Categorization of Tennis Swing Using a Recurrent Neural Network in Human Activity Recognition

Matthew Li
University of Chicago

Abstract

Human Activity Recognition (HAR) is critical in quantifying tennis player performance because of the lack of current quantifiable metrics of player performance. Most quantifiable models aim to summarize player performance based on the results of each action rather than the form of the player. Maximizing a classification network for basic similar tennis strokes can be used to analyze stroke frequency (beginner to professional) in tennis and other sports requiring different forms. Current models aiming at quantifying form detection only categorize forms for sports or activities that require slower movement (e.g., preventing injury during strenuous physical activity in the gym or categorizing behavior during housework) and require a large dataset for accurate results. By utilizing recurrent neural networks using transformed and augmented data, a model can detect and classify an athlete's tennis strokes to a high level of accuracy with significantly fewer amounts of training data required. This text details the use of different recurrent neural networks and types of data transformation to maximize the accuracy and confidence of the network in categorizing the data from a junior semi-competitive tennis player while minimizing the run time.

1 IMPLICATIONS

2 INTRODUCTION

Human activity recognition (HAR) can classify or quantify player movements and behaviors and has important applications in sports (Demrozi et al., 2020; Host & Ivašić-Kos, 2022), strength training (Ganesh et al., 2020; Hussain et al., 2022), and physiotherapy (Billiet et al., 2016). However, HAR requires a large amount of classified training data collected uniquely for each application, this is difficult and expensive to collect, which limits the application of HAR (J. Chen et al., 2021; Kim et al., 2010). Effective data compression can enable easier training of neural networks by pre-extracting the most important features which are relevant to predicting the relevant output (Florea & Roman, 2021).

Most applications of HAR rely on neural networks which extract features from input data and classify outcomes (Cruciani et al., 2020; Oniga & Sütő, 2014; Uddin & Soylu, 2021). This means, it categorizes the quality of the shot by sensor data on the speed, location of where the ball lands, spin, and other data that signals the exact location and description of the ball hit (Busuttil et al., 2022; Pedro et al., 2022) Other forms of determining player metrics and performance is through standardized ratings like Association of Tennis Professionals (ATP) standings or Universal Tennis Rating (UTR) (Hunt, 2020).

Alternatively, another source of input data could come from utilizing sensors via an Apple watch or wristband sensor on the body of a tennis player, although this could obstruct the full potential of a shot from a tennis player due to restriction of movement (Ganser et al., 2021). Fixing this issue of mobility while still tracking the movement of a player across the court could be analyzed using video data (Y.-C. Jiang et al., 2009; Lara et al., 2018). However, video data is not optimal due to the variability in player distance from camera, different color clothes a player might wear, etc. Additionally, although there are numerous cameras on every match in the ATP, equipment that allows for 4k resolution and 3D visualization at over 200 frames per second (FPS) is not viable to the average tennis player (Renò et al., 2017).

This study shows that application of a pre-trained OpenPose wireframe neural network (Cao et al., 2021; Qiao et al., 2017), allows over 100 fold compression of input training data (60 FPS, 720p resolution) by extracting the most important features. The data from the OpenPose model (Cao et al., 2021) is then used to train a use-case specific

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, TBD, USA. PMLR: Volume XXX. Copyright 2023 by the author(s).

neural network achieving a testing accuracy 99.62% across three tennis strokes (forehands, backhands, and backhand slice) with 28 minutes of video training data. By training the use-case specific network on data which is highly compressed the training time and difficulty is dramatically reduced.

Another challenge posed for the compression of data by over 100 fold came with the volatility from stroke to stroke in a practice session (Colomar et al., 2020; Knudson & Elliott, 2004). This study utilizes a long short-term memory (LSTM) model combined with convolutional neural networks to create a categorization model that can maintain the same level of accuracy even with high volatility datasets that seem like they resemble no pattern (Borovkova & Tsiamas, 2019; Lv et al., 2022). The first step taken to test this theory was by categorizing essential and common tennis strokes. This study then adds in strokes that look relatively similar based on keypoints and stroke patterns. The forehand and backhand were the two essential strokes chosen for the first set of testing because they make up the majority of strokes in a match (González-González et al., 2018; Muhamad et al., 2011). The backhand slice was deemed a similar stroke to the backhand, and therefore should pose as more of a challenge to the model (personal communication C. Anderson).

