Using Deep Learning for Alcohol Consumption Recognition

Joseph P. Bernstein
Dept. of Computer Science
UMBC
Baltimore, Maryland, United States

Brandon J. Mendez
Dept. of Computer Science
University at Buffalo
Buffalo, New York, United States

Peng Sun, Yang Liu, Yi Shang
Dept. of Computer Science
University of Missouri
Columbia, Missouri, United States

***Abstract*—As Convolutional Neural Networks continue to produce state of the art results, more types of data are being used to see the results that would be produced. Using the heart rate data that was collected using sensors from various subjects who consumed alcohol, we converted it from the 1D waveform into a set of spectrograms. The spectrograms were fed into two pretrained CNNs, CaffeNet and AlexNet, to determine whether or not the given spectrogram was an instance of alcohol consumption. Using 80 training images (40 positive, 40 negative) and 20 test images (10 positive, 10 negative), we achieved a test accuracy, after adjusting learning rate, number of iterations, and gradient descent algorithm, as well as the time window and coloration of the spectrograms, of 72% (n=20, 5 trials), which are promising results for non-audio waveforms.**

Keywords—Convolutional Neural Network; Spectrogram Identification; Machine Learning; Body Sensor; Alcohol Consumption; Alcohol Assessment

I. INTRODUCTION

Mobile systems based on wireless wearable sensors have been actively developed for a variety of applications in mobile health and physiological monitoring. They are capable of continuously collecting biosensor and self-report data to assess or predict physical and psychological conditions, such as alcohol consumption, in daily life. Automatically identifying patterns of interests based on various physiological signals and survey results for each individual remains a challenge.

In recent years, deep learning with Convolutional Neural Networks (CNNs) has proven to be very effective at various perception tasks. On image recognition and object detection, CNN has produced state of the art results, reaching as high as a 99.7% accuracy on the MNIST dataset [1]. The CNN has also achieved state of the art accuracy of 73.2% on object detection with PASCAL VOC 2007 [8]. This technology has also been applied to various 1-D waveforms, such as translating audio to words [2], recognizing musical chords [3] and other attributes to music [4], and even finding patterns in human interactions [5]. However, to our knowledge, no efforts have been made in trying to identify, using deep learning, certain attributes about one's current state of wellbeing based off of body sensor data, such as heart rate.

Based on sensor data collected from a newly developed mobile ambulatory assessment system for automatic detection of alcohol usage and craving, ADA[6], we propose to use a deep learning approach to predict whether or not someone has consumed alcohol based off their heart rate, by training and testing a CNN to create an effective classification model. The approach can be easily extended to other physiological sensor data, such as breathing rate, skin temperature, movement, etc. Furthermore, the approach allows for the various parameters of how the data is fed into the CNN, as well as the CNN parameters itself, to be configured based off user input, so more research can be conducted in this field without starting from scratch. Thus, parameters can be adapted to achieve the best results for a given type of data.

An accurate prediction model for alcohol consumption would be very useful and open avenues into research where self-reporting of alcohol consumption would no longer be necessary, giving more accurate prediction results. Furthermore, it acts as proof that waveforms outside of audio can be identified by CNNs. The findings in this paper are only the beginning of this area of research that will continue to expand.

This paper is organized as follows. Section II describes the technical method used, which combines the signal process method and deep learning method used for data analysis. Section III presents the findings of the methods used. Lastly, Section IV concludes the paper.

II. METHOD

The pipeline, as a whole, (**Fig. 1)** is designed to collect data from sensors, and determine whether or not the user has consumed alcohol. There are many different interchangable parts that have been combined to develop a configurable system that allows for parameters to be changed and tested as needed. The system was set up in this configurable manner so that various components of the pipeline can be changed and tweaked as needed to find the optimal parameters to be used for different types of input data.

