2017 16th IEEE International Conference on Machine Learning and Applications

Wearable Motion Sensor Based Analysis of Swing Sports

Akash Anand*, Manish Sharma, Rupika Srivastava, Lakshmi Kaligounder and Divya Prakash
Advanced Technology Labs
Samsung R&D Institute
Bangalore, India
*Email: *akashanand.iitd@gmail.com*

***Abstract*—Recent trends show that wearable devices with embedded motion sensors are being utilized for enriching user experience in health and fitness by tracking an individual's physical activities such as walking, running, cycling etc. Sports is another essential domain where these sensors can be used to provide valuable information. Particularly for swing sports like Tennis, Badminton and Golf, sensors on wrist-worn wearables can easily be used to give insights into players' games for improvement and preventing injuries that ensue from incorrect techniques. In this paper we propose the design of a sports analytics system with underlying methodologies that efficiently distinguish intricacies of players' hand movements for a given sport. Under this system, we discuss generalized approaches for detecting shots. We also propose and compare two novel techniques for shot classification, one using correlation based feature selection with mRMR, and another based on CNN and BLSTM neural networks. Our commercialized applications TennisTraq and ShuttleTraq, available on Samsung Galaxy Appstore, are based on the proposed system.**

***Keywords*-Badminton; Deep Neural Network; Inertial Measurement Units; Machine Learning; Shot classification; Swing detection; Swing sports; Tennis; Wearable sensors**

I. Introduction

Advances in motion sensing technologies have enabled the integration of inertial measurement units (IMUs) into smartphones and wearables like smartwatches and smart rings. Health and fitness applications like Samsung Health and Google Fit on these devices use data from IMUs for seamless tracking of user's activities. Even new devices like Gear Fit, FitBit, which are dedicated towards fitness tracking are increasingly becoming popular. Since most of these devices are wrist worn and light-weight, they can be used for effectively tracking swing based sports like Tennis, Badminton, Squash and Golf. In this paper, we focus on the design of a generalized system for shot detection and classification in swing sports. After shot classification, biomechanics based detailed analysis of kinetically complex shot types like Tennis serve and Badminton smash can be performed based on [1]. The significant contributions of this paper include:

1) Developing a generalized framework for shot detection in swing sports where we also introduce a novel concept termed as Jerk-effective for accurate calculation of point of impact in a shot sensor signal.
2) Physical motion based feature generation followed by correlation and mutual information based feature selection for effective shot classification in swing sports.
3) Novel CNN and BLSTM based architectures for time series analysis of sensor data in sports.

II. Related Work

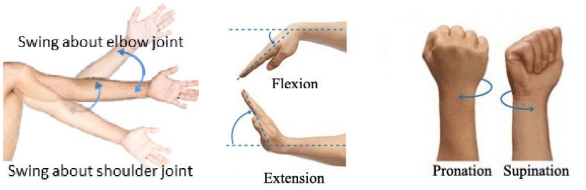
The use of IMUs for activity detection is a well-known field of research [2], [3]. There has also been significant research on the application of motion sensors in sports [4]. R. Srivastava et al. in [5], [6] use dynamic time warping (DTW) approach to classify swings in Tennis using IMU sensor data from wrist. It differs from our work as it uses an ideal dataset obtained from ball feeding. The authors in [7], [8] have proposed a threshold based criteria for Tennis shot detection and have explored learning based methods such as KNN and SVM for shot classification with accuracies ranging from 85-90%. In [9], P. Blank et al. used signal energy from the IMUs mounted on the racquet to detect strokes and feature based classifiers to classify them in Table Tennis. Another method proposed in [10] uses data at 100Hz frequency, from IMUs at multiple locations in the body to classify 14 different badminton stroke types with accuracy of 98%. The past research work on detection and classification of shots in swing sports is specific to one sport only. In our work, we propose a unified approach towards swing detection and classification for multiple swing sports using data from IMUs located **only** on wrist.

III. The swing motion

The motion of hand during shots in different swing-based sports can be seen as a combination of three basic component gestures [11]: 1) Pure-swing: Swing of the arm about shoulder joint or ankle joint 2) Wrist extension and flexion: Movement of wrist about the ankle joint 3) Pronation and supination: Rotation of hand about the forearm axis, shown in Fig. 1.

