Towards Ambient Intelligence in Smart Healthcare

John A. Stankovic
BP America Professor
Director, Link Lab
Department of Computer Science
University of Virginia

UVA ENGINEERING
LINK LAB

How I Got Here

- Teaching
  - High School Tutoring
  - Math and Engineering (lots of Math) – WHY?
- Research
  - ABM

My Work - Main Themes

- Wearables (Smart Watches) and In-situ
- Cognitive Assistance (ML and NLP)
- Conflict Detection (NLP)
- Acoustics (ML)
- Real Deployments

Current/Recent Projects

- Family Eating Dynamics for Obese Families - USC/Los Angeles – 23 families
- Smart Watch Reminder Systems – UVA Center for Telemedicine
- Alzheimer’s Patient – Caregiver Interaction – Ohio State and Univ. of Tennessee
- Smart Watch Handwashing Conformance, UVA PICU
- First Responders Cognitive Assistant – North Garden Fire and Rescue/ Richmond Fire, and Oxford Univ.

New Themes

- Use FM to integrate properties into ML models
- Address Uncertainties in ML Model Predictions

IOHT - Hype or Revolution

- Smart Watches
  - More and more sold; more and more sensors
- Smart Skin
- In-situ
- Smart Textiles
- …

UVA ENGINEERING
LINK LAB

University of Virginia

Vision

An ambient healthcare intelligence

WebMD
Big Data Collections
ML/Analytics
Data Mining

General Population

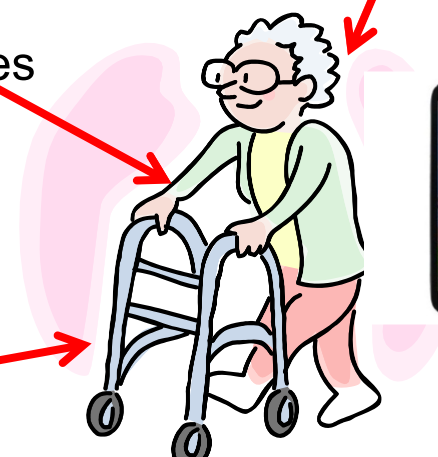
Actuations

Holistic

Nano-pills
Pacemaker

Wearables

Sensors



University of Virginia

Today's Main Themes

- Wearables (Smart Watches)
- Cognitive Assistance (ML and NLP)

Towards Ambient Healthcare Intelligence

- Acoustics (ML)
  - In situ => mood at distance
  - In situ => anxiety via MIL

Powerful Modality

UVA ENGINEERING
LINK LAB

University of Virginia

Cognitive Assistance (on a smart watch)

- Towards General Healthcare Intelligence (comprehensive services)
  - Interact with Internet Healthcare Services
  - Support Conversations (esp. for elderly)
  - Reminders, suggestions, alarms - explainable
  - Physical and Mental Health
  - Pandemic-aware
  - Privacy-aware – not discussed today

iAdhere – verbal medication and exercise reminder system

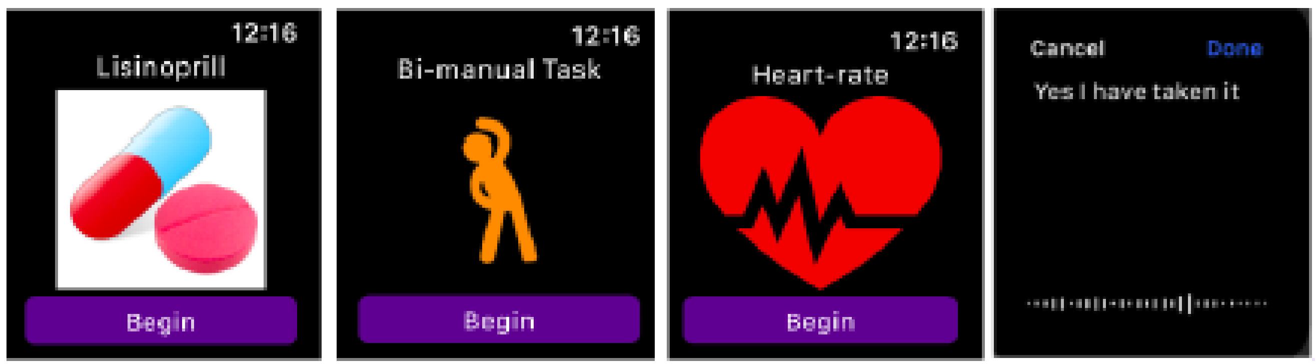


Figure 1: User Interfaces of the reminders to and the response from the users

For stroke patients

Using Apple Watch – with microphone and speaker

Appling in a Telemedicine setting

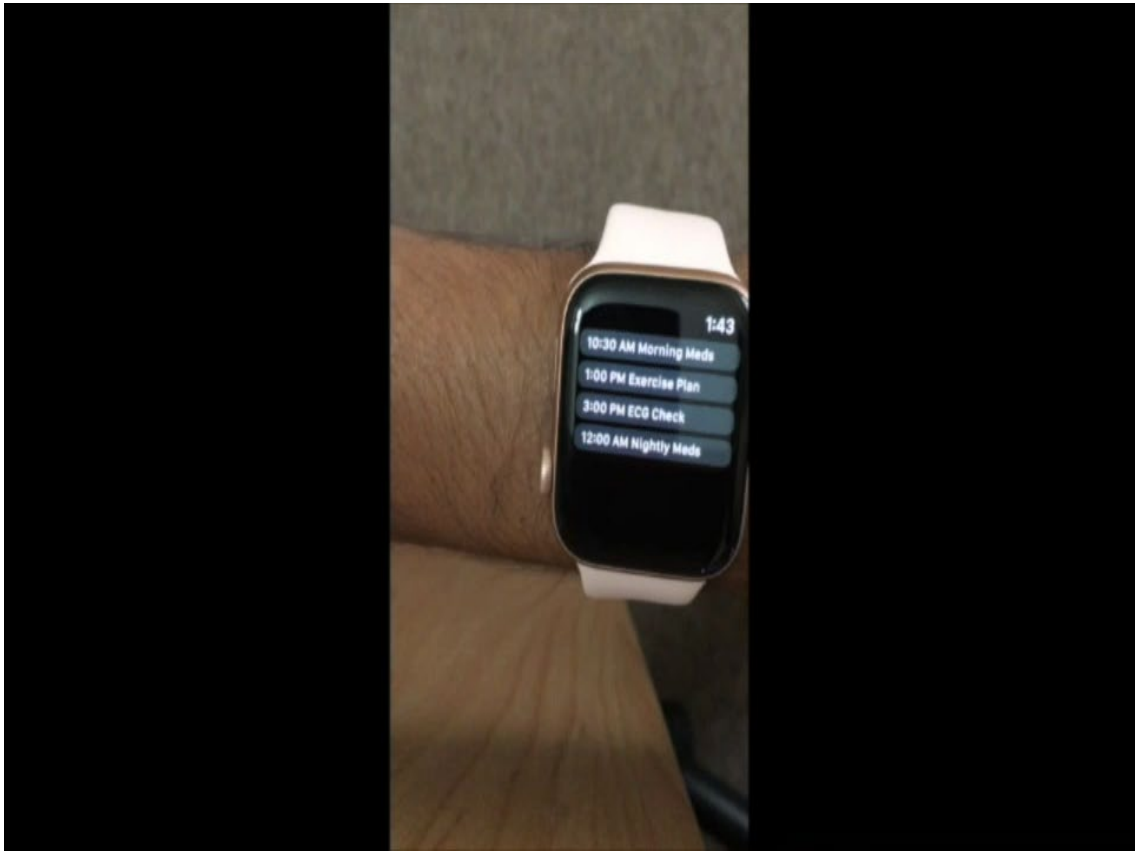
University of Virginia

iAdhere

- Medication and exercise reminders
  - Supports general verbal questions
  - Allows rescheduling
  - Current: quality of exercise and pain
- EKG
- General healthcare dialogue support

Towards comprehensiveness

Demo – A Few Features



Earlier version called Medrem

Services Expanded for Pandemics

- Collections of services on a smart watch
  - Handwashing (or general hygiene/elderly)
  - Mood/Depression/Anxiety/Loneliness
  - Voice based conversations
    - Pandemic info
    - Reminders/Alerts/Advice
  - Physiological parameters and more
    - Symptoms
  - …

