A Deep Learning Approach to Automatically Classify Ice Hockey Shooting Methods Using Inertial Sensors

Samuel Tremblay

Department of Kinesiology and Physical Education

McGill University

Montréal, Québec, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

June 2022

© Samuel Tremblay, 2022

Table of Contents

Abstract

Feedback from smart technology using body worn sensors has seen tremendous growth in various sports. Specifically in ice hockey, the demand for tracking body movements during games and practices with sensor data is growing. However, there has yet to be a feasible method to use this data to automatically classify shooting methods. Hence, the purpose of this study was to evaluate whether sensor data could be used with a deep learning model to classify various ice hockey shooting tasks and determine which sensor configuration would be ideal for a real-world application. A fully connected convolutional neural network (CNN) was used to classify seven ice hockey related task, with a focus on shooting. The testing tasks consisted of wrist, slap, backhand, and one-timer shots, along with a passing and a simple stick-handling task, ending with a resting period. 39 participants wore a 17-sensors IMU system to collect the data and only the three-dimensional free sensor acceleration data was used as the model input. Results showed that the proposed CNN successfully classified tasks and the all-sensor configuration (ASC) achieved the highest average F1 score of 97.1% compared to 93.6% for the hands-sensor configuration (HSC). The HSC achieved slightly lower F1 score compared to the ASC model but was most ideal due to a greatly reduced sensor count (2 vs 17, respectively). These results demonstrate the potential of using hands' 3D acceleration data alone to track player shooting skill execution in both training and game contexts, as well as the feasibility for hockey equipment manufacturers to secure sensors inside the gloves.

Abrégé

Les retours d'information générés par des technologies intelligentes utilisant des capteurs de mouvement portés sur le corps ont connu une croissance phénoménale dans divers sports. Plus précisément dans le hockey sur glace, la demande de suivi des mouvements du corps pendant les matchs et les entraînements avec des données de capteurs augmente. Cependant, il n'y a pas encore de méthode réaliste utilisant ce type de données pour classer automatiquement les méthodes de lancer. Par conséquent, le but de cette étude était d'évaluer si les données des capteurs pouvaient être utilisées avec un modèle d'apprentissage en profondeur pour classer diverses tâches de tir au hockey sur glace et déterminer quelle configuration de capteurs serait idéale pour une application dans le monde réel. Un réseau neuronal convolutif (CNN) entièrement connecté a été utilisé pour classer sept tâches liées au hockey sur glace, en mettant l'accent sur les lancers. Les tâches testées consistaient de tirs du poignet, frappés, du revers et sur réception, ainsi qu'une tâche de passe et une simple tâche de maniement du bâton, terminant par une période de repos. 39 participants portaient un système d'IMU à 17 capteurs pour collecter les données et seules les données tridimensionnelles d'accélération du capteur libre ont été utilisées comme entrée du modèle. Les résultats ont montré que le CNN proposé a classé avec succès les tâches et que la configuration contenant tous les capteurs (ASC) a obtenu le score F1 moyen le plus élevé de 97,1 % contre 93,6 % pour la configuration des capteurs de main seulement (HSC). Le HSC a obtenu un score F1 légèrement inférieur par rapport au modèle ASC, mais était le plus idéal en raison d'un nombre de capteurs considérablement réduit (2 contre 17, respectivement). Ces résultats démontrent le potentiel d'utiliser uniquement les données d'accélération 3D des mains pour suivre l'exécution des compétences de tir des joueurs dans les contextes

d'entraînement et de jeu, ainsi que la possibilité pour les fabricants d'équipement de hockey de sécuriser des capteurs à l'intérieur des gants.

Acknowledgments

As I am closing towards the end of my academic chapter at McGill, I feel extremely grateful and privileged to have been part of the Ice Hockey Research Group. First things first, Dr. David Pearsall, thank you for the thrust you had in me from the start, for believing in me taking a very audacious project and supporting my hockey career at the same time. You've been instrumental to my success, and I've learned so much from you. Thank you for your kindness and the time you invested in me, I wish you the very best in your well-deserved retirement. To our research assistant, Philippe Renaud; what a wild ride it has been! Figuring out a way to produce a project like this throughout the past two years with COVID-19 and dealing with my injury was the most challenging time of life, but you were with me all along and I could not thank you enough for the relentless work you did alongside me. Thank you, my friend! I want to also thank Dr. Philippe Dixon for being patient and available for me during this project. Coming from absolutely zero experience in machine learning and coding, I gained a ton of knowledge from your expertise that I will carry on in my future endeavors.

À ma collègue de laboratoire Gab, je me sens extrêmement chanceux de t'avoir eu à mes côtés durant ces deux dernières années. Je ne crois pas que ce projet aurait vu le jour sans ton aide et ton support morale, merci énormément pour ton temps et ta gentillesse envers moi. Finalement, à ma famille et ma copine, je vous remercie pour le support inconditionnel et le temps que vous avez passé à m'écouter parler de mon projet que vous ne compreniez pas trop. Vous avez toujours cru en moi, encore plus dans les moments difficiles que je vivais et quand je ne croyais plus y arriver, merci infiniment.

Contributions of Authors

Samuel Tremblay, MSc candidate, Department of Kinesiology and Physical Education, McGill University, was responsible for the research design, processing and analysis of data, and writing of this thesis. The candidate's supervisor, David J. Pearsall, PhD, Associate Professor, Department of Kinesiology and Physical Education, McGill University, contributed to the research design and analysis of the data. The candidate's co-supervisor, Philippe C. Dixon, PhD, Assistant Professor, School of Kinesiology, Université de Montréal, contributed to the research design and helped develop the machine learning models.

Philippe J. Renaud, MSc., Department of Kinesiology and Physical Education, McGill University, played a very important role in the research design and data collection, acting as the lab research assistant. Dr. Shawn Robbins, PhD, Assistant Professor, School of Physical and Occupational Therapy, McGill University and Dr. Richard Preuss, Assistant Professor, School of Physical and Occupational Therapy, McGill University were members of the thesis advisory.

List of Tables

List of Figures

1. Introduction

Performance analysis in sports has evolved significantly over the past 20 years from manual and subjective classification of game specific events to automatic recognition of sport specific movements by way of small and portable body worn sensors. Through the advancement of "smart" technology, it is now feasible to enhance the efficacy, accuracy, and development of real-time sport performance task identification. Currently, ice hockey analytics and task recognition use video-based approaches, accelerometers, and local positioning systems. However, adopting machine learning analysis from wearable sensor data inputs may be a more efficient way to automatically identify various hockey tasks. Body worn sensors such as Inertial Measurement Units (IMU) have the potential to provide meaningful and externally valid information about body segments' kinematics. In a high-speed team sport such as ice hockey, the environment, equipment worn, and rapid movements are challenges to collecting reliable body kinematic measures. The use of wireless, body worn IMUs has the potential to estimate kinematic data on multiple body segments that in turn can be used for skill task recognition. IMUs are not constrained to fixed set-up and environmental conditions like current optoelectronic systems (e.g., Vicon) used in hockey biomechanics research. IMUs can provide richer data detail about limb movements not available by use of local positioning systems (e.g., Catapult). Thus, IMUs may be a viable option to collect detailed multi-segment body data for input into machine learning algorithms to identify various hockey tasks and their transition sequences. Such real-time feedback to coaches and athletes could be used to augment skill-training development.

Hence, the primary aim of this study was to evaluate the performance of a deep learning model to identify seven different ice hockey tasks with a focus on the different type of shooting methods using three-dimensional kinematics data from a commercial IMU system as input. The secondary aim of this study was to determine the most efficient sensor configuration subset to generate the highest task recognition accuracy.

2. Review of Literature

2.1 Current State of Ice Hockey Shooting Research

Over the past 30 years, ice hockey researchers have mainly focused on two predominant skills: skating and shooting. With shooting being the essential skill to project the puck into the opponent's net to score goals, previous researchers have emphasized the analysis of the stick shaft properties during a shooting task, and the different shooting techniques such as the wrist and slap shots. For example, Pearsall and colleagues (1999) investigated the effect of stick stiffness on puck velocity when performing a stationary slap shot on an artificial ice surface. Stick deformation was recorded from high-speed video camera and puck velocity from a radar gun across four different stick stiffnesses used by the participants. Overall, the authors found that the sticks' stiffness did not significantly influence shot velocity. However, they did find a greater variability in shot velocities between subjects than between stick types, which suggests that the player's shooting technique and anthropometric differences may influence shot quality more than the stick stiffness. A similar study was conducted by Wu et al. (2003) wherein shot velocity was compared between skilled and unskilled players performing both stationary wrist and slap shots. Their findings verified that shot velocity was not affected by differences in stick stiffness but rather each player's technique. In contrast, Worobets, Fairbairn, and Stefanyshyn (2006) found that puck velocity was affected by stick shaft stiffness and how the participants loaded the shaft by storing more potential elastic energy in wrist shots. Villaseñor and colleagues (2006) explored the recoil effect of a hockey stick during stationary slap shots using high-speed digital video. Similar to Worobets et al. (2006), the investigators found that potential elastic energy stored during the loading of a shot can contribute to more puck velocity. They also identified that the more skilled group had a lower kick point (point where the predominant bending of the stick begins) along the shaft (Villaseñor et al., 2006).