A. The Sensory and Survey Data

The data from the ADA[5] was split up into 21 excel sheets, where each sheet represents one test subject. The sensors used, the Equivital EQ2 and Hexoskin smart shirt, collected the sensory data. The data collected was heart rate, breath rate, skin temperature, and activity level, at a frequency of one recording every 5 seconds. For this research, the only data used was the heart rate.

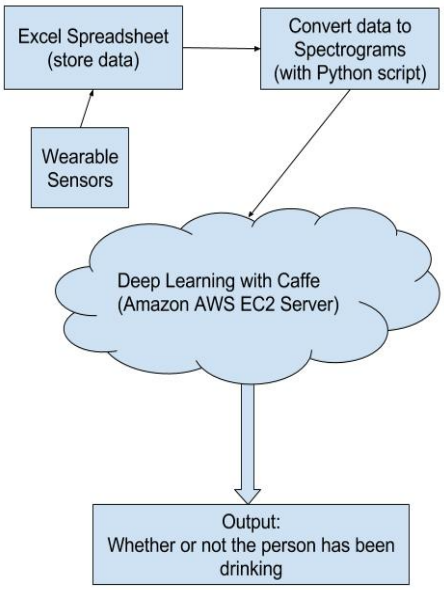

Fig. 1. The pipeline, from start to end, which goes through the processing path of the data to determine the output.

The next component of the data entered was the survey based data. Every time the user consumes a drink of alcohol, they fill out a survey which then indicates in the data when the drink was consumed. The data is set up so that each row has the associated date and time, followed by the sensor data and survey data. While the survey also records the mood the user is in, the only data used for this research was the time, date, heart rate, and instances of alcohol consumption.

B. The Spectrograms

The spectrograms were generated in python, using the Pandas library for data storage and manipulation, and Matplotlib library for the graphing of the data points. The parameters used for generating the spectrogram can be found in **Table 1.**

TABLE I. SPECTROGRAM PARAMETERS

Parameter	Value
NFFT	50
Fs	0.2
Pad To	MAX(256, $2^{\log_2 n}$), n = number of data points
Data Points (30 Minute Window)	361
Data Points (60 Minute Window)	721

For generating the drink spectrogram dataset, the script queried for each record of the drinking instances, and stored them separately. Then, going backwards through the untouched dataset, it went to each drink point, and removed that instance, as well as the window above or below it (either 15 minutes or 30 minutes in both directions, though could be set to any amount of time, in minutes), which resulted in the dataset removed of all the drink instances and their surrounding 15 or 30 minute time windows. The data was then plotted, and a spectrogram was created for each instance. After the drink spectrograms were created, the script then randomly selected data points, and, if there is a 15 or 30 minute window surrounding the random point without any data removed or missing data (missing data could be from the drink spectrogram data selection, or simply missing data from the sensors), then it would create a no-drink spectrogram based off the selected data (**Fig. 2**).

