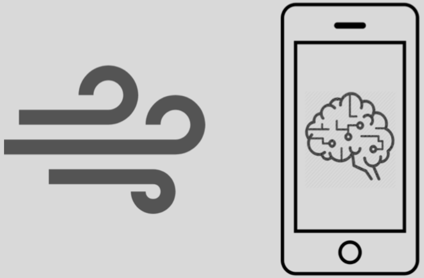
Spiro AI



Pulmonary Function Testing with a smartphone

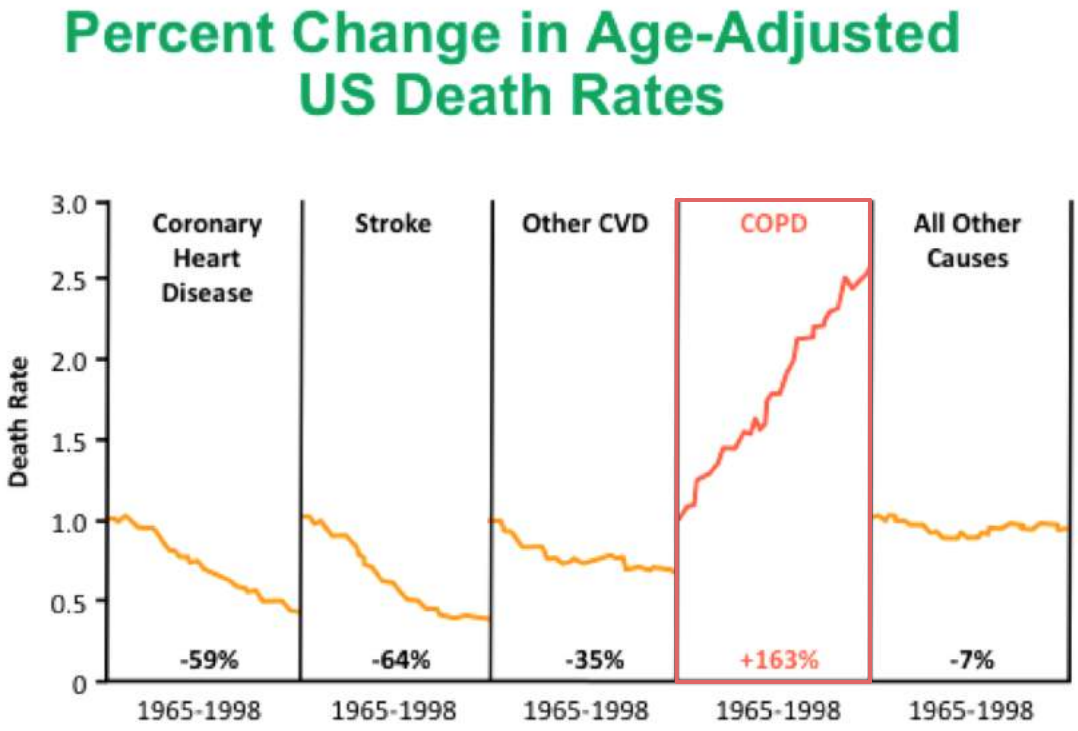
Jake Garrison

Motivation: Improve Spirometry

Spirometry is used to...

- **Diagnose:** asthma, COPD and other conditions that affect breathing

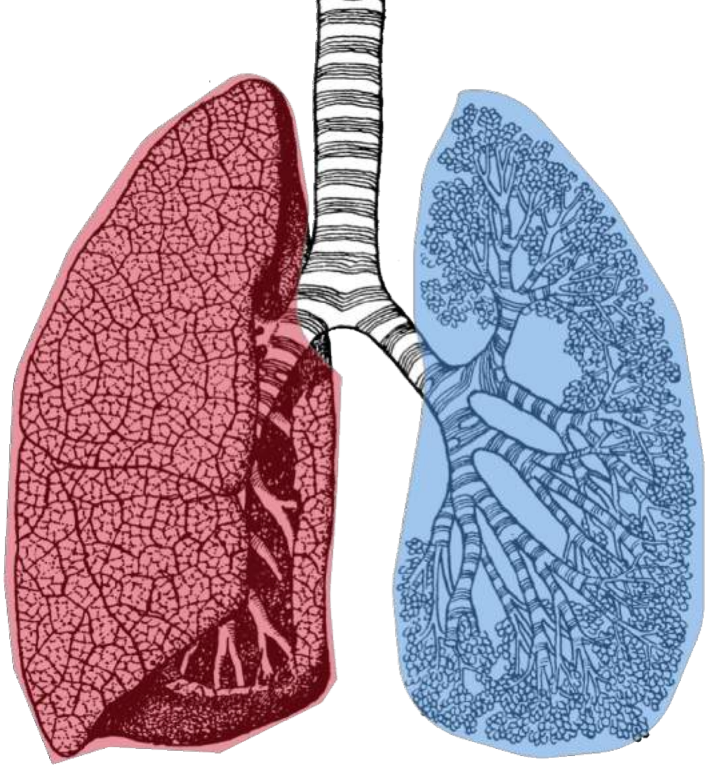
- **Monitor:** lung health and treatment



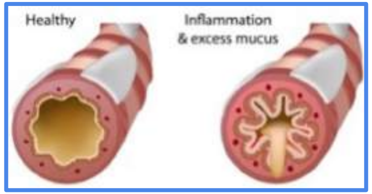
CVD = cardiovascular disease
GOLD. Available at: http://www.goldcopd.org/OtherResourcesItem.asp?l1=2&l2=2&intId=969.
Accessed May 18, 2009.

Lung Disorders

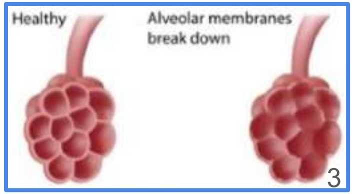
Problem	**Cause**
Restriction ***(restricts input)***	***Intrinsic:*** ● Pulmonary fibrosis ● Other scarring (interstitial) ***Extrinsic:*** ● Obesity ● Scoliosis ● Neuromuscular
Obstruction ***(obstructs output)***	● **COPD** ○ Chronic Bronchitis ○ Emphysema ● Asthma ● Bronchiectasis ● Cystic fibrosis*



Bronchitis / Asthma



Emphysema



Spirometry

[breathe meter]

History

portability

time

1900s

2010s

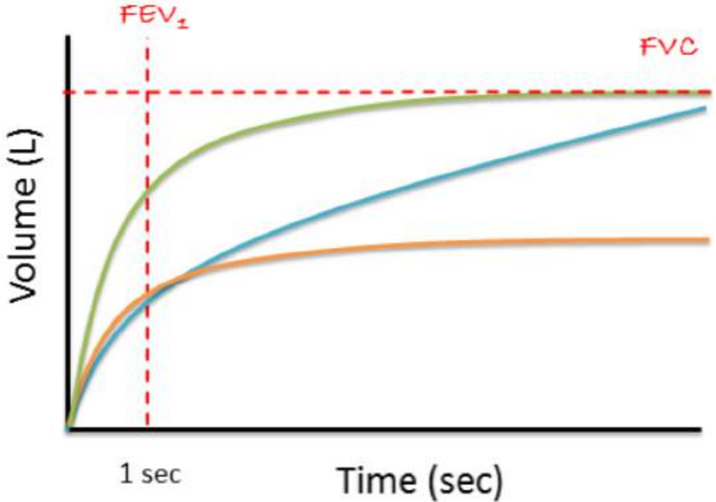
Metrics

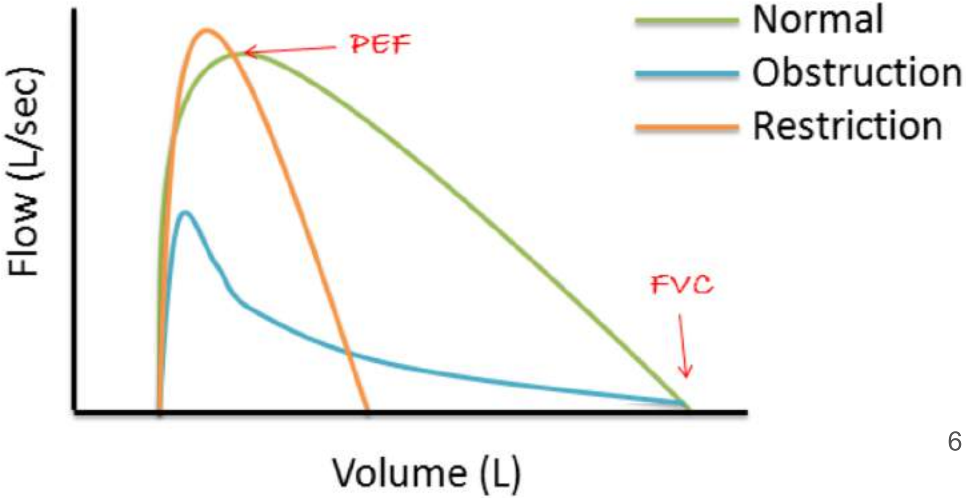
Metric	Units	Description
FEV1	Liters	Exhaled volume in first second
PEF	Liters/second	Peak flow
FVC	Liters	Total exhaled lung volume
FET	Seconds	Total exhale duration

Predicted Normal

Depends on:

- Age
- Height / Weight
- Gender / Ethnicity

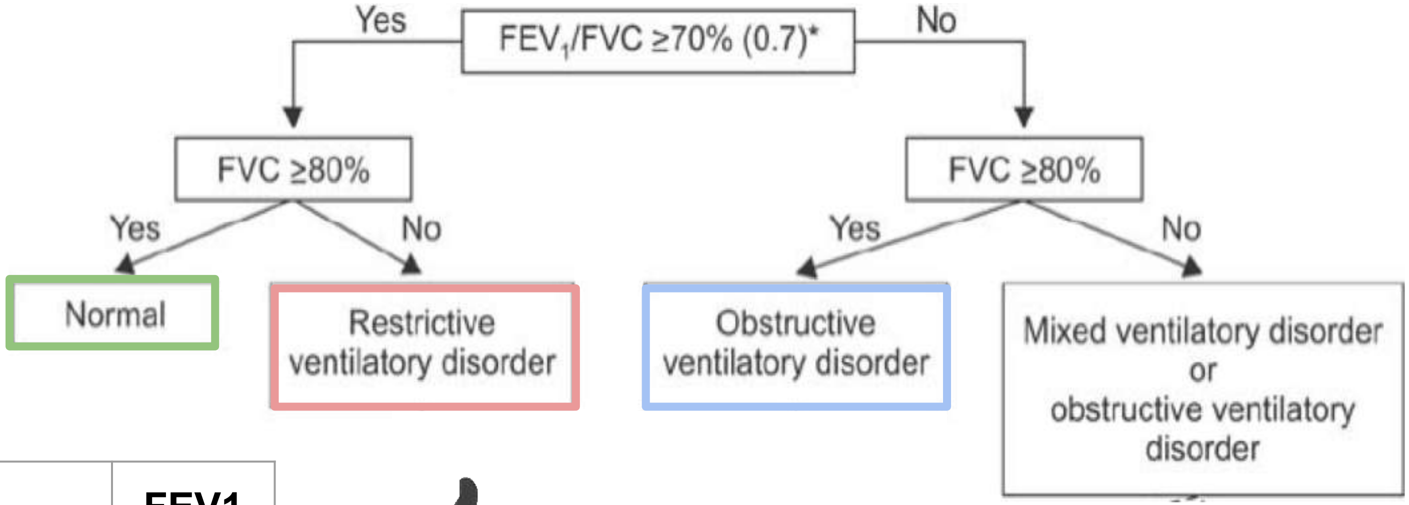




Interpretation

Given: *Spirometry results and patient information...*

Diagnosis Decision Tree



Condition	**FEV1**	**FVC**	**FEV1 / FVC**
Obstructive	⇩⇩	⇩	⇩
Restrictive	⇩	⇩⇩	⇧

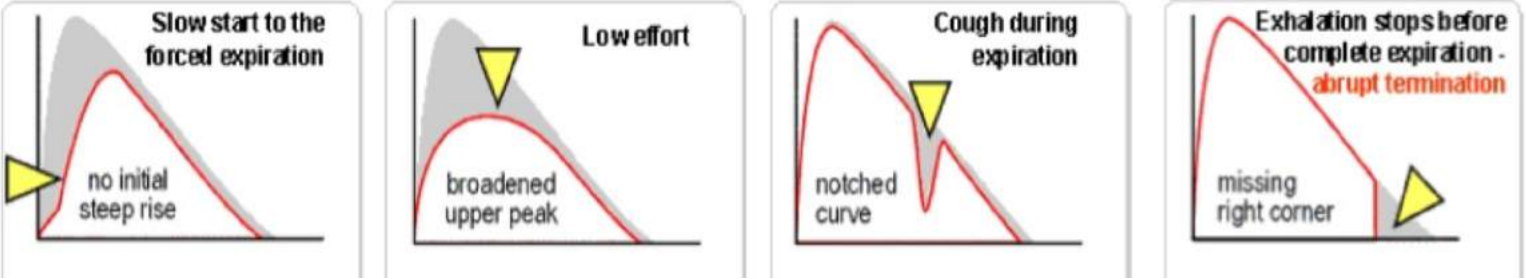
Healthy if…

- FEV1/FVC > 70%
- FEV1 > 80% predicted
- FVC > 80% predicted

Problem: Spirometry is great *but*...

- Inconvenient
- Hard to use
- Expensive
- Outdated
- Not used frequently enough
- Error prone without training

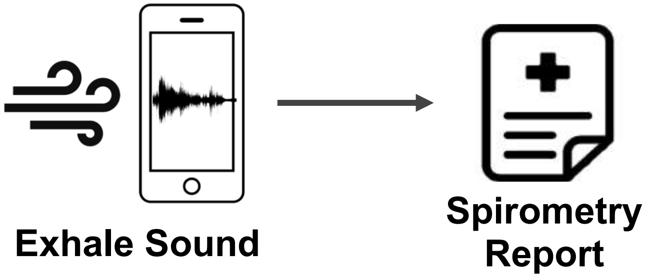
Example Errors:



Solution: Sound-Based Smartphone Spirometry

The good

- **Portable**, fits in pocket
- **Frequent** measurements
- **Easy** to use, no accessories
- Coaching feedback, error detection
- Powerful CPU, internet, easy to develop



The bad

- Sound is a **weak proxy** for airflow
- Requires **data**, complex modeling
- Background **noise**
- Lots of smartphones to support

Thesis Statement

Explore data driven methods for computing spirometry metrics from a smartphone recording of a spirometry effort

Contributions

- Spiro AI system
- CurveNet neural network
- Organized spirometry sound dataset
- Training and evaluation pipelines
- iOS app and backend

Vision

Research → Reality

*Measuring lung function is as **convenient** and **simple** as taking a selfie*

As a result people...

- Developing lung conditions **find out sooner**
- With lung conditions have a **better record** of their respiratory health
- In more of the world can **benefit** from spirometry

~ improved overall quality of life ~

Prior Mobile Spirometry Work

2011 — **SpiroSmart:** Measure lung function via smartphone microphone

mobileSpiro: Mobile Spirometry for asthma management

2013 — **mCOPD:** COPD sound based diagnosis and exercise system

SpiroCall: Measure lung function via phone call

2017 — **Automatic Characterization of Errors** via spirometry curves

SpiroConfidence: Error detection from sound of effort

Introducing: **Spiro AI**

Input: Sound of a forced expiratory effort

Output: Spirometry report with including curves

- *Quality Control*: Automatic improper effort rejection
- *Reproducibility metrics*: Automatic for multiple trials
- *Data Driven*:
  - Trained on over **5000 patients**
  - Evaluated on over **700 patients**
- Based on **SpiroSmart** concept and dataset

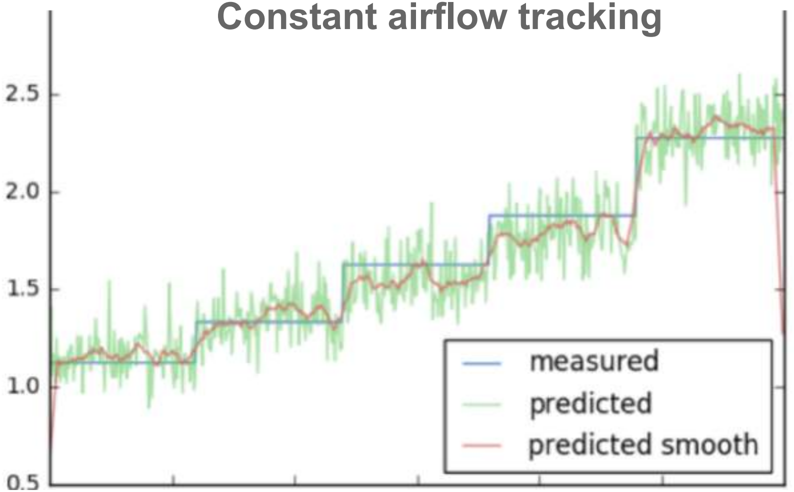
Physical Modeling

Two difficult tasks:

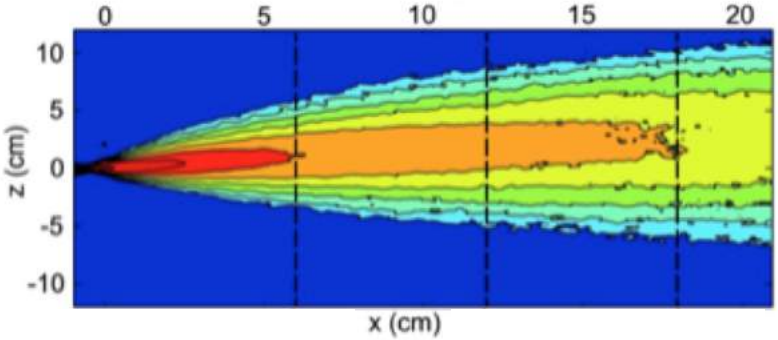
1. Measure airflow rate at position of phone microphone via sound
2. Transform to flow rate at mouth