The contribution of each component gesture in a swing depends upon the shot type and the sport. E.g. Shots in Tennis have dominant pure-swing gesture with some pronation and almost no flexion/extension whereas backhand short serves in Badminton have little pure swing component and dominant flexion/pronation motion.

0-7695-6321-X/17/31.00 ©2017 IEEE
DOI 10.1109/ICMLA.2017.0-149



(a) (b) (c)

Figure 1: Component gestures of swing (*Image source:*[12])

A swing can also be seen as a combination of sequentially occurring stages of hand movements [11]. These are based upon the direction of hand motion and can be defined as:

I) Backswing/Preparation: Player's hand moves from neutral position (initial position where arm is at rest) to the position where it is stretched backward.
II) Forward swing: Players move their hand in the forward direction and hit the ball/shuttle in this stage.
III) Follow-through: Stage from point of impact to the point where the hand comes to rest in forward direction.
IV) Retraction: This is the stage in which the player brings his hand back to the neutral position.

The retraction and backswing stages may overlap for consecutive shots in games due to the absence of a pause between the shots during fast paced games.

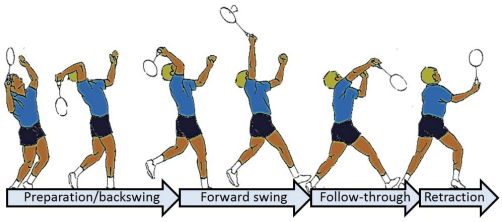
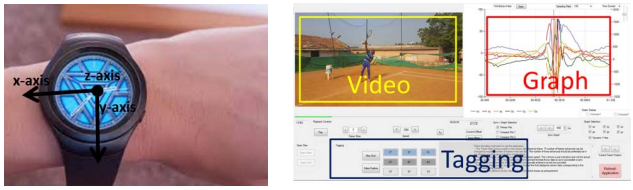


Figure 2: Badminton swing stages (*Image source:*[13])

Fig. 2 shows the stages in a Badminton swing. Such segregation of swing motion helps in distinguishing among shots of different shot types in sports. As an example, Forehand Slice and Forehand Topspin in Tennis are similar in preparation stage, but different in forward swing and follow-through stages. Hence, features generated for the aforementioned stages help in distinguishing among such shot types.

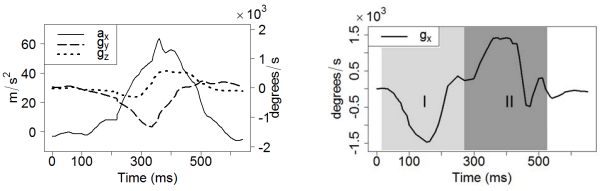
IV. Signal Analysis During Swing

The 3-dimensional accelerometer (a_x, a_y, a_z) and 3-dimensional gyroscope (g_x, g_y, g_z) data were collected from inertial sensors embedded in a wearable worn on wrist. Fig. 3a shows the wearable along with the axes. In order to understand the behavior of sensor signals during a shot swing, we used an in-house developed tool in Fig. 3b to study the shots captured in videos alongside corresponding sensor signals. Since, shot swing motion has a dominating effect on sensor signals, effects of gravity, body movement etc. have been



(a) (b)

Figure 3: (a) Motion sensor axes in the wrist worn wearable (b) Video tool used for sensor signal visualization



(a) (b)

Figure 4: Key sensor signals for various basic hand gestures: (a) Pure-swing (b) Supination followed by pronation

ignored. Due to the force required during shots to place the ball/shuttle, the flexion/extension gestures occur alongside pure-swing or pronation/supination for majority of the shots. Hence, it suffices to study the effects of Pure-swing and Pronation-supination gestures on the sensor output.

1) Pure-swing: It can be approximated as a circular motion with varying speed v and radius R. As the *x-axis* is along the forearm, a_x captures a significant component of the centripetal acceleration (v^2/R) during such motion as shown in Fig. 4a. The motion also results in a lobe in g_y or g_z depending on hand orientation during swing.
2) Pronation-supination: This gesture results in a rotation of the wrist worn wearable along the *x-axis*, leading to a lobe in the g_x sensor data as shown in Fig 4b.