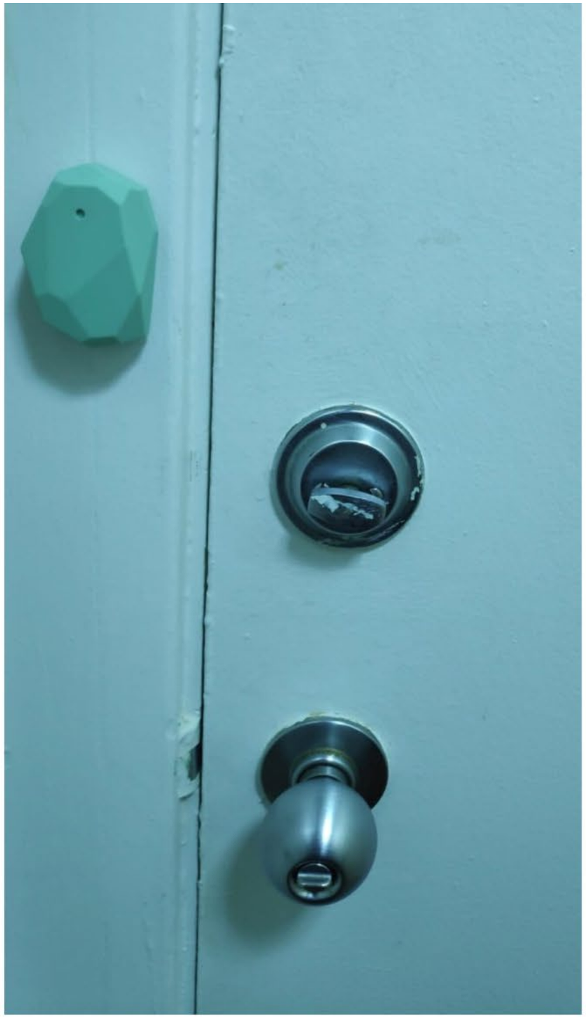
Quality of Handwashing

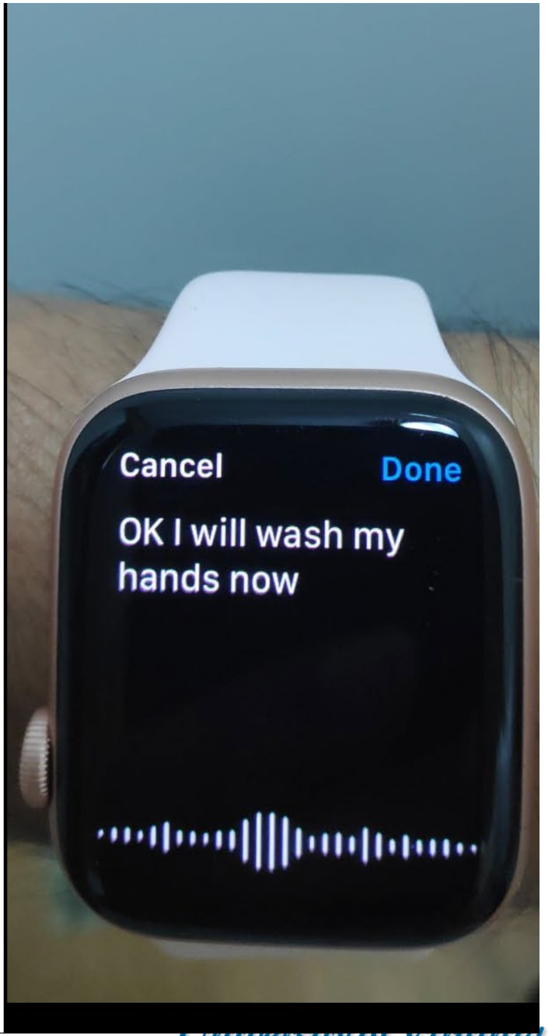
- WHO guidelines
- Quality
- Solution: Hybrid CNN-RNN

- Supports conversations
  - Reminders based on time and when return home (beacons)
  - Info on quality of handwashing

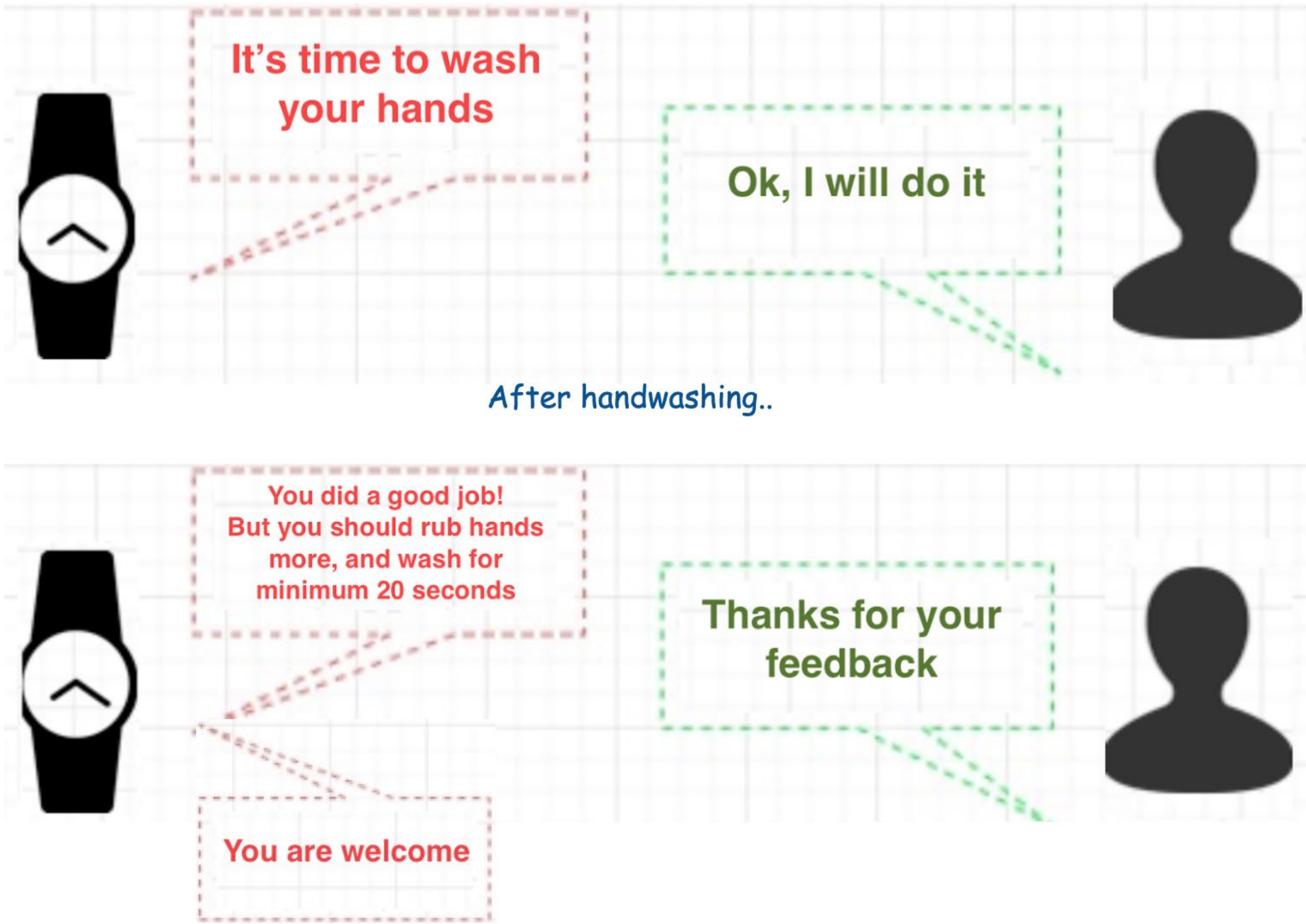

How to handrub?
WITH ALCOHOL-BASED FORMULATION
How to handwash?
WITH SOAP AND WATER
1a
1b
Apply a palmful of the product in a cupped hand and cover all surfaces.
0
Wet hands with water
1
apply enough soap to cover all hand surfaces.
2
Rub hands palm to palm
3
right palm over left dorsum with interlaced fingers and vice versa
4
palm to palm with fingers interlaced
5
backs of fingers to opposing palms with fingers interlocked
6
rotational rubbing of left thumb clasped in right palm and vice versa
7
rotational rubbing, backwards and forwards with clasped fingers of right hand in left palm and vice versa
8
rinse hands with water
9
dry thoroughly with a single use towel
10
use towel to turn off faucet