The computer-generated model is composed of 3 major components:

- Splitting the action for analysis into smaller, easier-to-analyze sizes using recursive partitioning-like methodology (Strobl et al., 2009)
- Feature extraction from the activity to analyze joint movement as well as extra data like wrist rotation, vertical displacement, etc.
- Classification process to identify the stroke in each frame.

This paper discusses the first steps to achieving a model that can detect and classify a player's movement and further steps to achieve this goal model. In the literature, numerous techniques are tested to determine whether creating a model to detect slight distinctions between strokes is possible with time to train and time to detect as limits to the model. Most research in improving athletes' form is tested through gym exercises where there are substantially fewer sudden movements, and the repetitions come in much smaller batches (K.-Y. Chen et al., 2022). The research done shows promising results that the model can, in fact, achieve a high level of accuracy in determining when a tennis player's strokes deviate from the correct form. The second objective of this paper is to determine the best network and methodology to analyze a tennis player's strokes efficiently.

The rest of the paper is organized as follows: Section 2 discusses related work using models to detect correct forms in tennis and other sports. Section 3 discusses the methodology used. Section 4 discusses the model's results. Section 5 is further discussion on the model and its applications.

3 RELATED WORK

3.1 Classification networks

Classification of tennis strokes has been performed previously utilizing a neural network based on varying data points. The most common is using sensor data through the use of a wristband on tennis players (Ganser et al. 2021; Silvia Vinyes Mora 2017). The data required for these models was nearly 5700 stroke repetitions, greater than 6 times as large as the 909 stroke repetitions used in the dataset for this model.

3.2 Data collection

Additionally, in large datasets, especially for Human Activity Recognition (HAR), finding accurate data sets is extremely hard. Previous models classifying tennis strokes utilized data from amateur tennis players where the form from player to player varied radically (Ganser et al. 2021). These models required each participant to attach the wearable sensors and record themselves going through numerous repetitions of each stroke. This strategy for gathering video data for the player(s)' strokes seems like the only viable and repeatable option.

The work in this study is unique because the player recorded in this dataset is a junior competitive player in the UTR and USTA system and the strokes recorded remained much more consistent with that of a professional tennis player. Additionally, each of the strokes was monitored and approved by a professional tennis coach. This paper also differs because the model implemented uses wireframe data rather than sensor data, removing the need for wearable obstructions for the tennis player and achieving a more reliable and applicable use for the model. Finally, this study utilizes less than 17% of the previously most efficient model whilst maintaining an accuracy of 99.62% in comparison to that of 96% in previous models through the use of sliding windows for more accuracy.

4 PREPROCESSING DATA

The proposed methodology consists of 4 main steps and specific implementations and testing for optimization of each step. The input to this model is the raw video captured from a widely available video recording device like an iPhone. The output of the model is a complete list of problems in relation to the stroke captured in the video.

	Forehand	Backhand	Backhand slice
Number of videos recorded	103	100	100
Strokes per video	3	3	3

Table 1: The table above describes the number of videos recorded for each stroke and number of strokes record per video.

For the sake of creating a model, given the limited computing abilities of the computer I have at the moment, the output was simplified down to identifying the difference between a forehand and a backhand, two basic strokes in tennis. The following steps would be to obtain a computer with higher computing power. Then, additional labeled input data would be recorded and added, then trained into a new model to detect imperfections in a tennis player's strokes.

4.1 Video format conversion

Due to the goal of implementation of this model to be widely available to different video cameras that record in different qualities and formats, the first step was to convert all of the videos into a standardized format: .mp4. The data was collected using an iPhone 11 camera. The import file was captured at 60 fps at 1080p and the video extension was .mov. The preprocessing done on the input vidoes was the conversion to .mp4. The fps and the quality of the videos remained the same.

4.2 OpenPose

Utilizing the OpenPose model provided by PyTorch (Cao et al., 2021), this project uses a region-based convolutional neural network. This pretrained RCNN model (Cao et al., 2021; Sherstinsky, 2020; Xiao et al., 2020) was utilized to gather the features used in identifying key parts of the body. The body points include the nose, left eye, right eye, and 14 other key points on the body.

