Stroke Classification 2.0 and Analysis of Swimming
Connor Morikawa

Motivation

The goal of this project was to build off my final project from the Sports,Technology and Learning course to find a more effective way to classify swim strokes and to analyze stroke technique. Motivated by the challenges athletes and coaches encounter while working to optimize stroke technique, the project objective was to harness cutting-edge sensor technology and data analysis methodologies to offer a more precise, real-time evaluation of swimming strokes. This endeavor was not just about refining classification accuracy but also how technology can converge with sports science to provide actionable insights that lead to tangible improvements in an athlete's training and performance.

Project Description

The project consisted of two parts: data collection and data analysis. I found the process of data collection to be the most tenuous aspect of this project.

For the data collection segment, I decided to use the Micro:bit v2. This was partly due to my familiarity with the device from prior classes as well as the ability to easily interact with the Microsoft MakeCode interface. This allowed me to create different prototypes quickly to acquire and log data. The Micro:bit's accelerometer and magnetometer provided me with a dataset to capture the intricate patterns of each swim stroke. The main features I used were the x,y,z and mx, my, mz. The Micro:bit was strapped to a swimmer's lower back, as we found a paper that ran a similar study and were able to obtain fairly accurate results.

For the data analysis, we utilized three classification models, Long Short Term Memory (LSTM), Convoluted Neural Networks (CNN), and Dynamic Time Warping(DTW), in order to compare how the accuracies of the models differed. The rationale for using LSTM was the model's ability to process sequential information over extended periods of time. In the context of swimming, each stroke represents a complex series of movements that are continuous, which is crucial to accurately classify the data for interpretation. LSTM is capable of learning long sequences of data without losing the relevant information from earlier step times. This allows the model to recognize the nuanced patterns that are inherent in different swim strokes, distinguishing between them based on the sequence of movements even if the movements are minute. LSTMs can handle variabilities in stroke execution between swimmers or within a single swimmer's performance over time where the model is able to adapt to diverse data.

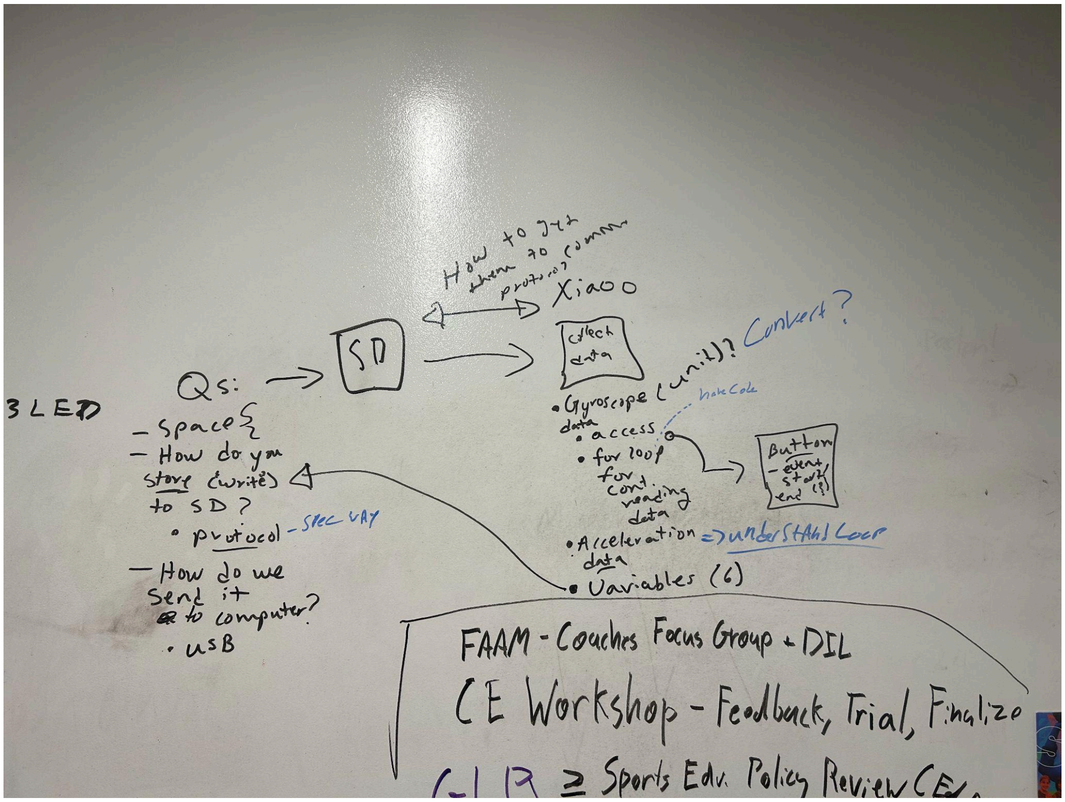
The reasoning for using CNN is its ability to adapt to time series data like repetitive swim strokes. The motivation to experiment with CNNs for this project stems from the model's ability to detect hierarchical patterns through the convolutional process, which can be analogous to identifying specific features within a sequence of several different movements. In swimming, each stroke consists of a distinct pattern of motions that can be abstracted as a "temporal image" where the convolutions effectively capture and highlight key features such as the stroke's rhythm, intensity, and recovery between stroke cycles. By applying filters over the time series data, CNNs can isolate and identify these unique signatures within the complex, multidimensional dataset. This capability allows CNNs to automatically learn the most prominent features that differentiate one stroke type from another.