To capture the effect of component gestures on gyroscope axis, we define a derived signal entity, g_{energy}, given as:

$$g_{energy} = g_x^2 + g_y^2 + g_z^2 \tag{1}$$

The g_{energy} signal is representative of the net rotational energy (E_ω) of the arm. E_ω can be written as $E_\omega = I\omega^2$ where I is the moment of inertia for the body about the axis of rotation, or $E_\omega \approx I(g_x^2 + g_y^2 + g_z^2)$ or $E_\omega \approx I \cdot g_{energy}$ or $E_\omega \propto g_{energy}$. Since rotational energy of the arm is relatively high during a swing, a dominant lobe occurs in g_{energy} signal. This serves as an important cue for detecting the swing motion.

Ball/shuttle impact is another distinct characteristic of a shot swing. The impulse transferred during such impact is captured at the wrist and appears as sudden high magnitude spikes or jerk in the accelerometer signal. This jerk (J) along

an axis is expressed as in Eq. 2.

$$J_{axis}[n] = |a_{axis}[n] - a_{axis}[n-1]|, axis \in \{x, y, z\} \quad (2)$$

$$J_{net} = \sum_{axis \in \{x,y,z\}} J_{axis} \quad (3)$$

To efficiently capture jerk in all axes, we define net jerk, J_{net} as given in Eq. 3. This formulation enables efficient capturing of jerks from all axes. Ball/shuttle impact is characterized by a significantly high J_{net} value and hence, it serves as an important cue for shot swing detection.

V. Block diagram

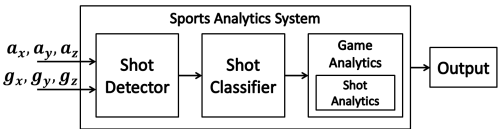


Figure 5: System design

The design of our proposed system is shown in Fig. 5. Sensor data from the wrist-worn wearable is input to the detector module for detection of shot swings. The detected shot swing regions are then classified by the classifier module into different shot types. Shots are further analyzed using shot analytics [1] and game analytics [6] modules. The following sections explain the proposed algorithms for detector and classifier units of the sports analytics system.

VI. Shot Detector

Based on the study of sensor behavior during shot swing, here we discuss 3 cues, a combination of which can be used to identify shot regions in sensor signals for different sports.

A. Threshold on a_x and/or g_{energy}

For sports like Squash and Tennis, a_x exhibits a dominant central lobe (Fig. 6b) due to presence of significant pure swing component in shots. g_{energy} signal also produces a lobe for all types of swings (Fig. 6a) as discussed in Section IV. Hence, a threshold on low passed a_x and g_{energy} signals can be used to detect swing shots.

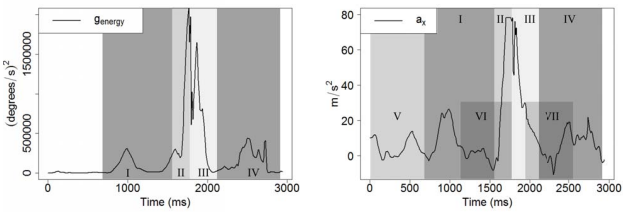
B. Jerk based detection

Impact region is characterized by multiple high-magnitude J_{net} peaks as the jerk sustains after impact and then dies quickly. Among the multiple peaks, it is observed and validated using the in-house tool that the first occurrence of high J_{net} is the impact point. In order to find this exact point of impact, we define a new measure, called Jerk-effective or $\mathcal{J}$,

$$\mathcal{J}[N] = |J_{net}[N]| * \prod_{k=1}^{k=K} |J_{net}[N-k] - J_{net}[N]| \quad (4)$$

where $N =$ sample number;

and $K =$ hyper-parameter based on the sport.



(a) g_{energy} signal for a tennis shot (b) a_x signal for a tennis shot

Figure 6: Stages of a shot are: I- Backswing, II-Forward-swing, III- Follow-through, IV- Retraction; Additional stages as per Section VII-B: V- Silent VI- Transition1 and VII-Transition2

Before impact, $\mathcal{J}[N]$ is low due to low $J_{net}[N]$. Beyond impact point, it is lowered by the term $|J_{net}[N-k] - J_{net}[N]|$ which suppresses high J_{net} values immediately following impact. Thus the expression of $\mathcal{J}[N]$ emphasizes the point of impact and reduces the weight of points before and after impact. Thresholding on $\mathcal{J}[N]$ can further be used to detect swing shots. The value of K is empirically chosen, e.g. in Tennis, $K = 2$ and in badminton, $K = 1$.