Further research has examined the ergonomics of how player's technique influence shot outcomes such as puck velocity and accuracy. Woo, Loh, Turcotte, and Pearsall (2004) compared elite and recreational hockey player stationary slap shot performance in-lab by recording the three-dimensional kinematics of the stick and upper body using electromagnetic tracking sensors tethered to the subject's body segments and stick. Higher (4.04 m/s) stick blade velocity was achieved by the elite players by way of both greater rotation and linear translational movements of the stick at ice contact (Woo et al., 2004). Technique (i.e. body and stick coordination) differences between the elite and novice players were predictive of slap shot speed (Woo et al., 2004). Villaseñor et al. (2006) further identified how slap shot technique translated to higher blade-puck velocities i.e., the longer the stick blade to puck acceleration, the greater impulse transmitted to the puck yielding high puck projection velocity. Shot success also depends on puck trajectory precision. Hence, the second predominant shot technique used is the wrist shot. Michaud-Paquette, Pearsall, and Turcotte (2008) set out to investigate what were the predictors of stationary wrist shot accuracy. Player's performance was assessed by recording movement of the stick shaft and blade, as well as the puck using a Vicon three-dimensional motion capture system. Key variables such as puck velocity at release, shaft bending, and blade orientations were identified as the best predictors of puck to target shot accuracy. Expanding on this study, Michaud-Paquette, Magee, Pearsall, and Turcotte (2011) then investigated both body and stick kinematics predictors of wrist shot accuracy and found that body kinematics are more important predictors of shooting accuracy than the stick construction alone. Hence, closer attention by coaches and players to better understand the body:stick mechanics of an accurate wrist shot is warranted (Michaud-Paquette et al., 2011).

A key limitation of the optoelectronic system used by prior cited ice hockey biomechanics researchers is that it is often constrained to an in-lab setting on artificial ice surface. Indeed, these constrained testing environments required by the camera's delimited field of view may not well replicate a hockey player's on-ice rink game environment. The lab setting is therefore strictly limited to standing shooting task compared to the on-ice setting which can allow for dynamic shooting task. Hence, Renaud et al. (2017) brought the lab camera set-up onto the ice and demonstrated the feasibility of collecting on-ice three-dimensional data by investigating the kinematics of the skating start movement technique. Similar optoelectronic camera on-ice set ups have been used by MacInnis (2019) who investigated slap shot performance, and Swarén, Söhnlein, Stöggl, and Björklund (2019) who studied on ice slap shots versus one-timer shots. Although feasible to use multiple cameras on the ice, it has limitations: (1) a limited field of view; (2) a time-consuming set up; and (3) extensive post-hoc post-processing to reconstruct virtual computer models of each player, stick and puck movements.

In summary, player's body and stick movement technique predominate in terms of predicting shot performance, followed by stick construction properties (Michaud-Paquette et al., 2011; Villaseñor et al., 2006).

2.2 Inertial Measurement Units

Considering the aforementioned limitations of the optoelectronic system, researchers have recently explored the potential of Inertial Measurement Units (IMU) to capture body movements. IMUs are wireless, lightweight, wearable sensors that require very low energy usage

(Cust, Sweeting, Ball, & Robertson, 2019). IMUs contain a combination of accelerometer, gyroscope and magnetometer sensors that quantify acceleration, angular velocity, and the magnetic field , respectively (Cust et al., 2019). There are many studies validating the use of IMUs for sport movement tracking (Ahmadi, Rowlands, & James, 2010; Fulton, Pyne, Hopkins, & Burkett, 2009; Krüger & Edelmann-Nusser, 2010) however, none before Denroche (2020) validated its use for hockey movements. Denroche (2020) evaluated the use of a 17-sensor Xsens MVN Link IMU system to measure joint kinematics in hockey. Denroche compared this IMU system with the "gold-standard" optoelectronic camera system. The concurrent measures of body movement were done of subjects during wrist and slap shots performed while standing on an artificial ice surface. The Xsens system demonstrated comparable body joint measures to those of the optoelectronic system, except for poor agreement with shoulder axial rotation measures (attributed to different shoulder models used in their respective analysis software). Hence, it is conceivable that future of ice hockey studies can use IMU measures (with attention to shoulder model adjustments) to assess hockey skills performed on an arena's ice surface. Testing with IMUs would afford greater external validity, be logistically more practical to execute and more time efficient in both data collection and analysis.

2.3 Movement Recognition in Sport

Athletic movement recognition can be achieved through machine learning and deep learning approaches to automatically identify the movement occurrence within a continuous data input signal (Bulling, Blanke, & Schiele, 2014). Thus, automatic recognition of sport specific movements using wearable sensors' data as inputs could potentially provide coaches with real-time analysis that may be applied to enhance both technical and tactical evaluation of athlete's performance, as well as preventing injuries (Cust et al., 2019). This would also enhance the objectivity of performance analysis conducted by coaches and analysts, which is prone to human errors, cognitive biases, and time consuming processes (Cust et al., 2019).

2.3.1 Deep Learning

One popular subfield of interest for task recognition in machine learning is supervised learning. A simple way to define supervised learning would be to train an algorithm using fully labelled (ground-truth) input data, so that it can make accurate predictions by automatically assigning an unknown input dataset to a specific label class (Preece et al., 2009). Deep learning supervised learning introduces an innovative approach on learning data representations that puts an emphasis on learning from successive layers transforming raw input data into increasingly meaningful representations to detect or classify patterns from the input (Chollet, 2018; LeCun, Bengio, & Hinton, 2015). Each layer can contain millions of parameters, which are adjusted to reduce the error between the model's output and the true label (e.g. accuracy) (LeCun et al., 2015). Chollet (2018) and LeCun et al. (2015) explained the learning process of a deep learning model: the values found for each parameter of the network's layers are used to 'map' the inputs to their associated targets. However, determining the right parameters cannot be done instantly. To measure how far the prediction outputs are from the actual targets, an objective function called the loss function is used to compute a distance score (loss score), which stores how good the network is at predicting the right targets for that specific iteration. Then, this score is used as a feedback signal to adjust the parameters' values to reduce the loss score (backpropagation), which is controlled by the optimizer (Figure 1). At the start of this training process, prediction outputs are usually far from the actual targets and the loss score is very high, but at every iteration of the training loop, the parameters are adjusted in such a way that the loss score decreases (gradient descent). The goal of the network is to achieve its minimal loss score, where the prediction outputs will be as close as they can get from the targets. This trained network is then ready to be tested on never seen before data to predict its outcome.

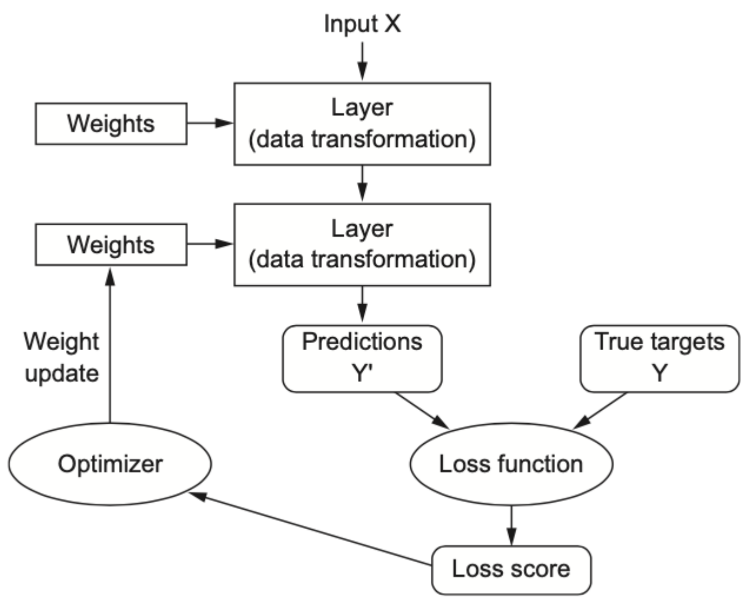

Figure 1 *A conceptual deep learning network training loop (Chollet, 2018)*

Deep learning has become very popular among data scientist for many reasons, one of them being that it simply offered better performance in solving problems that were previously challenging, such as dealing with high-dimensional datasets in science, business, and beating records in image and speech recognition (LeCun et al., 2015). It also presented a framework that was easier to implement, partly by automatically extracting important features compared to other machine learning models, such as Support Vector Machine and k-Nearest Neighbour classifiers, which had to be extracted manually (Ignatov, 2018). Meanwhile, handcrafted feature engineering became more challenging and introduced more bias than deep learning models as more complex problems were explored (Chollet, 2018). Overall, deep learning networks are a very attractive and performant methods to solve complex problems because of its simplicity, versatility and scalability (Chollet, 2018).

In the field of deep learning, convolutional neural networks (CNN) are one of the most commonly used algorithms (Alzubaidi et al., 2021). After revolutionizing the field of computer vision by winning the ImageNet competition in 2012 (Krizhevsky, Sutskever, & Hinton, 2012), CNNs have had tremendous success in many different fields such as image recognition and different language processing tasks (Ismail Fawaz, Forestier, Weber, Idoumghar, & Muller, 2019). Inspired by the success of these two-dimensional CNN models, many researchers have started adopting one-dimensional CNN to solve time series classification problems and have achieved or improved upon the state-of-the-art for classification accuracy (Hammerla, Halloran, & Plotz, 2016; Ismail Fawaz et al., 2019; Ruiz, Flynn, Large, Middlehurst, & Bagnall, 2021). Further, time series are present in many different real-world applications, such as in health care (ECG, EMG), sound classification and human activity recognition (Ismail Fawaz et al., 2019; Zheng, Liu, Chen, Ge, & Zhao, 2014). For the purpose of this research, only the topic of human activity recognition (HAR) will be explored. Collecting the data necessary to perform HAR can be done using body-worn sensors, such as IMUs. Hence, implementing a one-dimensional CNN with a time series dataset collected by IMUs seems to be an effective method to solve HAR problems (Gupta, 2021).

However, there seems to be a significant gap in the literature on the implementation process of different deep learning algorithms for the wide variety of scenarios in HAR (Hammerla et al., 2016). Researchers often report only the parameters and peak performances of various models but remain vague on the actual preprocessing and training stages, which makes it challenging to understand how model parameters were chosen (tuning), the decision process used behind the model architecture choice and the evaluation process to obtain their results (Hammerla et al., 2016). Thus, many questions remain unanswered, and it is difficult to replicate what other researchers have done for similar HAR problems.

