Modular Decision Support Networks (MoDN) to predict respiratory disease from lung sounds

Mariia Vidmuk, Daria Yakovchuk, Aleksandra Novikova
ML4Science Course project, EPFL

***Abstract*—The scope and combination of clinical data collected from patients is highly variable, depending on available resources and time. Often collecting all the data is not necessary to have a confident diagnostic prediction. Modular Clinical Decision Support Networks (MoDN) were created to deal with this flexible combination and number of inputs by encoding them in series. This project makes a real world implementation of the approach encoding a combination breath sounds for pneumonia classification. As a base model, we use an existing state-of-the art CNN-based model for the task called *DeepBreath* which encodes all recordings at once. We show that DeepBreath integrated into MoDN outperforms standard DeepBreath and validate performance by training these two networks on the same data and hyperparameters.**

I. Introduction

A clinical decision support system (CDSS) is intended to improve healthcare delivery by enhancing medical decisions with targeted clinical knowledge, patient information, and other health information [1].

In clinical contexts that are constantly changing, many clinical support systems use static, generic-based logic, which leads to unevenly distributed accuracy and inconsistent performance. The physician is often not aware of the inputs or processes the CDSS utilizes, creating a black box that degrades trust in the system.

Other important issues are that the static nature of the algorithms necessitates the acquisition of all inputs, which is not always possible in settings with variable resources. Indeed, it is often not necessary to collect all the inputs to have a confident prediction. There is thus a need for adaptable algorithms, able to handle a flexible number and combination of inputs. Previous attempts have used pre-defined feature clusters [2] to design a series of models with various combinations of inputs. However, this approach is laborious to train and lacks finer-grained flexibility desired and does not provide feedback to the clinician on what combination is most suitable for the given patient. Modular Clinical Decision Support Networks (MoDN) [3] were proposed by the intelligent Global Health Research group (hosting this project). This architecture proposes feature-wise *modules*, which encodes single features in a sequence of any length or combination. Encoding can stop after any number of features providing the physician interpretable, continuous predictive feedback.

MoDN has been successfully tested on tabular data with a series of MLP encoders [3]. It is the aim of this project to expand the implementation to audio signals for the diagnosis of pneumonia. Pneumonia is often diagnosed from between 1 and 12 auscultations on the chest on various anatomical positions of a a patient recorded for variable lengths. By using MoDN for this task, we could keep the flexibility of variable durations and combinations of positions and provide continuous feedback to the clinician to guide their acquisition.

II. Related Work

MoDN. To expand on the introduction of MoDN above, we refer to the architecture of MoDN in (Fig. 1) from [3].

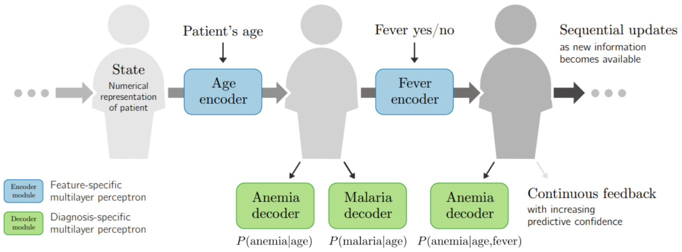

Fig. 1: Schematic architecture of MoDN, obtained with permission from [3]

MoDN is a novel decision tree composed of neural network *modules* specific to each feature. The MoDN model consists of sequentially connected encoders and decoders (Fig. 2). A state - vector-representation of a patient, is transmitted through the entire network from encoder to encoder. Each encoder takes as input the current state from the previous encoder and data. Encoder updates the state based on the value of new features. Decoders extract predictions from the state. Decoders can be applied to the state after any module thus offering comprehensible ongoing feedback on the propensity for prediction of each question in a CDSS questionnaire.

MoDN was previously applied to tabular CDSS-derived data [3]. Here each feature was encoded by an individual MLP module. In addition to creating interpretable feedback, MoDN also outperformed ”monolithic” baselines (i.e. where all features are added to a single encoder). Expanding on this, we would like to test MoDN on non-tabular data and deeper neural networks. The use case and data are described below.

