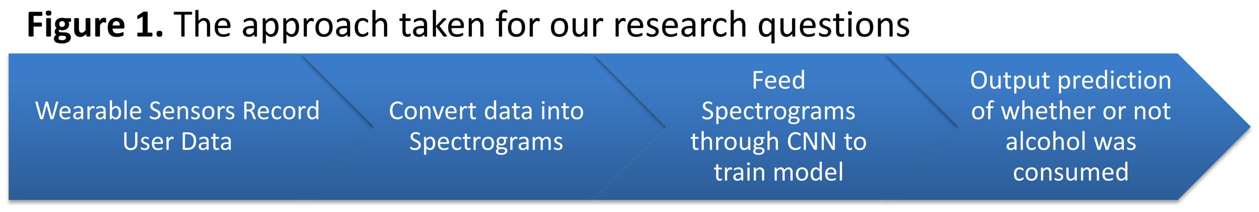
Using Deep Learning for Alcohol Consumption Recognition

Joseph Bernstein[1] , Brandon Mendez[2] , Peng Sun[3] , Yang Liu[3], Yi Shang[3]
UMBC[1] , University at Buffalo[2] , University of Missouri[3]

Introduction

- With machine learning, computers have proven effective at
  - Recognizing objects in images (99.7% accuracy)[1]
  - Translating audio to words and sentences
  - Recognizing musical chords
- Research Questions
  - Can alcohol be detected based off body signals such as
    - Heart Rate
    - Breath Rate
    - Skin Temperature
    - Activity
  - Can machine learning be used to solve the task?
- Our Approach
  - See **Figure 1**
- Impact
  - With an automatic alcohol detection system, self reporting is no longer necessary
    - Allows for better research that does not rely on self reporting
  - Help monitors addiction, potentially applicable to other substances
  - Proof of concept for using body signals and machine learning as prediction
  - Allows for an easy way to input an entire data set into machine learning techniques

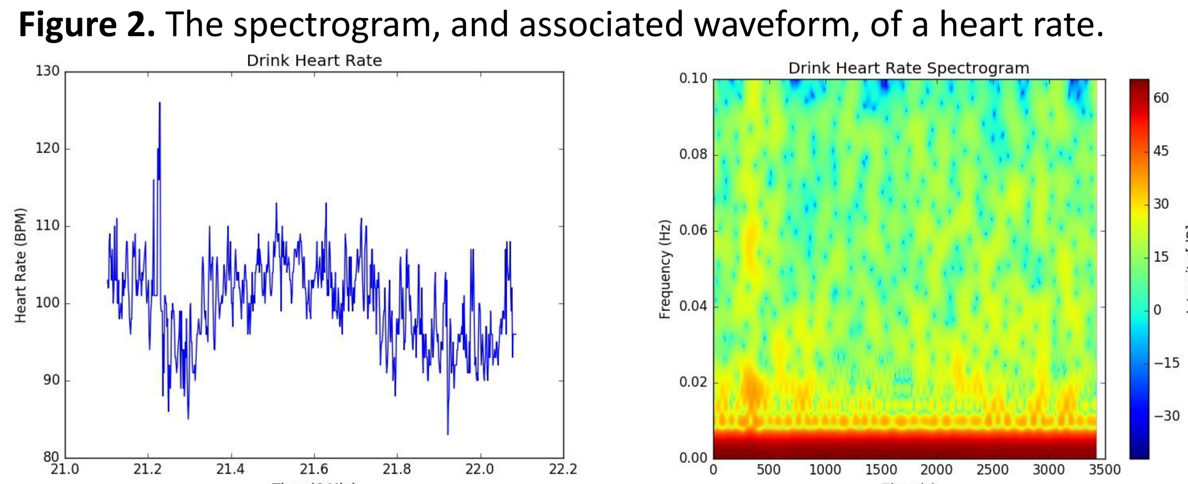
Figure 1. The approach taken for our research questions



Method and Technology

- Sensor Data Source
  - Two types of data, comes from ADA[2], survey and sensor data
    - Sensor Data , measures
      - Heart Rate
      - Skin Temperature
      - Breath Rate
      - Activity
    - Survey Data, measures
      - Points where alcohol is consumed
      - Number of drinks
- Spectrogram
  - Data is converted to a spectrogram (**Figure 2**)
  - Spectrogram generation is configurable, allowing for all parameters to be adjusted as necessary
  - Algorithm to generate spectrograms (**Figure 3**)
- Convolutional Neural Network (CNN)
  - Pretrained and finetuned pre-existing CNN, AlexNet (**Figure 4**)
    - Located on AWS EC2 Server w/ GPU access
    - Pretrained on ILSVRC 2012 data

Figure 2. The spectrogram, and associated waveform, of a heart rate.



