Categorizing Data Acquisition System for Head Response (DASHR) Head Impact Data with Deep Learning

Duke PRATT SCHOOL of ENGINEERING
Duke Institute for Brain Sciences

Sarah Glomski, Patrick Liu, Audrey Kline, Jason F. Luck

Injury Biomechanics Laboratory, Department of Biomedical Engineering, Duke University, Durham, N.C.

Duke UNIVERSITY BASS CONNECTIONS
Duke BIOMEDICAL ENGINEERING

Introduction

Head Injury

- Mild traumatic brain injury (mTBI), or concussion, resulting from an impact or inertial loading of the head or sub-concussive head loading is a common and complex problem in youth contact sport athletes.
- Football athletes are disproportionately affected by head injuries and concussions, making up 12.8% of all sports-related head traumas and 19.4% of all sports-related concussions [1].
- Quantifying head injury in high school tackle football is essential for determining a player's injury risk per athletic exposure. This involves the use of wearable devices that measure kinematic head impact data during football practices and games.
- The high prevalence of false positives in recorded head impact data means that data processing is critical to drawing accurate conclusions about athletic exposure and injury risk.

Deep Learning

- 1-D Convolutional Neural Networks (CNNs) are a common deep learning model used to classify temporal activity data, such as valid and invalid head impacts.
- After the model has been trained and tested, the accuracy of the valid impact predictions is measured by the model's precision, whereas the completeness of these predictions is measured by the model's recall.

Data Collection

Valid impacts

- Head impact data was simulated via drop tests with post-mortem human surrogate (PMHS) heads in lacrosse helmets.
- Drop tests were performed on 2 PMHS for 7 common impact locations and 3 drop heights, each with 4 repeat trials.

Table 1. Drop test conditions for helmeted PMHS heads.

Impact Location	Drop Height
Facemask	8 cm
Frontal	
Frontal oblique right	50 cm
Parietal left	
Parietal right	
Occipital	90 cm
Vertex	

Fig 1. Parietal right impact location (as indicated by the crosshair) on the lacrosse helmet in which the head was placed.

Fig 2. The Data Acquisition System for Head Response (DASHR) earpiece on a dummy head.

- Two Data Acquisition System for Head Response (DASHR) earpieces (left and right) were used to record the head's linear acceleration and rotational velocity at 1000 Hz.
- The valid impact data was pre-processed to exclude all tests that did not meet a 10g linear acceleration threshold (LAT) and 3 ms duration threshold (DT).

Invalid impacts

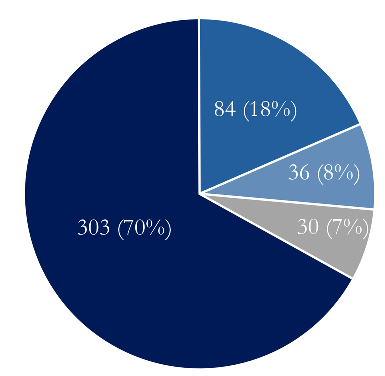
- Running/walking and standing still data were simulated while wearing the DASHR earpiece.
- High-g non-impact data consisted of flicking, pressing, and re-situating the DASHR in the ear while either standing still or walking to simulate background noise.
- The high-g non-impact data was then processed to exclude all tests that did not meet the 10g LAT and 3 ms DT. This was not repeated for the running/walking and standing still data.

Methods

Training Set

- 4 classes of data:
  1) Valid impact
  2) High-g non-impact
  3) Running/walking
  4) Standing still
- Training set data was simulated via:
  1) Drop test data from helmeted PMHS heads
  2-4) Activity data from worn DASHR

Distribution of Training Set



- In total, 453 samples were used in the training set with the distribution shown above.
- 100 ms segments of linear acceleration and rotational velocity data were used as model inputs (see Fig. 3).
- A 10g LAT and 3 ms DT were used to isolate valid impacts and high-g non-impact events. These data segments were shifted to begin 10 ms prior to the peak linear acceleration.