Non-idealities*: human, phone, environment*

Constant airflow tracking



Airflow



Airflow decay models

Ideal jet

$$u_{max} = U\frac{5d}{x}$$

Fan Blade

$$u_{max\ fan} = U\frac{ld}{\sqrt{x}}$$

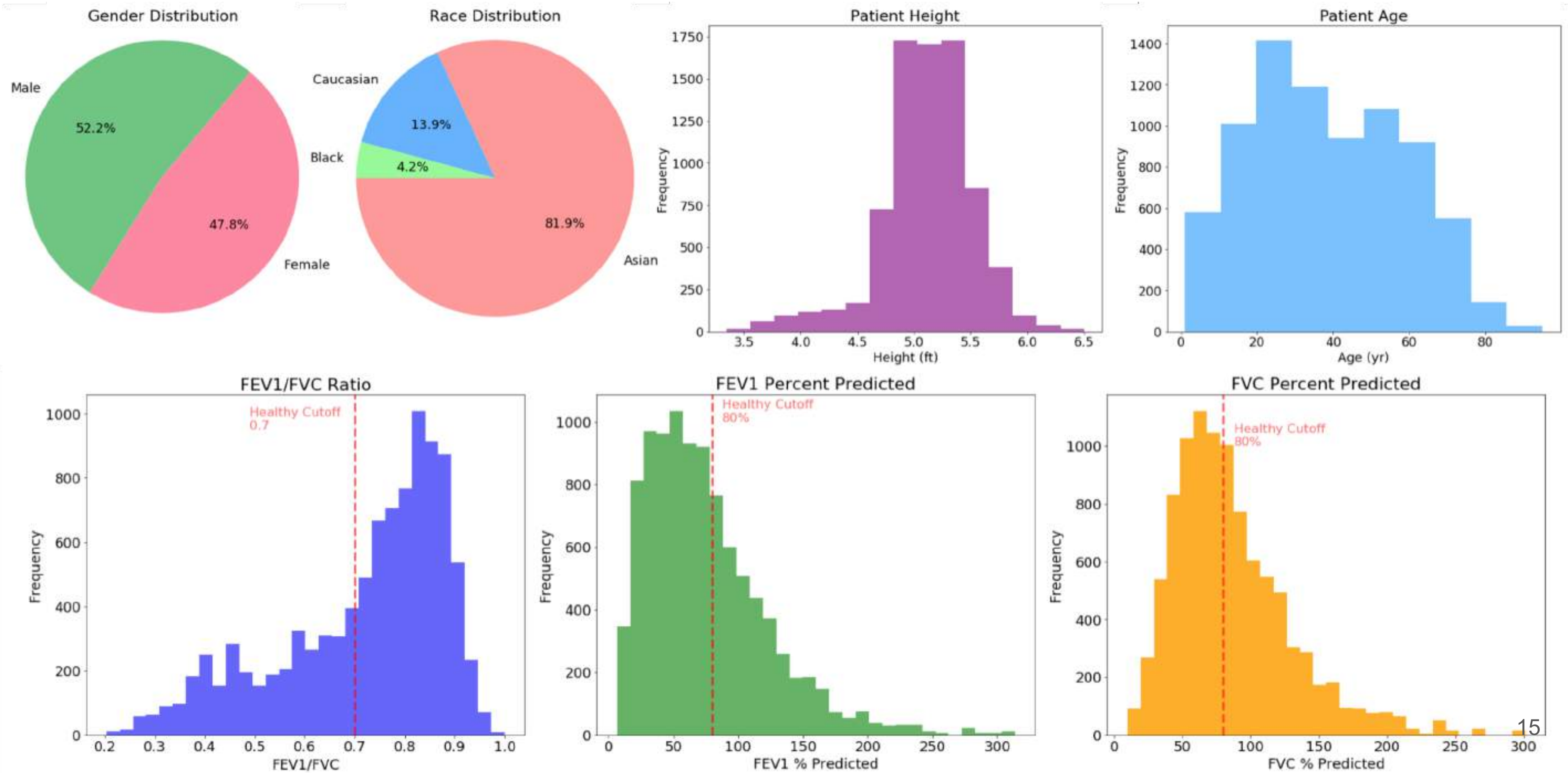
Human Exhale

??