Utilizing DTW presents a valuable approach for analyzing swim data due to its exceptional ability to measure the similarity between temporal sequences, which may vary in speed or duration. This characteristic is particularly beneficial for swim stroke analysis, where the timing and pace of strokes can differ significantly between swimmers or even within a single swimmer's laps due to fatigue, speed changes, and technique variations. DTW accommodates these variations by aligning sequences in a non-linear manner, allowing for a comparison that can accurately match similar movements occurring at different times. This flexibility makes DTW adept at identifying and classifying swim strokes based on their inherent temporal patterns, regardless of variations in stroke execution. Leveraging DTW can overcome one of the primary challenges in swim data analysis which is the variability of stroke signatures across different conditions and individuals.

Process

At the start of the project, Jacob and I used two Micro:bits that would be connected through bluetooth to one another. A Micro:bit would be attached to the swimmer and the coach would have the ability to start and stop tracking the data. Ideally, this code is written in micropython and would graph the x,y,z, pitch, and roll in real time. However, it proved to be difficult to effectively track clean data and to manage the distance of the bluetooth connection between the two Micro:bits. After running a series of tests, we later discovered that when submerged-underwater, the bluetooth Micro:bit connection would fail. This was unfortunate because our current setup allowed for data to be easily synced to the computer. A significant challenge we encountered was the irregular timing intervals recorded by the Micro:bit, leading to inconsistencies within our dataset. For instance, the data points were unevenly spaced in time, with examples like 10.8, 11.2, and 12.0 seconds.

In pursuit of cleaner and more precise data for our analysis, we transitioned to utilizing the XIAO Seeed Studio board as an alternative data collection tool. This decision was motivated by the board's superior capabilities, notably its high-precision 6-Axis Inertial Measurement Unit (IMU) which included a 3-axis accelerometer and a 3-axis gyroscope and more memory storage, which offers enhanced accuracy in tracking movement and orientation. This upgrade was promising in order to overcome the limitations observed with the Micro:bit, ensuring that our dataset achieved the level of granularity and precision necessary for accurate and nuanced analysis of swim techniques. However, I ran into difficulties with the XIAO board because of my limited experience with this specific hardware or for that matter hardware in general and the lack of documentation available for the Seeed Studio board. I was able to meet with Hermino to map out each individual component to get a better understanding of how to build out the XIAO board. He recommended that we not use micropython, and suggested that we use C instead. Unfortunately, I have limited experience with C and this represented another challenge. During this time working with the XIAO board, I was required to experiment with soldering to connect the battery pack to the board itself. Soldering was something that I had never done before and it was difficult. I decided to stop working with the XIAO board due to the lack of progress and moved back to the Micro:bit.



[XIAO board components mapping with Herminio B]

As the project progressed, my emphasis shifted towards meticulously logging data specific to each of the four distinct swim strokes. This pivot enabled me to develop code that was capable of capturing high-quality data, which effectively distinguished between the different strokes. To enhance the dataset's analytical value, I incorporated measurements of magnitude force, utilizing the tracked magnetometer readings (mx, my, mz) on to the Micro:bit. Having secured reliable data with all four strokes, I transitioned to the next crucial phase of the project: deploying the models for analysis.

[Graph revealing all four strokes- x axis: time , y-axis: range of x,y,z coordinates]

In the initial phase of model experimentation, my approach was focused on leveraging data exclusively from the accelerometer, encapsulating time (seconds), the three spatial dimensions (x, y, z), and the stroke label. The labels were defined as ‘1’ for butterfly, ‘2’ for backstroke, ‘3’ for breaststroke, ‘4’ for freestyle. This dataset provided a preliminary understanding of each swim stroke's unique characteristics through the lens of motion intensity and directionality. To explore the potential of capturing the more nuanced aspects of swim dynamics, I shifted my attention to incorporating data from the magnetometer, analyzing mx, my, and mz values. This phase aimed to discern whether magnetic orientation interaction offers additional insights into stroke differentiation. Finally, in pursuit of a more holistic analysis, I integrated both accelerometer and magnetometer readings, creating a comprehensive six-dimensional feature set for model training. To determine the most effective model for stroke classification, I executed Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Dynamic Time Warping (DTW) models across all datasets, assessing each model's accuracy in identifying different swim strokes.