DeepBreath. Lower Respiratory Tract Infection (LRTI), also known as pneumonia, is a major global cause of preventable mortality and the main factor in improper antibiotic prescriptions. Because no one test can be used to confidently identify it, there is a significant rate of misdiagnosis, which is widespread.

Commonly, clinicians use stethoscopes to auscultate the chest and interpret the breath sounds. However, their interpretations are notoriously inconsistent and limited [4] Heitmann et al. of our host lab (intelligent Global Health) at EPFL created the DeepBreath model to automate the interpreation of breath sounds [4]. DeepBreath is a composite binary classifier individually trained for several diagnostic categories (e.g. pneumonia vs not pneumonia etc.). It is composed of a CNN audio classifier, which generates predictions for individual recordings (i.e. at various anatomic positions on the chest), and a logistic regression, which aggregates the predictions of the recordings corresponding to the eight anatomical sites. Then, it produces a single prediction for each patient. An overview of the architecture can be seen in Fig. 3 (adapted from [4] Heitmann et al.

It is the hypothesis of this work that MoDN could more flexibly and even better aggregate the sound recordings. This project thus aims to explore the efficiency of MoDN for the flexible aggregation of breath sound recordings. While out of the scope of this project, it could also be further extended to include multi-modal inputs (i.e. aggregating sound, images and tabular inputs from various combinations of models).

III. Contributions

The goal of this project is to create a single model able to flexibly integrate diverse modalities of data for the task of pneumonia diagnosis. That is, to integrate the existing DeepBreath model into the MoDN architecture. In more detail there are 2 main tasks before us:

1) We adapt the DeepBreath model to accept and encode a MoDN state vector at the level of each digital auscultation of a patient. (decoders are a binary classification for ”pneumonia or not”). This model is called *MoDN-DeepBreath* The representation of this architecture is depicted in figure Fig. 2 below.

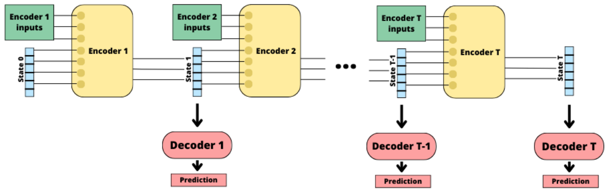

Fig. 2: Schematic architecture of MoMoNet (realization of MoDN), created by Thierry Bossy, co-supervisor of this project in iGH, EPFL

2) We compare the results with *MoDN-DeepBreath* to those obtained from the ”monolithic” baseline *standard-DeepBreath*model.

IV. Methods

A. Dataset

We have 5 datasets collected at several locations and described in Table 1 below. Patients are recruited at pediatric emergency outpatient departments in several countries (GVA: Geneva, Switzerland; POA: Porto Alegre, Brazil; DKR: Dakar, Senegal; MAR: Marrakesh, Morocco; RBA: Rabat, Morocco; YAO: Yaoundé, Cameroon). The inclusion criteria were children aged 0 to 60 months presenting for the first time with a history of acute respiratory illness.

Location	Control	Pnemonia	Other diseases	Total
GVA	23	27	70	120
YAO	13	44	22	79
POA	80	38	112	283
DKR	17	5	9	31
MAR/RBA	32	0	27	59

Table 1. Number of patients used to train the models. GVA: Geneva, Switzerland; POA: Porto Alegre, Brazil; DKR: Dakar, Senegal; MAR: Marrakesh, Morocco; RBA: Rabat, Morocco; YAO: Yaoundé, Cameroon.