U: velocity at x=0 *d: tube diameter* *l: fan blade length*

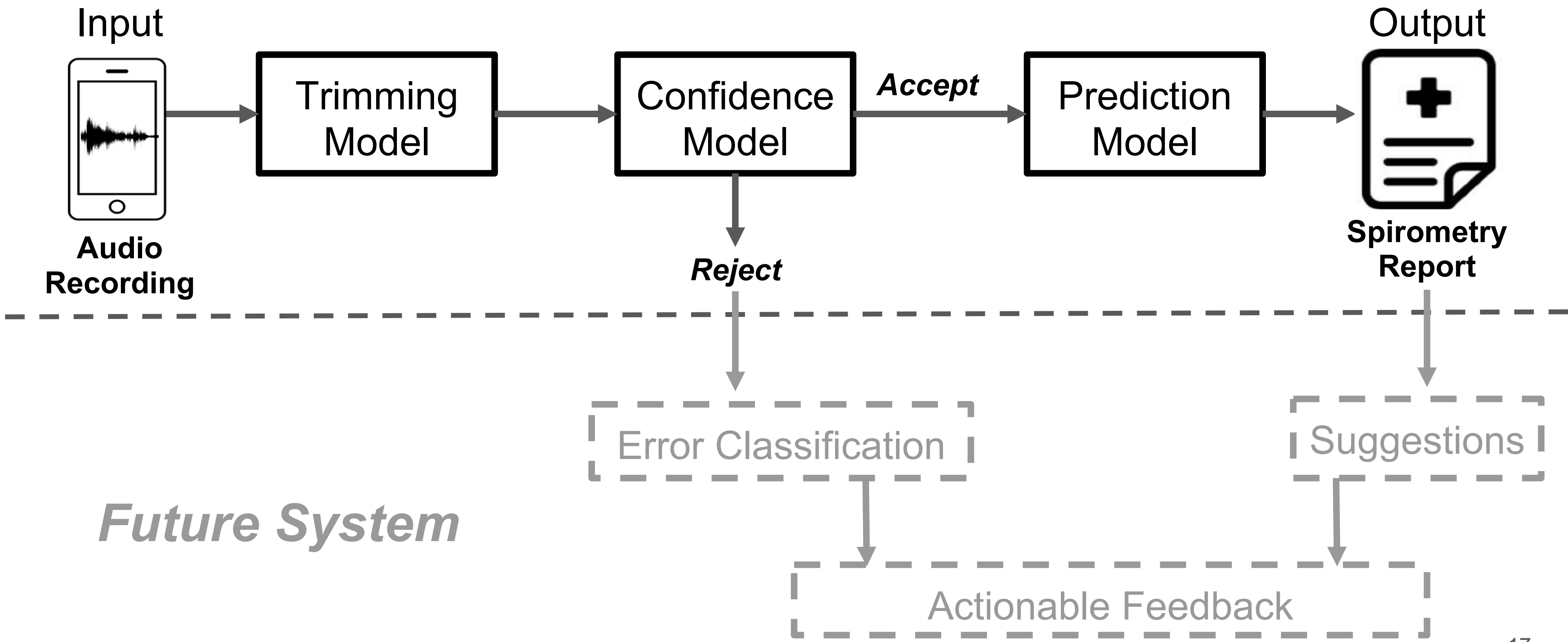
SpiroSmart Dataset

~5000 total patients



System

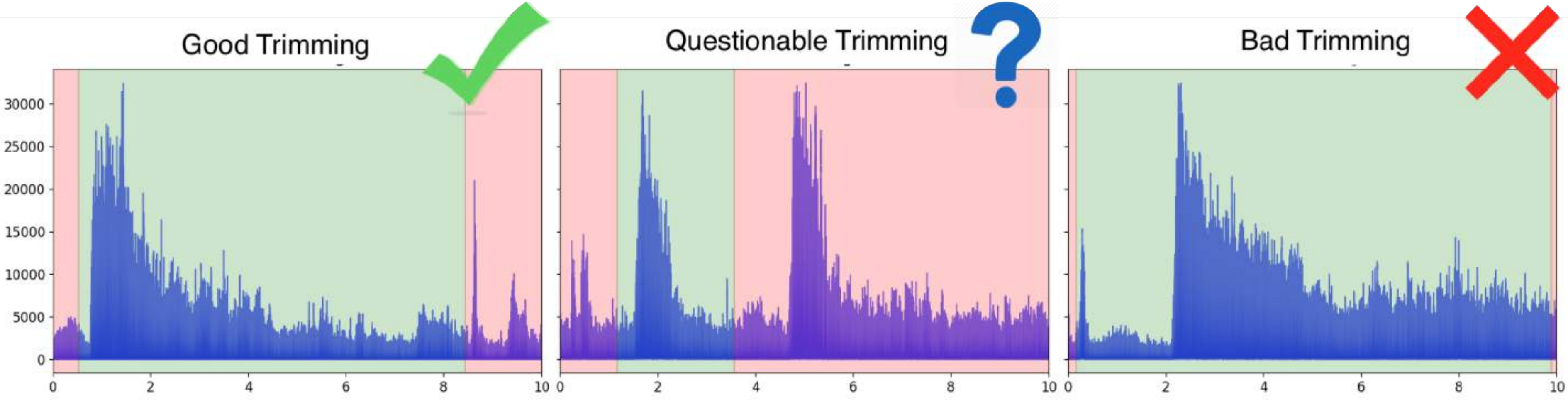
Spiro AI System



Trimming Algorithm

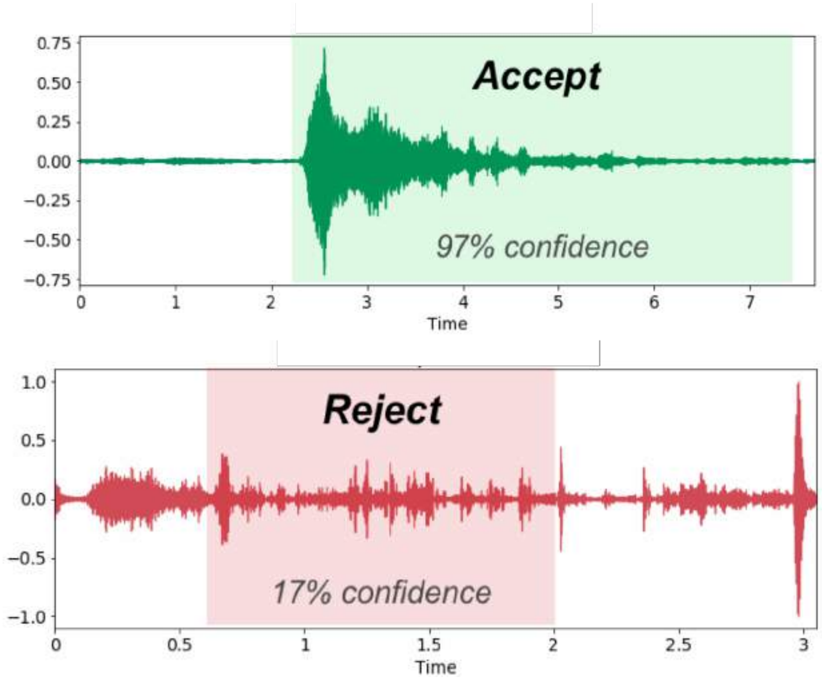
- Decode audio as 16kHz mono integer array
- Find room **noise**: *quiet* region near beginning or end
- Find **start**: 1st *impulse* relative to room noise
- Find **end**: 2nd *impulse* relative to room noise or end of array

Passing > 95% of the time *(based on 2000 or 10% random sampled)*



**segmented spirometry audio, plotted as waveform absolute value*

Confidence Model: Binary



Goal: Quality control

Reject low quality audio efforts such as:

- Coughing, speech, background noise
- Too far away, too close, incomplete

Dataset:

- Binary labeled audio examples from SpiroSmart set and other sources*
- 40k = 20k *accepted* + 5k *rejected* + 15k external *negatives**

Output: Confidence percentage (*accept* >70% confidence)

Prediction Model: Regression

Goal: Generate spirometry report for accepted audio effort

Dataset:

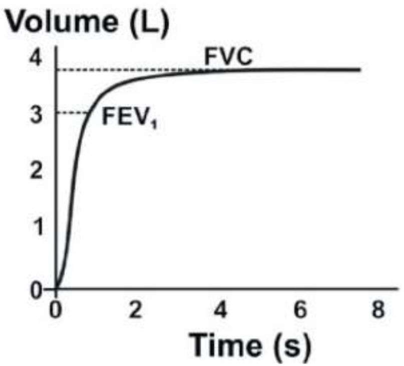
- Exhale audio with paired groundtruth (FEV1, FVC, curves ect)
- 14k *accepted* audio entries from ~4k patients with r*eproducible* groundtruth

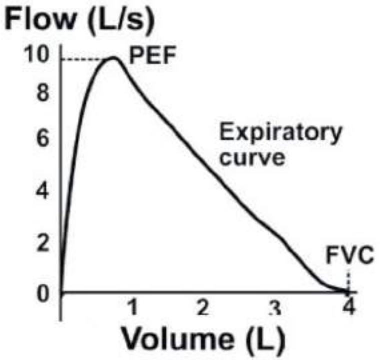
Output: Spirometry test report...