2.3.2 Shooting Task Recognition

Previous literature using IMUs have demonstrated the feasibility of using sensor technology for the classification of shooting tasks. Hardegger et al. (2015) conducted a study on eleven amateur and eight professional hockey players to automatically differentiate wrist shots from slap shots and recognize a player's skill level. The hockey sticks used were instrumented with an IMU into the cavity at the upper end of the stick, two potentiometers measured the player's hand position, while four strain gauges measured stick flexion. A rule-based classifier was used to identify the type of shot while a Support Vector Machine (SVM) classifier assessed the player's skill level. The rule-based classifier detected shot types with an accuracy of 100% and the SVM achieved 88.5% accuracy for wrist shots and 92.9% for slap shots. Although these results may appear robust, only 64 shots (43 wrists, 21 slaps) by amateur players and 36 shots (26 wrists, 10 slaps) for the professionals were recorded due to data loss at the time of impact, which may have affected the scores. Nonetheless, this study demonstrated a feasible method to identify shot types and skill level based on a sensor mounted stick.

In the same article, Hardegger et al. (2015) conducted another study on gameplay events classification in ice hockey players. A total of three IMUs were placed separately on both gloves and on the chest protector to identify slap and wrist shots, hits, turns, time on ice, and puck possession time. Four professional hockey players performed 10 rounds of a circuit that included the tasks previously mentioned. Using a random forest classifier, they found an overall F1 score (combination of precision and sensitivity) of 0.7 for all events. However, the classifier detecting puck possession performed too poorly to be included in their final approach, which would have reduced the F1 score. Overall, slap and wrist shots were identified with high accuracy, but the sample size for this machine learning approach was very small. Therefore, more research with larger sample sizes is required to confirm these results.

To the author's knowledge, no other studies in ice hockey have examined shooting method recognition using sensor data. However, studies in sports with similar striking motions as in hockey have been published. For example, Shahar, Ghazali, As'ari, and Swee (2020) compared four different classifiers to identify specific tasks in field hockey. They used a combination of four sensors on the chest, waist and both wrist and achieved the highest performance with the Cubic SVM classifier at 96.7%. Similarly, Jiao et al. (2018) used a deep CNN and a SVM to classify various golf swing. Two strain gauges, an accelerometer, and a gyroscope were instrumented directly onto the golf club. The deep CNN achieved an overall 95% accuracy while the SVM classifier achieved 86.8% across nine different golf swings. In a different study, Whiteside, Cant, Connolly, and Reid (2017) attempted to quantify tennis players' hitting workload by classifying various striking techniques. Participants wore a single IMU on their wrist and performed a total of nine different shot type. Six machine learning models were compared. They found that an SVM classifier achieved the highest accuracy with 93.2%. Finally, Kautz et al. (2017) used a single sensor on the wrist of volleyball players to classify nine different tasks. They implemented a deep CNN with their dataset and achieved 83.2% of overall accuracy.

Despite achieving high classifying accuracies across most studies, researchers do not seem to have a clear consensus on which is the right method and algorithm combination to use to classify striking motion tasks. Moreover, there are many differences in sensor configuration used in the literature for sport movement recognition. However, Jang et al. (2018) used a unified CNN with a long-short term memory deep learning model to classify skating and classical styles of cross-country skiing. They used a 17 sensor IMU system and compared accuracy of various sensor configurations. Their so-called “sport biomechanics configuration”, comprised of IMUs placed on the pelvis and both wrists and feet, obtained the highest accuracy when classifying skiing sub-technique on flat and natural courses with 91.15%. These findings are relevant for the current study considering that cross-country skiing technique has similar movement pattern as in hockey skating.

3. Objectives and Hypotheses

The literature does not propose a consistent framework to automatically classify sport movements. In general, studies have a relatively low sample size, large range of sensor configurations and sports analysed, which means that comparisons between studies should be treated with caution (Preece et al., 2009). For hockey specifically, providing a reliable framework to automatically identify shooting techniques could increase the feedback quality from coaches to athletes and enhance performance assessments, potentially in real-time.

Hence, the primary objective of the proposed study is to evaluate the performance of a CNN deep learning model to automatically recognize seven ice hockey tasks performed by elite hockey players using three-dimensional sensor free acceleration data from a commercial IMU system (Xsens, Enschede, Netherlands) as input. The secondary objective of this study is to determine the optimal sensor configuration subset of the 17-sensor system.

Based on previous studies, it is first hypothesised that a CNN model will be able automatically identify the different tasks. It is also hypothesised that an all-sensor configuration will show strongest performance.

4. Methods

4.1 Participants

43 elite hockey players (27 men, 16 women) from the McGill men's and women's varsity team, and from the surrounding Montreal hockey community participated in the current study (see table 1). "Elite" was defined as a player having played in either the Canadian Hockey League (Junior) and/or the USports hockey league or any other higher-level league. Both male and female from all skater positions (except goaltending) and shot handedness were included. All players were injury free at the time of data collection. All task requirements were fully explained to the participants prior to signing an information and consent form in accordance with the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (McGill REB II, file # 375-0216).

Table 1 *Descriptive statistics of participants*

	Age (yrs) Mean ± SD	Height (cm) Mean ± SD	Weight (lbs) Mean ± SD	Experience (yrs) Mean ± SD
Men	29.3 ± 2.1	182.7 ± 6.0	185.1 ± 17.0	18.0 ± 2.7
Women	21.1 ± 2.2	168.5 ± 4.4	148.4 ± 15.6	14.4 ± 2.6
All	22.9 ± 2.5	177.4 ± 8.8	171.5 ± 24.2	16.7 ± 3.2

4.2 Testing Instrumentation

The IMU system used for this study was the Xsens MVN Link system (Xsens, Enschede, Netherlands). The MVN Link system consists of 17 IMU sensors connected through wires along the entire body. Sensors were placed on the head, sternum, pelvis and along all four limbs of the body (Figure 2). Through the wire leads, all sensors communicate back to a body pack placed on the participant's back which in turn communicates through a Wi-Fi connection access point. The MVN link system has a sampling frequency of 240 Hz and a reported indoor range of 50 m for data collection (Xsens, 2018). To collect a visual representation of the trials, a Bonita video camera (Vicon Motion Systems Ltd., Oxford, UK) set at a 30 Hz frequency positioned as shown in Figure 3. The Bonita camera was synced with the Xsens system through the Xsens SyncStation (Xsens, Enschede, Netherlands) for simultaneous data collection and future labelling.

Prior to data collection, the system had to be calibrated. Similar to Denroche (2020), the calibration process required the participant to stand in a static N-pose (neutral pose) for approximately 2 seconds before walking forward in a straight line for approximately 4 meters, then walking back to the initial point to resume the static posture. The Xsens software then presents quality of the calibration process as either "good", "acceptable", "poor" or a "failed" calibration. As proposed in Denroche (2020), the N-pose was used and only "good" calibrations was accepted during the study. Denroche also suggested using the "No-Level" scenario that the Xsens MVN software (Xsens, Enschede, Netherlands) provided in which the user's avatar was fixed in one global position at the pelvis. This scenario seemed to be best suited for the current study as it was not concerned with global positioning or environment interactions. It was also best for tasks such as skating where contact events between the foot and the ground are not precisely defined (Denroche, 2020).

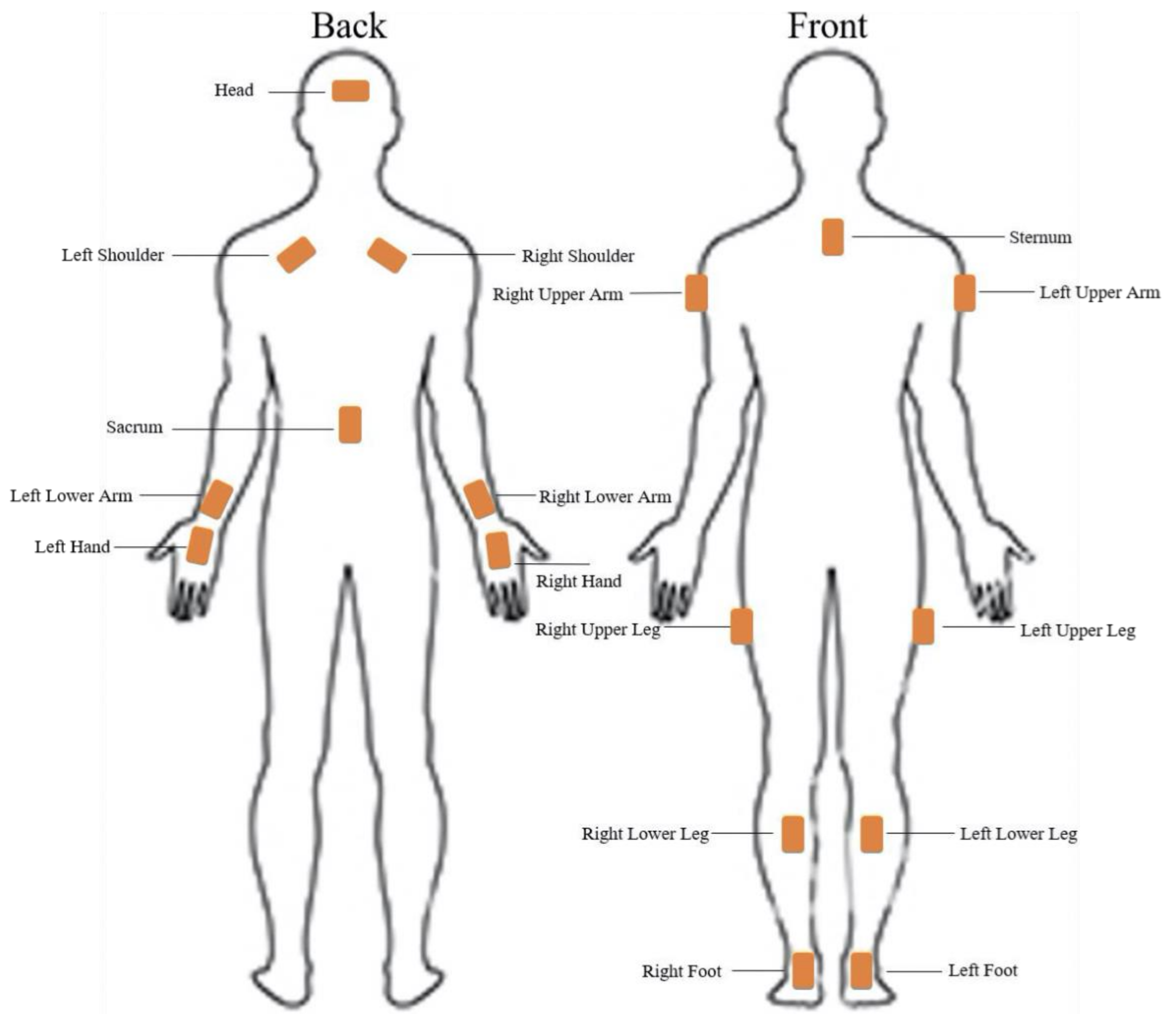

Figure 2 *IMU sensor configuration in accordance with Xsens MVN model (Denroche, 2020)*

