Uncertainty-Aware Boosted Ensembling in Multi-Modal Settings

Utkarsh Sarawgi* Rishab Khincha*, Wazeer Zulfikar*, Satrajit Ghosh, Pattie Maes
Massachusetts Institute of Technology, BITS Pilani Goa
{utkarshs, rkhincha, wazeer, satra, pattie} @mit.edu

International Joint Conference on Neural Network (IJCNN 2021)

BITS Pilani
K K Birla Goa Campus

Massachusetts Institute of Technology

Summary of Contributions

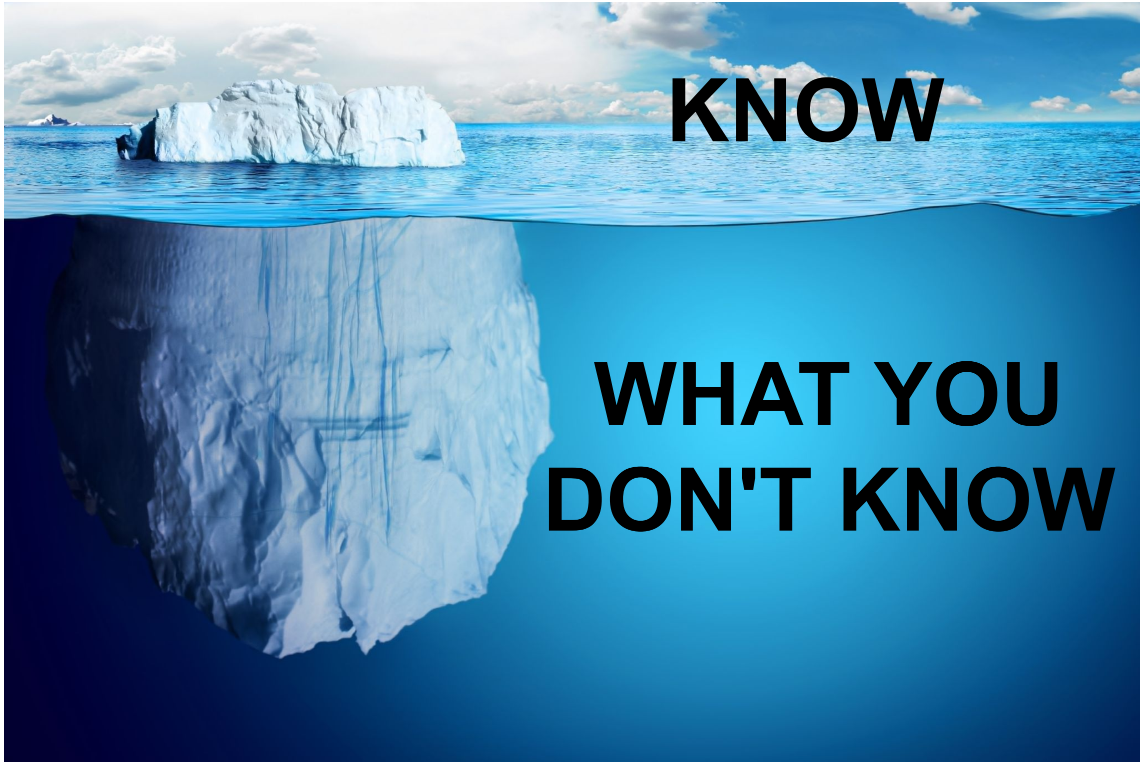
- We propose and formulate an ‘uncertainty-aware ensemble’
- We evaluate our method on multi-modal speech and text datasets on healthcare tasks using different ML models and uncertainty estimation techniques
- We perform further analyses to highlight the significance of introducing uncertainty-awareness into the ensemble

ML Models when training

ML Models in deployment

Introduction

- Reliability crucial in safety-critical applications
- Confidently incorrect predictions
- Poor performance during deployment due to distribution shifts