Outputs

FEV1	*FET (duration)*
FVC	*Percent Predicted*
PEF	*Confidence Percent*

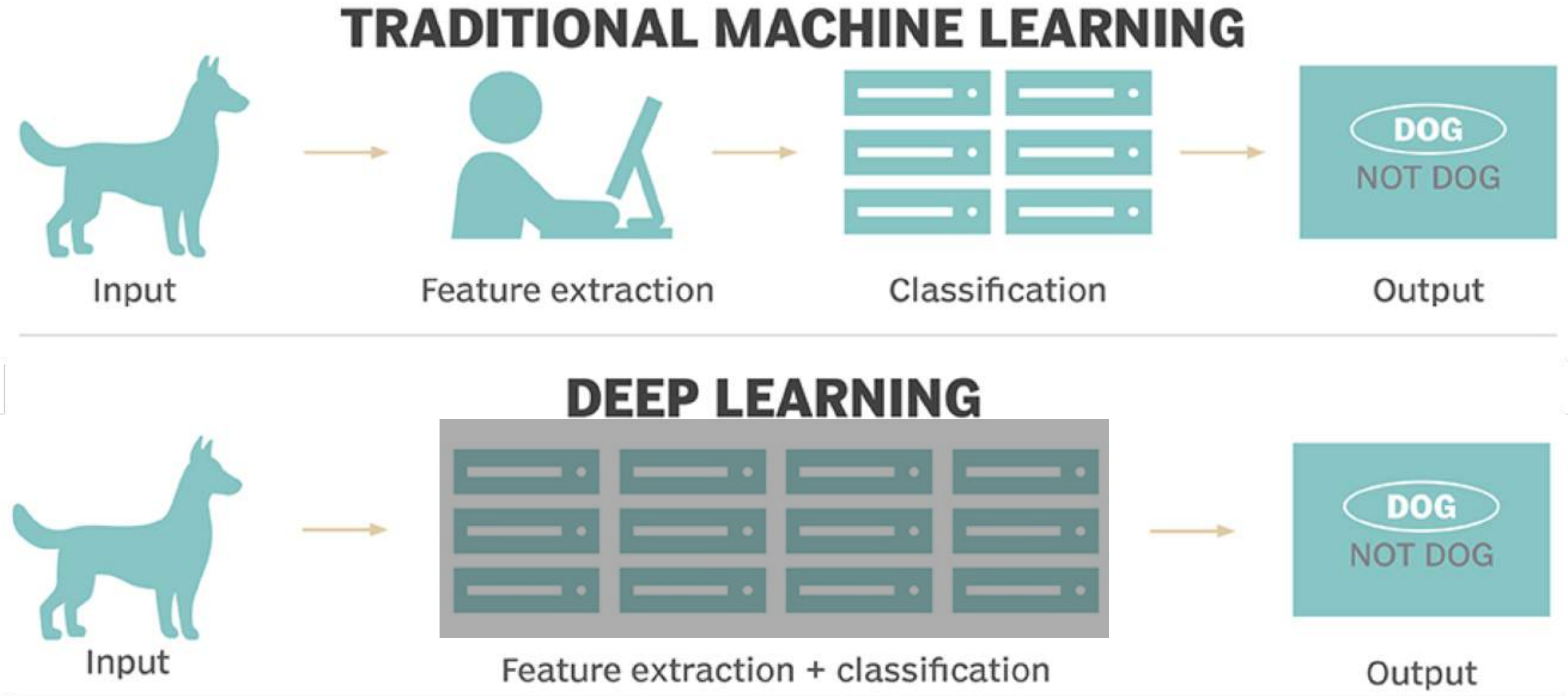
Bonus Output: Curves





Methods

Glass Box

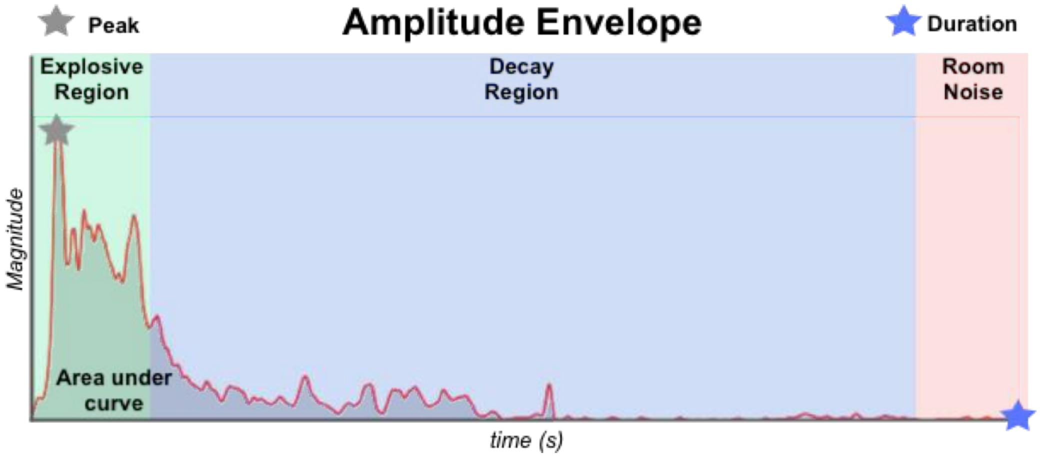


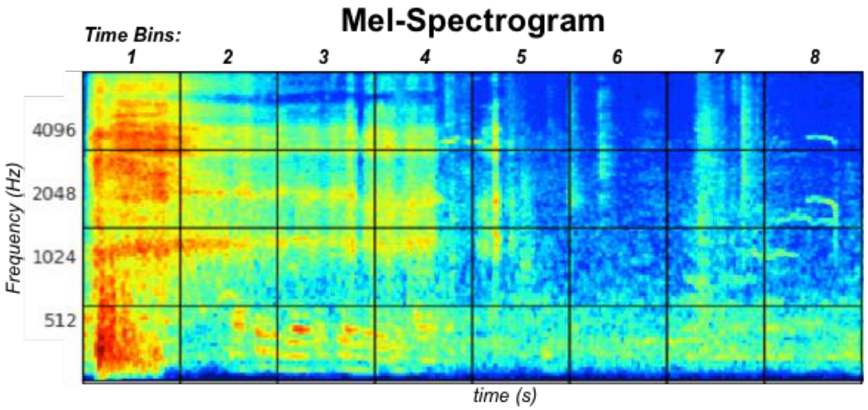
Black Box

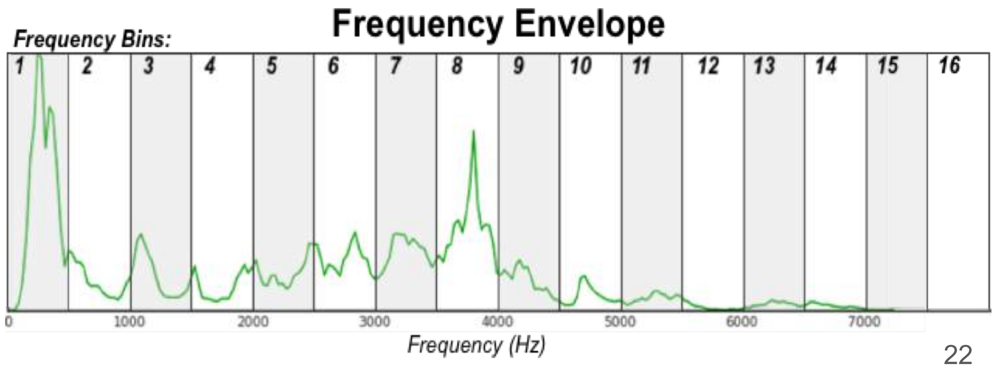
Manual Feature Extraction

- 64 Amp-based features
- 16 Freq-based features
- 80 Mel-based features

= 140 total features







Classical Machine Learning

Only as good at the manually defined input features are...

- Use different subsets of the 140 **manual features** as inputs
- Train binary for Confidence output, regression for Prediction output(s)
- Trained with **5-fold cross validation** and optimized with **gridsearch**

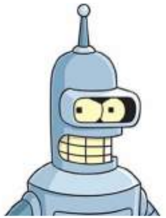
Type	Binary Model	Regression Model
Decision Tree	Gradient boosting (GBM)	Gradient boosting (GBM)
Linear	Logistic regression L2 norm	Linear regression L2 norm
Sanity Check	Naive Bayes (NB)	Guessing mean

Also investigated:

Random forests, K-nearest neighbors, SVM and L1 regularization

Deep Learning via Neural Network

Hope that it extracts meaningful features and does not find a way to cheat...



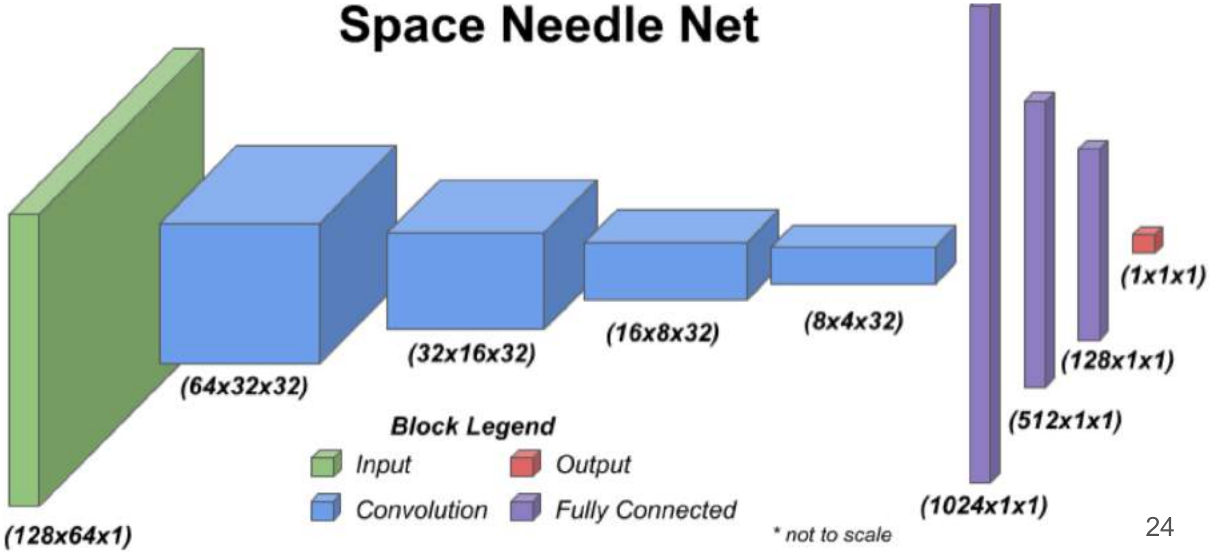
- Raw audio input, **automatic** feature extraction learned through training
- Easy to **parameterize** (used for Confidence and Prediction)

Convert raw audio to spectrogram

Trainable feature extractor filters (CNN)

Predict output from extracted features (FCN, RNN, ect)

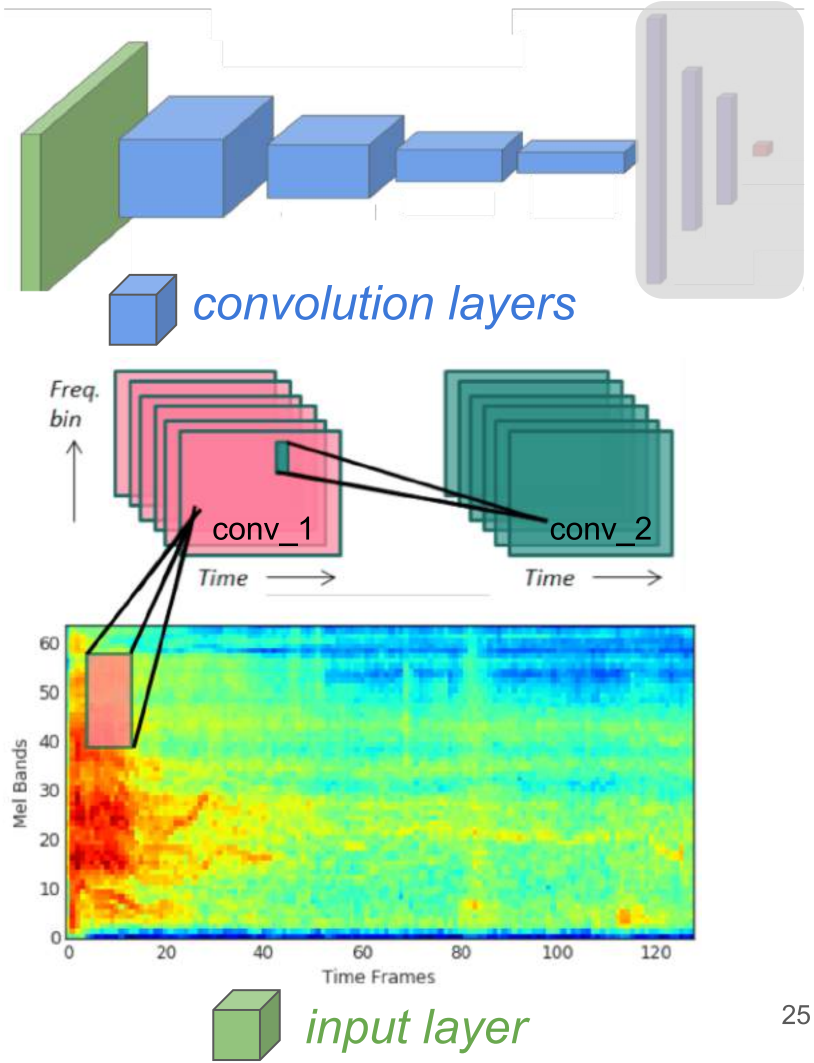
Output value, or sequence



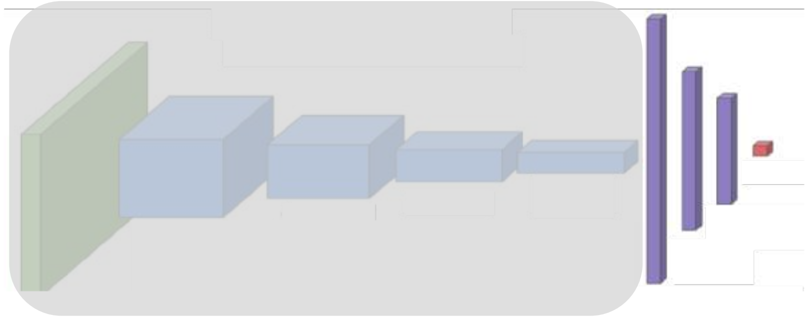
Feature Map Extraction

Data represented as a 3D block volumes

- Initialized as Mel spec
- Spectrogram can be tuned in training
- Convolution layer creates **variants of the input** via tuned filters
- These variants are passed to the next convolution layer
- More **specific** feature maps **deeper** in the net
- Yields a stack of **small feature maps**