Smartwatch App



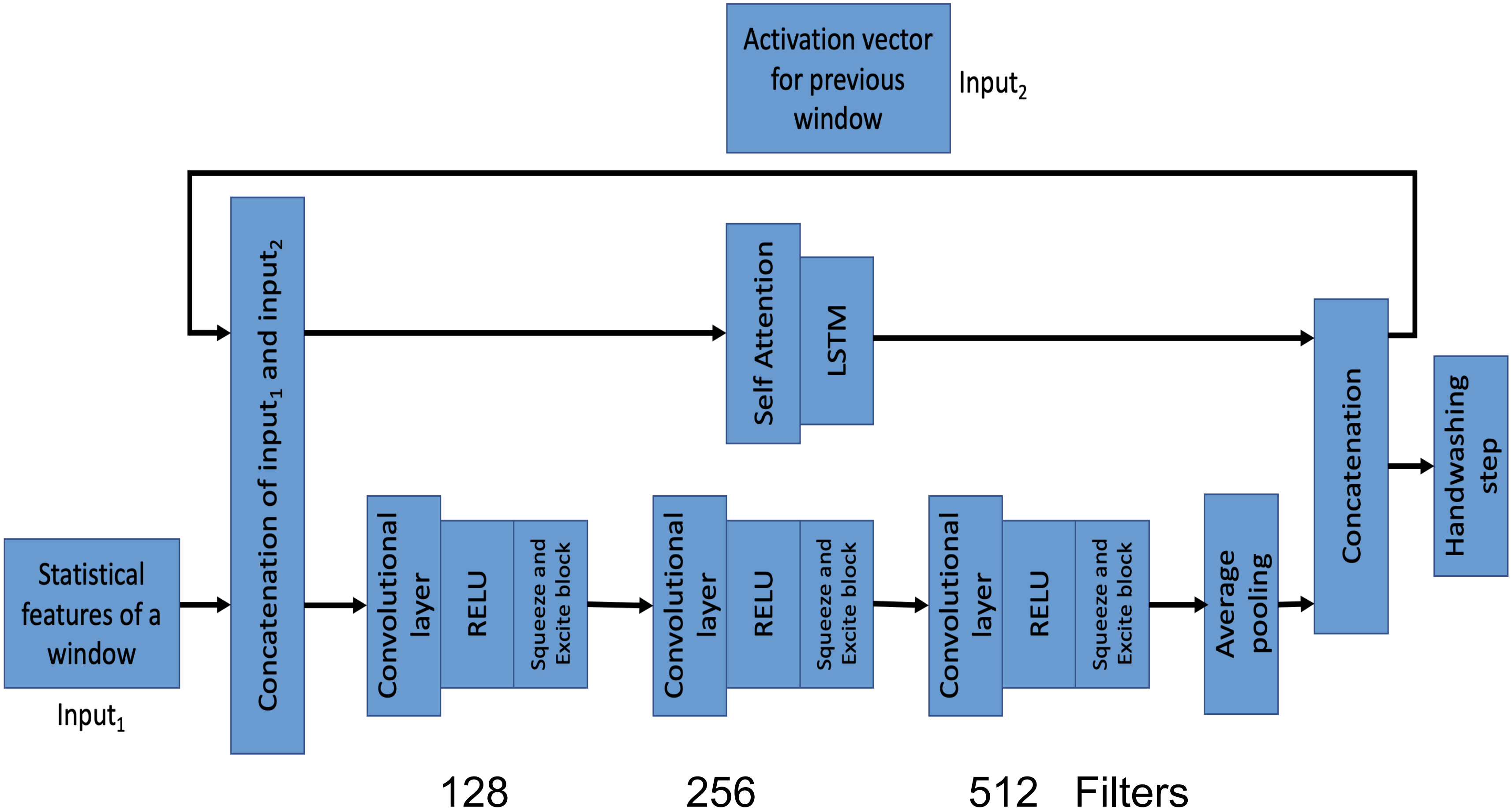


Dialogue



University of Virginia

Solution – A Hybrid DNN

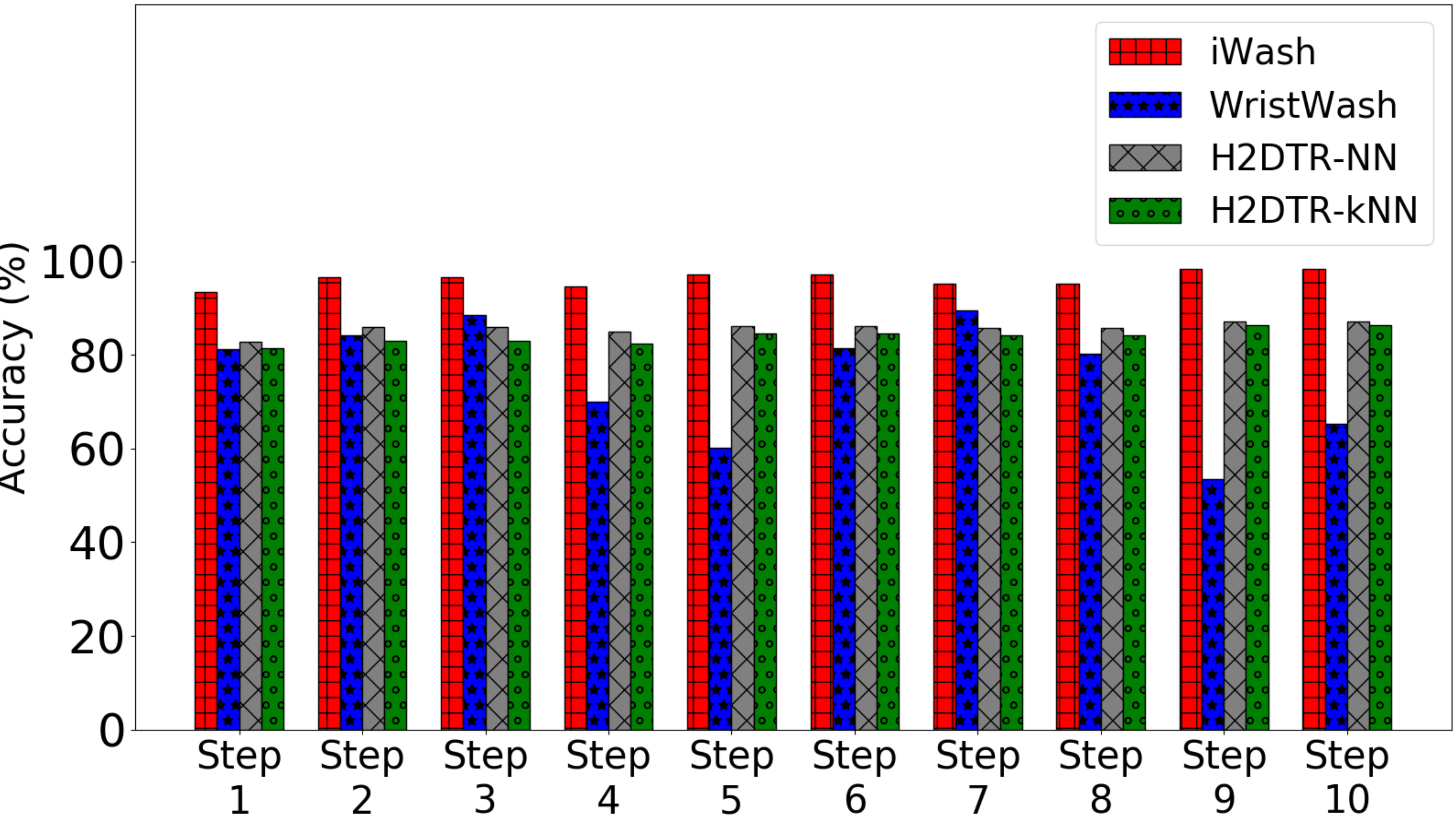


S. Samyoun, S. Shubba, A. Mondol, and J. Stankovic, iWash, A Smart Handwashing Quality Assessment and Reminder System with Real-Time Feedback in the Context of Infectious Diseases, *CHASE*, Dec. 2020.