KNOW
WHAT YOU
DON'T KNOW

Uncertainty Estimation

- Predict a distribution rather than a single value
- Aleotoric - Uncertainty in the data
- Epistemic - Uncertainty in the model
- Distribution shifts quite common

Note: Uncertainty = Aleotoric uncertainty (for this presentation)

Ensembling Techniques

- Combining decisions from multiple models

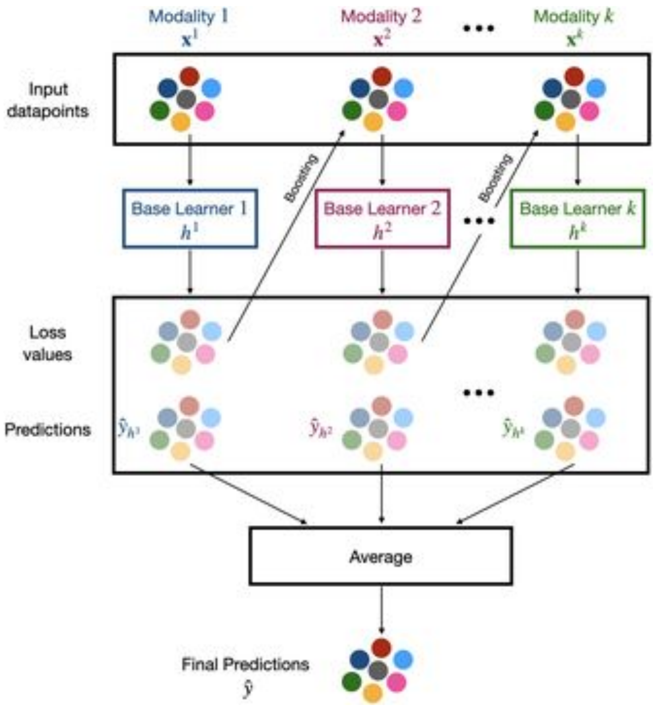
- Bagging: Parallely training with different training sets

- Boosting: Sequentially training by iteratively re-weighting training examples

Multi-modal Ensembling - Setup

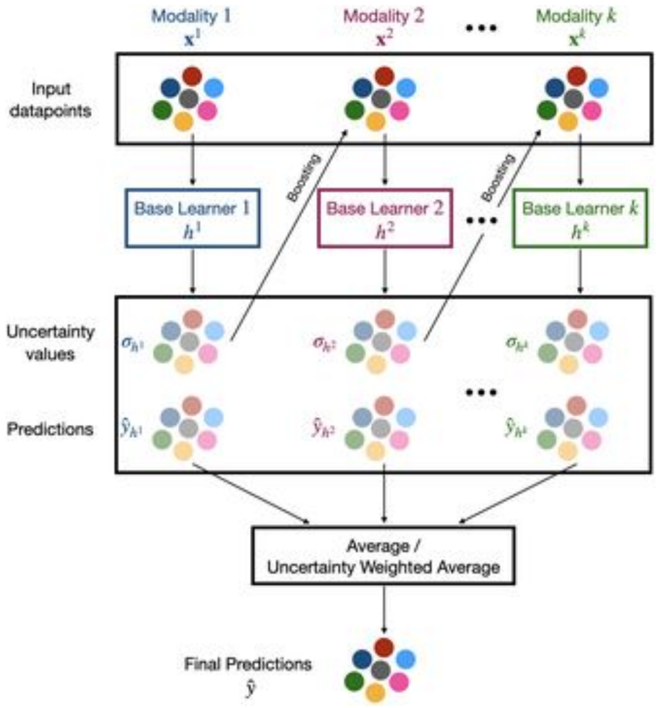
- Given multi-modal data $x^1, x^2, x^3 \ldots x^k$ with k modalities
- Given base learners $h^1, h^2, h^3 \ldots h^k$ for each modality
- Final prediction y which is a function of $y_{h1}, y_{h2}, y_{h3} \ldots y_{hk}$

Vanilla Ensembling



Boosting done using loss values!