C. Shape based detection

Shots having dominant Pure-swing gesture (e.g. shots in Tennis and Squash) exhibit a central high-magnitude lobe with two low-magnitude side lobes in a_x and g_{energy} signals as shown in Fig. 6. We calculate reference signal sequences that capture this shot specific shape. Dynamic Time Warping or DTW [14] distance of candidate shots with reference signals is then used for shot detection.

To find the reference or ideal signal sequence, Ref_{signal}, corresponding to a particular signal type ($signal \in \{g_{energy}, a_x\}$), we take a collection of N training shot sequences (S_{signal}) corresponding to $signal$. Let $DTW(a, b)$ represent the DTW distance between sequences a and b. The first step is the generation of DTW cost matrix C_{DTW} as:

$$C_{DTW}(i, j) = DTW(S_{signal}(i), S_{signal}(j)); i, j \in [0, N-1] \quad (5)$$

We calculate a column matrix C^{sum}_{DTW} which represents average distance of $S_{signal}(i)$ from all the other shot signals, as given in the following equation.

$$C^{sum}_{DTW}(i) = \frac{1}{N-1} \sum_{j=0, j \neq i}^{N-1} C_{DTW}(i, j) \quad (6)$$

The reference signal sequence is one with the least average distance from all other sequences, as shown in Eq. 7.

$$Ref_{signal} = S_{signal}(\hat{i}); \text{where } \hat{i} = argmin(C^{sum}_{DTW}) \quad (7)$$

For a candidate shot c, the DTW distance against reference sequence for a particular $signal$ is evaluated as:

$$d_{signal} = DTW(c_{signal}, Ref_{signal}), signal \in \{g_{energy}, a_x\} \quad (8)$$

If the distances d_{signal} for both the signal types are within the empirically determined thresholds for a sport, we consider the candidate shot as a true shot.

The shot detector module uses a combination of multiple signal cues to detect a shot swing in different sports. In tennis, we use a_x and g_{energy} based threshold and shape based detection, while in badminton we use jerk and g_{energy} based threshold for detection. Once a shot swing is detected, fixed number of samples before and after the impact point are together considered as a shot region. The motion sensor signals and derived signals like g_{energy} corresponding to shot regions, as detected by shot detector, are used to develop classifiers for different sports. These regions are labeled with various shot types in a sport and used as input in the two classification approaches discussed below.

VII. Feature based shot classification

Before generating the relevant features for classification, each shot is segregated into the various stages described in Section III. Features generated in these stages are then used to develop classifiers for different sports.

A. Shot stage boundary estimation

The start and end of shot stages are referred to as boundaries and can be calculated empirically, called static boundaries, or using signal characteristics in different stages, termed as dynamic boundaries. For a new shot, static boundaries can be set based on pre-determined average duration of each stage from the point of impact. These durations are estimated by manually tagging different stages for a collection of shots in a sport. Alternatively, for dynamic boundaries, the ends of backswing and follow-through stages are determined by locating appropriate minima in the low passed g_{energy} signal, which occur due to slowing down of player's hand motion.

Depending on the distinction between shot types in a sport, suitability of using static or dynamic boundaries is decided. E.g. in badminton, the shot types like clear, drive, drop differ mostly in the length of the signal. Here static boundaries prove more useful as in using dynamic boundaries, the differentiating features average out over the length of the signal.

B. Feature generation

In addition to the four stages mentioned in Section III, we consider three more stages, namely 'Silent', 'Transition 1' and 'Transition 2', for feature generation. These stages are shown in Fig. 6b and hereafter referred to as 'windows'. 'Silent' window captures the persistent hand stability just before serving the ball/shuttle. Transition windows also hold important cues for differentiating among shot types. E.g. in Tennis, clockwise and anti-clockwise hand rotation is captured in 'Transition 1' for forehand slice and topspin shot respectively.