As can be seen above, GVA and YAO datasets have the largest fraction of positive pneumonia diagnoses, so we decided to use them for training and testing our models. For a pneumonia classification problem, we set the target value as 1, if the disease is pneumonia, and 0 if it is the control group or other diseases. In total, we have 199 patients with 35% pneumonia. Each patient has eight 30-second recordings of breath sounds recorded with a digital stethoscope at the anatomic positions indicated in Fig. 3. 14 Patients didn't have all 8 recordings, and we removed them.
For dataset partitioning, nested 5-fold stratified cross-validation (CV) was used to create a range of performance estimates, which are then provided as a mean with standard deviation, in order to make sure that performance was independent of chance data partitioning (SD). To preserve class balance, restrictions are placed on the random fold compositions (i.e. preserving the percentage of samples for each class).After that, we conducted five experiments in which the train set is four of the partitions (80% of the data), and the test set is the fifth partition (20% of the data). We rotated through the partitions, so that each partition acts as a test set exactly once.

B. Model

1) Standard-DeepBreath: The *standard-DeepBreath* uses the adapted architecture from the PANN [5] paper codebase, which consists of 5 convolutional blocks that consist of 2 convolutional layers with a kernel size of 3 × 3. Batch normalization is used between each convolutional layer to speed up and stabilize the training. After each convolutional layer and Batch normalization, it is used ReLU nonlinearity.

For down-sampling, there is an application of average pooling of size 2 × 2 after the first 4 convolutional blocks. After the last convolutional layer, the frequency dimension is reduced with average pooling. Average and max pooling of size 3 and stride 1 are applied over the time dimension and summed to get smoothed feature maps. Then, a fully connected layer followed by a ReLU nonlinearity is applied to each of the time-domain feature vectors.

In *MoDN-DeepBreath*, we started to train the integrated model for sound event detection on cnn6 and cnn10, the difference between them is in the number of convolutional layers that consists in each block. cnn6 consists of 4 convolutional blocks with 1 convolutional layer for each with a kernel size of 5 × 5 (which is also a distinction between these two models). The cnn10 model contains 2 convolutional layers with a kernel size of 3 × 3 for every 4 convolutional blocks. It applied average pooling of size 2 × 2 for cnn6 and average pooling of size 1 × 1 for cnn10 after each convolutional block for down-sampling. Batch normalization is used between each convolutional layer for both models. After each convolutional layer and Batch normalization, it is used ReLU nonlinearity.

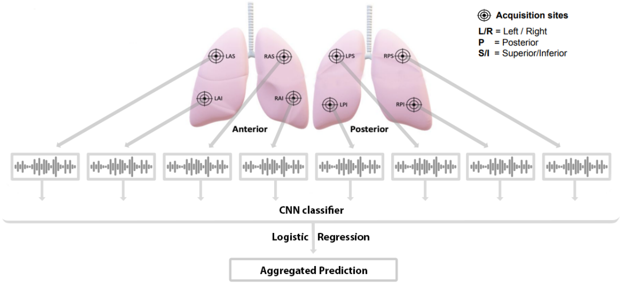

Fig. 3: *Standard-DeepBreath* architecture. Lung sounds acquired from several anatomic positions are fed into a CNN classifier then aggregated to predict respiratory disease. Adapted from Heitmann et al [4]

2) MoDN-DeepBreath: MoDN-DeepBreath consists of 8 sequential encoders that take an input of an individual recording from a single antomical position and a state. They output an updated state of the patient as depicted in Fig. 4. To create the encoder, we have taken the DeepBreath model but without the last layer, so that we get just the output of convolutional blocks. Then, we concatenate these features with a state vector and pass it to a fully connected layer to get the output of the state size. The decoder predicts the diagnosis based just on the state vector. The initial state is generated randomly and we picked a state size equal to 100 for our experiments.

Standard-DeepBreath predicted every position independently and aggregated them using a logistic regression for final prediction. But in *MoDN-DeepBreath*, every encoder gets information from the previous one and makes a prediction based on it and the new position. Compare Fig. 3 above with Fig. 4 below.