Uncertainty-Aware Ensembles



Boosting done using uncertainty estimates!

UA Ensemble Predictions

- UA Ensemble: $y = \text{average}(y_{h1}, y_{h2}, y_{h3} \ldots y_{hk})$

- UA Ensemble weighted: Weigh each of the prediction with the inverse of the predictive uncertainty for the particular modality

$$\hat{y}(\mathbf{x}_n) = \frac{\sum_{j=1}^{k} \frac{1}{\sigma_{hj}(\mathbf{x}_n)} \hat{y}_{hj}(\mathbf{x}_n)}{\sum_{j=1}^{k} \frac{1}{\sigma_{hj}(\mathbf{x}_n)}}$$

UA Ensembles - Note

- Sequentially boost across base learners, each of the corresponding to a different input modality

- Base learners need not be weak learners!

DementiaBank Pitt

- Speech recordings and transcripts
- 242 samples from 99 control healthy subjects and 255 samples from 168 AD subjects
- MMSE scores, ranging from 0 to 30

BITS Pilani
K K Birla Goa Campus

Massachusetts
Institute of
Technology

Dementia - Feature Sets

- Disfluency: Word, intervention, and different kinds of pause rates reflecting upon impediments like slurring and stuttering
- Acoustic: ComParE 2013 acoustic feature set (6,373 features) normalized and with dimensionality reduction using PCA
- Interventions: Sequence of speakers from the transcripts categorizing it as subject or the interviewer

Multimodal Inductive Transfer Learning for Detection of Alzheimer's Dementia and its Severity. *Sarawgi et. al.* https://arxiv.org/abs/2009.00700

Parkinson’s Telemonitoring

- Biomedical voice measurements
- 5875 samples collected from 42 subjects with early stage PD
- UPDRS scores, ranging from 0 to 199.

Parkinson's - Feature Sets

- Amplitude: Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA, NHR, HNR, RPDE, DFA

- Frequency: Jitter(%), Jitter(Abs), Jitter:RAP, Jitter:PPQ5, Jitter:DDP, PPE

DementiaBank - Results

TABLE I
COMPARISON OF INDIVIDUAL MODALITIES I.E. BASE LEARNERS AND ENSEMBLE METHODS ON TEST SET RESULTS OF THE ADRESS DATASET.

Model	RMSE
Disfluency	5.71 ± 0.39
Interventions	6.41 ± 0.53
Acoustic	6.66 ± 0.30
Vanilla Ensemble	5.17 ± 0.27
UA Ensemble	$\mathbf{5.05 \pm 0.53}$
UA Ensemble (weighted)	$\mathbf{4.96 \pm 0.49}$

TABLE II
COMPARISON OF UNCERTAINTY-AWARE ENSEMBLE METHODS WITH STATE-OF-THE-ART RESULTS ON THE ADRESS TEST SET.

Model	RMSE
Pappagari et al. [55]	5.37
Luz et al. [50]	5.20
Sarawgi et al. [15]	4.60
Searle et al. [56]	4.58
Balagopalan et al. [57]	4.56
Rohanian et al. [58]	4.54
Sarawgi et al. [17]	4.37
UA Ensemble	**4.35**
UA Ensemble (weighted)	**3.93**

Multimodal Inductive Transfer Learning for Detection of Alzheimer's Dementia and its Severity. *Sarawgi et. al.* https://arxiv.org/abs/2009.00700
Simple and scalable predictive uncertainty estimation using deep ensembles. *Lakshminarayanan et. al.* https://arxiv.org/abs/1612.01474

DementiaBank - Results

TABLE III

COMPARISON OF INDIVIDUAL MODALITIES I.E. BASE LEARNERS AND ENSEMBLE METHODS ON TEST SET RESULTS OF THE ADReSS DATASET.