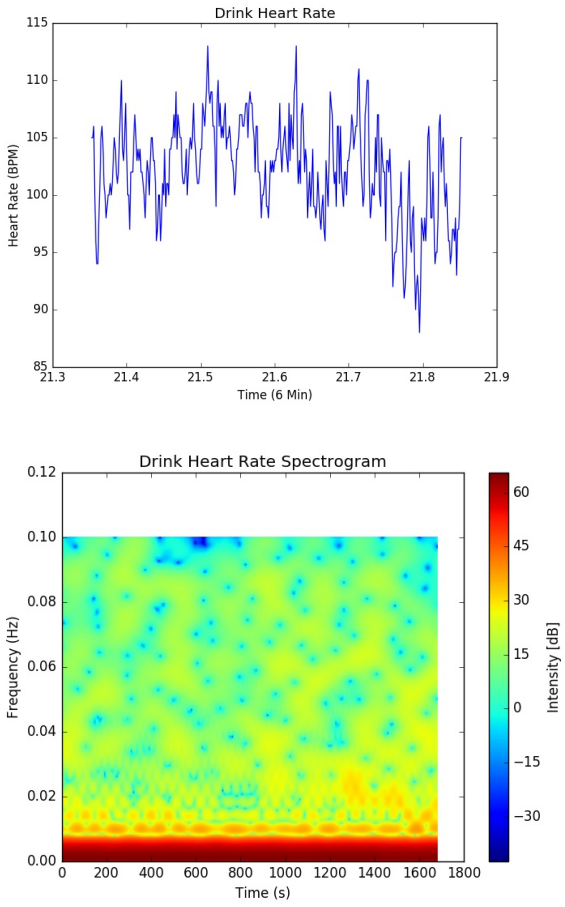

Fig. 2. One of the drink spectrograms an its associated waveform generated over a 30-minute window.

This algorithm operates with two O(n) operations in the drink spectrogram generation, and then, due to the danger with using random numbers which may dramatically increase run time, every unsuccessful search for a complete dataset of no-drink spectrograms decrements the total number of data points required to create a spectrogram. In other words, if the 30 minute window that requires 361 points cannot find a data range with 361 points, then it will search for 360 (and continue to decrease down until the proper size is found). That way, in the event that there are no more data sets the same size as the drink spectrograms, slightly (or, if necessary, much smaller) sizes can be used instead.

The NFFT is the number of points for to perform the FFT. It provides the size of the window that the FFT will use when transforming the data points into the spectrogram. The Frequency Sampling, Fs, is based off of the calculation of the

data. Because the sample rate of the sensors collected information once every five seconds, the Fs is set to 0.2 Hertz. The Noverlap, or number of overlap, is how many datapoints can be overlapped between frequency bins when generating the spectrogram. By setting Noverlap = NFFT – 1, it creates more frequency bins for the spectrogram, which allows for more detail to be generated in the spectrogram. The Pad To formula was suggested by MATLAB and Matplotlib, and was used accordingly. The data points for the 30- and 60-minute window was calculated by going forward and then backward in time 15 and 30 minutes respectively from the drinking instance.

These parameters are in no way "optimal," and it is very possible that altering them will result in higher testing accuracy for the CNN. The parameters used were successful in producing promising results and acts as a sound proof of concept.

C. *The Deep Learning*

The selection of Caffe for the deep learning component was based off of its preexisting pretrained CNNs. The two we used, CaffeNet and AlexNet, were very similar. **Fig 3**[7] shows the two CNNs, and the differences can be found in their normalizing and pooling layers' ordering.

We opted for a pretrained CNN and fine-tuned our dataset due to the limited amount of training and testing data available from the subjects. Both CaffeNet and AlexNet were pretrained on ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 data. The use of the pretrained CNN would allow for the preexisting weights to be adjusted faster, and result in higher accuracy despite the small data size. Caffe was run on an AWS EC2 Ubuntu server with access to a GPU.

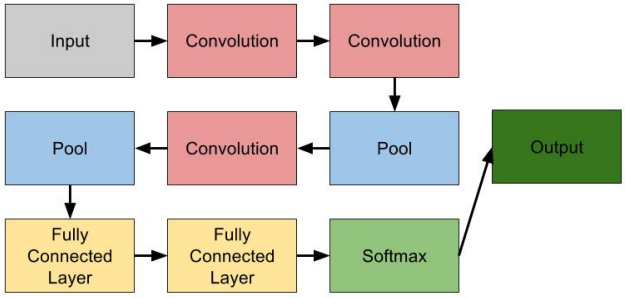

Fig. 3. A diagram of AlexNet. The difference between AlexNet and CaffeNet is the ordering of their Normalizing and Pooling layers.

The specific parameters that were adjusted in CaffeNet and AlexNet can be found in **Table 2**.

