Neural Network for Detecting Head Impacts from Kinematic Data

Michael Fanton, Alissa Ling, Nicholas Gaudio

CS 229 Project Milestone Update

Contributions:

Nicholas Gaudio: Lead the team understanding of Tensorflow and Keras. He further preprocessed the data in Python so that it was properly formatted to be inputted into the networks. He lead the Net_2 investigation.

Michael Fanton: Developed Net_1, analyzed performance of classifiers, and helped compile dataset. Helped write all sections.

Alissa Ling: Performed statistical analysis and preprocessed data in matlab. Wrote the first draft of sections.

Motivation:

Mild traumatic brain injury (mTBI), more commonly known as concussion, has become a serious health concern with recent increase in media coverage on the long term health issues of professional athletes and military personnel. Acute symptoms include dizziness, confusion, and personality changes which can remain for days or even years after injury [1]. Further, recent studies have shown that repetitive mTBI can lead to long-term neurodegeneration and increase the risk of diseases such as Alzheimer's and Chronic Traumatic Encephalopathy [2]. Although the mechanisms of this injury are not well understood, studies have shown that one of the biggest risk factors for mTBI is a history of prior mTBI [8]. Further, concussion symptoms are more severe with a longer recovery time if an individual does not rest after injury [9]. Therefore, it is imperative that individuals who have been suspected to have received a TBI be immediately removed from risky situations.

According to the CDC, contact sports such as football are one of the leading causes of mTBI. In these sports, mTBI is diagnosed by a sideline clinician through subjective evaluation of symptoms and neurological testing. Because of the large variance of symptoms within different individuals, and the pressure of athletes to return to play, mTBI can often be missed by these tests [11]. In efforts towards developing an objective diagnostic tool for concussion prevention, Professor David Camarillo's lab at Stanford University created an instrumented mouthguard that rigidly connects to the upper dentition to record the linear acceleration and angular velocity of head impacts in six degrees of freedom. Whenever the linear accelerometer measures a signal of over 10g of acceleration, the device will trigger and record 100 ms of impact data. However, one of the primary challenges of this device is that it is prone to false positives, with non-impact events such as chewing, spitting, or dropping the mouthguard often triggering the device. In order for this device have promise to be used as a diagnostic tool in the future, it must be able to accurately classify between real impacts and false positives. Currently, this is done after the game; a research assistant will tediously watch hours of video footage time synced with the mouthguard, and each impact is manually labeled as a real impact or false positive. However, a machine learning classifier should be able to automatically differentiate between real and false impacts to a high degree of accuracy, as the kinematic data between these two impact types typically look distinct, as shown in Figure 1 in both the time and frequency domain.

In a previous study, the Camarillo Lab developed an impact classifier using a sequential feature selection to determine the most important classifier features (e.g. time domain features, power spectral density features, etc.), and used these features to train a support vector machine [3,4]. In this project, our goal is to instead train a neural net to do this, which will automatically extract the features it deems relevant. Currently, due to the lengthy process of manually labeling the dataset, we only have 527 examples, half of which are true impacts and the other half are false impacts. However, as the Camarillo Lab begins to disseminate their device around the country to different colleges and high schools, it is expected that they will soon have a significantly larger dataset of real and false positive impacts to work with, and a neural net should perform better than SVM in predicting a non-labeled impact or nonimpact on a large dataset.

Methods:

Our data set is 527 examples of which half are labeled impact and the other half are labeled non-impact. We split our dataset up into 70% for training, and 15% for both evaluation and testing. Each example has dimension 199x6 comprised of 6 time traces of length 199. The six time traces are the linear acceleration at the head center of gravity in the x, y, and z axes, and angular velocity of the head in the x, y, and z anatomical planes. The data is sampled with a time step of 1000 Hz, with 100 ms recorded pre-trigger and 100 ms post-trigger for 299 data points. One representative impact and non-impact each were plotted to show the linear acceleration and angular velocity in the x,y, and z direction and the magnitude vs. time, shown in Figure 1a,b. Power spectral density estimates were also plotted for impact and non-impact to show the frequency of impact vs. non-impact, shown in Figure 1c, d. As shown by Figure 1, true impacts generally are comprised of lower frequency content, while false impacts have much higher frequency content. Intuitively, this makes sense, as biting or dropping the mouthguard would likely result in a high frequency noisy signal, while football head impacts typically have frequency content in the 20-30 Hz range [10].