Output Prediction



- Receives **stack of feature maps**
- Decides how to use them to **predict** desired output
- Either **fully connected** (FC), **recurrent** (RNN) or **pooling**
- Either output single value (scalar), or sequence

Space Needle based Models

Confidence Models

- ConfidenceNet
- ConfidenceCRNN

Prediction Models

- ScalarNet (single output)
- CurveNet (sequence output)

Results

Confidence Model Results

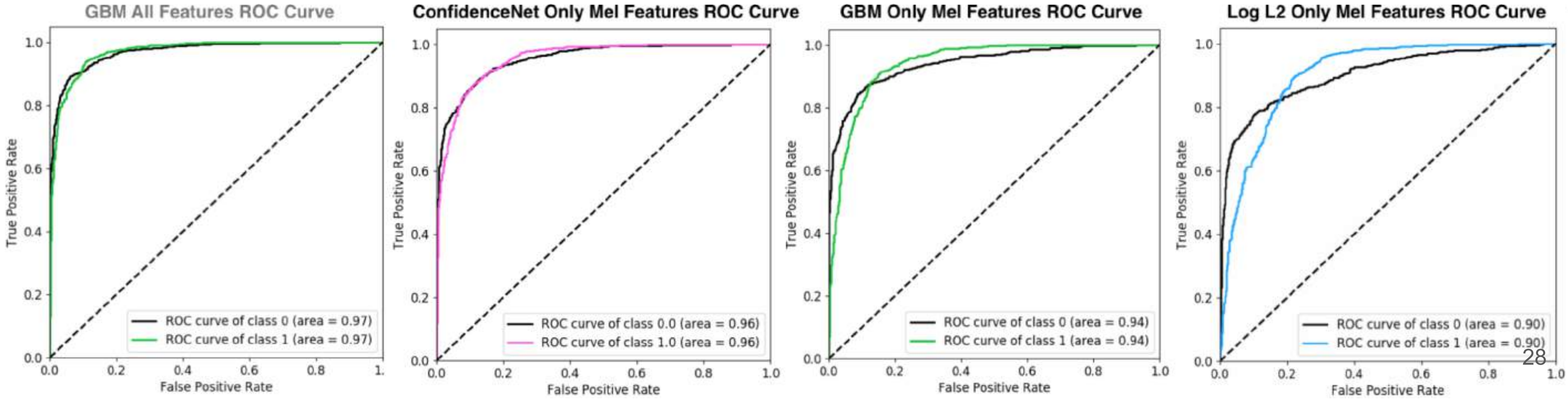
Evaluation Set: 200 difficult accept/reject (balanced)

Input Feature Dimension:

- ***only mels:*** *80 Mel-spec derived features*
- ***all:*** *140 Amp + Freq + Mel derived features*

important

Model	Features	Accuracy	Fscore	Precision	Recall
ConfidenceNet	only mels	0.852	0.850	0.814	0.888
GBM	only mels	0.843	0.837	0.813	0.863
Log L2	only mels	0.821	0.821	0.775	0.873
GBM	all	**0.895**	**0.891**	**0.870** ★	**0.914**
Log L2	all	0.850	0.845	0.819	0.873



Prediction Model Results

Evaluation Set:

700 unseen patients somewhat evenly distributed health wise

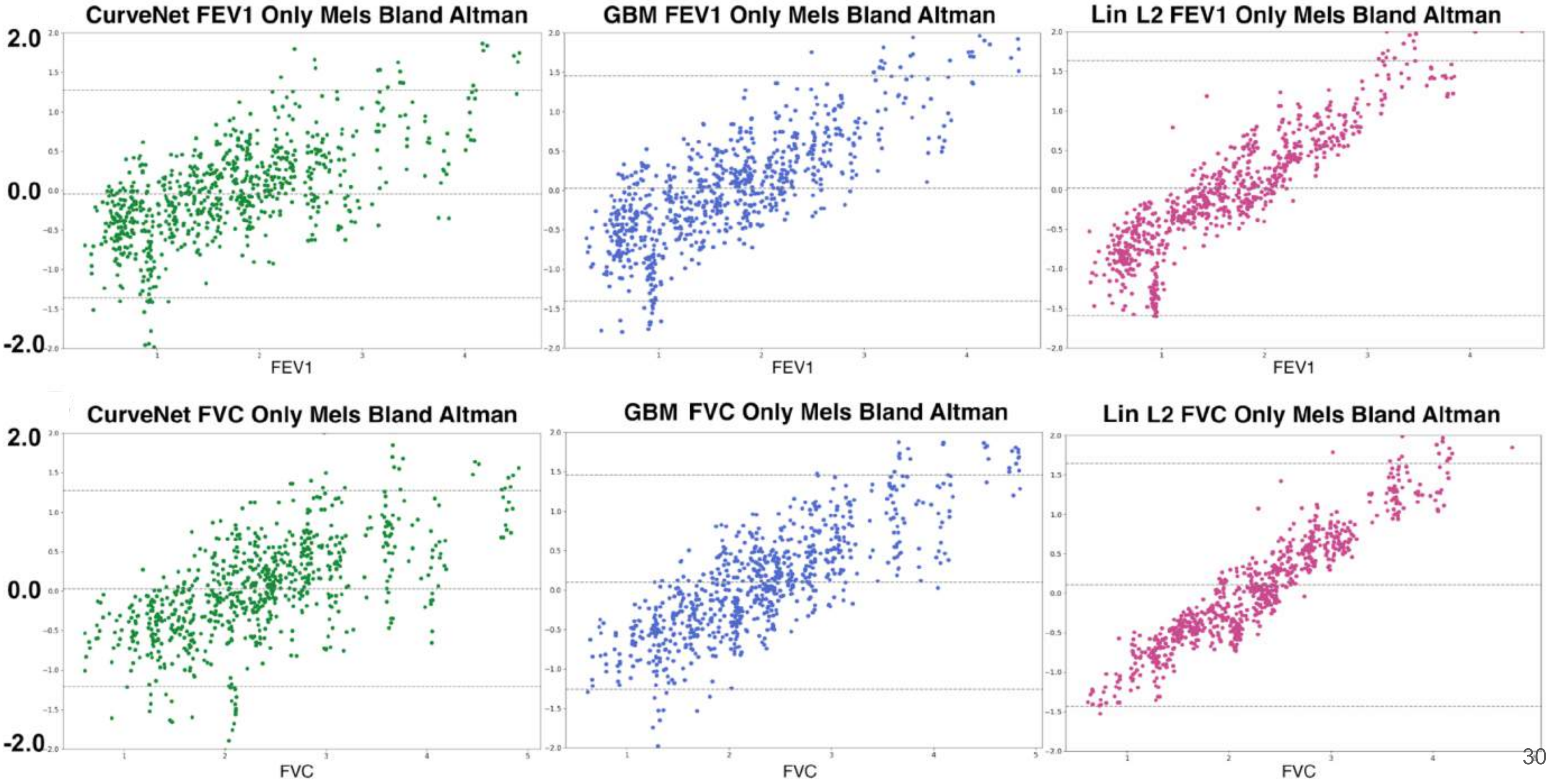
Absolute Error

Model	Features	FEV1	FVC	PEF	FET
Mean Tracking	NA	0.78	0.68	1.90	1.73
★ CurveNet	only mels	**0.48**	**0.50**	1.39	1.72
ScalarNet	only mels	0.50	0.52	**1.33**	1.79
Lin L2	only mels	0.63	0.62	1.69	1.70
GBM	only mels	0.56	0.54	1.51	1.69
Lin L2	all	0.60	0.57	1.69	1.65
GBM	all	0.52	0.54	1.53	1.61

Metric Information

Metric	Unit	Mean	Min	Max
FEV1	L	1.8	0.25	6.2
FVC	L	2.5	0.41	6.8
PEF	L/s	4.7	0.75	11.5
FET	s	7.4	1.6	15.2

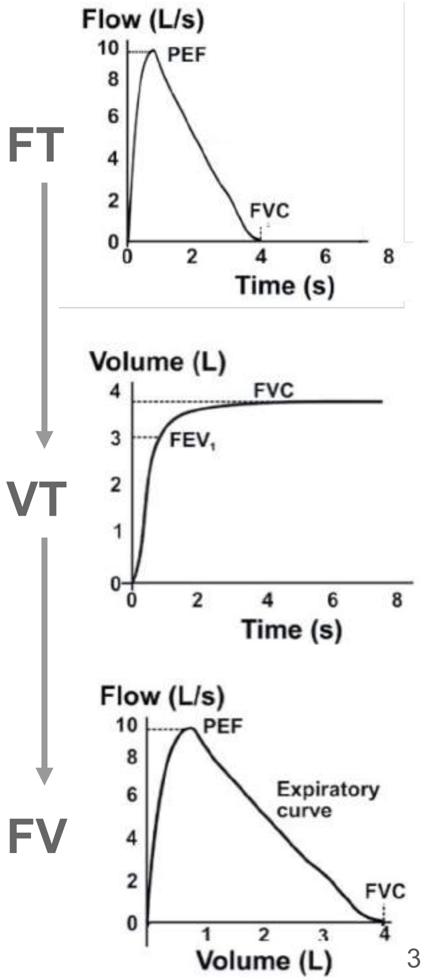
CurveNet: Bland Altman plots



CurveNet Custom Cost Function

- **Output:** Flow vs Time (FT), not clinically useful
- **Derive:** clinical metrics with discrete operations
- **Cost Function:** Subjective combination of each error