TABLE II. ALEXNET AND CAFFENET PARAMETERS

Parameter	Value
Number of Iterations	500
Base Learning Rate	0.2
Gradient Descent	MAX(256, $2^{\log_2 n}$), n = number of data points

The number of iterations was set to 500, lower than the predefined AlexNet and CaffeNet, due to system constraints- in order to test the system as well as run through simulations, the number could not be too large. By decreasing the number of iterations to 500, promising results were still attainable. Futhermore, for AlexNet, the test loss and test accuracy curves still converge despite the limitation. For the base learning rate, due to the small number of test images, the initially defined learning rate of 0.001 was too fast, and no features could be picked up. However, by setting it to 0.0001, the results ended up improving dramatically. Likewise, the gradient descent changed from stochastic gradient descent to Adam's Algorithm also produced superior results.

III. RESULTS

After setting the parameters in the pipeline, the testing and training data was classified based off of the total available spectrograms. In this experiment, there was 80 training spectrograms with 40 drink spectrograms and 40 no-drink spectrograms. The remaining 20 spectrograms were used for testing, with 10 drink and 10 no-drink spectrograms. Due to the small dataset, each CNN was trained 4 times, and the average of the 4 testing data accuracies is what will be reported during the discussion of the results for all the CNNs. The different parameters that were tested were AlexNet versus CaffeNet, color spectrograms versus gray scale spectrograms, and a 60-minute window versus 30-minute window. The initial results (**Table 3**) show that AlexNet, as a whole, outperforms CaffeNet.

TABLE III. MEAN (N = 5) OF TESTING ACCURACY OF THE CNNS

CNN	Window (Minutes)	Color	Gray Scale
AlexNet	30	61%	63%
	60	72%	59%
CaffeNet	30	56%	57%
	60	62%	60%

The cause of AlexNet's superior results comes from the difference in the order of the normalizing and pooling layer. CaffeNet has a training loss and training accuracy that is more erratic and does not converge, unlike that of AlexNet's (**Fig. 4**). As a result, CaffeNet is not as effective at learning features, as some of the feature vectors **(Fig. 5)** are blank and do not learn any characteristics as a result. The one area that CaffeNet is superior is the 60 Minute gray scale which may be due to not having enough runs through the testing accuracy, rather than the CNN itself.

Anothing finding is that every 60-minute instance outperforms the 30-minute performance except the AlexNet gray scale, which is indicative again that the test results for this were, by luck, low. However, it makes sense, otherwise, for the 60-minute windows to outperform the 30-minute windows, as processing alcohol through the body is a gradual processes, so a larger time window can pick up more attributes.

The AlexNet colored 60-minute trials significantly outperforms the gray scale counterpart, as well as all the other results. The cause for this may be attributed to the colored spectrogram having more details to be read, and more learning

can take place due to the contrast in colors as well as spread of time. Overall, despite the small number of run throughs for the system, the results are promising for many reasons.

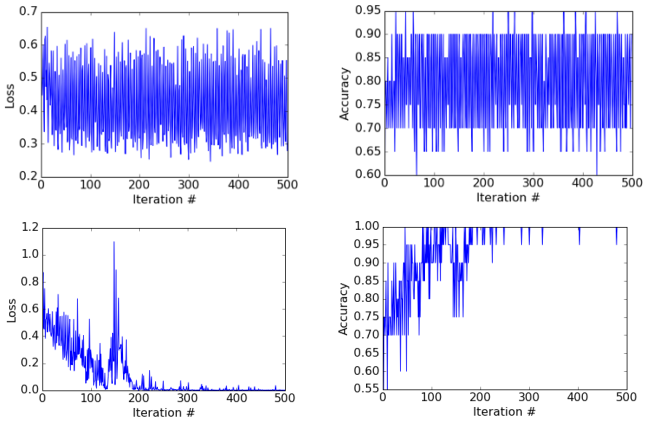

Fig. 4. The training loss and training accuracy of the 60 minute color spectrograms from CaffeNet (top) and AlexNet (bottom)

The color 60-minute windows all are superior to the respective gray scale, however, the 30-minute window gray scales slightly outperform the respective colored spectrograms. However, these numbers, except for the AlexNet 60 minute color, are such a small difference, that no conclusions can be drawn. However, because the 60-minute colored AlexNet perfomed significatnly better than any of the other configurations, it is very likely that results can be improved based off of tweaking different aspects of the input data and CNN.