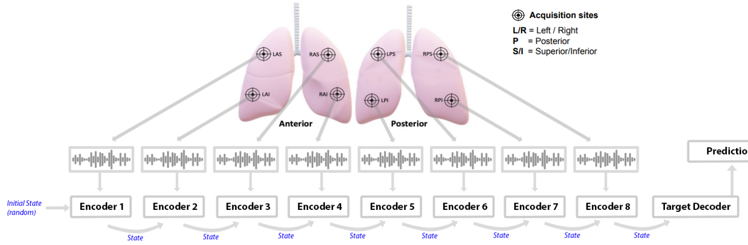

Fig. 4: *MoDN-DeepBreath* architecture. Instead of a logistic regression aggregation, all recordings are aggregated using the MoDN state.

C. Training

To be able to compare results, we trained and tested our models on the same datasets, the same testing partition, and hyperparameters. We trained our models on 15 epochs.

V. Results

The main results are presented in the table I for cnn6 architecture and in the table II for cnn10 architecture. The results separately by folds are presented for the cnn6 version of DeepBreath in the Fig. 5 and cnn10 version in the Fig. 6.

	Cnn6		
	AUC	Sensitivity	Specificity
DeepBreath	**0.472** ±0.058	0.314 ±0.352	0.633 ±0.386
DeepBreath into MoDN	**0.568** ±0.025	0.366 ±0.257	0.825 ±0.197

TABLE I: Comparison of results of DeepBreath into MoDN vs DeepBreath for cnn6 architecture for metrics AUC, Sensitivity and Specificity

	Cnn10		
	AUC	Sensitivity	Specificity
DeepBreath	**0.530** ±0.075	0.340 ±0.200	0.673 ±0.192
DeepBreath into MoDN	**0.591** ±0.023	0.187 ±0.158	0.952 ±0.073

TABLE II: Comparison of results of DeepBreath into MoDN vs DeepBreath for cnn10 architecture for metrics AUC, Sensitivity and Specificity

Since we are predicting a disease (pneumonia), sensitivity is more significant for us than specificity, since we want to minimize the number of negative examples that are incorrectly classified. But in our case, the plots are presented for a particular $threshold = 0.5$ of the binary classification, and therefore a more revealing metric is AUC, which is independent of the classification threshold (unlike sensitivity and specificity).

Overall, the results of the prediction are very low, averaging near random. There are many reasons for this low performance, the most important of which is the difficulty of the task vs the size of the data set and the poor consensus of what constitutes a diagnosis of ”pneumonia” in the two countries selected. In the future an easier task on more homogenous labels will be selected. Regardless, it is not the aim of this project to have a highly performant model, but rather

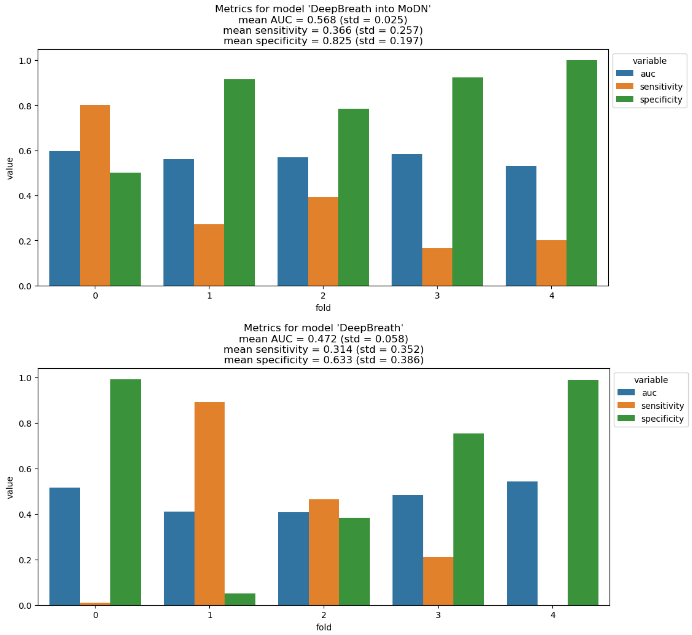

Fig. 5: Comparison of results of DeepBreath into MoDN vs DeepBreath for cnn6 architecture