The implementation of the wireframe was through an OpenPose network trained utilizing a dataset from UCI (Cao et al., 2021).

4.3 Pre-processing for prediction model

The information given by the OpenPose (Cao et al., 2021) wireframe was formatted as follows:

- X-coordinate
- Y-coordinate
- Visibility

Each of the 17 points contained these three features for a total of 51 data points per frame. However, the data needed

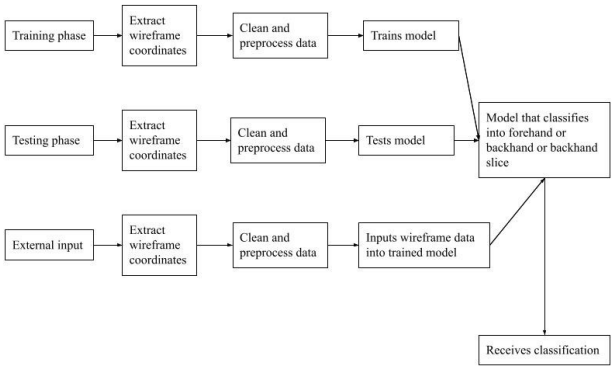

Figure 1: Flowchart for stroke classification process

to be normalized in order to ensure a consistent starting point in the data. The data was then normalized to the nose (Singh & Singh, 2020) by setting the nose as (0,0) and subtracting the (x,y) coordinate values from each other keypoint. Since the nose point was always the value (0,0), the preprocessing omitted this singular data point.

The visibility data point (how well the keypoint was exposed to a camera; eg. if the keypoint was hidden behind the body it would be 0 and if the keypoint was fully exposed it would be 1) did not offer anything valuable to the prediction model so each of those points was cut out. After preprocessing, each frame resulted in 32 data points.

4.4 Stroke recognition

This is a supervised machine learning algorithm (T. Jiang et al., 2020). Strokes video composition consists of around 100 videos of each forehand and backhand to ensure the model had enough samples to distribute into test and train.

The supervised learning algorithms I used to classify the strokes include:

- LSTM (Hochreiter & Schmidhuber, 1997; Sherstinsky, 2020) - the LSTM is able to accurately predict time-series data, the type of data that this model used to analyze. This recurrent neural network at its base is perfect for video analysis.
- CNN LSTM (Mutegeki & Han, 2020). - adding a convoluted neural network (CNN) allows for more spacial correlation of the data.
- Conv LSTM (Ge et al., 2019; Shi et al., 2015) - the Conv LSTM is one order of magnitude faster than a regular LSTM, allowing for faster classification and creation of the model.

Creating the model and testing it includes three phases:

- Training: the training phase consists of using the training videos (80% of the data) and the training labels to feed into the training step of the model. This data was split up in the preprocessing step and the corresponding labels were made by the player who hit and recorded the shots.
- Testing: the testing phase consists of using the testing videos (20% of the data) and using the testing labels to check if the model would assume the correct classification. This data was split up in the preprocessing step. This step was not manually done. Instead, it was completed by a built-in function in the keras model library.
- Using the mode: the user phase consists of using a new input video without a label in the data set and running the model for the user to see what the classification of the stroke is.

4.5 Sliding windows

Video processing includes the analysis of multiple frames. The LSTM model (Majd & Safabakhsh, 2020; Sherstinsky, 2020) would achieve maximum accuracy if the input data included overlapped data to have built-in memory when creating the model.

- The original data shape was: (total frames, all vars)
- The goal shape was: (total sliding frames, total frames per sliding window, all vars)

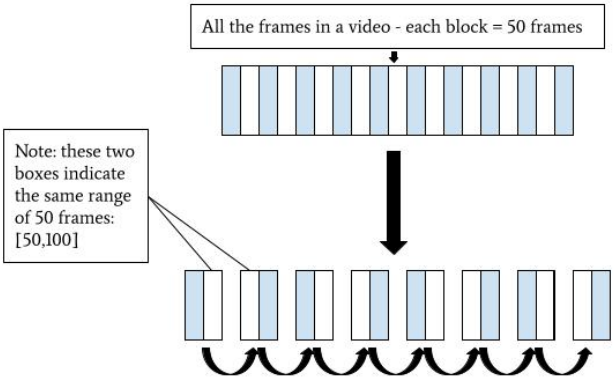

Figure 2: The figure above denotes the data augmentation into sliding frames.