Results:

	Accelerometer (x,y,z)	**Magnetometer (mx,my,mz)**	**Combined**
LSTM	**83%**	**80%**	**94%**
CNN	**64%**	**n/a**	**n/a**
DTW	**78%**	**93%**	**95%**

Current Status/Future Works

The swim project was able to achieve 95 percent accuracy of swim stroke classification using both accelerometer and magnetometer data. The model runs are located in both the ipynb files which contain the dataset for each stroke as well as the accelerometer and magnetometer data.

I would love to be able to explore the XIAO board more in depth and be able to compare the results to the Micro:bit. In addition, I would like to find a way to better understand the details of a single stroke cycle in terms of stroke rate and distance per pull. Because the model is solely based on one swimmer, future works would entail incorporating datasets that represent a wider range of swimming skills and styles. This would create an inclusive model that can accurately classify strokes and analyze performance metrics for swimmers at all levels of expertise.

Challenges/Limitations

Throughout this quarter, I faced various challenges. The first issue that I faced was finding an efficient way to waterproof the Micro:bit The strategy we ended up using, even though not the most efficient, was just placing it in two plastic bags and then using athletic tape to attach it to the swimmer's body. The subsequent dilemma was discovering that the bluetooth connection between two Micro:bits would fail when placed under water. This was frustrating because it meant that in order to transfer the current data from the Micro:bit we had to take the Micro:bit off the swimmer and plug it into the computer which proved to be a tedious task.

I was also very inexperienced with the programming language C and hardware which made working with the XIAO board challenging. Working with the XIAO board required soldering which again was something I had little experience with and was not comfortable with. This in theory, made attaching the battery onto the XIAO board a difficult task.

One of the biggest obstacles for the entire project was finding more accurate ways to track data. Data collection always proved to be tedious because during recording sessions the Micro:bit's memory would become full and I would not realize until the end of the swimming trial. The memory limitation of the Micro:bit at times was frustrating to work with as the XIAO would have been able to solve this given we were able to get it to work.

What I Learned

This was the first time I conducted research and I learned the invaluable lesson that failure is not a setback but a pivotal component of the learning process, offering unique insights and directing me towards more effective solutions. Navigating the complexities of the Micro:bit and XIAO Seeed Studio boards not only improved my technical skills but also deepened my appreciation for the intricacy of software and physical devices in data collection. Furthermore, this project served as a comprehensive introduction to deep learning models. I delved into the intricacies of Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Dynamic Time Warping (DTW), which allowed me to gain a profound understanding of their applications, strengths, and limitations in analyzing time-series data.

File Structure

SwimData_X.ipynb - This notebook is dedicated to exploring models solely using accelerometer data (x, y, z), providing insights into stroke classification accuracy

SwimData_MX.ipynb- This notebook delves into the analysis utilizing magnetometer data (mx, my, mz), as well as a comprehensive evaluation of models that integrate **both** accelerometer and magnetometer readings

AllStrokes.xlsx: file containing all of the raw data from testing

Individual sheets:

- allStrokesData: This sheet compiles data captured during a continuous session encompassing all four swim strokes
- Fly: This sheet is exclusively dedicated to data pertaining to butterfly strokes
- Trimmed Fly: Features data adjusted through an educated analysis to indicate the start and end points of swimming activity, specifically focusing on butterfly strokes.
- FlyX: data derived solely from accelerometer readings, augmented with stroke labels to facilitate the classification of butterfly strokes.
- FlyMX: Similar to the FlyX sheet, this dataset focuses exclusively on magnetometer readings, each entry tagged with a stroke label.
- Back: This sheet is exclusively dedicated to data pertaining to backstroke
- Trimmed Back: Features data adjusted through an educated analysis to indicate the start and end points of swimming activity, specifically focusing on backstroke strokes.
- BackX: data derived solely from accelerometer readings, augmented with stroke labels to facilitate the classification of backstroke.
- BackMX: Similar to the BackX sheet, this dataset focuses exclusively on magnetometer readings, each entry tagged with a stroke label.
- Breast: This sheet is exclusively dedicated to data pertaining to breast strokes

- Trimmed Breast: Features data adjusted through an educated analysis to indicate the start and end points of swimming activity, specifically focusing on breaststroke.

- BreastX: data derived solely from accelerometer readings, augmented with stroke labels to facilitate the classification of breaststroke.

- BreastMX: Similar to the BreastX sheet, this dataset focuses exclusively on magnetometer readings, each entry tagged with a stroke label.

- Free: This sheet is exclusively dedicated to data pertaining to freestyle strokes

- Trimmed Free: Features data adjusted through an educated analysis to indicate the start and end points of swimming activity, specifically focusing on Freestyle strokes.

- FreeX: data derived solely from accelerometer readings, augmented with stroke labels to facilitate the classification of freestyle strokes.

- FreeMX: Similar to the FreeX sheet, this dataset focuses exclusively on magnetometer readings, each entry tagged with a stroke label.

- Combined: contains all of the trimmed data for both the accelerometer and magnetometer data