4.3 Testing Protocol

Participants took part in a 30-minute on-ice testing session. They performed a series of six different tasks for a total of 60 trials and a five-minute resting period at the end. All data collection took place on the ice rink surface at the McConnell Arena, at McGill University. Participants were asked to bring their own stick, gloves, and skates for the data collection. Prior to on-ice testing, participants’ body anthropometrics were collected for use as inputs in the three-dimensional model associated with the Xsens system. The required measurements were the heights of the entire body, ankle, sole, knee, hip and shoulder, foot length, hip width, shoulder width and finally, the arm span. Afterwards, the Xsens sensors were placed directly on the participants' body using double-sided tape and secured with additional medical tape placed over top of each sensor, similarly as in Denroche (2020). This process was used for all sensors with the exceptions of the head sensor (placed inside of a headband), hand sensors (placed inside of gloves) and the foot sensors (taped on top of the skate laces) (Denroche, 2020). Lastly, participants put on a tight fitting Xsens T-shirt containing pockets for the Xsens body pack, battery pack, and sensors' wires. Participants then put on their gloves and skates. Once fully equipped, the remaining Xsens sensors for the feet were be placed on top of the skate laces and secured with tape.

Subsequently, participants moved onto the ice and were allowed time to warm-up and get accustomed with the setup by skating and taking shots at one end of the ice. Once warm-up was completed, participants headed to the other end of the ice to calibrate the system, as described in Section 4.2. Note, since the walking calibration process was done on the ice with skates on, a thin carpet was place on the ice to permit walking.

The testing tasks consisted of wrist, slap, backhand, and one-timer shots, along with a passing and a simple stick-handling task, ending with a resting period of five minutes as non-movement data input. For all shooting tasks, participants were asked to aim anywhere at the net and to shoot for maximum velocity. Additionally, they were asked to skate, stick-handle, pass and shoot as they normally would in a game situation and did not receive specific instructions with respect to their technique. After completion of each trial, participants were granted up to a thirty-second break to skate back to the respective starting position without fatiguing. Individual task instructions are listed below along with a summary of all tasks in Table 2 at the end:

1) Wrist and slap shots: Each task was divided into three positions (Figure 3). Five trials for each starting position were completed for a total of 15 trials per shot type. Participants started skating forward from the designated starting position on the red line towards the net. They were instructed to start their shooting motion as soon as they crossed the top of the circle area.

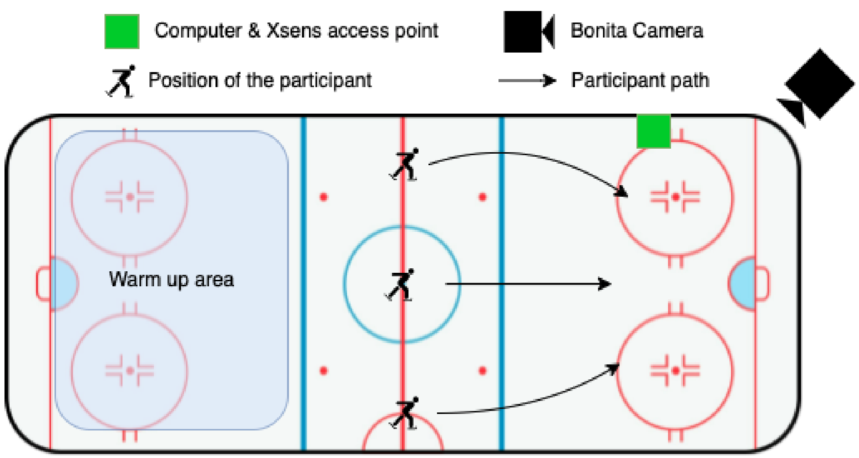

Figure 3 *Top view, on-ice data collection layout for wrist and slap shots*

2) Backhand shots: Participants were positioned at a single location on their backhand side (Figure 4). This means that a right-handed participant was positioned on the left side of the ice, and a left-handed shooter was positioned on the right side of the ice. Participants skated towards the net and started their shooting motion after crossing the top of the circle area. A total of 10 backhand shots were performed by each participant.

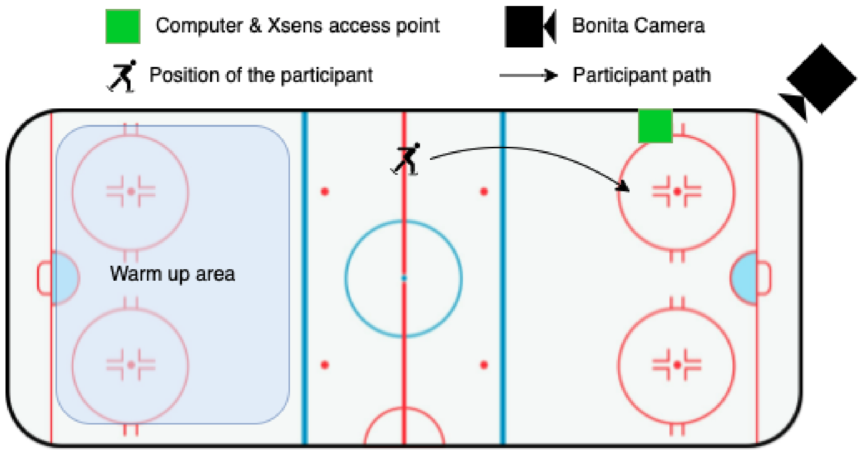

Figure 4 *Top view, on-ice data collection layout for backhand shots (right-handed participants)*

3) One-timer shots: For this task, the principal investigator (PI) of this study was the puck passer as he was also considered an elite hockey player who could deliver good and reliable passes to the participants. Both the participant and PI remained still between the face-off dot and the top of the circle on their respective side of the ice (Figure 5). Right-handed participants were positioned on the left side of the net while left-handed participants were positioned on the right side of the net. Each trial consisted of a pass of the puck going from the PI to the participant, so that they could in turn perform a one-timer. A total of 10 one-timer shots were performed by each participant.

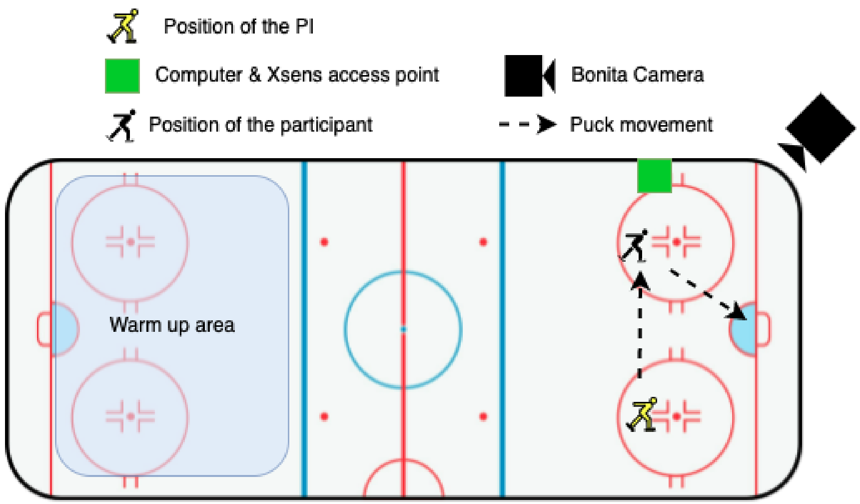

Figure 5 *Top view, on-ice data collection layout for one-timer shots (right-handed participants)*

4) Passing task: For this task, the participant was positioned at the second blue line and the PI was positioned on the goal line on the opposite side of the participant's handedness (Figure 6). Thus, for a right-handed participant, the PI was be positioned on the left side of the net and for a left-handed participant, the PI was positioned on the right side of the net. For each trial, the participant skated forward with the puck and performed a pass to the PI when reaching the first blue line. A total of 5 trials were recorded.