Model	Modality	MPIW	PICP (%)		
			$\Delta = 1\sigma$	$\Delta = 2\sigma$	$\Delta = 3\sigma$
Vanilla Ensemble	Disfluency	4.47 ± 0.39	61.66 ± 8.29	95.83 ± 2.63	97.50 ± 0.83
	Interventions	7.27 ± 0.58	$\mathbf{87.50 \pm 5.43}$	$\mathbf{99.17 \pm 1.02}$	$\mathbf{100.00 \pm 1.18}$
	Acoustic	$\mathbf{4.50 \pm 0.73}$	59.58 ± 12.54	94.58 ± 2.12	98.75 ± 1.02
UA Ensemble	Disfluency	6.29 ± 0.81	82.91 ± 6.37	97.91 ± 1.31	100.00 ± 0.00
	Interventions	$\mathbf{5.46 \pm 1.57}$	73.75 ± 14.47	93.33 ± 5.17	97.91 ± 1.86
	Acoustic	5.31 ± 1.30	$\mathbf{75.41 \pm 11.21}$	$\mathbf{96.25 \pm 3.06}$	$\mathbf{99.16 \pm 1.02}$
UA Ensemble (weighted)	Disfluency	6.29 ± 0.81	83.33 ± 6.58	97.91 ± 1.31	100.00 ± 0.00
	Interventions	$\mathbf{5.46 \pm 1.57}$	76.25 ± 13.85	92.50 ± 5.98	96.66 ± 3.11
	Acoustic	5.31 ± 1.30	$\mathbf{75.83 \pm 10.59}$	$\mathbf{95.00 \pm 3.86}$	$\mathbf{99.16 \pm 1.02}$

DementiaBank - Calibration

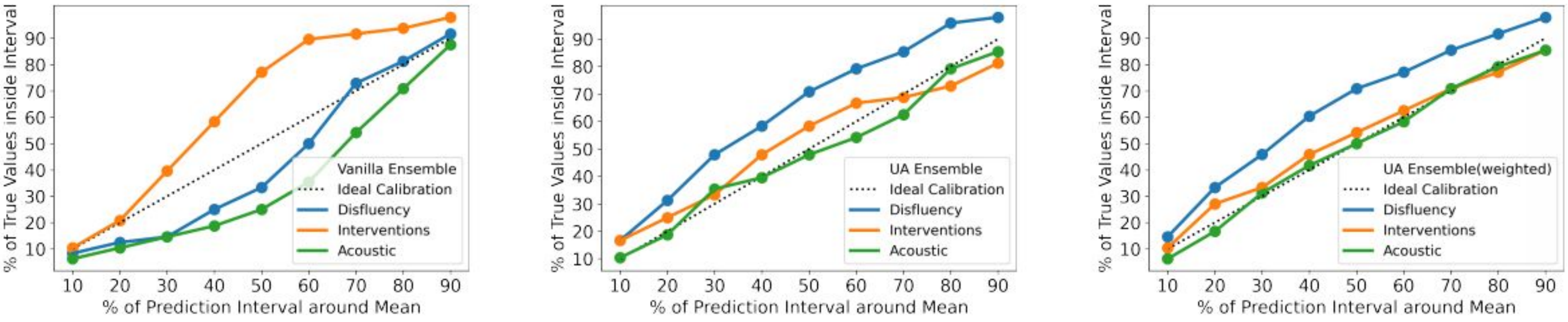


Fig. 1. Calibration curves for the ensemble techniques on the ADReSS dataset.