The set of shot windows is represented as R ={Backswing, Forward-swing, Follow-through, Retraction, Silent, Transition 1, Transition 2}. Prior to feature generation, we consider if shot types in a sport are independent of shot speeds as in Tennis. For such sports, accelerometer (a) and gyroscope (g) signals are min-max normalized to introduce speed invariance. Let a_i^r and g_i^r, $i \in \{x, y, z\}$ and $r \in R$, represent the accelerometer and gyroscope data (normalized or otherwise) for window r and axis i. Three types of feature sets are generated from sensor data for all shot windows, r.

1) Statistical features e.g. mean, median, skew, kurtosis, minimum, maximum, range for a_i^r and g_i^r .
2) Pairwise correlation coefficients between elements of the set $\{a_x^r, a_y^r, a_z^r, g_x^r, g_y^r, g_z^r\}$.
3) Shape based features determined by the correlation coefficient between a predefined linear and parabolic sequence with the min-max normalized sensor data a_i^r and g_i^r.

The above feature set comprehensively represents the swing motion. Feature selection for different classifications is further performed on this extensive list of $\sim$2000 features.

C. Feature selection

Feature selection is performed on training dataset by maximizing correlation of features with the output classes (shot types) and minimizing correlation among selected features [15]. In order to calculate correlation between output classes and a particular feature, we train a logistic model on the output classes using that feature and consider the training accuracy as a measure of correlation. For correlation between two features, X and Y, absolute value of the Pearson correlation coefficient (ρ) [16] is used.

For feature selection, the features are sorted in decreasing order of their correlation measure with the output classes. Features are then selected in order, provided their correlation with the already selected features is less than a correlation threshold (ranges from 0.5 to 0.8 depending on desired number of features). This selected feature set of around 300-500 features is further reduced by using the mRMR technique. The discussed method of feature selection falls into the filter approach [15] of selecting features and thus, the output feature set can be applied to a wide variety of classification algorithms.

D. Model creation

The relevant features obtained from the feature selection step are used to develop models for performing multiclass classification of shot types. Depending on the sport, one of the two classifiers namely, hierarchical classifier and all vs. all (AVA) classifier [17], is used. For games like tennis, we use a hierarchical model because the distinct shot type definitions follow an innate hierarchy. E.g. we can label a shot as forehand/backhand/serve in the first step and then

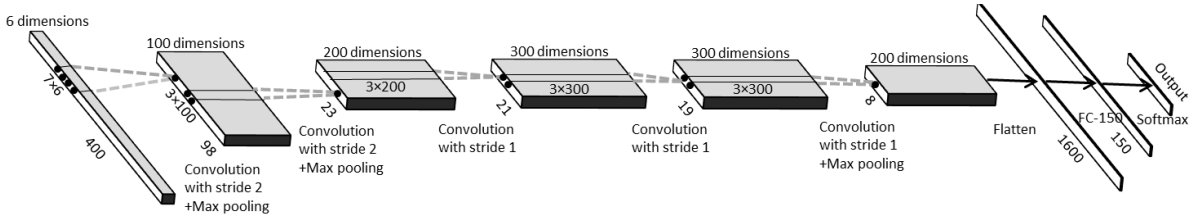


Figure 7: CNN architecture used in shot classification for swing sports

for a forehand, we can subclassify it as topspin or slice. For games like badminton, such hierarchy does not exist and hence AVA classifier is used. The component classifiers in the multiclass classification model can be constructed using SVMs, logistic regression, naive bayes etc. In our case, we use logistic regression to build them.

VIII. Neural network based shot classification

Over the past few years, neural networks have emerged as the most widely used approach for many machine learning problems. Convolutional neural networks (CNNs) have been successful in performing computer vision tasks e.g. image classification, image segmentation etc. [18], [19] and are proven to outperform feature based classifiers that use SIFT, SURF etc. On the other hand, Long Short Term Memory (LSTM) networks [20] have become state-of-the-art for modeling sequential or temporal information in the domain of time series like speech recognition, handwriting recognition, video captioning, etc. For shot classification in sports too, we use these two methods.

A. CNN based classification

We begin by drawing an analogy between image classification and shot classification in sports. An image is a 2-D signal in three channels of red, green and blue colors respectively, whereas a detected shot is a 1-D signal along six channels namely a_x, a_y, a_z, g_x, g_y and g_z. In case of image, the convolution layers learn representations like texture, shape, average intensity etc. [21] which are analogous to attributes such as mean, standard deviations, signal correlation, for a shot. Also CNNs efficiently capture and combine spatially located features in an image, hence it is expected to capture and combine the relevant features in different shot windows to yield the desired classification output.