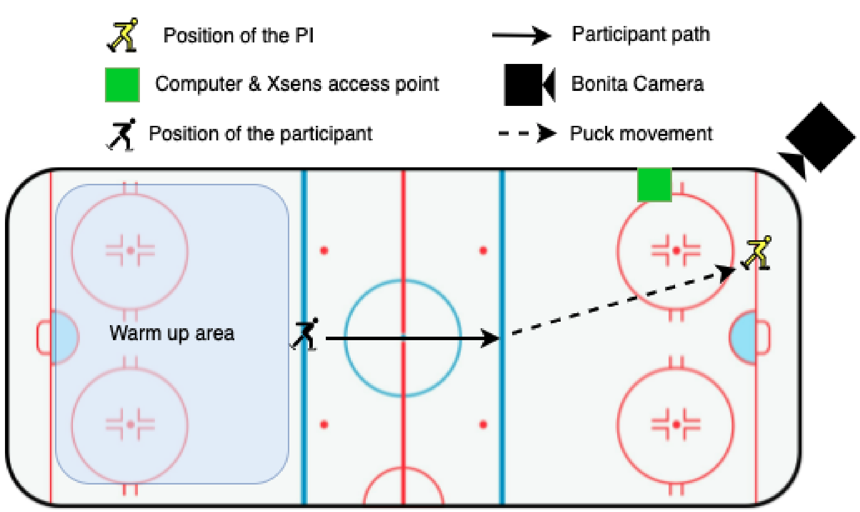

Figure 6 *Top view, on-ice data collection layout for the passing task (right-handed participants)*

5) Stick-handling task: For this task, the participant started in the same position as in the passing task. The participant was asked to skate forward and stick-handle the puck at the same time either on the forehand or in front (however they felt comfortable), replicating a similar stick-handling pattern as in the start of a wrist or a slap shot trial. Participants skated while stick-handling all the way to the next blue line (Figure 7). A total of 5 trials were recorded.

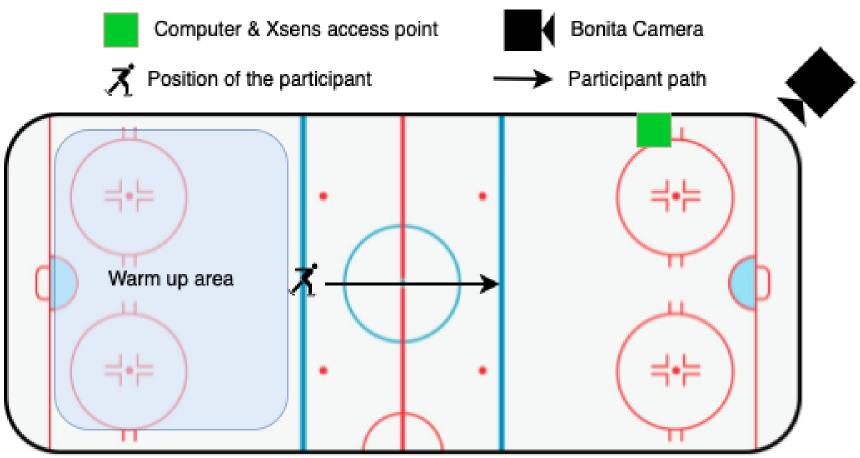

Figure 7 *Top view, on-ice data collection layout for stick-handling task (right-handed participants)*

6) Rest: At the end of the 60 trials, the participants had a resting period of five minutes, sitting on the bench.

Table 2 *Summary of the various task involved*

Task	Number of trials	Approximate duration (min)
Wrist shot	5/position = 15	7.5
Slap shot	5/position = 15	7.5
One-timer	10	5
Backhand shot	10	5
Passing	5	2.5
Stick-handling	5	2.5
Rest	1	5
Total	61	~ 30 - 35

4.4 Data Analysis

The Xsens system can automatically process data using the MVN Analyze software. It was suggested by an Xsens product specialist to use the normal quality reprocessing as it was 10x less time consuming than the HD method. It would also not make a significant difference since the HD reprocessing affects mostly foot contact data which was not relevant to this project since it was recorded in the No-Level scenario (no foot contacts). For this study, only the sensor free acceleration data (i.e., the acceleration sensed by the IMU, without the Earth’s gravity acceleration component (Xsens, 2018)) was selected for the analysis.

As a result, the sensor free acceleration data from 39 out of 43 participants were exported to a Google (Google LLC, Mountainview, USA) Drive for storage and future use. Unfortunately, participants 6’s data could not be exported as the file was corrupted and data from participants 30, 31 and 38 were left aside due to sensors falling off during data collection which polluted the signals. The synced video camera footage from the Bonita camera was exported in the Vicon Nexus software (Vicon Motion Systems Ltd., Oxford, UK) and analysed afterwards to mark events from the on-ice trials, as marking events directly from the Xsens three-dimensional model was more complex and time consuming.

Thus, for every time frame found from the video footage at which a participant started a trial, the label “Pre” was recorded to identify the sequence that preceded the stick-related task. When the start of the task was occurring, the label “WS” (wrist shot), “SS” (slap shot), “OT” (one-timer), “BH” (backhand), “Pass” (passing) or “SH” (stick-handling) was recorded. At the end of that motion, the label “Post” was recorded to identify the post-shot sequence (participant skating back to it’s the original position). Lastly, the final task was labelled “Rest” to identify the participant’s resting period. The specific frame at which the tasks started and the trial length were also recorded. Since the Xsens system sampled data at 240 Hz and the Bonita camera was recorded at 30 Hz, the time frame entered from the camera footage was multiplied by 8 to match the Xsens frame rate.

The deep learning model preprocessing manipulations and its implementation were done using the Python software (Python Software Foundation, https://www.pyton.org/) on Google's Colaboratory Pro+ GPU (GPU: 1xTesla P100, 54.8GB RAM). Specific code instructions and details can be found on the PI's GitHub repository dedicated to this project.

4.4.1 Data Preprocessing

The first preprocessing step was to select the sensor configuration. The first one was the all-sensor configuration (ASC), which included all 17 sensors, and the second was the hands-sensor configuration (HSC), which included only the hand sensors.

Each participant's data contained approximatively 400,000 frames due to the continuous data collection session. The next step was to group all the "Pre", "Post" and "SH" tasks into one single label: "Other". These labels were grouped because their kinematics were very similar and distinguishing each was not relevant to this project.

Afterwards, each trial had to be reframed to the same length to fulfill the model's input shape requirements. For the stick-related trials (WS, SS, OT, BH, Pass), the first step was to find the maximum trial length (576 frames) out of all participants. Then, for any stick-related trials having less than 576 frames, the difference in frames was added from the task preceding it. For example, if a WS trial had a length of 400 frames, 176 frames labelled Other preceding the trial were assigned to the label WS, so that the final WS length was 576 frames. This reframing procedure was chosen instead of a resampling method to not modify the original signal at the expense of adding some signal that was prior to the stick-related trial. This added signal was also justified as there could possibly be some relevant information from the pre-shot sequence that could further help with the task identification.

Each participant's dataset had a major class imbalance due to the number and length of Other trials. To help counter this imbalance, only 15 trials of 576 frames were randomly selected from the pool of Other trials for each participant.

Following was the amplitude standardization of the datasets, which was used to make the model's learning process less difficult by reducing the wide range of values (Chollet, 2018). For each value, the mean was subtracted and then divided by the standard deviation, so that each value was centered around 0 with a standard deviation of 1 (Chollet, 2018).

A label encoder was then used to encode the labels into unique integers, which is one of the encoded formats required to feed a deep neural network (Chollet, 2018). For this project, BH was encoded as 0, OT as 1, Other as 2, Pass as 3, Rest as 4, SS as 5, and WS as 6.

For the next step, each participant's dataset was transformed into $n \times 576$ frames $\times f$ tensor, where n and f represent the number of trials and the number of features, respectively (Figure 8). For each participant, the total number of trials were 85 with some participants having plus or minus 1-2 trials. In those 85 trials, there were 15 WS, 15 SS, 10 BH, 10 OT, 5 Pass, 15 Other, and the resting period was divided into 15 trials. The 576 frames' axis was for the trial length while the number of features was the number of sensors in their three axes. For the ASC, there were 51 features (17 sensors * 3 axes) and for the HSC, there were 6 features (2 sensors * 3 axes). Labels had their own one-dimensional tensor with a shape corresponding to the participants' trials.

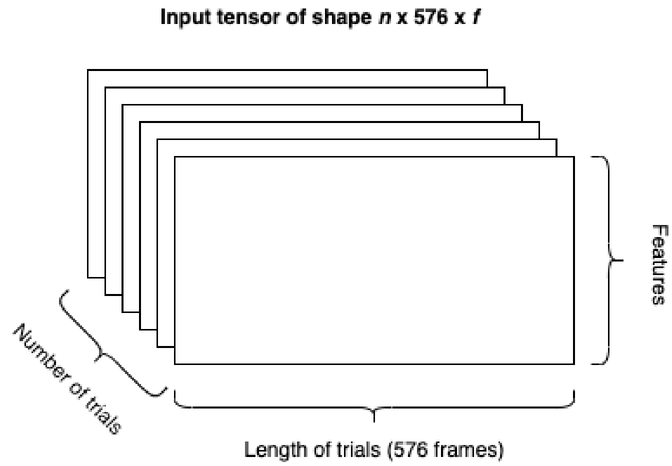

Figure 8 *Illustration of a* $n \times 576 \times f$ *tensor*

The final step was to split the participant's tensor into training, validation, and testing datasets for the sensor values and for their respective targets. The training dataset was used to train the model, while the validation set was used for tuning and evaluating, and finally, the test set was kept aside until the model was fully trained and ready to test on participants' data that it had never seen before (Chollet, 2018). The train set contained 70% of the participants for a total of 27 participants, the validation set contained 10% for a total of 4 and the test set contained 20% of the participants for a total of 8. To accomplish that, the participant id's were randomly shuffled and split subject-wise (inter-subject scheme) into a train and test set with a ratio of 80/20. The subject-wise shuffle and split was performed to avoid leaking information of a participant from the train dataset to the test dataset. Then, the train set was split subject-wise once more using a K-fold cross-validation approach (Figure 9) into five train and valid datasets with a ratio of 70/10. The purpose of the K-fold validation was to split the training set into K-partitions of equal size, train the model on K-1 partitions and evaluate it on the one remaining (Chollet, 2018). This process was repeated K-times after each partition was evaluated and the final score was averaged from each K scores obtained (Chollet, 2018).