DementiaBank - Entropy Plots

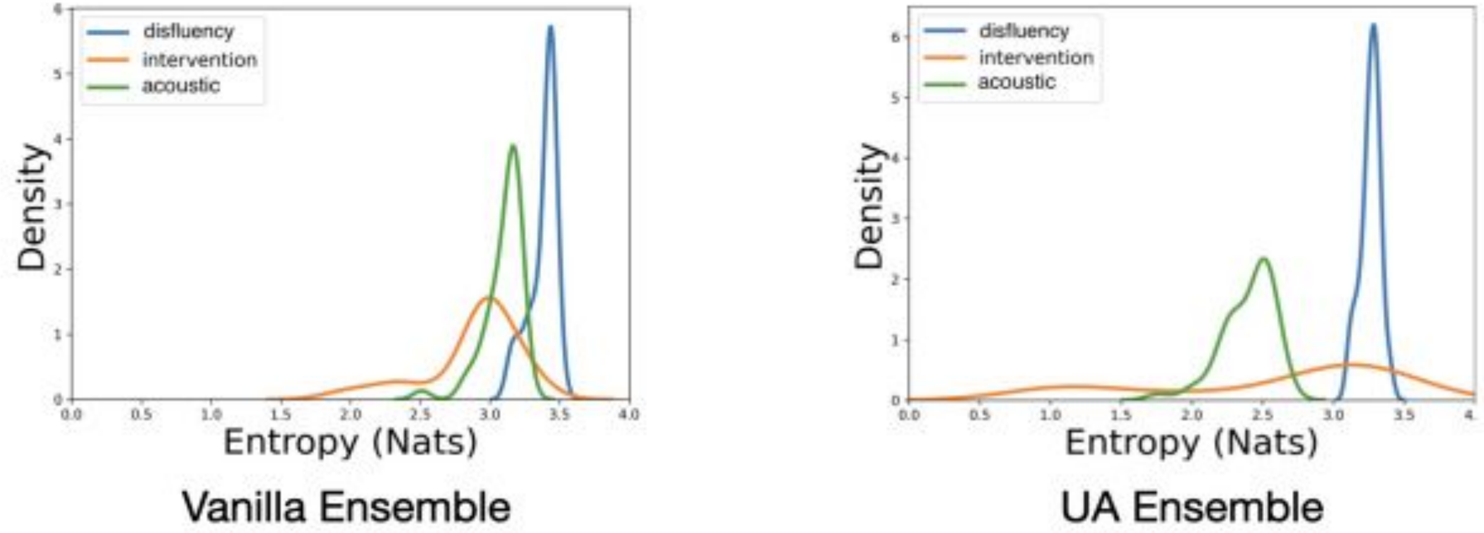


Fig. 2. Entropy analysis, using kernel density estimation plots, of the base learners in a vanilla ensemble (left) and UA ensemble (right).

Parkinson's - Results

TABLE IV
COMPARISON OF INDIVIDUAL MODALITIES I.E. BASE LEARNERS AND ENSEMBLE METHODS ON 5-FOLD CROSS VALIDATION RESULTS OF THE PARKINSON'S TELEMONITORING DATASET.

Model	**RMSE**
Amplitude	3.21 ± 0.06
Frequency	3.32 ± 0.10
Vanilla Ensemble	3.18 ± 0.05
UA Ensemble	$\mathbf{3.04 \pm 0.04}$
UA Ensemble (weighted)	$\mathbf{3.05 \pm 0.05}$

Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife. *Wager et. al.* https://jmlr.org/papers/v15/wager14a.html

Parkinson's - Results

TABLE V

COMPARISON OF INDIVIDUAL MODALITIES I.E. BASE LEARNERS AND ENSEMBLE METHODS ON 5-FOLD CROSS VALIDATION RESULTS OF THE PARKINSON'S TELEMONITORING DATASET.