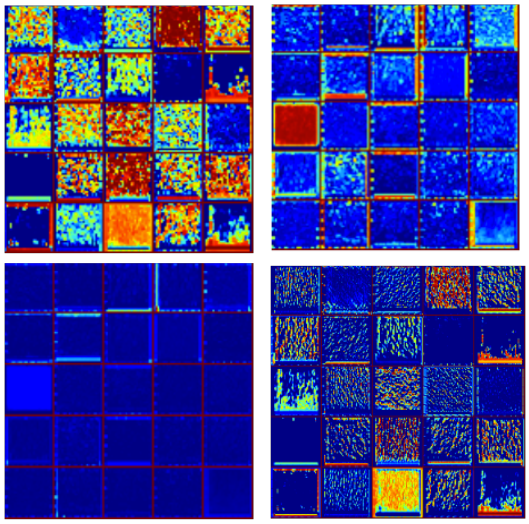

Fig. 5. The feature vectors for the first pooling and normalizing layers of CaffeNet (top) and AlexNet (bottom).

Furthermore, all but one trial outperformed random guessing (50%) accuracy. Of the 40 trials across both CNNs, the only one that did not was the AlexNet 30-minute Colored spectrogram trial with a test accuracy of 45% (all other four trials were 65%). While one could argue that the 60% results were luck, then it would also indicate there should be more 40% through 50% results, which did not happen. Finally, the spectrograms predicted 75% (15/20) images properly in 3/5 of the trials for the 60 minute AlexNet color spectrogram trial.

IV. Conclusion

The primary goal was to predict the occurrence of alcohol consumption which, overall, had promising results. The results attained shows that CNNs can be used for identification of alcohol based off of heart rate. There are more body signals to be used, such as activity, skin temperature, and breath rate, which can be used to not only improve accuracy of alcohol detection, but also indicate other states the body may be in, such as unconsciousness, or under the influence of other drugs.

Not only does it give hope for future progress in using body sensors for detection of substances in the body, it also shows the wide ranged use of the CNN in identification and classification. Furthermore, it opens up this avenue of research to test other types of sensor data, as well as wave forms in general outside of speech and audio. These results were achieved with spectrograms that had only minor differences. Different spectrograms that focus on different areas of frequency, scaling, and other time windows may achieve even better results; these results are just preliminary and only the beginning.

The configurable nature of the pipeline will allow for new data to be tested, and find what parameters achieve superior results. One final point to make is that due to the separate nature of the data, the spectrogram generation, and the CNN, each piece can be replaced as long as the input and output is the same, allowing for components to be changed as necessary future technology progresses. Overall, the findings in this paper are only the beginning of what can be discovered about the body based off of various sensors, using deep learning.

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. Swietojanski, A. Ghoshal, and S. Renals, "Convolutional neural networks for distant speech recognition," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1120–1124, Sep. 2014. [Online]. Available: http://dx.doi.org/10.1109/lsp.2014.2325781. Accessed: Jul. 24, 2016.

[3] E. J. Humphrey and J. P. Bello, "Rethinking Automatic Chord Recognition with Convolutional Neural Networks," *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, Boca Raton, FL, 2012, pp. 357-362.

[4] J. Schlüter and S. Böck, "Improved musical onset detection with Convolutional Neural Networks," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 6979-6983.

[5] E. P. Ijjina and K. M. Chalavadi, "Human action recognition using genetic algorithms and convolutional neural networks," *Pattern Recognition*, Jan. 2016.

[6] 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.

[7] Bibliography: [1] H. Kataoka, K. Iwata, and Y. Satoh, "CaffeNet vs VGGNet," 2015. [Online]. Available: http://arxiv.org/abs/1509.07627. Accessed: Jul. 24, 2016.

[8] Shaoqing Ren and Kaiming He and Ross Girshick and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems (NIPS), 2015.