The transcription of the data from the original to the goal shape was done by looping through the data and choosing a set interval of 100 frames per sliding window with a 50-frame overlap from window to window. The extra frames would be discarded if there were not enough to create another full sliding window of 100 frames.

5 Results

A CNN-LSTM model (Mutegeki & Han, 2020) was trained with 303 videos which were up to eight seconds to classify between forehand, backhand, or backhand slice tennis swings with an accuracy of 100% for forehand, 99.86% for backhand, and 99.91% for backhand slice.

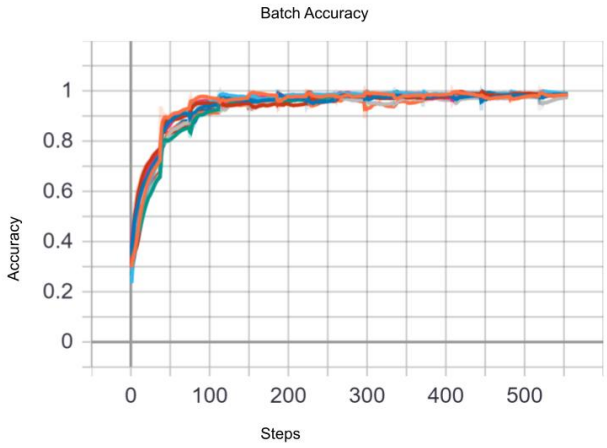

Figure 3: Batch Accuracy for all 10 experiments across 550 steps.

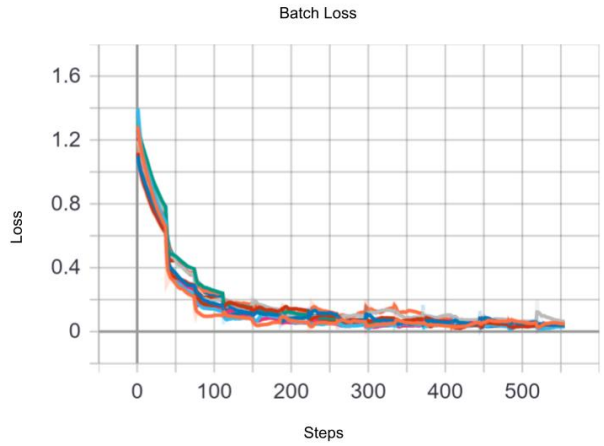

Figure 4: Description: Batch Loss for all 10 experiments across 550 steps.

Figures 3 and 4 indicate that the number of steps required to achieve a comfortable threshold of 95% accuracy was only, on average across the 10 experiments, around 100 steps. Additionally, to achieve 99% accuracy, on average across the 10 steps, the model only required around 150 steps.

Figures 5 and 6 indicate that the number of epochs chosen allowed for a perfect logarithmic graph of results for epoch accuracy where the top of the curve just flattens at 12 epochs, achieving an accuracy of 95%, on average, 4 epochs, and 99% 12 epochs. Additionally, epoch loss was

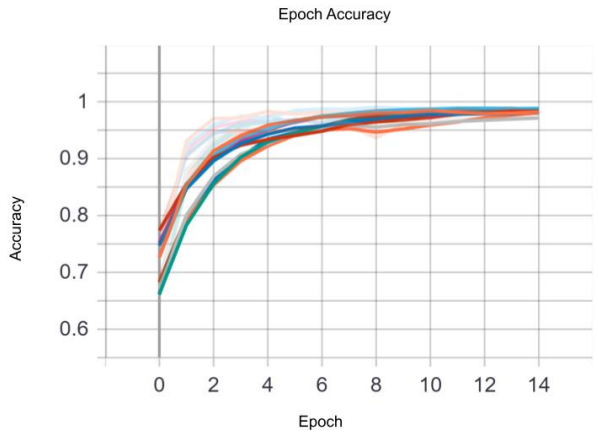

Figure 5: Epoch Accuracy for all 10 experiments across 14 epochs.