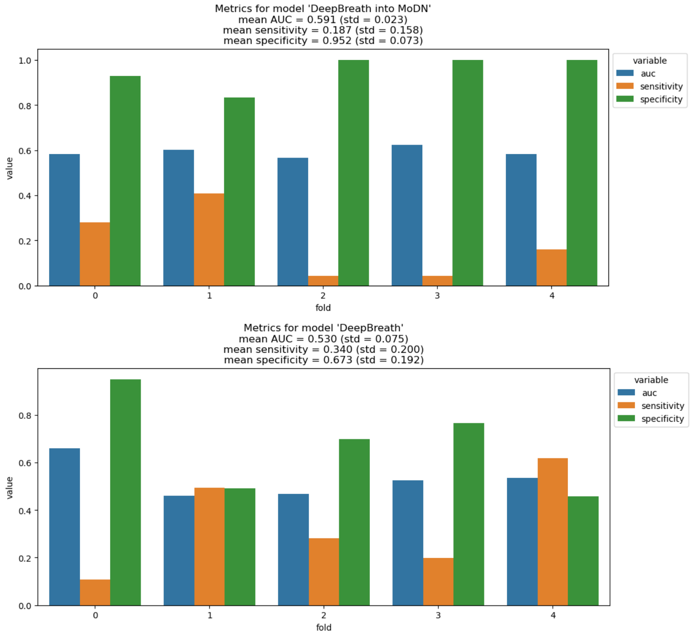

Fig. 6: Comparison of results of DeepBreath into MoDN vs DeepBreath for cnn10 architecture

to compare the performance between standard and MoDN implementations. DeepBreath in MoDN in both architectures (cnn6 and cnn10) shows a significant improvement in mean AUC and a reduction in variance. Thus, using the MoDN architecture in this task is reasonable.

A. *Future work*

The fact that we trained our model on a little number of epochs and with randomly picked state size means that we don't have the whole picture of the potential of this approach. It would be useful to look for optimal state size, train for more epochs, and test our results on external datasets. Training on several disease labels would also allow a better evaluation of the approach. One of the most important next steps is to mix these modules for audio encoding with encoders for tabular data (e.g. age, sex, symptoms), thus creating a multimodal modular network.

VI. Conclusion

In conclusion, this project integrated the DeepBreath model into MoDN and applied it to the pneumonia classification task. Compared to a standard implementation of DeepBreath, we obtained promising results that show that MoDN is significantly superior within the limited epochs for which it was trained.

VII. Acknowledgements

We would like to express our gratitude to Mary-Anne Hartley for providing us with the opportunity to work on an interesting research, as well as to our supervisor Thierry Bossy for his valuable support.

References

[1] Osheroff. *Improving Outcomes with Clinical Decision Support: An Implementer's Guide*. HIMSS Publishing, 2012.

[2] German Prado-Arechiga Martha Pulido, Patricia Melin. Blood pressure classification using the method of the modular neural networks. 2018.

[3] Kristina Keitel-Alexandra V Kulunkina Rainer Tan Ludovico Cobuccio Martin Jaggi Mary-Anne Hartley Cécile Trottet, Thijs Vogels. Modular clinical decision support networks (modn)—updatable, interpretable, and portable predictions for evolving clinical environments. 2022.

[4] Julien Heitmann, Alban Glangetas, Jonathan Doenz, Juliane Dervaux, Deeksha M Shama1, Daniel Hinjos Garcia, Mohamed Rida Benissa, Aymeric Cantais, Alexandre Perez, Daniel Müller, Tatjana Chavdarova1, Johan N. Siebert, Laurence Lacroix, Martin Jaggi, Alain Gervaix, and Mary-Anne Hartley. Deepbreath—automated detection of respiratory pathology from lung auscultation in 572 pediatric outpatients across 5 countries. 2022.

[5] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D. Plumbley. Panns: Large-scale pretrained audio neural networks for audio pattern recognition, 2019.