Model	Modality	MPIW	PICP (%)		
			$\Delta = 1\sigma$	$\Delta = 2\sigma$	$\Delta = 3\sigma$
Vanilla Ensemble	Amplitude	6.79 ± 1.28	84.56 ± 1.46	98.51 ± 0.58	99.89 ± 0.12
	Frequency	8.69 ± 0.59	74.17 ± 8.25	94.28 ± 3.37	98.60 ± 1.18
UA Ensemble	Amplitude	6.50 ± 1.76	74.09 ± 9.15	93.70 ± 4.11	98.23 ± 1.47
	Frequency	$\mathbf{6.91 \pm 0.85}$	$\mathbf{77.90 \pm 5.28}$	$\mathbf{95.64 \pm 2.40}$	$\mathbf{99.33 \pm 0.51}$
UA Ensemble (weighted)	Amplitude	6.50 ± 1.76	74.24 ± 8.59	93.71 ± 4.13	97.97 ± 1.66
	Frequency	$\mathbf{6.91 \pm 0.85}$	$\mathbf{77.65 \pm 5.67}$	$\mathbf{95.45 \pm 2.56}$	$\mathbf{99.18 \pm 0.70}$

Parkinson's - Calibration Curves

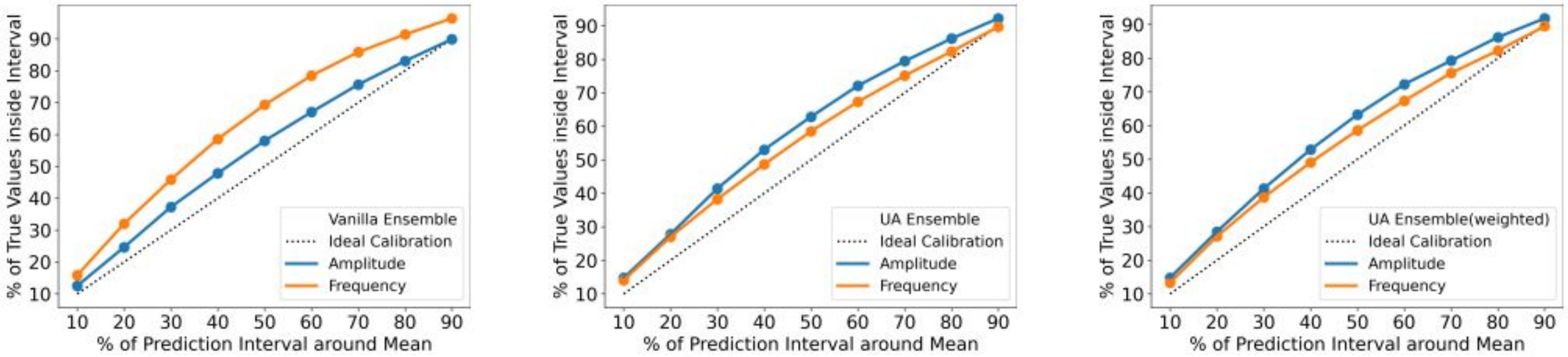


Fig. 3. Calibration curves for the ensemble techniques on the Parkinson's Telemonitoring dataset.

BITS Pilani
K K Birla Goa Campus

Massachusetts Institute of Technology

Discussion

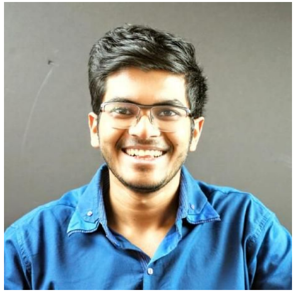
- Outperform state-of-the-art methods
- Reduce the overall entropy of the system
- Well calibrated predictions with high quality prediction intervals

Future Work

- Account for uncertainty as well as loss values when boosting
- Experiment with other ML models and architectures
- Experiment with other uncertainty estimation methods
- Actively learn from the uncertainty estimates at deployment time

Meet the team!

Rishab
Khincha



Utkarsh
Sarawgi

Wazeer
Zulfikar

Satrajit
Ghosh

Pattie
Maes

BITS Pilani
K K Birla Goa Campus

Massachusetts
Institute of
Technology

Questions?

Paper

Code