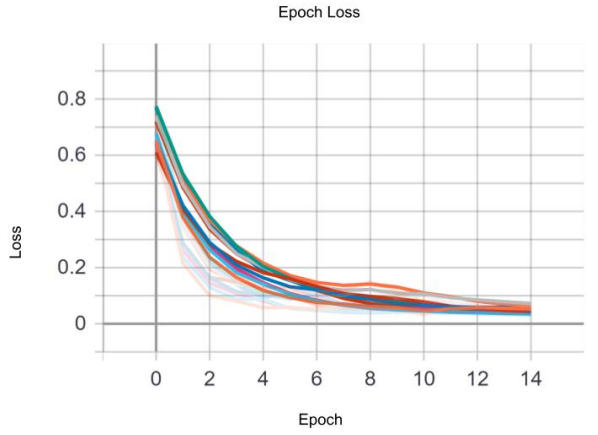

Figure 6: Epoch Loss for all 10 experiments across 14 epochs.

minimized across the 14 epochs and reached a low at an average of 0.05% loss.

The model created utilizing a CNN-LSTM (Mutegeki & Han, 2020) could categorize at an extremely high confidence between a forehand, backhand, and backhand slice with 303 8 seconds or less videos containing 3 iterations of each stroke with less than 1 minute of training per iteration of model and less than 15 seconds per categorization from raw video footage. This proved the following:

5.1 LSTM with 4 minutes of training data distinguishes between forehand, backhand, and backhand slice with 99.62% testing accuracy

Across a collection of 242 training and 61 testing videos wireframes including 17 points were constructed from 2D pose time series dataset with the Openpose algorithm (Fig. 3) (Cao et al., 2021). The dataset included 103, 100, 100 occurrences of forehand, backhand, and backhand slice tennis strokes respectively. Out of the 17 points, the nose was used as the center of the body and was set as the origin.

Figure 7: The diagram above labels each of the 17 key points given by Openpose on a single frame taken from a sample backhand video.

The LSTM was trained with 2 hidden layers (Salman et al., 2018) using the key points from all 242 training videos and trained a model capable of distinguishing the 3 strokes at an average 99.62% accuracy across 10 trials in under 4 minutes of training per trial. In Figure 7, trial 1 of of the experiment created a model with an accuracy of 100% across all 3 strokes and correctly identified this video and sliding frame as a forehand.

In Figure 7, the model was given the task of classifying between a forehand, backhand, and backhand slice. The model successfully classified the stroke as a backhand slice, while the confidence of the model actually rose to an average above 99.999%. This result is surprising because despite the added augmented data, the confidence of the model to detect between the three strokes increased. According to the expert in the field, professional coach Craig Anderson, the backhand and backhand slice are relatively similar strokes. The model was still easily able to detect that this was a backhand slice in all 8 of the 100 frame sliding windows.

	Forehand	Backhand	Backhand Slice
Test Precision	100.0	99.857	99.906
Test Recall	100.0	99.928	99.811
Test Fscore	100.0	99.892	99.858

Table 2: The table above gives accuracy metrics averaged across 10 trials for the LSTM model detecting between forehand, backhand, and backhand slice

5.2 No loss in accuracy when key features are hidden behind other parts of the body

In Figure 7, the wireframe model was able to detect 15 out of 17 of the keypoints reliably and was forced to use an estimation algorithm to detect the location of the last 2 points. These 2 points were still included in the dataset presented to the CNN-LSTM network and proved to be valuable data points. The maximum number of keypoints not visible at any frame during this entire experiment was 7 when the body was fully turned and the entire left side of the body was not visible to the camera.

Additionally, including these points was one of the unique aspects of this study. Keeping these points in still allowed for a 99.62% total accuracy and helped with categorization as the predicted locations of the points were still fed into the model and after examining frames where many key points were hidden, the model was still accurate in classifying the stroke.