Results

- Training Set
  - 80 images, parameters found in **Figure 5**
    - 40 positive (drink)
    - 40 negative (no-drink)
- Testing Set
  - 20 images
    - 10 positive (drink)
    - 10 negative (no-drink)
- Performance Metrics
  - Training Loss, **Figure 6a**
    - Descriptor of how effective the model is at learning features of the spectrograms. Lower loss is better.
    - Our training loss tended to 0, indicative of learning many features
  - Training Accuracy, **Figure 6b**
    - Descriptor of the prediction rate of the model based off the training dataset
    - Our training accuracy tends to 100%, implying that the model is over fit to the data.
  - Testing Accuracy, **Figure 7**
    - Descriptor of the prediction rate for data the model has never seen. Tested after the model is finished training.
    - **Overall accuracy of 75% (15/20 images)**
  - Feature vectors (**Figure 8**)
    - Features that are picked up from spectrogram set
    - Useful in determining where the model believes the significant information is.

Figure 5. The spectrogram parameters used for the spectrograms that were used for the listed results.

Spectrogram Data	
NFFT	50
Fs	0.2
Noverlap	49
Pad To	256
Number of Data Points	721

Figure 3. The algorithm used to generate the spectrograms used.

The 'Drinking' Spectrogram Algorithm

```
GET data FROM excelSheet
PUT data INTO table
DROP EACH row IN table WHERE there is no instance of drinking
FOR EACH row IN table of drinking instances
    GET all points IN original dataset
    REMOVE all in dataset EXCEPT those within 60 minutes of the drink point.
    PLOT points
```

The 'No Drinking' Spectrogram

```
GET data FROM excelSheet
PUT data INTO table
DROP FROM table WHERE row was used for Drinking Spectrograms
WHILE(number of 'No Drink' Spectrograms ≠ number of 'Drink' Spectrograms)
    GET random date/time, select random 60 minute time window
    IF (random date/time NOT missing data)
        PLOT points
```

Figure 4. AlexNet is a pretrained CNN. "Pretrained" means it has already been fed hundreds of thousands of data points. As such, it takes less time to train, as well as fewer example images.

Figure 6. The confusion matrix that breaks down the predictions of the CNN, to see how dispersed the predictions are. 75% accuracy.

	Drink	No-Drink	Total
Drink	9	1	10
No-Drink	4	6	10
Total	13	7	20

Figure 5. Each square is a feature the computer has read from various spectrograms.

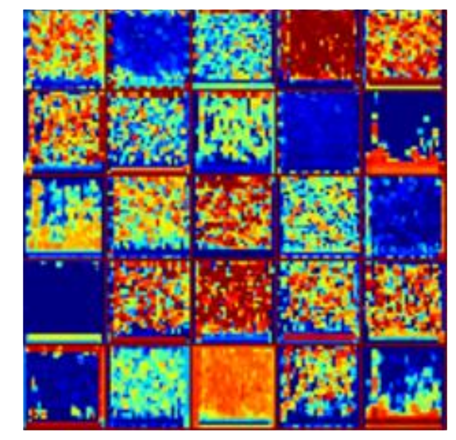


Figure 7a. The accuracy of the model as it trains on the given images. There is a total of 500 iterations through the images where the CNN tries to learn new details.