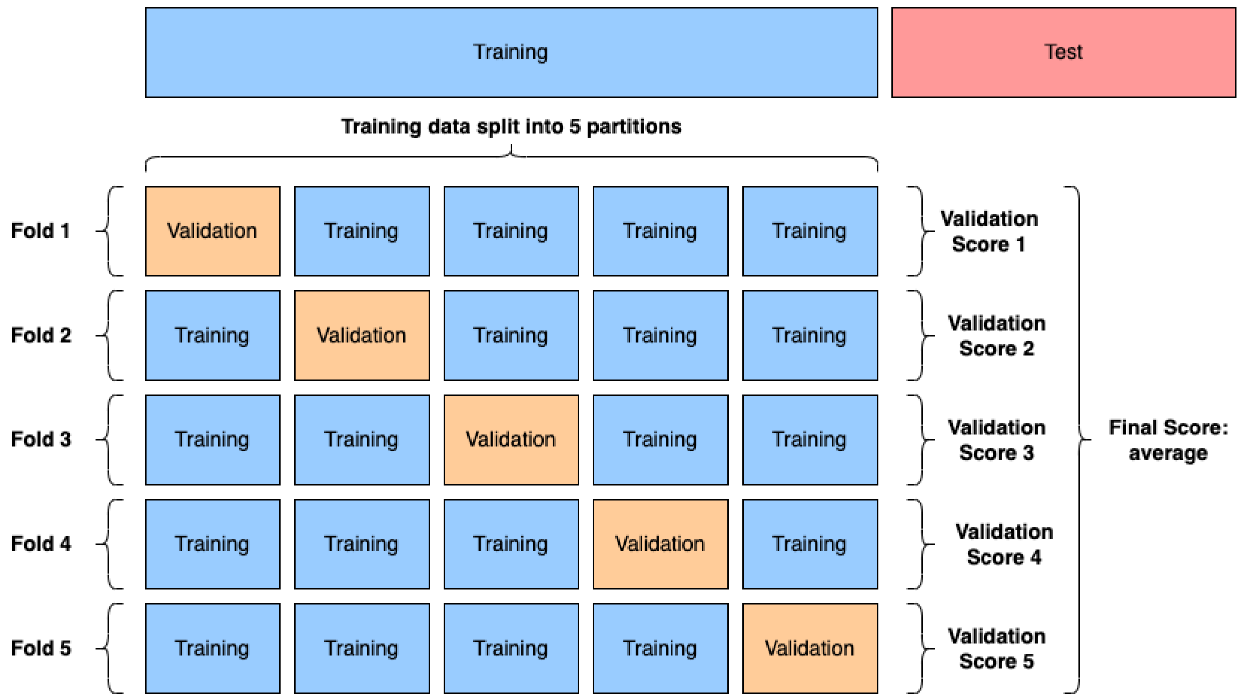

Figure 9 *Illustration of a 5-fold cross-validation*

4.4.2 Deep learning model architecture

The deep learning model developed in this study was inspired by the Fully Convolutional Network (FCN) architecture originally proposed in Wang, Yan, and Oates (2016). This architecture was modified using some of Chollet's methods to reduce overfitting by adding dropout layers and kernel regularizers (Chollet, 2018). Finally, the FCN hyperparameters were tuned for both sensor configuration and for each fold via random search using the KerasTuner (O'Malley et al., 2019) to optimize the model. The tuned hyperparameters were the filter sizes (64 to 256 with a step of 64), the kernel sizes (3 to 5), the type of regularizers (L1, L2 and L1_L2), the dropout probability (0 to 0.5 with a step of 0.1), and the choice of optimizers (Adam, SGD and RMSprop). Tuner results for each fold were examined and the combination of hyperparameters that produced the overall best classification performances were selected for the final models.

Both tuned FCN models each received $n \times 576 \times f$ tensors as input layers. These input layers then went through three convolutional blocks consisting of a 1-D convolutional layer followed by a batch normalization layer, a ReLU activation layer and a dropout layer (Figure 10). For the ASC specifically, each block had convolutional layers with filter sizes of [192, 192, 256], kernel sizes of [3, 3, 4] and regularizers of type [L1_L2, L1_L2, L2]. The dropout probability was also set at 0.4 in each block.

For the HSC, the convolutional layers had 192 filters of size 4 and L1_L2 regularizers while the dropout probability was set at 0.5 in each block. Following their respective three convolutional blocks, both models fed the features into a global average pooling layer instead of a fully connected layer, which is the main difference between an FCN and a typical CNN. The global average pooling layer was used over the fully connected layer to reduce the number of parameters and to avoid overfitting (Wang et al., 2016). Finally, a softmax layer produced the final output which represented the probability of the 7 different tasks.

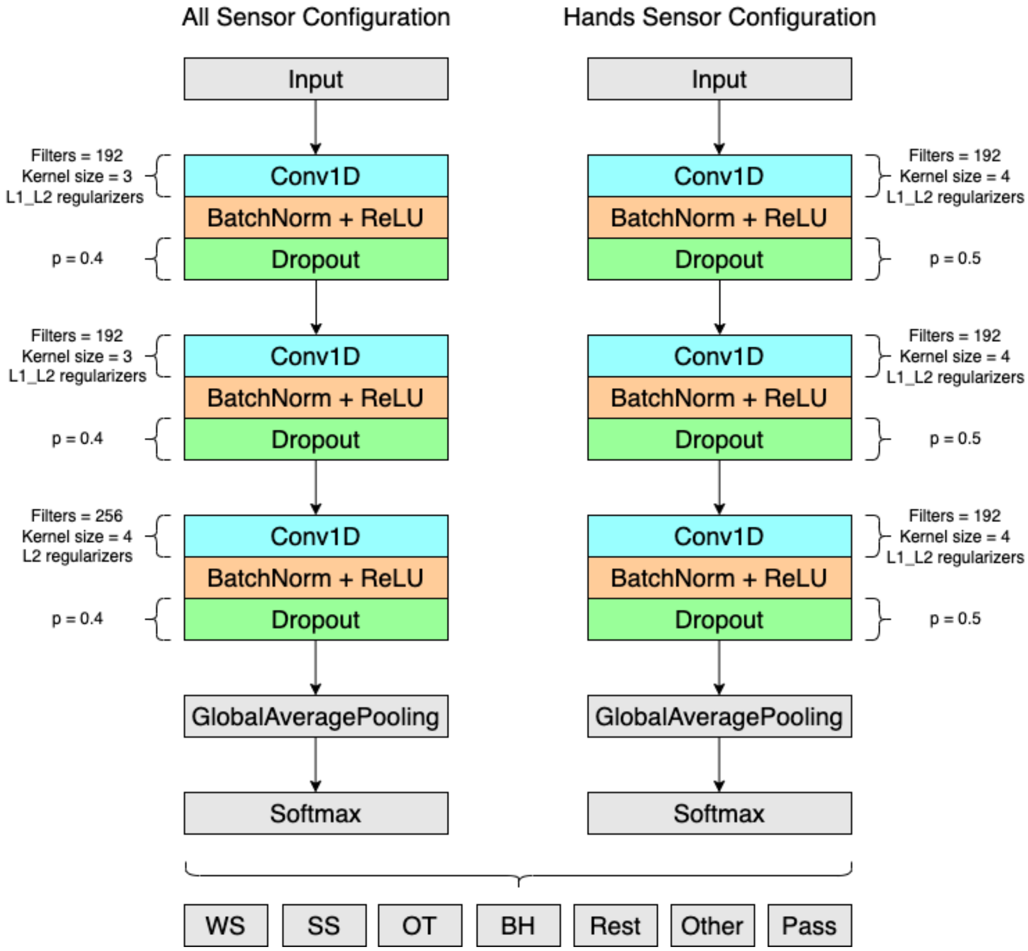

Figure 10 *A schematic view of the Fully Convolutional Network architecture for the ASC on the left, and the HSC on the right*

4.4.3 Model evaluation

From the KerasTuner results, the Adam optimizer was chosen to train the model. The sparse categorical cross entropy was used for the loss function and the accuracy was used to evaluate the model training. The learning rate (0.0001) was monitored by the validation loss and adjusted by a factor of 0.5 through a Keras callback with a patience of 20.

The FCN models were trained for a maximum of 500 epochs with a batch size of 32. A Keras early stopping callback with a patience of 50 epochs was implemented to avoid overfitting. Once fully trained, the model performances were evaluated on the test sets using the following metrics: Precision (Equation 1), Recall (Equation 2), F1 score (Equation 3), and Specificity (Equation 4). The Precision score can be interpreted as the proportion of task predictions that were actually correct (Pedregosa et al., 2011). On the other hand, recall (sensitivity) can be interpreted as the proportion of actual tasks being correctly predicted (Pedregosa et al., 2011). Meanwhile, F1 score can be interpreted as the weighted average of the precision and recall (Pedregosa et al., 2011). Therefore, a high F1 score would demonstrate high precision and recall, which would correlate to a high model performance. The F1 score was the most relevant metric to evaluate model performance for this study over accuracy due to the imbalanced number of trials per task (Pedregosa et al., 2011). Specificity can be interpreted as the rate at which the model identified a given task successfully as not another task (Pedregosa et al., 2011). In other words, a model with high specificity would have few false positives predictions. Finally, 7×7 confusion matrices showing true and predicted labels for both sensor configurations were obtained to visualize and compare model performances (Pedregosa et al., 2011).

$$(1)\ Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$(2)\ Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$(3)\ F1\ score = \frac{2\ x\ (Precision\ x\ Recall)}{Precision + Recall}$$

$$(4)\ Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

5. Results

5.1 Model performance for the All-Sensor Configuration

Overall, the FCN model had strong performances for all tasks for ASC, demonstrated by high performances of F1, Precision, Recall and Specificity scores, with all scores ranging from 90.0% up to 100.0%. For all tasks, the model revealed an average F1 score 97.1%. The task that generated the highest F1 score was the BH with 99.4%. Of note, the passing task had the lowest score overall with a 90.0% recall, which means that it correctly identified 90% of all passing tasks. When looking at the confusion matrix, the Other task had the most false positive and negative predictions with 7 and 6 respectively and was mostly confused by the resting task (Table 3 and Figure 11a).

