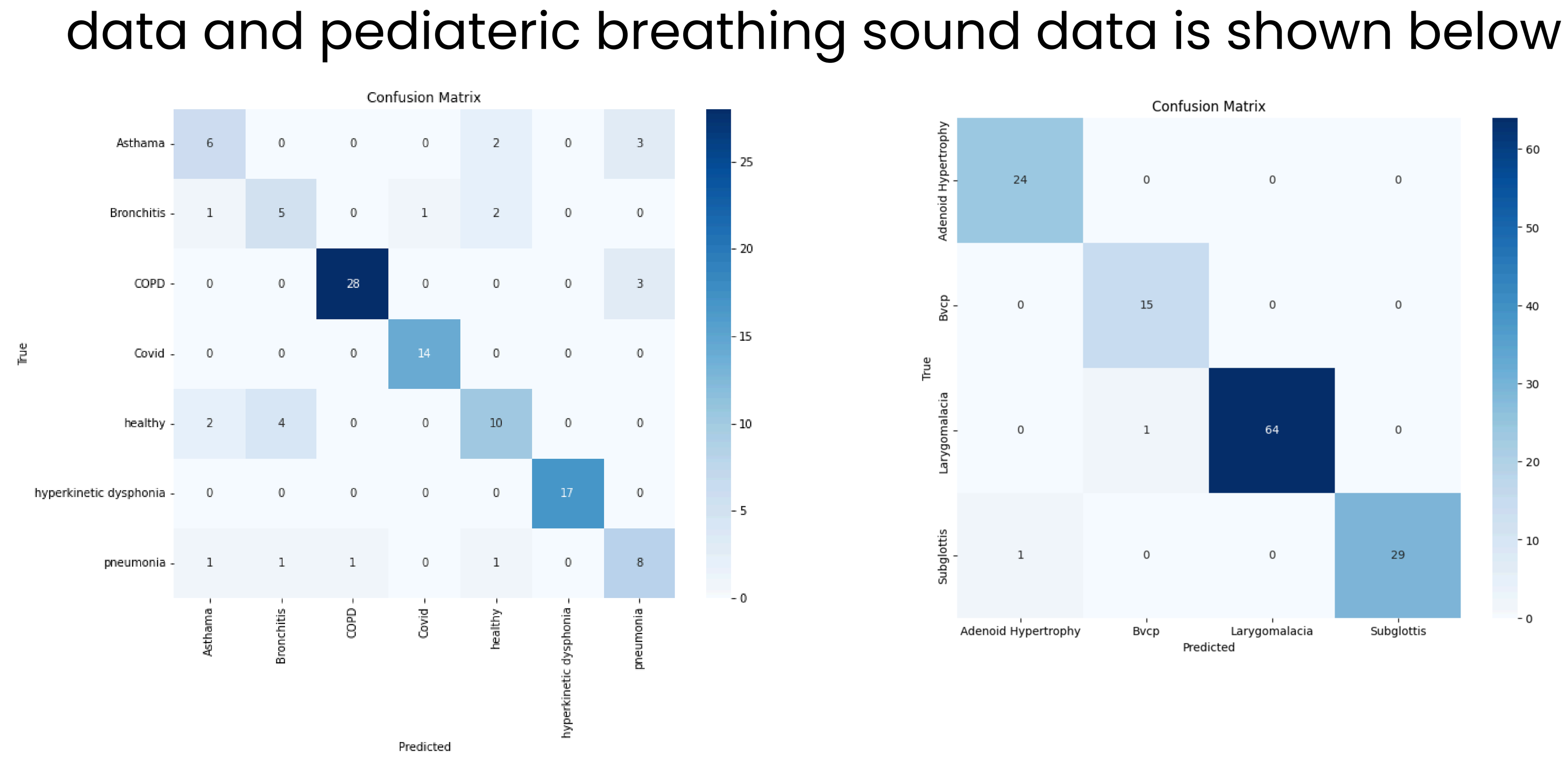
Results

t-SNE visualisation on Adult Respiratory data and pediateric breathing data



As observed in the the first plot classes sharing similar recording instruments are clustered very close to each other, therefore application of domain adaptation technoques becomes necassary

Confusion Matrix obtained on test data of composite adult respiratory data and pediateric breathing sound data is shown below



Conclusion

In conclusion, the proposed methodology shows promise in the diagnosis and detection of audio related disease diagnosis. From the analysis we can easily observe the heavy influence of device-based domain similarity in the data. To mitigate the influence of such factors and to increase the similarity between the samples of same class contrastive loss and domain adversarial training was introduced which shows remarkable promise on such tasks. One of the remarkable contribution of this research work is the introduction of the dataset of adult breathing sounds that contains data from diverse populations and clinical settings. Training and testing on this dataset not only helped us gain insight on the various target areas, real life identification of audio related diseases suffer but also gave us direction for future work.

References

- Kim, June-Woo, et al. "Stethoscope-Guided Supervised Contrastive Learning for Cross-Domain Adaptation on Respiratory Sound Classification." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.
- Song, Wenjie, Jiqing Han, and Hongwei Song. "Contrastive embeddind learning method for respiratory sound classification." ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021.

Dr. Tushar Sandhan
Assistant Prof. (EE)
Email: sandhan@iitk.ac.in

Hemansh Shridhar
Student: B.Tech ECE
Email: shridharhemansh@gmail.com