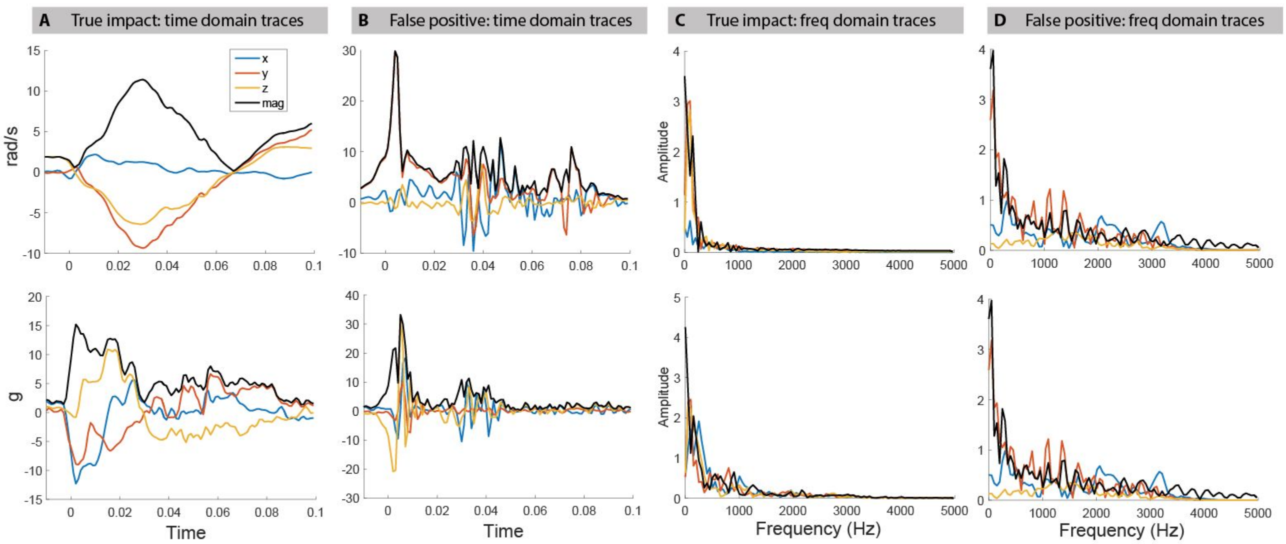

Figure 1: ***Example time and frequency domain traces from true and false impacts.*** *A single representative true and false impact time trace is plotted. Real impacts tend to be comprised of lower frequency signals, while false positive impacts (due to biting, spitting, chewing, etc.) tend to have higher frequency content.*

As the goal of this project is to implement a neural net to classify data, our focus thus far has been getting familiar with Tensorflow and Keras to implement simple neural nets on our data. As a starting point, we focused on a sequential model to be used as a "baseline" test because the sequential model is a simple and standard first approach. In developing these simple neural nets, we make no assumptions about the features and have every input affect every output. As an input, we feed in the six time traces vertically stacked. This takes advantage of the topology of the input data; for a real impact, we would expect an increase in signal around the same time in all six time traces. Further, as a starting point, we experimented with convolutional neural nets, as CNNs generally work well to classify data with spatial (in our case temporally spaced) relationships. Indeed CNNs are commonly used for classifying accelerometer motion data, with input data consisting of stacked channels of the sensor signals [5,6].

Our first implemented neural net, "Net_1" is comprised of a single 1D convolutional layer with 10 filters with a kernel size of 3 with ReLU activation, followed by a dropout layer with a cutoff of 0.25, and a dense output layer with a sigmoid activation function to classify between impact and non-impact. We also developed a more complex neural net with 2D convolutional layers. The second model, "Net_2" has a 2D convolutional input with 32 filters and a kernel size of (3,3), followed by a ReLU activation function. This feeds into a second 2D convolutional layer, with 32 filters an a kernel size of (1,1), followed by a ReLU activation function. This feeds into two consecutive 2D convolutional layer with 64 filters size (1,1) and a ReLU activation, followed by a 2D max pooling operation of size (2,2) and a dropout with threshold of 0.25. This is fed into a dense layer of size 256, followed by a dropout operation with 0.5 threshold, and another dense layer of size 2. Both models were compiled with binary cross entropy loss and the adam optimizer. As of now, hyperparameters and layers for these models were chosen somewhat arbitrarily, with the primary goal being getting a qualitative “feel” for how neural nets are implemented and how net properties affect training and test accuracy. Knowing that we might have to employ even deeper network architectures, we are currently looking into recursive model architectures (not to be confused with recurrent neural networks which we are also investigating) and denser networks such as the DenseNet or U-Net as architectural starting points [5,6].

Preliminary experiments:

The performance of the two preliminary neural net classifiers is summarized in Figure 2, as shown by the receiver operating characteristic (ROC) curves. The ROC curve plots the true positive rate against the false positive rate at different threshold values for a binary classifier; a perfect classifier has an area under the curve (AUC) of 1 while random guessing has an AUC of 0.5. Net_1 had an AUC of 0.929, while Net_2 had an AUC of 0.946. Net_1 had a training accuracy of 82.1%, and a test accuracy of 78.8%. Net_2 had a training accuracy of 83.4%, and a test accuracy of 88.1%.