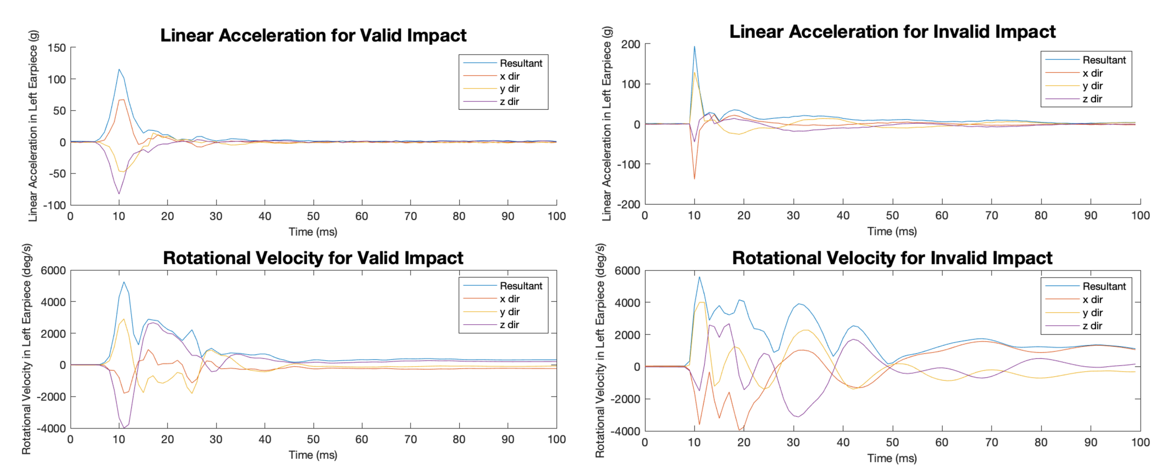
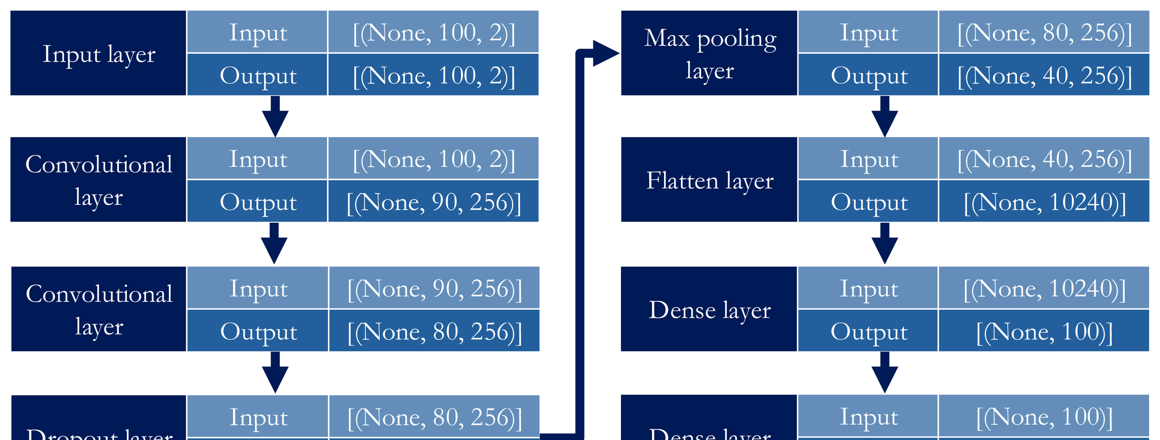


Fig 3. Linear acceleration and rotational velocity over time for a valid impact (left ear DASHR data from frontal drop test from 50 cm) (left) and invalid impact (flicking DASHR while standing still) (right).

Deep Learning

- A 1-D Convolutional Neural Network (CNN) was trained in TensorFlow to classify segments of linear acceleration and rotational velocity data into 4 classes.
- The model was trained for 10 epochs with a categorical cross entropy loss function, using a mini-batch gradient descent training algorithm with a batch size of 32.
- The kernel size and number of features were varied to determine the effects on model accuracy. Accuracy values were averaged over 10 trials.

Model Architecture



Test Set

- 25% of the training set was randomly set aside as the test set for model evaluation (n=114).
- Model precision, recall, and overall accuracy were calculated as follows:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:
TP = true positive,
FP = false positive,
TN = true negative,
FN = false negative.

Results

Model Performance

- Using a kernel size of 5 and 128 features, the model achieved an overall accuracy of 84.6% ± 2.0.
- A permutation test revealed that the trained model performed significantly better ($p<0.01$) than an uninformed model (62.3% ± 0.2 accuracy).

Hyperparameters

- The model performed best with a kernel size of 11 and 256 features, achieving an accuracy of 85.8% ± 2.0.
- However, varying the kernel size and number of features did not significantly affect model accuracy.

Confusion Matrix

- Overall, the model performed best for valid impacts, with a precision of 90.1% and a recall of 92.8%.
- On average, the model had the most difficulty predicting running/walking, with a precision of 87.5% and a recall of only 46.7%.
- The model was most often confused between running/walking and standing still, with a binary accuracy of 66.7%.
- Valid impacts and high-g non-impacts had a binary accuracy of 89.5%.

Table 2. Overall model accuracy for different kernel sizes and numbers of features.