5.2 Model performance for the Hands' Sensor configuration

Overall, the FCN model had strong performances for most tasks for HSC, with the majority of scores ranging from 90.2% to 100.0%. For all tasks, the model averaged an F1 score of 93.6%, while the task that generated the highest F1 score was the WS with 99.6%. Similar to the ASC, the passing task had the lowest score with a recall of 80.0%. Looking at the confusion matrix, the Other task had again the most false negative predictions with 18, where 17 of them were against the resting task. Also of note, the OT task had the second most false negative predictions, all against the SS task with 10 (Table 4 and Figure 11b).

Table 3 *Model performance in % for the All-Sensor Configuration. Highest scores are in bold and standard deviations in parentheses.*

FCN All Sensor Configuration

Task	Precision	Recall	F1 score	Specificity
BH	**100.0**	98.8	**99.4**	**100.0**
OT	98.8	98.8	98.8	99.8
Other	94.2	95.0	94.6	98.8
Pass	**100.0**	90.0	94.7	**100.0**
Rest	95.1	96.7	95.9	98.9
SS	99.2	96.7	97.9	99.8
WS	96.0	**100.0**	98.0	99.1
Weighted average	97.1 (2.4)	97.1 (3.3)	97.1 (1.9)	-

Table 4 *Model performance in % for the Hands' Sensor Configuration. Highest scores are in bold and standard deviations in parentheses.*

FCN Hands' Sensor Configuration

Task	Precision	Recall	F1 score	Specificity
BH	98.8	**100.0**	99.4	99.8
OT	**100.0**	87.5	93.3	**100.0**
Other	90.3	85.0	87.6	98.0
Pass	**100.0**	80.0	88.9	**100.0**
Rest	86.9	94.2	90.4	97.0
SS	90.2	**100.0**	94.9	97.7
WS	99.2	**100.0**	**99.6**	99.8
Weighted average	94.0 (5.7)	93.7 (8.3)	93.6 (4.8)	-

(a) *(b)*

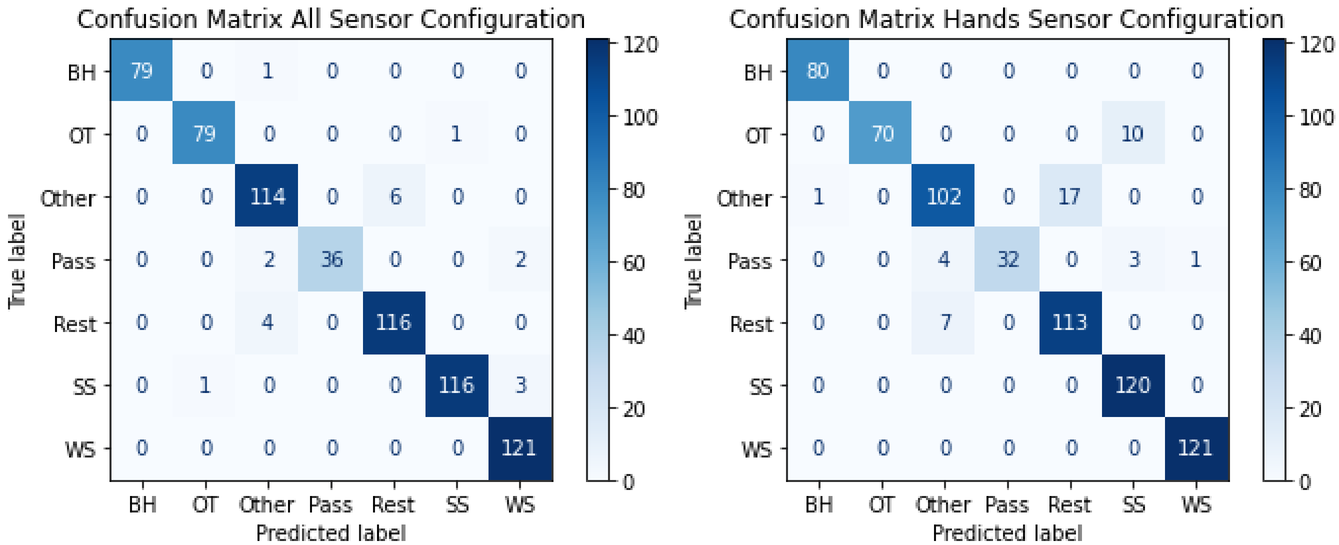

Figure 11 *Confusion matrices of the FCN model for the ASC (a) and the HSC (b)*

6. Discussion

As hypothesised, the proposed FCN model successfully classified tasks and the all-sensor configuration achieved the highest average F1 score of 97.1% compared to 93.6% for the hands-sensor configuration. From these results, the HSC configuration achieved slightly lower F1 score compared to the ASC model, but with a greatly reduced sensor count (2 vs 17, respectively).

This project used a CNN deep learning model to build a simple and reusable framework for future research in ice hockey classifying tasks. The FCN model was inspired by a baseline deep learning model first created by Wang et al. (2016) for end-to-end time series classification tasks and some of Chollet's method to reduce overfitting by adding dropout layers and adding weight regularizers to the convolutional layers. Optimization of the hyperparameters was done for both sensor configuration models using the KerasTuner to maximize classification. This method was chosen because it has been proven to reach near state-of-the-art performance (Wang et al., 2016), it was simple to adapt, and it provided a strong baseline for future research in this domain. Another advantage of using a deep learning model over feature-based classifier was CNNs ability to automatically extract features from the raw data, removing some heavy preprocessing and hand-crafted feature selection procedures. Compared to feature-based algorithms, CNNs can automatically extract and learn from features through the convolutional layers, reducing the likelihood of losing important information from the signal (Dixon et al., 2019; Wang et al., 2016).

One of the strengths of this study was the sample size. Previous work from Hardegger et al. (2015) had only four elite hockey players to classify various gameplay events including slap and wrist shots using a feature-based machine learning algorithm. Such low sample sizes make it difficult for the algorithm to produce meaningful results and generalize them over a larger population. In fact, predictive power and generalizability of a machine learning model is largely dependent on the size of the training sample (Figueroa, Zeng-Treitler, Kandula, & Ngo, 2012). The sample size of 39 participants in the current study from players of different handedness and genders provided substantial variety in shooting and skating patterns to make the task identification challenging. Given the great performance of the model for both sensor configurations, the model showed that the results can be generalized to an even greater variety of players' shooting method.

The ability to differentiate between subtle differences in similar movement tasks is challenging; for example, prior research by Swarén et al. (2019) noted minimal differences in stick blade velocity or stance between slap shots and one-timers. Hence, classification learning from acceleration data alone on tasks showing little differences could be difficult, particularly for the hands-sensor configuration of only two sensors. Previously, Hardegger demonstrated the ability of stick sensors and SVM classifiers to predict shot tasks (wrist vs slap) and player skill level (elite vs amateur) but did not attempt to classify difference in low acceleration tasks. Numerous studies (MacInnis, 2019; Villaseñor et al., 2006; Woo et al., 2004; Worobets et al., 2006; Wu et al., 2003) have shown that there are significant acceleration differences between a slap and a wrist shot, as well as between the player's skill level. Hence, it may have been expected that Hardegger and colleagues achieved 100% accuracy using the rule-based classifier to predict shot types and achieved 88.5% and 99.7% accuracy on predicting elite vs amateur players for wrist and slap shots identification, respectively. Thus, a strength of the current study was the ability to differentiate between similar dynamic skills (e.g., a one-timer and a slap shot; a wrist shot and a long-distance pass).

Overall, sensor configurations differences affected F1 scores. In particular, using only two sensors significantly reduces the size of the input signal (51 features for the ASC vs 6 for the HSC), reducing the amount of information the model could be trained on. Arguably, more sensor information from different body regions is required to better predict shooting task. Previous studies have shown that full body measures including the upper and lower limbs and torso could substantially improve task identification. For example, Michaud-Paquette et al. (2011) found accurate stationary wrist shot accuracy predictors and part of their findings required input from the lower limbs and torso kinematics, as these influenced shooter’s stability, base of support, and trunk orientation. Thus, the HSC model was trained without important information coming from the lower limbs and other body parts. However, it was chosen instead of the biomechanics sensor configuration (hands, feet, and pelvis) for example, to analyse the simplest, most feasible and most replicable sensor combination for a hockey product design company while still achieving high classification performance.

Surprisingly, the classification of wrist shots did not suffer from having only two sensors, achieving an F1 score of 99.59% vs 97.98% for the ASC. However, when examining the HSC’s precision for one-timers and recall for slap shots, they both achieved 100%, which means that no other tasks were predicted as one-timers, and all slap shots performed were predicted accurately. But One-timers’ recall achieved 87.5% and Slap shots’ precision achieved 90.23% which are mostly due to the model being confused when participants performed one-timers and misclassified some as slap shots. Other tasks that could influence recall and precision by the HSC were the Pass, Other, and Rest’s tasks. The FCN model achieved the lowest scores on the Pass recall for both sensor configurations with 90% for the ASC and 80% for the HSC, leading to confusion in discriminating between a Pass and Other, Slap and Wrist shot tasks. On review of participants' passing trials, differences in puck velocities and techniques were easy to observe; for example, some participants were very careful in their approach and made soft passes, while other participants skated in faster and made quicker passes. These variable pass approaches and the low number of pass trials contributed to difficulty in pass classification. Meanwhile, the bulk of false positive and false negative predictions were related to Other tasks, e.g. concurrent stick-handling and/or skating before and after shooting. On the other hand, the post shot included a lot of different movements since participants were not specifically told to do anything except to skate back to the starting position of the following trial. During these post shot trials, some participants were often playing randomly around with the puck, some were passing it on the board, others were simply skating slowly towards the next trial, and often participants were standing still to rest or to wait for the research tester to fix or re-secure a sensor. All of these instances were still recorded and potentially included as a trial into the Other task during the preprocessing step explained in the method section. This may have contributed to some of the Other trials' misclassification with the resting task where the acceleration was close to none, or to the passing task when participants were playing around with the puck, for example. Lastly, the model performed almost perfectly in specificity scores for both sensor configurations, which suggests that it was simple to predict which class the trials did not belong to.