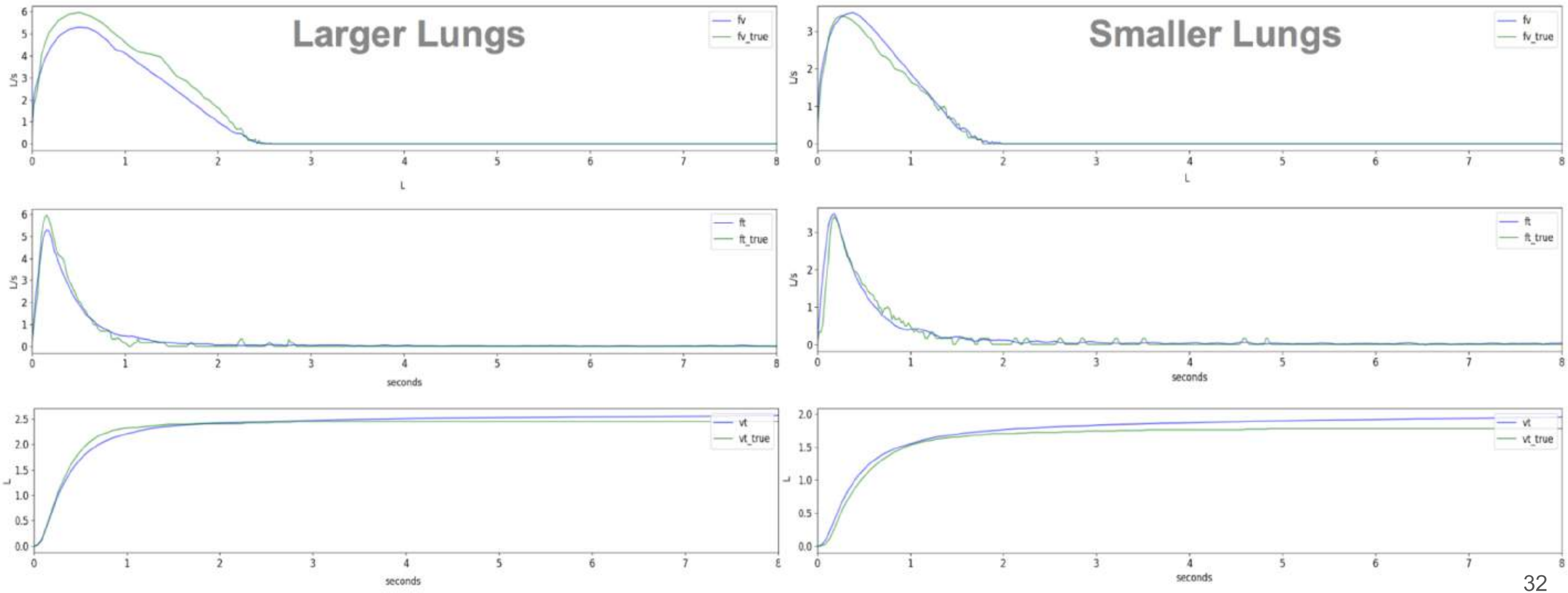
Weight	Metric	Operation
0.2	**FT curve**	*Outputted by CurveNet*
0.2	**VT curve**	*Cumulative sum of FT*
0.3	**FEV1**	*VT value at 1 second*
0.2	**FVC**	*Max of VT*
0.1	**PEF**	*Max of FT*