Evaluation

- Our own dataset
- 14 participants
  - Each 19 HW sessions
- 3 practice runs
- Video for Ground Truth

Evaluation



iWash
WristWash
H2DTR-NN
H2DTR-kNN
Accuracy (%)
0
20
40
60
80
100
Step 1
Step 2
Step 3
Step 4
Step 5
Step 6
Step 7
Step 8
Step 9
Step 10

Acoustics: Exploiting Speech

- Distance Emotion Recognition
  - Happy, sad, angry, neutral
- Anxiety and Depression

Mental Health

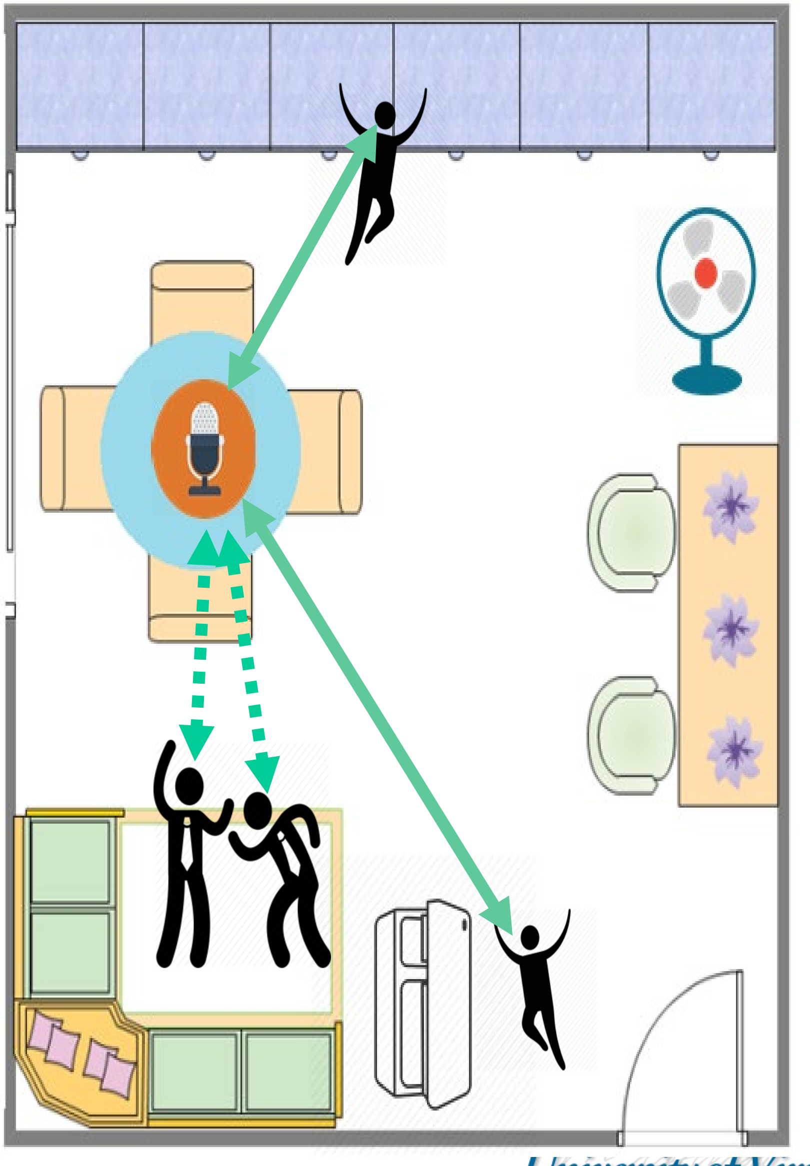
Distance Emotion Recognition

Close to microphone

Fixed distance

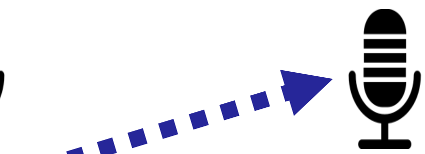
24/7

A realistic indoor speech emotion recognition system



- *Reverberation*
- *Ambient noise*
- *De-amplification of speech*
- *Overlapping of speech*

Solution

1. *Distance Agnostic Features/code words*

2. *Feature Modeling: Emo2vec*

3. *Classifier: LSTM*

Select Robust Features

Consider 231 LLD features

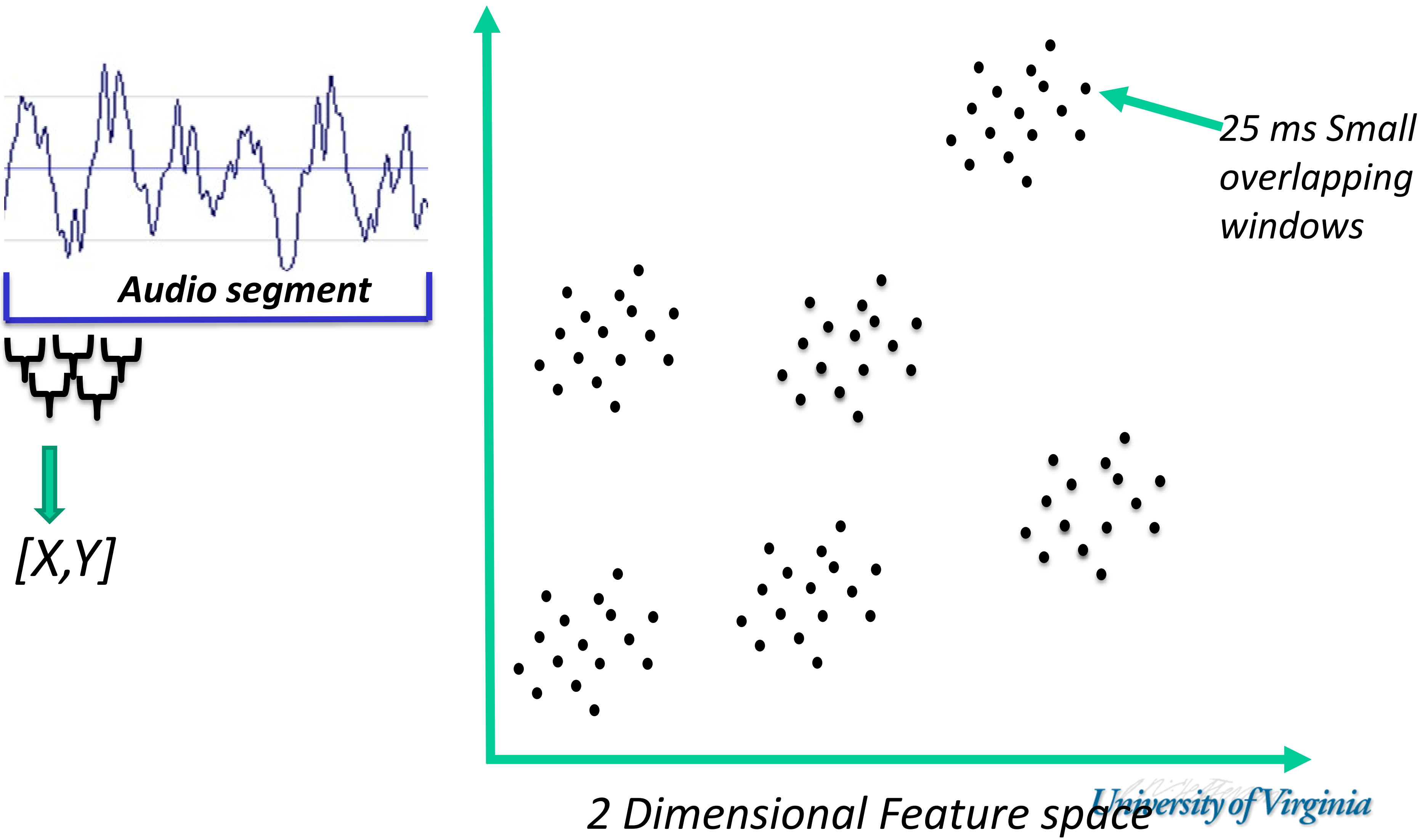
Feature	Count
Mel-Frequency cepstral coefficients (MFCC) 1-25	25
Root-mean-square signal frame energy	1
The voicing probability computed from the ACF	1
The fundamental frequency computed from the Cepstrum	1
Pitch	1
Harmonics to noise ratio (HNR)	1
Zero-crossing rate of time signal	1
PLP cepstral coefficients compute from 26 Mel-frequency bands	6
The 8 line spectral pair frequencies computed from 8 LPC coefficients	8
Logarithmic power of Mel-frequency bands 0 - 7	32

Select 48 LLD features:

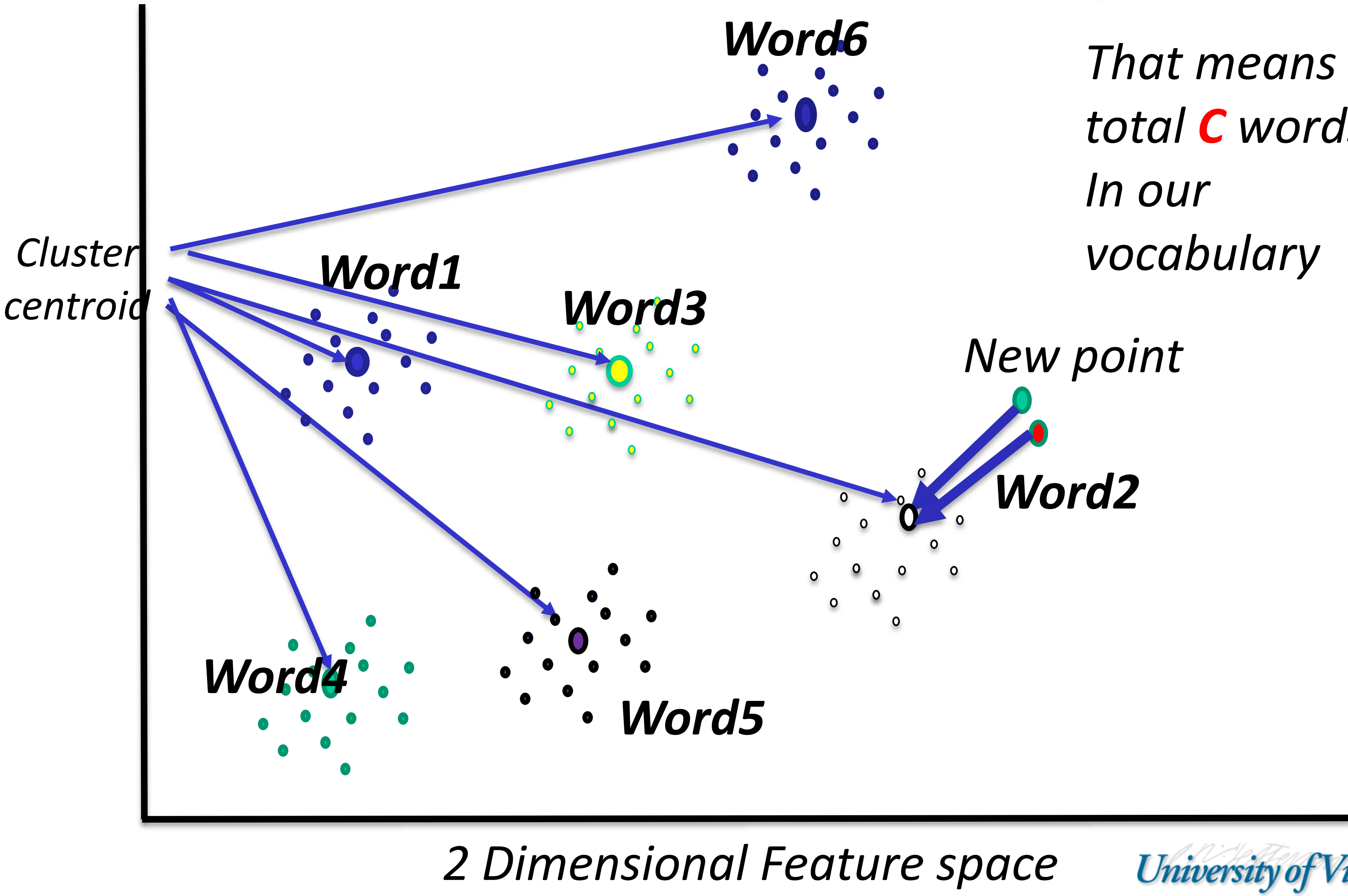
- *5 MFCC*
- *Voice probability*
- *Fundamental frequency*
- *Zero crossing rate*
- *8 line spectral pair frequencies*
- *32 logarithmic power of Mel-frequency bands*

Delta and delta-delta of these 77 features

Audio (Code) Words



Audio Word



C clusters

That means total C words In our vocabulary

Code Book Sizes

- Tested 500 to 2500 in increments of 500
  - K-means clustering

- Interesting Result: Different code book sizes for different emotions

Adaptation of Word2Vec : Emo2vec

- Convert audio words into vectors
- Words which occur in similar context (that means with similar neighbor words), for **a specific emotions** have similar vector representations.

Words A and C in similar Context but Not for Happy

(word,{Neighbour set})	(word,{Neighbour set})
(A , {P,Q,R,S,T,U,V,W,M})	⋮
(A , {P,R,Q,S,T,U,V,W,N})	(C , {P,Q,R,S,T,U,V,W,X})
⋮	(C , {P,R,Q,S,T,U,V,W,N})
(B , {O,P,Q,R,S,T,U,V,W})	(C , {P,R,Q,S,T,U,V,M,N})
(B , {P,Q,R,S,T,N,U,V,W})	⋮
(B , {P,R,S,T,U,V,W,M,Q})	(E , {F,X,P,Y,Z,S,T,W})
⋮	(F , {A,P,E,G,H,H,J,J})
(D , {E,F,E,G,H,E,B,C})	(J , {E,F,J,M,M,K,N,P})
(D , {G,H,F,E,J,I,GW})	⋮
(D , {F,O,X,D,K,M,N,J})	
Input corpus of happy D_H	Input corpus of Not happy D_N

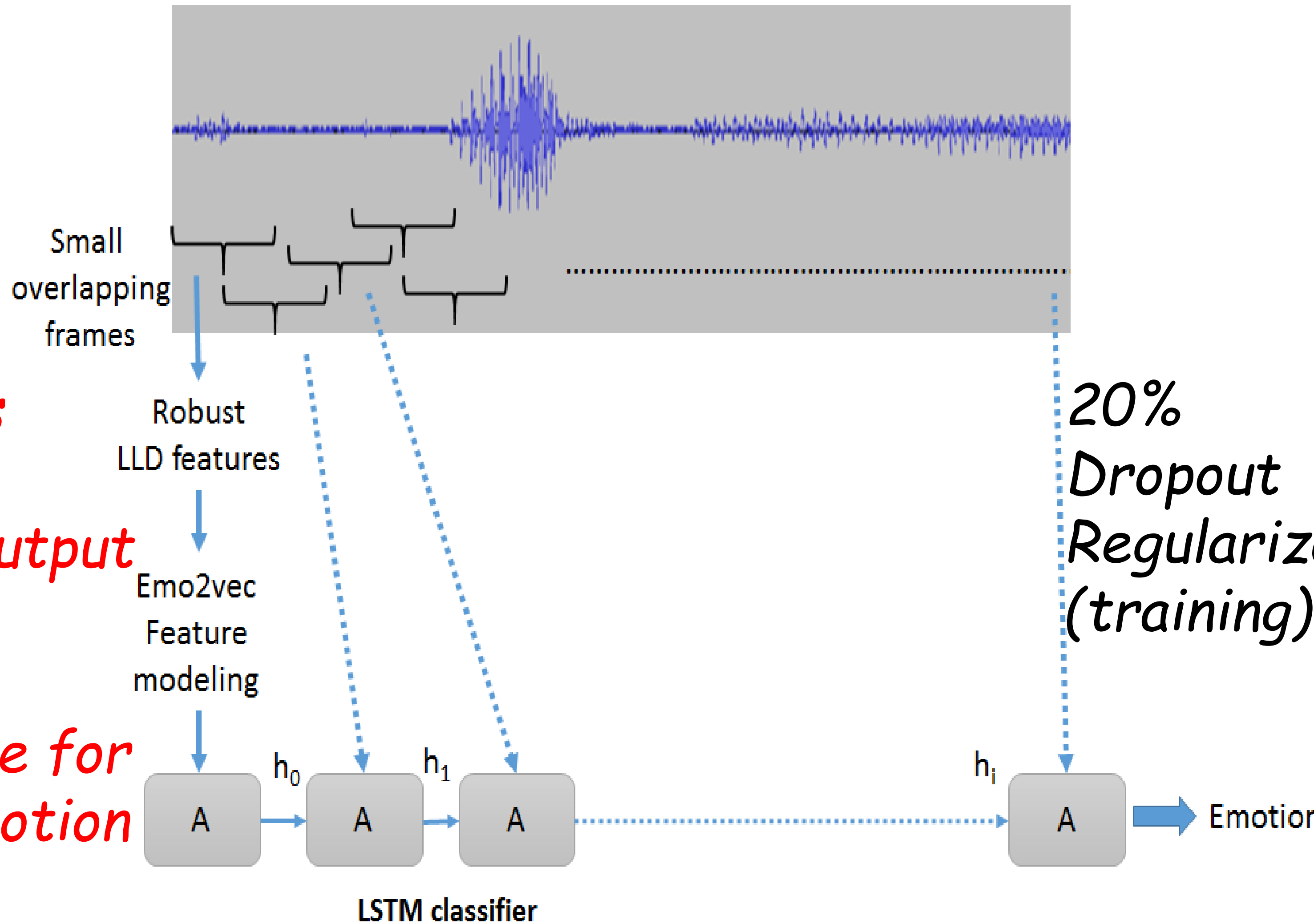
Words A & B, appear In similar context (with similar neighbors) for Emotion Happy