Fig. 7 shows the CNN architecture used for sports which is built on similar lines as the image classification work presented by Krizhevsky et al. in [22]. The activations of all the neurons was set to be RelU [22] and Xavier initialization [23] was used to initialize all the network layers. Input to the CNN is motion sensor data sequence of length 400 across six sensor axes, thus having net dimension of 400×6. The first and second convolutional layers have 100 and 200 kernels of sizes 7×6 and 3×100, each with stride of 2, respectively. Both the layers are also max pooled by factor of 2 before feeding to the next layer. The third, fourth and fifth convolutional layers have 300, 300 and 200 filters having sizes of 3×200, 3×300 and 3×300 respectively. The output of fifth layer is max pooled by a factor of 2 and then flattened to apply dense layers. The resulting sequence of length 1600 is fed to following dense/fully-connected layer with output length 150. This output is passed through a dropout layer with dropout factor of 0.5 and further fed to the final softmax layer which returns the output classes. The dropout layer reduces chances of overfitting during the training phase.

The training data of shot signal sequences are generated in two steps. Firstly, shot regions constituting 300 and 200 samples before and after the impact point respectively are considered. It is found that a signal sequence of length 500 is enough to capture shots in all swing sports. In the second step, multiple subsequences of length 400 are generated by running a 400 length window in steps of 10 over the previously obtained sequence. These resultant sequences serve the purpose of increasing the training data set by a factor of 10 and increasing the robustness of CNN classifier against inaccuracies in impact point estimation. This is similar to the data augmentation step in case of image classification. The CNN network is trained using ADAM backpropagation algorithm [24].

B. BLSTM based classification

Sensor data from wearable are a temporal sequence of data from 6 axes of sensors. However, a shot sequence can also be visually decomposed into sequentially occuring motion sub-sequences (motion within a small time-frame). Therefore, rather than using the stacked vector of values from each of the axis as a timestep for LSTM cell, we learn signal based features within a time-frame and use this set of learnt features in frames as timestep input. This step is analogous to the preprocessing step of MFCC coefficients extraction in framewise phoneme classification [25]. However, the difference is that we learn these features in our network using a convolution layer rather than using a pre-defined preprocessing algorithm.

We use Bi-directional Long Short Term Memory network (BLSTM) [25] for our architecture. The basic idea of BLSTM is to join two layers of LSTM, both operating in opposite directions on the input sequence, using a summation operation on the output of each cell. In contrast to uni-directional LSTM, which has the context of just the past

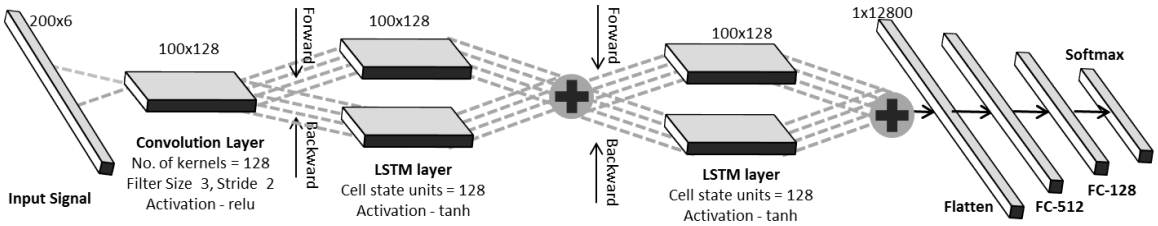


Figure 8: BLSTM architecture for shot classification

information for each cell, BLSTM uses the context of the whole input sequence, containing information from the past and future. In [25], BLSTM has been shown to give about 4% improvement over the uni-directional counterpart for phoneme classification. For shot classification problem also, BLSTM shows 1-2% improvement over LSTM architecture.