CurveNet: Good Output

1. flow volume
2. flow time
3. volume time

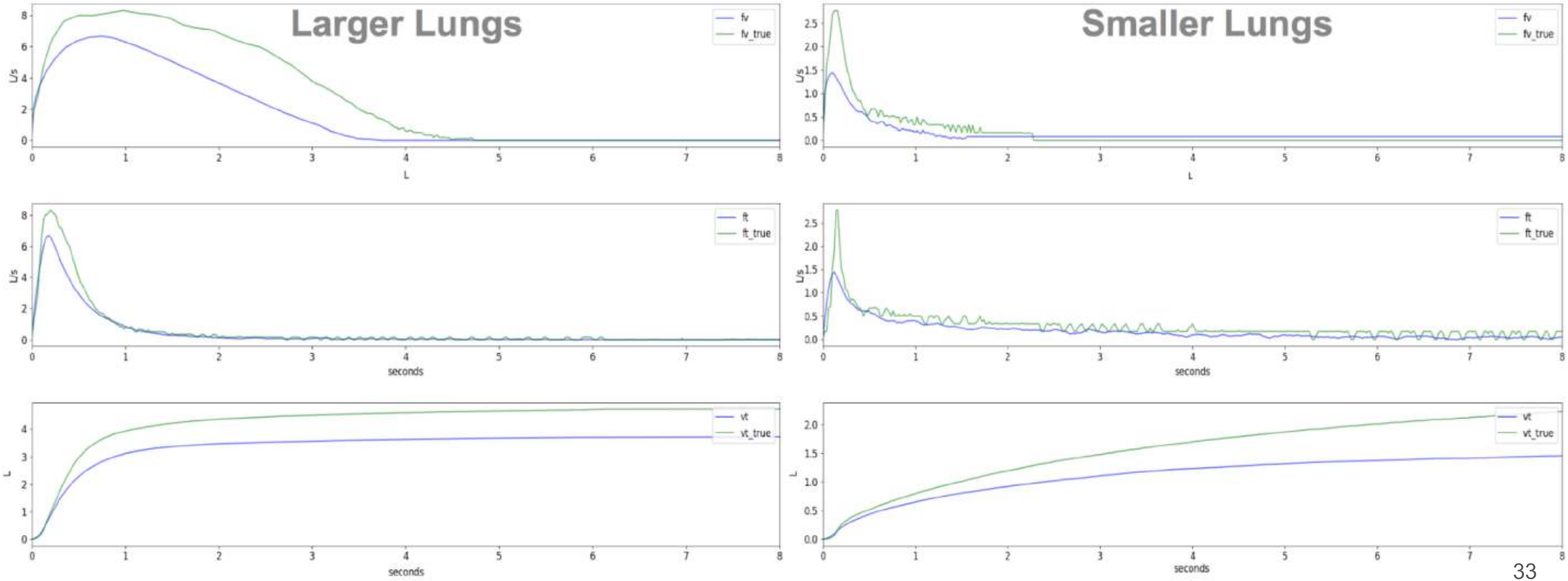
groundtruth
predicted



CurveNet: Underestimated

1. flow volume
2. flow time
3. volume time

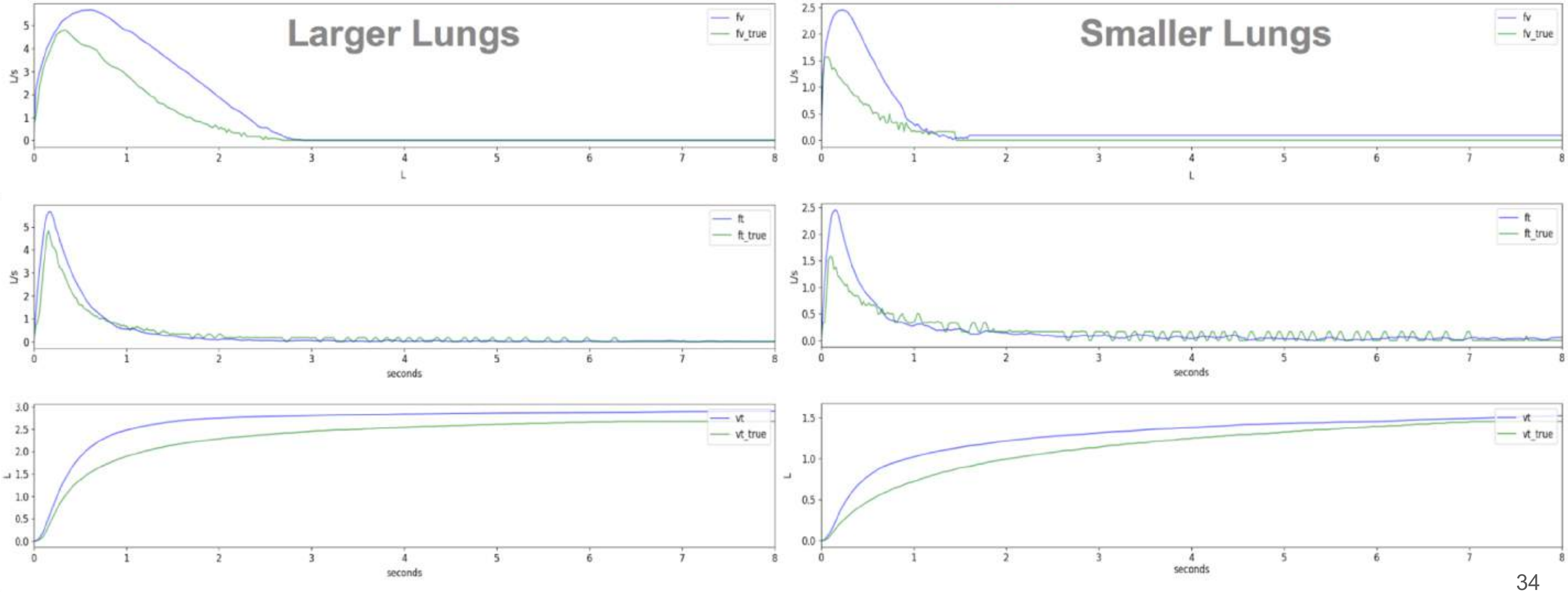
groundtruth
predicted



CurveNet: Overestimated

1. flow volume
2. flow time
3. volume time

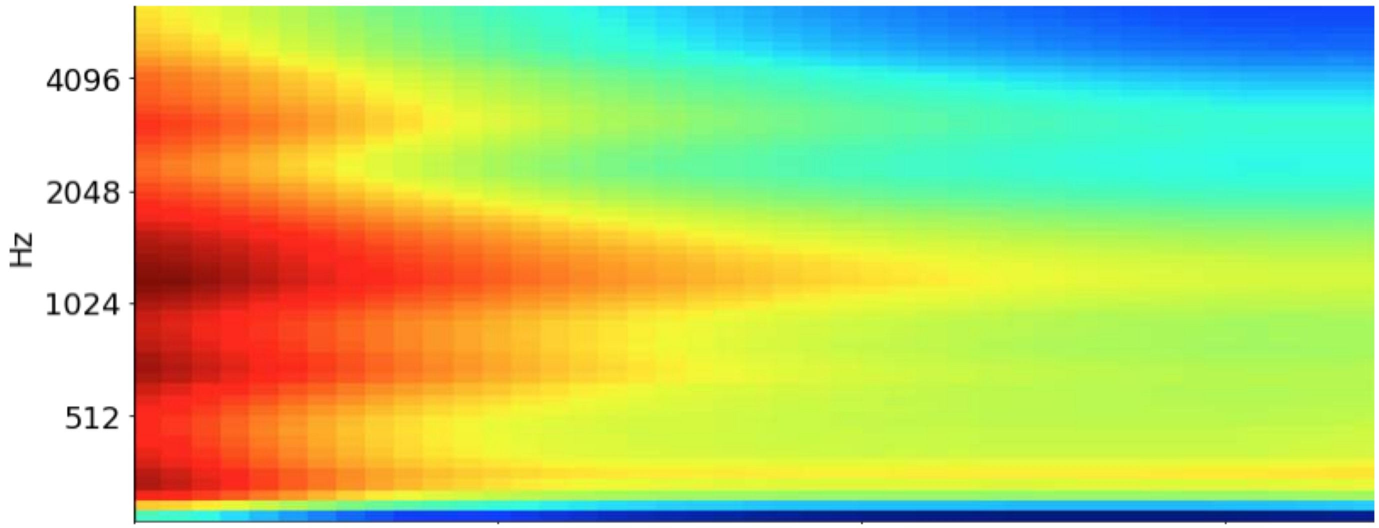
groundtruth
predicted



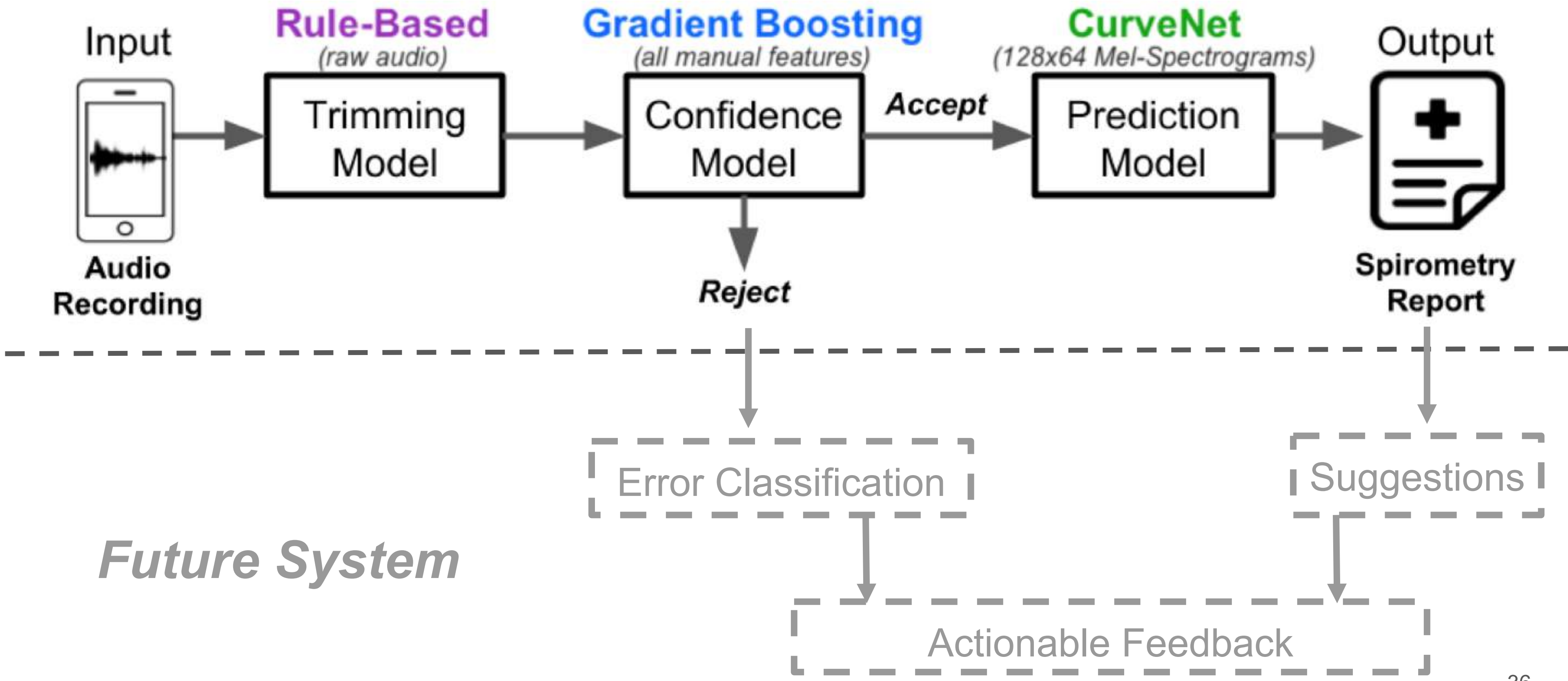
CurveNet Dissection

- Neural nets are a **black box**, even more so with a *sequence to sequence* model
- Post-training feature map analysis shows areas the model *attends* to

CurveNet regions of interest in first 2 seconds of exhale



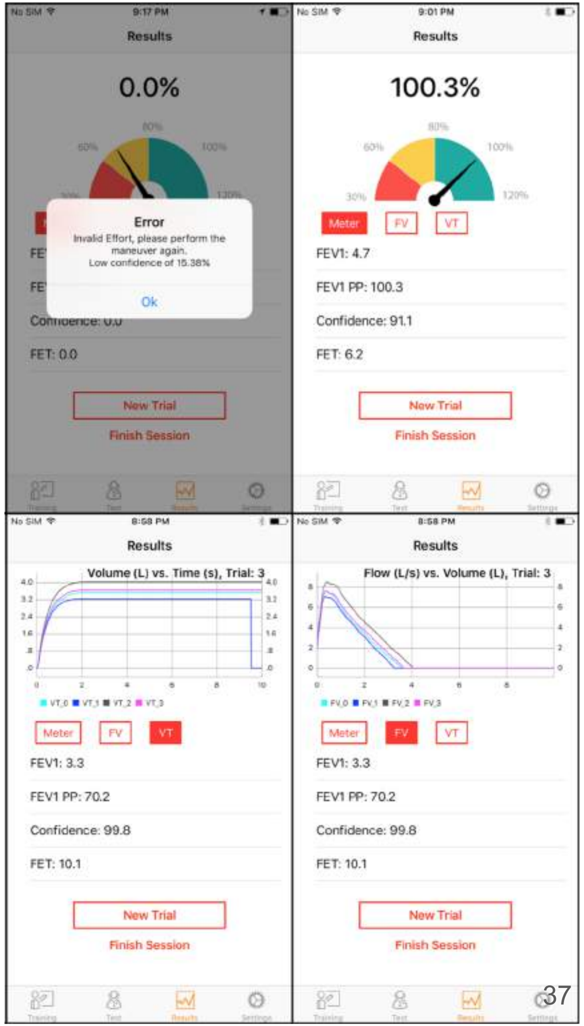
Current Spiro AI System: Revisited



Deployment

Purpose: Data collection, user experience research, live demos, collaboration, validation

- Production **backend** powered by Spiro AI + docker *(thanks Elliot and Dennis)*
- FreshAir iOS **frontend** app *(thanks Varun)*
  - Patient input, clinic selection
  - Coaching videos, multiple languages
  - Capture multiple audio trials
  - Reject effort or return Spiro AI report
  - Compute reproducibility of session
  - Store results for trend reporting
- Available on Ubicomp Lab demo phones



What is next...

Data Collection

- Targeted, controlled data collection
- Longitudinal validation study (track trends)
- FreshAir app usability studies (clinical and home-use)
- Pursue FDA regulation

Technical

- Unify trimming and confidence model
- Sound based error classification
- On-device Sprio AI system

Final Pitch

Spiro AI:

*“Make spirometry as **EASY** and **FUN** as taking a selfie”*

BEFORE

AFTER

(dramatization)

Thanks