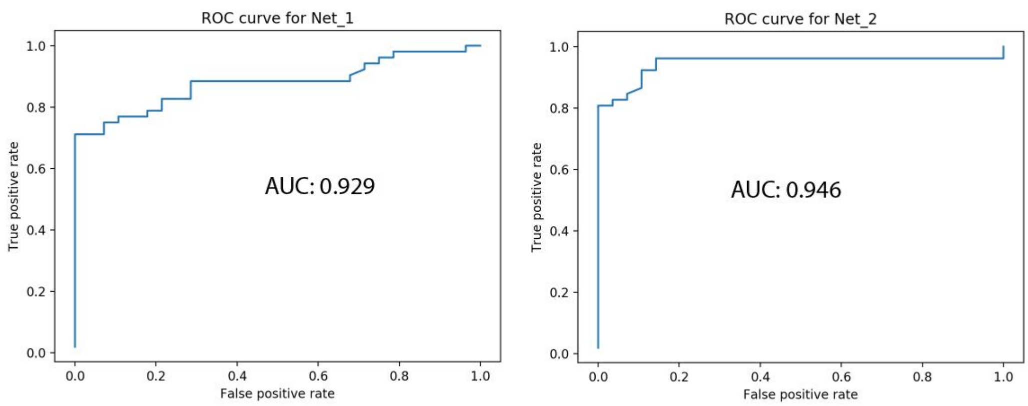

Figure 2: ***ROC curves for Net_1 and Net_2.***

Next Steps:

Moving forward, we plan to gain a more quantitative understanding of how neural net layer topology and hyper parameter choices will affect training and test accuracy. As an immediate next step, we want to develop a net that will fit the training data as well as possible, likely overfitting the data. Further, we plan to implement the Net described by Kasnesis P., et al. [7], we will try making a deep convolutional neural network that applies a late 2D convolution to multimodal time-series sensor data. After analyzing the sequential neural net to see how the data breaks down hierarchically, we will vary and optimize the number of 2D convolutional layers, the number of filters per 2D convolution, the filter size, and the recurrent architecture of our network. Adding more 2D convolutional layers we give us a higher level of specific features we see that are emphasized by the model. We will also vary the epoch number and the batch number, or simply stop training upon convergence. By changing the number of runs and iterations, the neural net will learn more or less about the data. Finally, a recurrent neural network and hold-out cross validation will be implemented. For the recurrent neural network, we will use the feedback to reinforce different weights. We will implement feedback into loop for the features we want to emphasize will. To ensure that the samples are representative of the entire field, we will use hold-out cross validation will validate our assumption that the data came from the same distribution. As a final goal, we hope to optimize the neural net to have 97% test accuracy.

References

1. Langlois, Jean A., Wesley Rutland-Brown, and Marlena M. Wald. "The epidemiology and impact of traumatic brain injury: a brief overview." *The Journal of head trauma rehabilitation*21.5 (2006): 375-378.
2. Ramos-Cejudo, Jaime, et al. "Traumatic Brain Injury and Alzheimer's Disease: The Cerebrovascular Link." *EBioMedicine* (2018).
3. Wu, Lyndia C., et al. "A head impact detection system using SVM classification and proximity sensing in an instrumented mouthguard." *IEEE Transactions on Biomedical Engineering*61.11 (2014): 2659-2668.
4. Wu, Lyndia C., et al. "Detection of American football head impacts using biomechanical features and support vector machine classification." *Scientific reports* 8.1 (2017): 855.
5. Huang, Gao, et al. "Densely connected convolutional networks." *CVPR*. Vol. 1. No. 2. 2017.
6. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
7. Kasnesis, Panagiotis, Charalampos Z. Patrikakis, and Iakovos S. Venieris. "PerceptionNet: A Deep Convolutional Neural Network for Late Sensor Fusion." *Proceedings of SAI Intelligent Systems Conference*. Springer, Cham, 2018.
8. Lasry, Oliver, et al. "Epidemiology of recurrent traumatic brain injury in the general population: A systematic review." *Neurology* 89.21 (2017): 2198-2209.
9. Williams, Richelle M., et al. "Concussion recovery time among high school and collegiate athletes: a systematic review and meta-analysis." *Sports medicine* 45.6 (2015): 893-903.
10. Laksari, Kaveh, et al. "Mechanistic Insights into Human Brain Impact Dynamics through Modal Analysis." *Physical review letters* 120.13 (2018): 138101.
11. Albicini, Michelle, and Audrey McKinlay. "A review of sideline assessment measures for identifying sports-related concussion." Journal of concussion 2 (2018): 2059700218784826.