5.3 High confidence interval in multiple sliding frames even when strokes are considered "similar" and "hard to differentiate" by expert tennis coach

The backhand and backhand slice are relatively similar strokes. In the initial algorithm testing stage, the forehand was compared to the backhand to prove model viability. The second test of the model added a backhand slice where the takeback (first part of the stroke) looks very similar between both the backhand and backhand slice. The distinct difference between the strokes is the wrist rotation and grip on the racket, as well as direction of the finish of the full swing on the ball. Despite the similarities in the first half of the stroke, the model was able to detect the difference between a backhand and backhand slice, even during the first sliding windows of 100 frames with a test precision of 99.86% for backhand and 99.91% for backhand slice as shown in Table 2.

6 DISCUSSION

The results in this paper show that a LSTM model with limited amounts of data can accurately and confidently predict the classification of a tennis stroke between a forehand, backhand, and backhand slice. These three strokes have a relatively large amount of differences, but with no drop in accuracy and confidence of the model when the backhand slice was added to the data set, this model can confidently classify between two similar strokes (backhand and backhand slice). Future applications of this model will be to determine the error in a player's strokes. Because this model can classify between two strokes where the difference in arm, leg, and body positioning is just a couple inches apart, with further testing, this model should be able to classify the errors of a tennis player. By collecting the same amount of data or more for this error classification model, this model could be outputted for wide use by all tennis players.

Another step in this study would be to create a 3D wireframe (Rani et al. 2021; Kaneko et al. 2019) by adding another camera angle during each iteration of the stroke. This would involve rerecording the entire dataset to ensure that each of the multiple camera angles is recorded during the same repetition of the stroke.

This model also shows promise because of how fast it reached the 99%+ accuracy threshold over the total number of steps, first hitting that threshold at around step 150 on average. The results show that reducing the number of steps per epoch would still allow the model to have an astounding accuracy, reducing the training time without significant repercussions on the model's ability to classify.

The third result of this model indicated that it had no difficulty in classifying between the two similar strokes. The next step for this model would be for form error detection. This form error detection is applicable in any action/sport/activity that requires accurate form or technique (Gajjala & Chakraborty, 2021). Some examples of these applications, but not limited to are: violin, baseball, running, typing on a keyboard efficiently, etc.

The International Tennis Foundation (ITF) recently reported 89 million tennis players internationally in 2017, a player count that continues to grow. Most of these players do not have access to a coach capable of properly identifying these areas for growth due to a limited number of qualified coaches and funds for regular lessons. Coaches are essential to the consistent improvement of a tennis player (Anderson et al., 2021; Keller et al., 2022). In addition to growth, since tennis is a very physical intensive sport, injuries are quite common (McCurdie et al., 2017; Pluim et al., 2006) and limiting the maximum number of repetitions of a stroke consecutively or measuring stroke consistency can help eliminate the chance of injury (Hunt, 2020; Lambrich & Muehlbauer, 2022).

7 Acknowledgements

Thank you to professional tennis coach Craig Anderson for training the player used for collecting the data in the 303 video dataset across all 3 strokes, for verifying that each of the frames contained the correct form at every critical point, and confirming that the backhand and backhand slice are similar strokes in tennis.

References

Anderson, E., Stone, J. A., Dunn, M., and Heller, B. (2021). Coach approaches to practice design in performance tennis. *Int. J. Sports Sci. Coach.*, 16(6):1281–1292.

Billiet, L., Swinnen, T., Westhovens, R., de Vlam, K., and Van Huffel, S. (2016). Activity recognition for physical therapy: fusing signal processing features and movement patterns. In *Proceedings of the 3rd International Workshop on Sensor-based Activity Recognition and Interaction*, number Article 5 in iWOAR '16, pages 1–6, New York, NY, USA. Association for Computing Machinery.

Borovkova, S. and Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *J. Forecast.*, (for.2585).

Busuttil, N. A., Reid, M., Connolly, M., Dascombe, B. J., and Middleton, K. J. (2022). A kinematic analysis of the upper limb during the topspin double-handed backhand stroke in tennis. *Sports Biomech.*, 21(9):1046–1064.

Cao, Hidalgo, Simon, Wei, and Sheikh (2021). OpenPose: Realtime Multi-Person 2D pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43:172–186.

Chen, J., Sun, Y., and Sun, S. (2021). Improving human activity recognition performance by data fusion and feature engineering. *Sensors*, 21(3).