LSTM Classifier

100
Neurons

Dense output
Layer

Separate for
each emotion



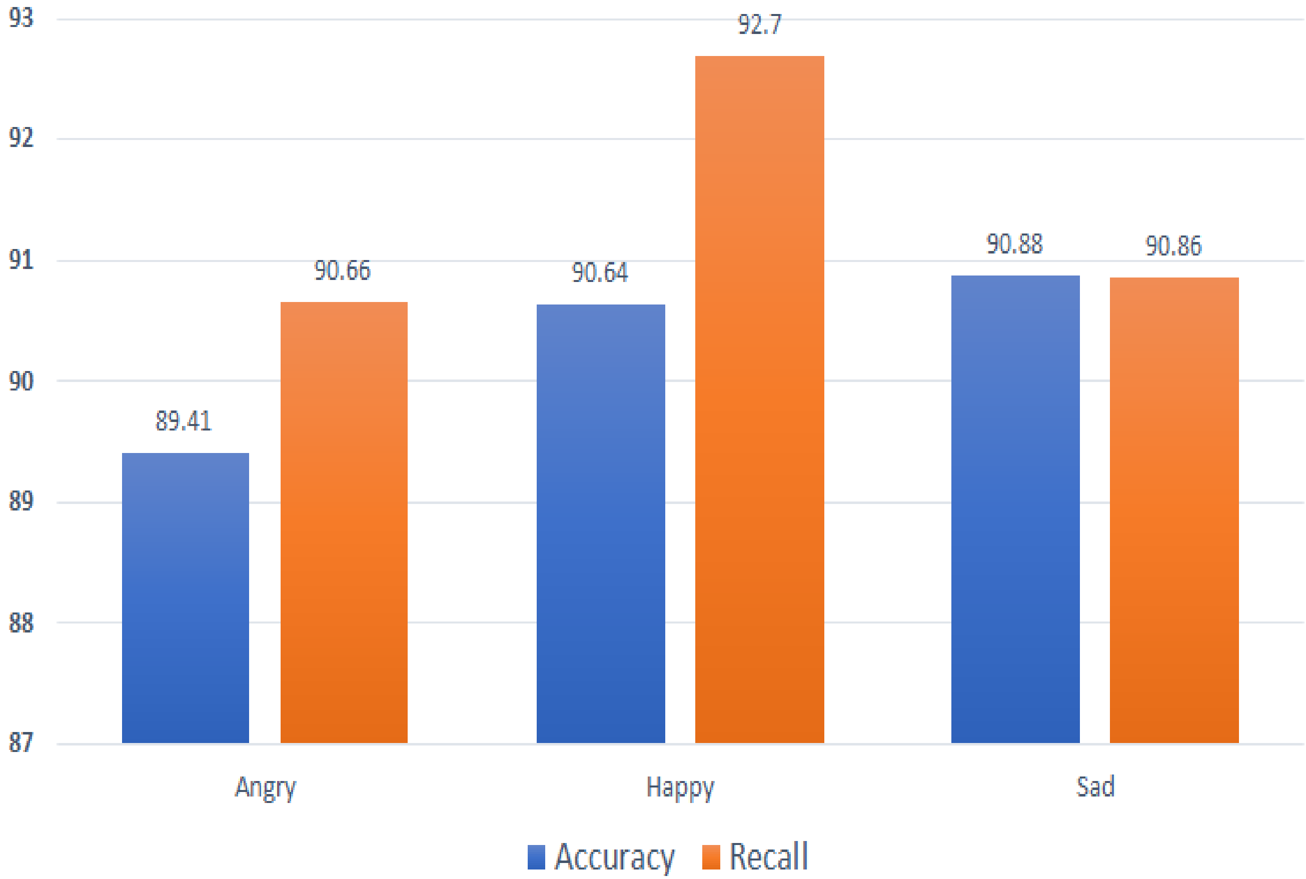
20%
Dropout
Regularization
(training)

Evaluation

- 2 literature datasets
- Our own family-discussion experiments
  - 12 families; 28 people
  - Spontaneous discussions
  - Similar performance to literature datasets
- 4 Baselines
  - Approximately 16% better than best baseline

A. Salekin, Z. Chen, M. Ahmed, J. Lach, D. Metz, K. de la Haye, B. Bell, and J. Stankovic, Distance Emotion Recognition, *ACM Interactive, Mobile, Wearable, and Ubiquitous Technologies*, Vol. 1, Issue 3, Sept. 2017, 96:1-96:24.

Angry, Happy, Sad

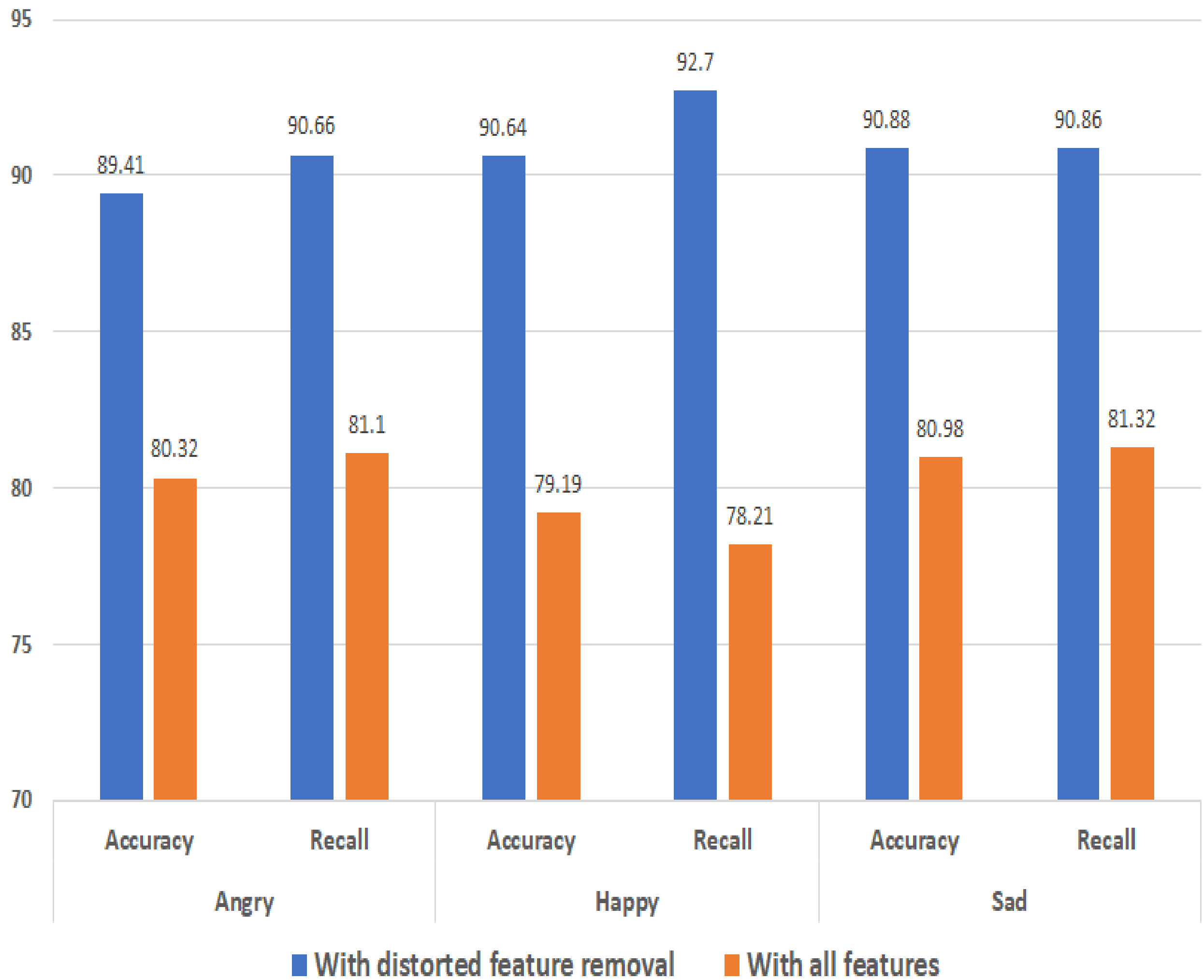


Emo2vec approach on datasets

Observation

- In the past – Happy versus Angry difficult (acted datasets)
  - Why? - Use of Energy based features
  - In real setting: Laughter helps discriminate

Elimination of distorted features helpful?