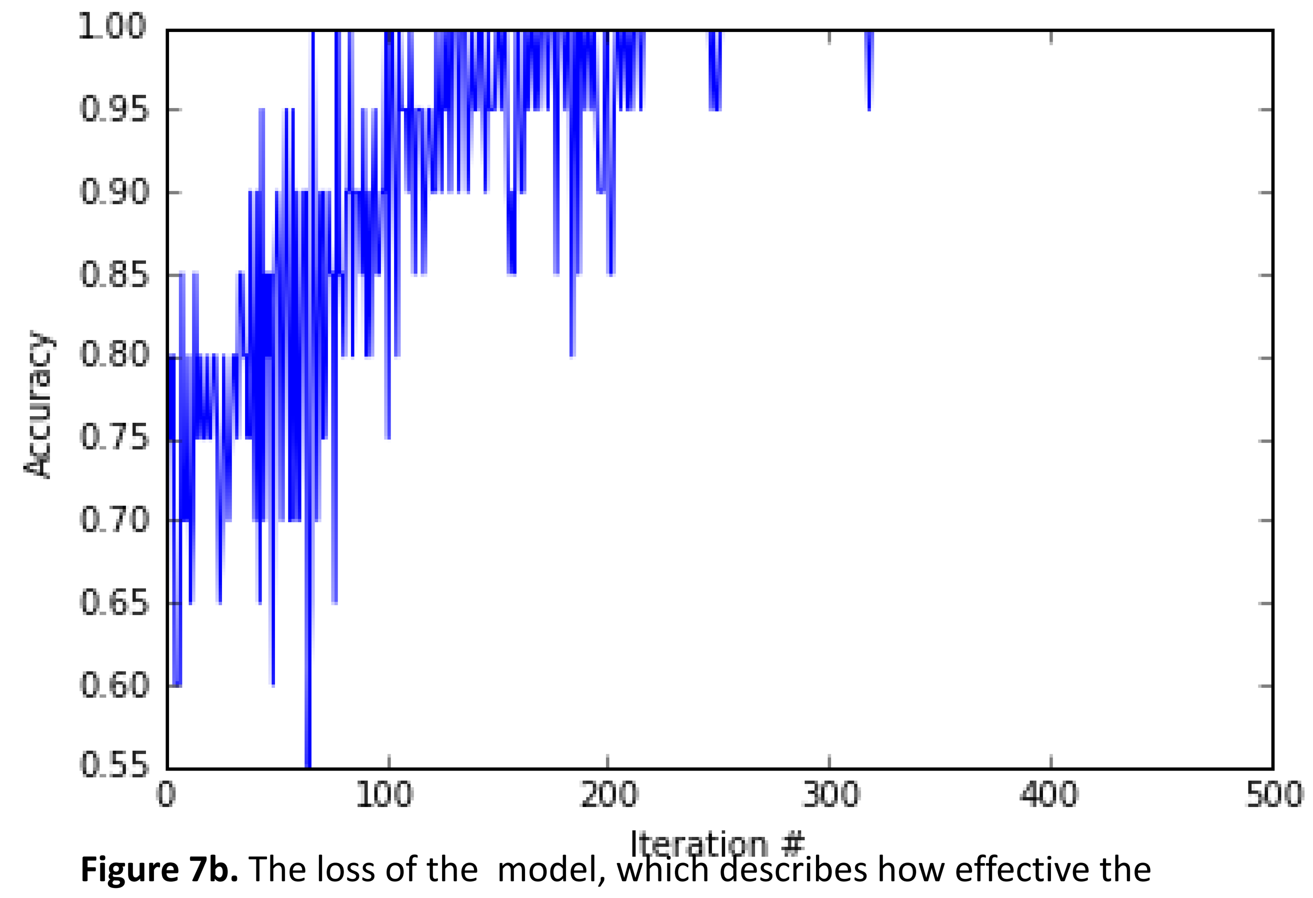
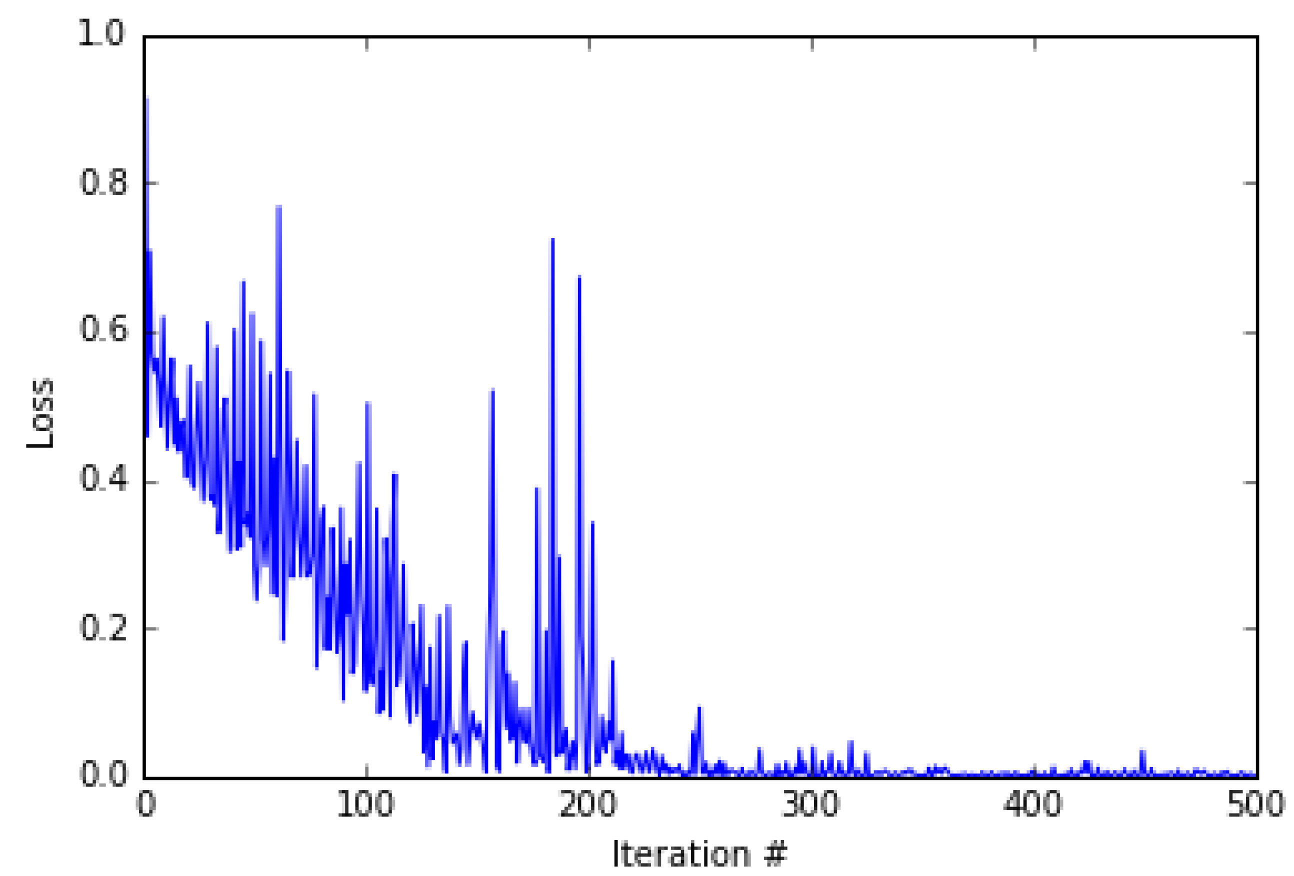