Chen, K.-Y., Shin, J., Hasan, M. A. M., Liaw, J.-J., Yuichi, O., and Tomioka, Y. (2022). Fitness movement types and completeness detection using a Transfer-Learning-Based deep neural network. *Sensors*, 22(15).

Chitnis, A. and Vaidya, O. (2014). Performance assessment of tennis players: Application of DEA. *Procedia - Social and Behavioral Sciences*, 133:74–83.

Colomar, J., Baiget, E., and Corbi, F. (2020). Influence of strength, power, and muscular stiffness on stroke velocity in junior tennis players. *Front. Physiol.*, 11:196.

Cruciani, F., Vafeiadis, A., Nugent, C., Cleland, I., McCullagh, P., Votis, K., Giakoumis, D., Tzovaras, D., Chen, L., and Hamzaoui, R. (2020). Feature learning for human activity recognition using convolutional neural networks. *CCF Transactions on Pervasive Computing and Interaction*, 2(1):18–32.

Demrozi, F., Pravadelli, G., Bihorac, A., and Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8:210816–210836.

Florea, A. R. and Roman, M. (2021). Artificial neural networks applied for predicting and explaining the education level of twitter users. *Social Network Analysis and Mining*, 11(1):112.

Gajjala, K. S. and Chakraborty, B. (2021). Human activity recognition based on LSTM neural network optimized by PSO algorithm. In *2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII)*, pages 128–133.

Ganesh, P., Idgahi, R. E., Venkatesh, C. B., Babu, A. R., and Kyrarini, M. (2020). Personalized system for human gym activity recognition using an RGB camera. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, number Article 21 in PETRA '20, pages 1–7, New York, NY, USA. Association for Computing Machinery.

Ganser, A., Hollaus, B., and Stabinger, S. (2021). Classification of tennis shots with a neural network approach. *Sensors*, 21(17).

Ge, H., Yan, Z., Yu, W., and Sun, L. (2019). An attention mechanism based convolutional LSTM network for video action recognition. *Multimed. Tools Appl.*, 78(14):20533–20556.

González-González, I., Rodríguez-Rosell, D., Clavero-Martín, D., Mora-Custodio, R., Pareja-Blanco, F., García, J. M. Y., and González-Badillo, J. J. (2018). Reliability and accuracy of ball speed during different strokes in young tennis players. *Sports Med Int Open*, 2(5):E133–E141.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term memory. *Neural Comput.*, 9(8):1735–1780.

Host, K. and Ivašić-Kos, M. (2022). An overview of human action recognition in sports based on computer vision. *Heliyon*, 8(6):e09633.

Hunt, S. S. (2020). Physiological performance characteristics of universal tennis. *Dissertations and Theses @ UNI*, (1061):48.

Hussain, A., Zafar, K., Baig, A. R., Almakki, R., AlSuwaidan, L., and Khan, S. (2022). Sensor-Based gym physical exercise recognition: Data acquisition and experiments. *Sensors*, 22(7).

Jiang, T., Gradus, J. L., and Rosellini, A. J. (2020). Supervised machine learning: A brief primer. *Behav. Ther.*, 51(5):675–687.

Jiang, Y.-C., Lai, K.-T., Hsieh, C.-H., and Lai, M.-F. (2009). Player detection and tracking in broadcast tennis video. In *Advances in Image and Video Technology*, pages 759–770. Springer Berlin Heidelberg.

Kaneko, T., Takahashi, J., Ito, S., and Tobe, Y. (2019). A hybrid map with permanent 3D wireframes and temporal line segments toward Long-Term visual localization. *SICE Journal of Control, Measurement, and System Integration*, 12(4):149–155.

Keller, M., Schweizer, Jonas, and Gerber, M. (2022). Pay attention! the influence of coach-, content-, and player-related factors on focus of attention statements during tennis training. *EJSS*, pages 1–9.

Kim, E., Helal, S., and Cook, D. (2010). Human activity recognition and pattern discovery. *IEEE Pervasive Comput.*, 9(1):48.

Knudson, D. and Elliott, B. (2004). Biomechanics of tennis strokes. In Hung, G. K. and Pallis, J. M., editors, *Biomedical Engineering Principles in Sports*, pages 153–181. Springer US, Boston, MA.