Effect of Distance

- As we move from mic to 6m away from mic, drop in accuracy is about 5%
- State of art: the drop is about 12%

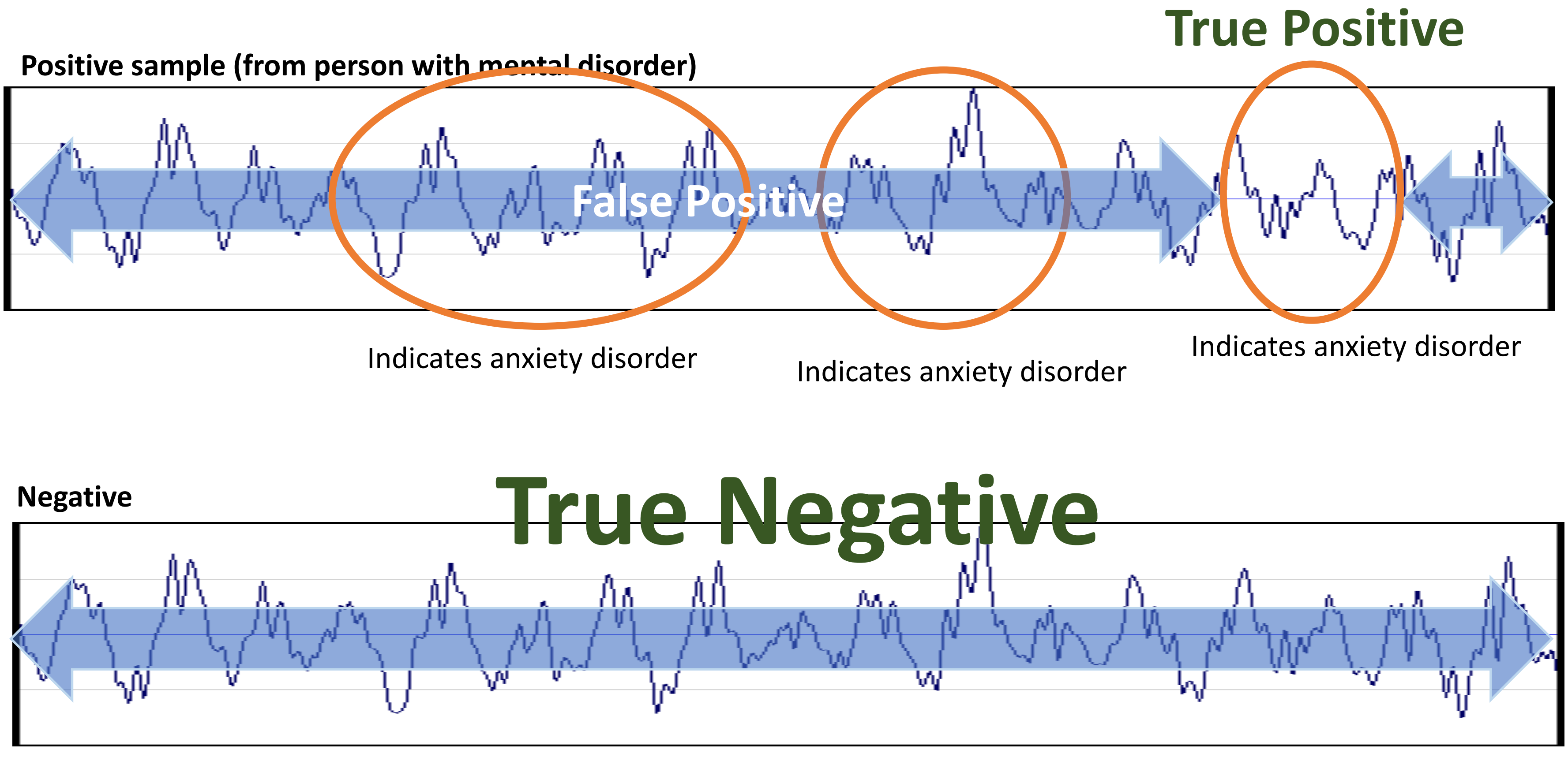
Mental Disorder - Anxiety

- 11% of Americans suffer from Anxiety
- Prior work – most use fully supervised learning
  - But what parts of the speech are representing anxiety
  - Very difficult to label

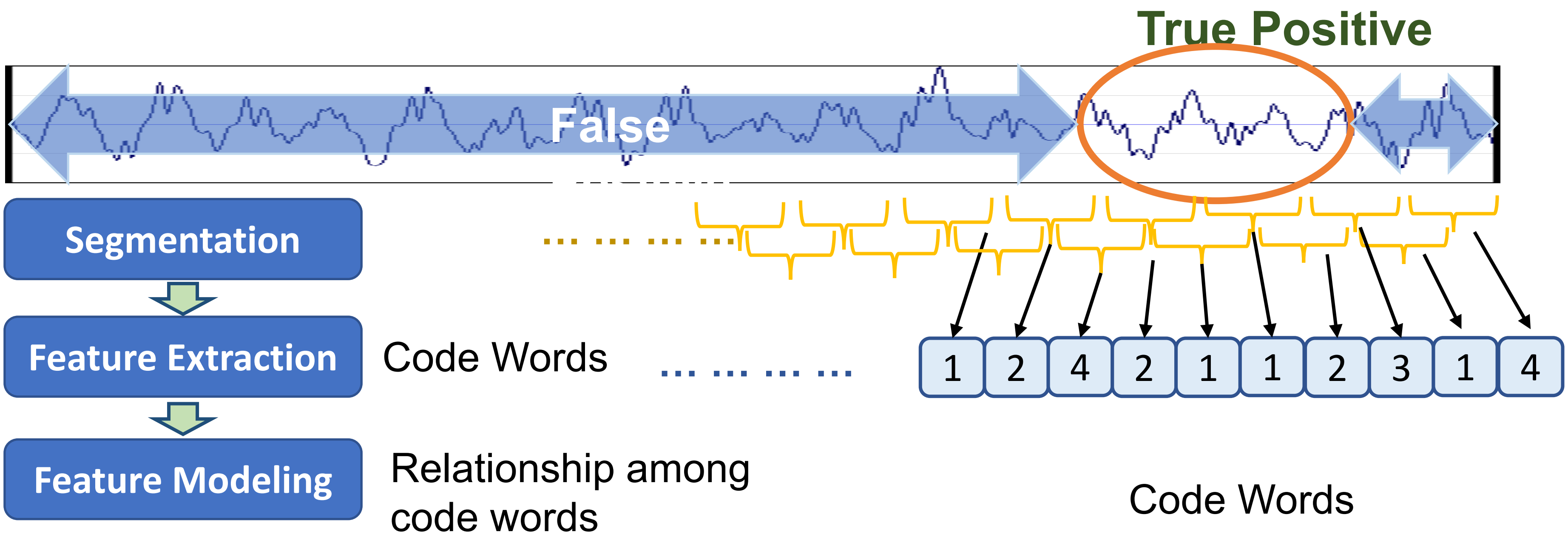
Contributions

- MIL => weakly supervised learning
- Novel feature modeling => NN2vec
- New classifier => BLSTM-MIL
- 90% F-1 and accuracy
- 17% better than baselines

Weakly Labeled Data



Positive Audio Clip



NN2VEC (vector embedding)