		Number of Features (f)		
		64	128	256
Kernel Size (k)	3	83.9% ± 2.6	83.0% ± 2.6	83.9% ± 2.4
	5	83.5% ± 3.3	84.6% ± 2.0	78.7% ± 6.4
	7	81.7% ± 2.2	81.8% ± 3.8	82.6% ± 3.7
	9	83.9% ± 3.2	83.5% ± 2.6	83.3% ± 1.9
	11	82.7% ± 2.6	84.5% ± 2.1	85.8% ± 2.0

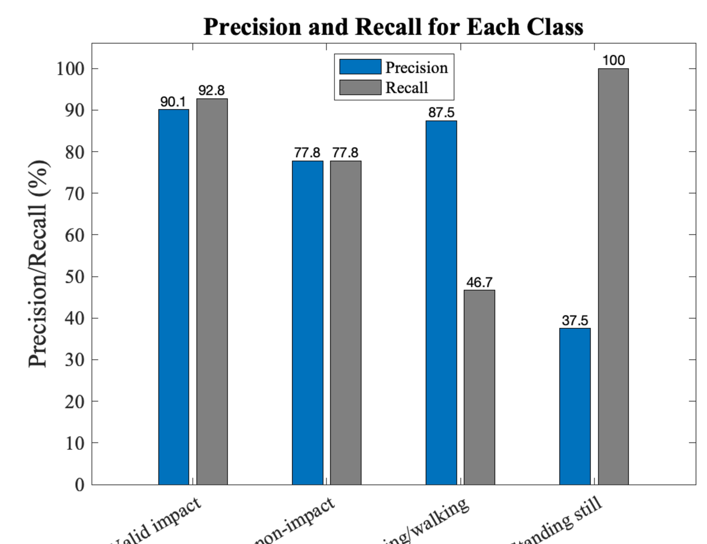


Fig. 4. Precision and recall values calculated for each class (k=5, f=128).

Table 3. Confusion matrix showing the number of samples predicted for each class (% of actual class) for k=5, f=128.

		Predicted Class			
		Valid impact	High-g non-impact	Running/ walking	Standing still
Ground Truth Class	Valid impact (n=69)	64 (92.8%)	4 (5.8%)	1 (1.4%)	0 (0%)
	High-g non-impact (n=27)	6 (22.2%)	21 (77.8%)	0 (0%)	0 (0%)
	Running/ walking (n=15)	1 (6.7%)	2 (13.3%)	7 (46.7%)	5 (33.3%)
	Standing still (n=3)	0 (0%)	0 (0%)	0 (0%)	3 (100%)

Conclusions and Future Work

Conclusions

- Deep learning has the potential to be a valuable tool for classifying head impact data as valid and invalid. This is a critical step in the process of analyzing head impact data prior to reporting athletic exposure rates.
- The overall model performance was good with an accuracy of 84.6% ± 2.0.
- The model performed best for valid impacts and worst for running/walking. This trend may be related to the relative training set sizes for these classes.
- Varying the model hyperparameters of kernel size and number of features did not significantly influence the accuracy of the model.

Limitations

- All training set data was collected via simulated activities.
- The training set was unbalanced, including 66.9% valid impacts versus a combined 33.1% invalid impacts. This may have skewed the accuracy of the model.
- The training set data was not augmented and therefore may have resulted in overfitting.
- The uninformed model achieved a 62.3% ± 0.2 accuracy, which was above average. This indicates that the pre-processing techniques and/or model conditions were prone to overfitting.

Ongoing & Future Work

- Investigation of the relevant features in the current model through saliency mapping.
- Augmentation of the current dataset to correct for skewed sample sizes and potential overfitting. This will improve model robustness.
- Addition of more data—both to the existing classes and to new classes—for increased model robustness and advanced activity classification. This includes low-g behavioral activities associated with head impact.
- Implementation of "out of set" classification for increased model generalizability and precision.
- Collection of data from athletes during practices and games. This will improve the validity of the study.
- Continued exploration of other deep learning models (ie. General Adversarial Network, or GAN, Long Short-Term Memory, or LSTM) or potential combinations of these models (ie. CNN-LSTM) for activity classification.

Acknowledgements & References

We acknowledge Duke Bass Connections - Brain & Society, Duke Institute for Brain Sciences (DIBS), USA Lacrosse, Duke Pratt School of Engineering, and the Department of Biomedical Engineering and thank them for their aid and support of this study.

References

[1] C. E. Gaw and M. R. Zonfrillo, "Emergency department visits for head trauma in the United States," *BMC Emergency Medicine*, vol. 16, no. 1, Jan. 2016.