Figure 7b. The loss of the model, which describes how effective the computer is at picking up new features. A lower loss indicates that the computer is learning from the images.



Conclusions and Looking Forward

- Successfully created deep learning pipeline to take in sensor information and predict whether or not alcohol was consumed
- Created configurable system to allow various parameters to be changed, so that different input can be tested to find optimal results
- Achieved prediction accuracy of 75% which outperforms random guessing
- Looking forward, create more spectrogram options such as
  - Change scaling from linear to log based
  - Zoom in on certain frequency bands
- Search for which frequency bands have the best results
- Combine predictions to incorporate different signals at one time

Acknowledgements

We would like to give thanks to the NSF for funding all of the costs to stay in Mizzou for our research.

References

1. J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, Miami, FL, 2009, pp. 248-255.
2. P. Sun, N. M. Wergeles, C. Zhang, L. M. Guerdan, T. Trull and Y. Shang, "ADA - Automatic Detection of Alcohol Usage for Mobile Ambulatory Assessment," *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, St. Louis, MO, 2016, pp. 1-5.
3. "3 Convolutional neural network (LeNet) - shahriarinia," in Mshahriarinia, 2012. [Online]. Available: http://www.mshahriarinia.com/home/ai/machine-learning/neural-networks/deep-learning/theano-mnist/3-convolutional-neural-network-lenet. Accessed: Jul. 22, 2016.
4. T. N. Sainath, A. r. Mohamed, B. Kingsbury and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 8614-8618.