Lambrich, J. and Muehlbauer, T. (2022). Physical fitness and stroke performance in healthy tennis players with different competition levels: A systematic review and meta-analysis. *PLoS One*, 17(6):e0269516.

Lara, J. P. R., Vieira, C. L. R., Misuta, M. S., Moura, F. A., and Barros, R. M. L. d. (2018). Validation of a video-based system for automatic tracking of tennis players. *Int. J. Perform. Anal. Sport*, 18(1):137–150.

Lv, P., Wu, Q., Xu, J., and Shu, Y. (2022). Stock index prediction based on time series decomposition and hybrid model. *Entropy*, 24(2).

Majd, M. and Safabakhsh, R. (2020). Correlational convolutional LSTM for human action recognition. *Neurocomputing*, 396:224–229.

McCurdie, I., Smith, S., Bell, P. H., and Batt, M. E. (2017). Tennis injury data from the championships, wimbledon, from 2003 to 2012. *Br. J. Sports Med.*, 51(7):607–611.

Muhamad, T. A., Rashid, A. A., Razak, M. R. A., and Salamuddin, N. (2011). A comparative study of backhand strokes in tennis among national tennis players in malaysia. *Procedia - Social and Behavioral Sciences*, 15:3495–3499.

Mutegeki, R. and Han, D. S. (2020). A CNN-LSTM approach to human activity recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 362–366.

Oniga, S. and Sütő, J. (2014). Human activity recognition using neural networks. In *Proceedings of the 2014 15th International Carpathian Control Conference (ICCC)*, pages 403–406.

Pavai, A. T. (2018). *Sensor-based human activity recognition using bidirectional lstm for closely related activities*. PhD thesis, California State University, San Bernardino.

Pedro, B., João, F., Lara, J. P. R., Cabral, S., Carvalho, J., and Veloso, A. P. (2022). Evaluation of upper limb joint contribution to racket head speed in elite tennis players using IMU sensors: Comparison between the Cross-Court and Inside-Out attacking forehand drive. *Sensors*, 22(3).

Pluim, B. M., Staal, J. B., Windler, G. E., and Jayanthi, N. (2006). Tennis injuries: occurrence, aetiology, and prevention. *Br. J. Sports Med.*, 40(5):415–423.

Qiao, S., Wang, Y., and Li, J. (2017). Real-time human gesture grading based on OpenPose. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6.

Rani, S., Ghai, D., and Kumar, S. (2021). Construction and reconstruction of 3D facial and wireframe model using syntactic pattern recognition.

Renò, V., Mosca, N., Nitti, M., D'Orazio, T., Guaragnella, C., Campagnoli, D., Prati, A., and Stella, E. (2017). A technology platform for automatic high-level tennis game analysis. *Comput. Vis. Image Underst.*, 159:164–175.

Salman, A. G., Heryadi, Y., Abdurahman, E., and Suparta, W. (2018). Single layer & multi-layer long Short-Term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Comput. Sci.*, 135:89–98.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long Short-Term memory (LSTM) network. *Physica D*, 404:132306.

Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., and Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *NIPS*.

Silvia Vinyes Mora, W. K. (2017). Computer vision and machine learning for In-Play tennis analysis: Framework, algorithms and implementation. *University of London Imperial College of Science, Technology and Medicine Department of Computing*, page 235.

Singh, D. and Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.*, 97:105524.

Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods*, 14(4):323–348.

Uddin, M. Z. and Soylu, A. (2021). Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. *Sci. Rep.*, 11(1):1–15.

Vonstad, E. K., Vereijken, B., Bach, K., Su, X., and Nilsen, J. H. (2021). Assessment of machine learning models for classification of movement patterns dur-

ing a weight-shifting exergame. *IEEE Transactions on Human-Machine Systems*, 51(3):242–252.

Wang, Y., Yang, X., Wang, L., Hong, Z., and Zou, W. (2022). Return strategy and machine learning optimization of tennis sports robot for human motion recognition. *Front. Neurorobot.*, 16:857595.

Xiao, Y., Wang, X., Zhang, P., Meng, F., and Shao, F. (2020). Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information. *Sensors*, 20(19).