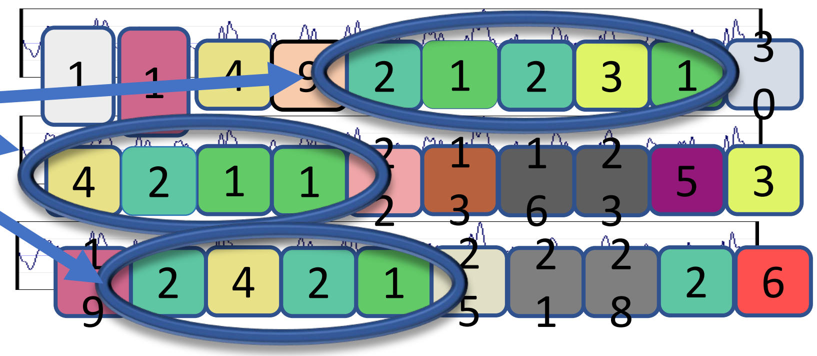
and

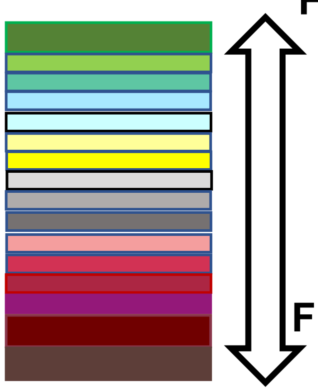
Learned via DNN

NN2VEC Feature Modeling

Region indicative to Positive class: mental disorder

Positive Example

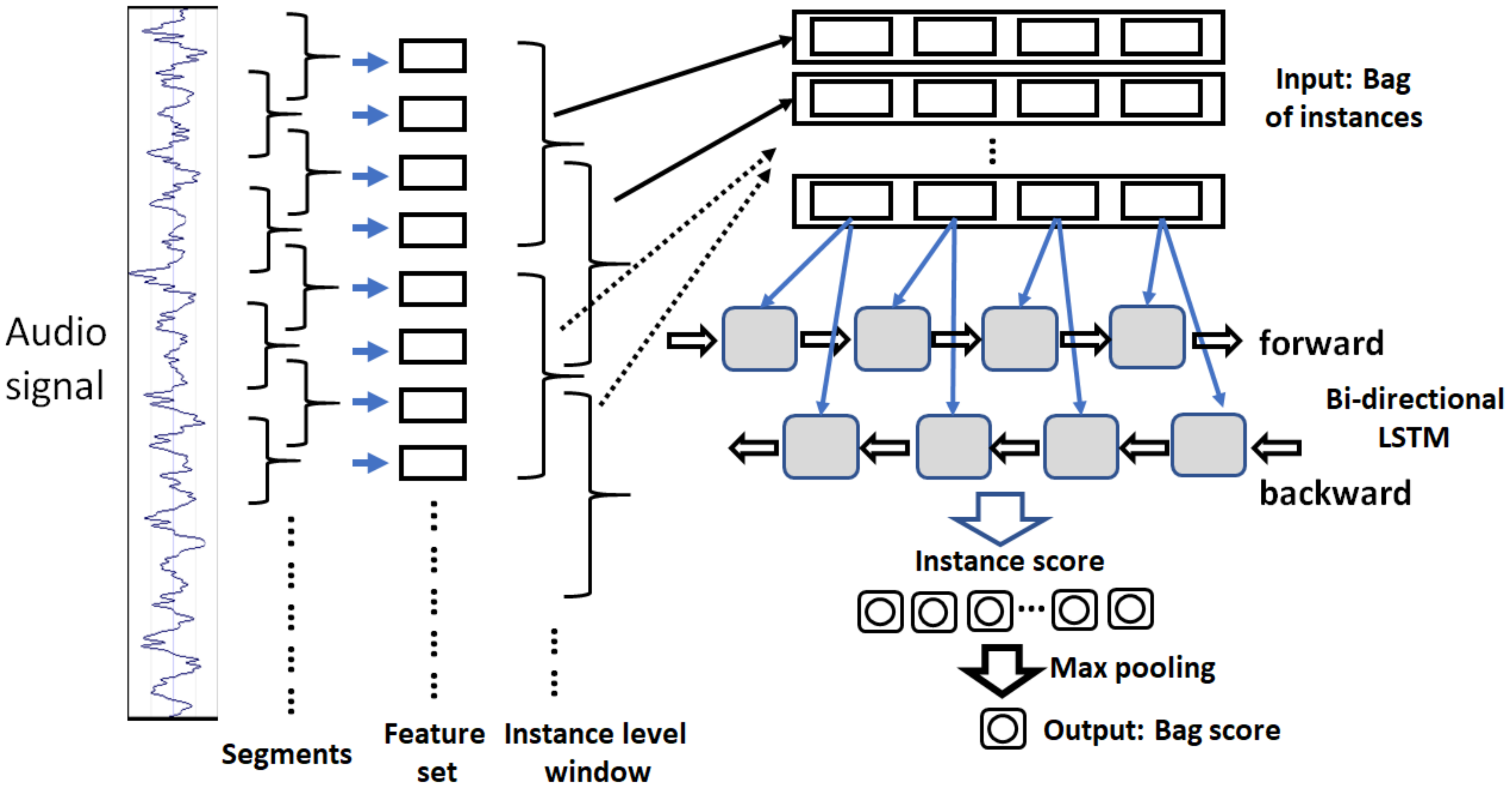




Classifier

- Prior Work (not RNNs)
  - SVM – MIL
  - DNN – MIL
  - Fail to account for temporal dynamics in speech segment

BLSTM - MIL

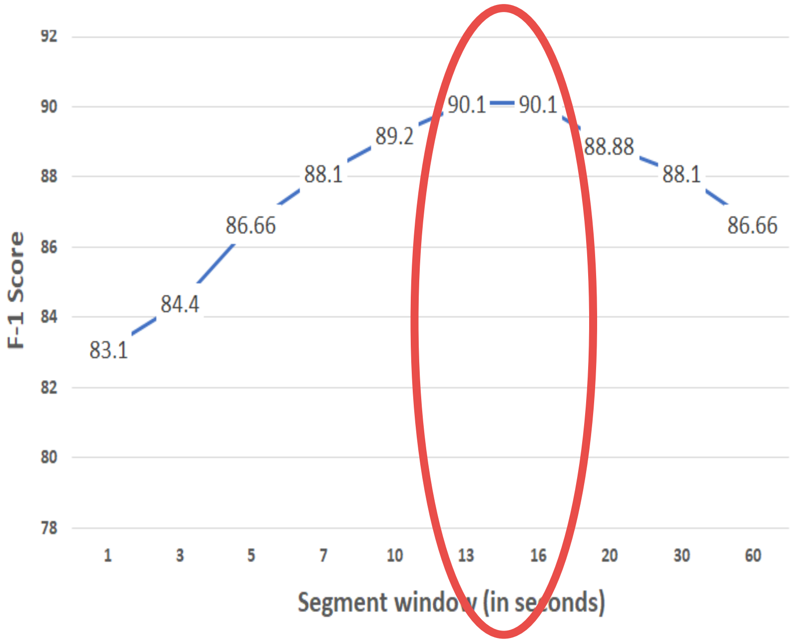

Audio signal
Segments
Feature set
Instance level window
Input: Bag of instances
forward
Bi-directional LSTM
backward
Instance score
Max pooling
Output: Bag score

Evaluation

Social Anxiety	Depression
• 105 Participants • Mean Age: 19.24, SD: 1.84 • Mean audio clip length: 3 minutes • Labeled by Licensed Clinical Psychologists • Social Interaction Anxiety Scale (SIAS) and Social Phobia Scale (SPS)	• Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ) • 142 participants • Mean audio clip length: 12 minutes

A. Salekin, J. Eberle, J. Glenn, B. Teachman, and J. Stankovic, A Weakly Supervised Learning Framework for Detecting Social Anxiety and Depression, *ACM Interactive, Mobile, Wearable, and Ubiquitous Technologies*, Vol. 2(2), Article 1, June 2018, 26 pages.

Social Anxiety



Evaluation on Length of Instance

Audio States	F-1 Score	Accuracy
500	77.2	78.9
1000	82.8	84.1
2000	87.9	89.1
2500	88.17	89.1
3000	89.13	90.1
3500	90.1	91
4000	90.1	91
4500	89.1	90
5000	88.17	89.1

Evaluation on Number of Audio States

Social Anxiety

Feature	F-1 Score	Accuracy
NN2Vec	90.1	90
Emo2vec	[illegible]	[illegible]
I-vector	74.7	79
Audio word	55.55	68
Raw features	56.82	62.4

15.4% higher

Feature Modeling Baselines
With BLSTM classifier

Algorithm	F-1 Score	Accuracy
BLSTM-MIL	90.1	90
DNN-MIL	[illegible]	88.11
mi-SVM	[illegible]	85
BLSTM	[illegible]	88.1
CNN-BLSTM	83.5	85.1
CNN	83.5	85.1
DNN	68.3	73

3.5% higher

Weakly Supervised Learning baselines

Supervised Learning baselines

Classifier Baselines
with NN2Vec features

15.4% improvement F-1 score compared to the best baseline in literature: *I-vector with BLSTM*

Summary

- Smart and Connected Health Based on Wearables and in-situ systems
- Towards Ambient Intelligence: Cognitive Assistance for Healthcare
- Key modality: Acoustics/Speech

Thanks

- Abu Mondol, now at Amazon
- Sirat Samyoun, at UVA
- Asif Salekin, now at Syracuse
- Meiyi Ma, now at Vanderbilt
- Sarah Preum, now at Dartmouth

Real Deployment Observations

- 70% non-research software development
- Significant effort at daily monitoring of deployed systems
- Publication lag
- Where/what to publish

Future Work

- Multiple emotions at the same time
- Rare emotions
  - Frustration versus anger
- Longer distances from microphone
- Use of linguistics content
- Anxiety with elderly
- Chronic stress

Future Work

- Greater and greater intelligence
- Better support for self-help, e.g., dealing with conflicts
- Privacy