Despite the strong performances of the models, there were some limitations within the current study. While external validity was increased by using IMUs and not having space limitations within the ice rink, the data collected did not completely represent a real-game scenario. First, the Xsens did not allow participants to wear their full protective equipment, they could only wear their gloves, skates, and stick. However, it's important to note that not wearing their full equipment did not modify the participants' task execution. Participants were also restricted to distinct and delimited tasks in which skating was purposely deemphasized to focus specifically on shooting-related classification. This was done by creating an Other task, where the method used significantly reduced the overall impact the skating could have had for the classification problem at hand. Additionally, the stick-handling and passing tasks were very simple and did not represent the full range of possibilities available. Therefore, it is unknown how well the models will perform in a dynamic game situation where players wear their full equipment, where there are many more skating instances and where there is more kinematic variability in the tasks performed.

Another limitation of the study was the extensive post-hoc processing steps necessary prior to train the deep learning model. For example, the labelling process required to manually label each task using the video camera footage would not be practical for future stakeholders. Therefore, optimizing the event identification process from the raw data could significantly reduce the amount of time required to process the data. This could be done by finding the acceleration peaks for each trial and selecting a predetermined time frame before and after to capture the shot.

Finally, the participants' sample consisted of elite hockey players only, raising the concern of generalizability of the models for a wider population including male and female players of lower calibre. Future studies should attempt to increase the sample size, increase the number of female players, and have a wider range of player calibre.

7. Conclusion

This study provides a strong baseline method in ice hockey shooting-related tasks classification by using a deep learning approach with input data from IMUs' three-dimensional free acceleration. The FCN model proposed successfully classified the seven different on-ice tasks (wrist, slap, backhand, one-timer, pass, rest and other). Compared to the 97.1% F1-score of the 17-sensors configuration, the hands' sensor configuration achieved 93.6% F1-score and was most ideal due to having only two sensors. Thus, the hands' 3D acceleration data alone becomes a practical option in terms of tracking player shooting skill execution in both training and game contexts, as well as feasible for hockey equipment manufactures to secure sensors inside the gloves.

Future research should expand the scope of this project and explore a greater variety of hockey tasks by including more skating and different type of players. This would increase the knowledge and efficacy in hockey event classification and close the gap towards a real-time feedback analysis. Researchers should find a solution to secure and use only sensors in the gloves, which would enable them to collect data on players wearing their full gear and recreate a more realistic game-like environment. It would also enable the possibility to pair the gloves' sensor system with a sensor-instrumented stick (strain gauges or IMUs) to evaluate players' shot performance. This along with the development of a live streaming pipeline that would classify event in real-time has the potential to provide athletes and coaches with valuable information about the players' performance in real-time and further enhance training methods towards the player's needs.

8. References

Ahmadi, A., Rowlands, D. D., & James, D. A. (2010). Development of inertial and novel marker-based techniques and analysis for upper arm rotational velocity measurements in tennis. *Sports Engineering, 12*(4), 179-188.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., . . . Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data, 8*(1), 1-74.

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys, 46*(3), 1-33. doi:10.1145/2499621

Chollet, F. (2018). *Deep learning with Python*: Simon and Schuster.

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *J Sports Sci, 37*(5), 568-600. doi:10.1080/02640414.2018.1521769

Denroche, S. K. (2020). *Evaluating the use of inertial measurement unit technology during an ice hockey shooting task.* (Master of Science Thesis). McGill University, Montreal, Quebec.

Dixon, P. C., Schutte, K. H., Vanwanseele, B., Jacobs, J. V., Dennerlein, J. T., Schiffman, J. M., . . . Hu, B. (2019). Machine learning algorithms can classify outdoor terrain types during running using accelerometry data. *Gait Posture, 74*, 176-181. doi:10.1016/j.gaitpost.2019.09.005

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC medical informatics and decision making, 12*(1), 1-10.

Fulton, S. K., Pyne, D. B., Hopkins, W. G., & Burkett, B. (2009). Variability and progression in competitive performance of Paralympic swimmers. *Journal of sports sciences, 27*(5), 535-539.

Gupta, S. (2021). Deep learning based human activity recognition (HAR) using wearable sensor data. *International Journal of Information Management Data Insights, 1*(2). doi:10.1016/j.jjimei.2021.100046

Hammerla, N. Y., Halloran, S., & Plotz, T. (2016). Deep, Convolutional, and Recurrent Models for Human Activity Recognition using wearables.

Hardegger, M., Ledergerber, B., Mutter, S., Vogt, C., Seiter, J., Calatroni, A., & Tröster, G. (2015). *Sensor technology for ice hockey and skating*. Paper presented at the 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN).

Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing, 62*, 915-922. doi:10.1016/j.asoc.2017.09.027

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery, 33*(4), 917-963. doi:10.1007/s10618-019-00619-1

Jang, J., Ankit, A., Kim, J., Jang, Y. J., Kim, H. Y., Kim, J. H., & Xiong, S. (2018). A Unified Deep-Learning Model for Classifying the Cross-Country Skiing Techniques Using Wearable Gyroscope Sensors. *Sensors (Basel), 18*(11). doi:10.3390/s18113819

Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Multi-sensor golf swing classification using deep CNN. *Science Direct, 129*, 59-65.

Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a Deep Convolutional Neural Network. *Data Mining and Knowledge Discovery, 31*(6), 1678-1705.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*.

Krüger, A., & Edelmann-Nusser, J. (2010). Application of a full body inertial measurement system in alpine skiing: A comparison with an optical video based system. *Journal of Applied Biomechanics, 26*(4), 516-521.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

MacInnis, N. R. (2019). *Does X-factor exist in ice hockey slap shots? Comparison of elite versus recreational players thorax to pelvis coordination.* (Master of Science Thesis). McGill University, Montreal, Quebec.

Michaud-Paquette, Y., Magee, P., Pearsall, D. J., & Turcotte, R. A. (2011). Whole-body predictors of wrist shot accuracy in ice hockey: a kinematic analysis. *Sports Biomech, 10*(1), 12-21. doi:10.1080/14763141.2011.557085

Michaud-Paquette, Y., Pearsall, D. J., & Turcotte, R. A. (2008). Predictors of scoring accuracy: ice hockey wrist shot mechanics. *Sports Engineering, 11*(2), 75-84. doi:10.1007/s12283-008-0009-9

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., & others. (2019). KerasTuner. Retrieved from https://github.com/keras-team/keras-tuner

Pearsall, D. J., Montgomery, D. L., Rothsching, N., & Turcotte, R. A. (1999). The influence of stick stiffness on the performance of ice hockey slap shots. *Sports Engineering, 2,* 3-12.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The journal of machine learning research, 12*, 2825-2830.

Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors--a review of classification techniques. *Physiol Meas, 30*(4), R1-33. doi:10.1088/0967-3334/30/4/R01

Renaud, P. J., Robbins, S. M., Dixon, P. C., Shell, J. R., Turcotte, R. A., & Pearsall, D. J. (2017). Ice hockey skate starts: a comparison of high and low calibre skaters. *Sports Engineering, 20*(4), 255-266.

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. (2021). The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min Knowl Discov, 35*(2), 401-449. doi:10.1007/s10618-020-00727-3

Shahar, N., Ghazali, N., As'ari, M., & Swee, T. (2020). *Wearable Inertial Sensor for Human Activity Recognition in Field Hockey: Influence of Sensor Combination and Sensor Location.* Paper presented at the Journal of Physics: Conference Series.

Swarén, M., Söhnlein, Q., Stöggl, T., & Björklund, G. (2019). *Using 3D Motion Capture to Analyse Ice Hockey Shooting Technique on Ice*. Paper presented at the Proceedings of the 7th International Conference on Sport Sciences Research and Technology Support.

Villaseñor, A., Turcotte, R. A., & Pearsall, D. J. (2006). Recoil effect of the ice hockey stick during a slap shot. *Journal of Applied Biomechanics*(22), 201-210.

Wang, Z., Yan, W., & Oates, T. (2016). *Time series classification from scratch with deep neural networks: A strong baseline.* Paper presented at the 2017 International joint conference on neural networks (IJCNN).

Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring Hitting Load in Tennis Using Inertial Sensors and Machine Learning. *Int J Sports Physiol Perform, 12*(9), 1212-1217. doi:10.1123/ijspp.2016-0683

Woo, T., Loh, J., Turcotte, R. A., & Pearsall, D. J. (2004). The ice hockey slap shot, elite versus recretional. *ISBS*, 511-514.

Worobets, J. T., Fairbairn, J. C., & Stefanyshyn, D. J. (2006). The influence of shaft stiffness on potential energy and puck speed during wrist and slap shot in ice hockey. *Sports Engineering, 9*, 191-200.

Wu, T.-C., Pearsall, D. J., Hodges, A., Turcotte, R. A., Lefebvre, R., Montgomery, D., & Bateni, H. (2003). The performance of the ice hockey slap and wrist shots: the effects of stick construction and player skill. *Sports Engineering, 6*(1), 31-39.

Xsens. (2018). Xsens MVN User Manual. In.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). *Time series classification using multi-channels deep convolutional neural networks.* Paper presented at the International conference on web-age information management.