Fig. 8 shows the proposed architecture. The network contains two stacked BLSTM layers followed by two dense layers to give softmax output for n-class classification, where n is the number of shot classes for a sport type. Each BLSTM layer consists of 2 halves, which operate on the input sequence in forward and backward directions respectively. First layer following the input is a convolution layer of 128 kernels with stride 2 and kernel-size 3, whose output is fed into BLSTM layer. For the first BLSTM layer, one LSTM reads the convolution output in forward direction, while the second reads in backward direction. Each LSTM cell has 128 output units. Output from two halves is combined and added at each cell output to form a BLSTM layer. Another BLSTM layer is stacked on top of this output, with 128 output units for each cell. At the output of second BLSTM layer, each BLSTM cell unit is flattened and concatenated with units of other cells. This is then fed to the fully connected layers of 512 and 128 neurons respectively. Network is trained with $AdaDelta$ optimizer. Dropout of 0.2 is added after each convolution and LSTM layer, and dropout of 0.5 is added after every dense layer to avoid overfitting. Weight initialization was Xavier normal for the whole network.

IX. Results

We used Samsung smartwatch Gear S2 to capture and store data at 100Hz from 3-axes accelerometer and gyroscope sensors embedded in the watch. The ranges of accelerometer and gyroscope sensors were $\pm 8g$ and $\pm 2000 degrees/s$ respectively. The analysis of this data was performed on the smartphone both in real-time and after game, using our developed analytics system. The procedure of collecting data involved each player being asked to wear the smartwatch in the playing hand and play singles game against an opponent, while we video recorded the player using Samsung S6 smartphone at 30fps. Using the in-house developed tool, the video with the corresponding sensor data was looked at and each shot was tagged with its shot type. The training and testing dataset are collected from different players and include: a) Tennis : training set of ≈ 4500 shots by 15 players; testing set of ≈ 5000 shots by 16 players, b) Badminton : training set of ≈ 3500 shots by 20 players; testing set of ≈ 2000 shots by 14 players, c) Squash : training set of ≈ 500 shots by 3 players; testing set of ≈ 100 shots by 2 players.

The tagged data serves as the ground truth or true positive for our detector and classifier analysis. The detector accuracies for different sports are shown in Table I. The detected shot swings are then fed into the classifier to calculate classifier accuracies. The shot types considered for classification in Tennis are Forehand Topspin, Forehand Slice, Backhand Topspin, Backhand Slice, Serve; in Badminton are Serve, Clear, Drop, Smash; and in Squash are Forehand, Backhand, Serve. Table II shows the classification accuracies of our proposed methodologies of feature-based (F-Based) classification and, CNN and BLSTM discussed under neural network based classification for each sport.

Table I: Shot detection accuracies

Sports	Precision	Recall	F1 Score
Tennis	0.91	0.92	0.92
Badminton	0.83	0.93	0.88
Squash	0.95	0.96	0.96

Table II: Shot classification accuracies(%)

Sports	F-based	CNN	BLSTM
Tennis	90.2	93.8	93.3
Badminton	76.1	77.2	78.9
Squash	92	93.2	94.6

The classification accuracies obtained are comparable for all three proposed techniques, with neural networks slightly outperforming the feature-based classification. The lower accuracies for Badminton can be attributed to the fact that its shot types are characterized less by hand gestures and more by location of the player and shuttle drop. Also, since CNN and BLSTM neural networks are memory and execution time intensive using a set of $\approx 10^6$ parameters as compared to $\approx 10^3$ being used in feature based classifier, they are more suited for cloud based systems. For real-time analysis on low power processors, feature-based classification is preferred.

X. Conclusion

In this work, we have described and demonstrated the Sports Analytics System for a comprehensive analysis of swing-based sports. The proposed system entails a shot detection framework based on various signal cues effectually applied in different sports and introduces concept of Jerk-effective for accurate impact point calculation. Novel techniques for shot type classification using feature engineering based on correlation between physical motion and sensor data, and CNN and BLSTM neural network architectures for sports data are discussed. The accuracies obtained for three swing sports show promising trends. TennisTraq and ShuttleTraq applications on Samsung apps store use the discussed techniques. Our proposed Smart Sports System will enrich sporting experience by letting players use wearable device/smartphone for analysis, without any additional device as is the case with competitive solutions like Zepp, Sony Smart Tennis and Coollang.

References

[1] M. Sharma, R. Srivastava, A. Anand, D. Prakash, and L. Kaligounder, "Wearable motion sensor based phasic analysis of tennis serve for performance feedback," in *2017 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, March 2017, pp. 5945–5949.

[2] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst., Man, Cybern., Part C (Appl. and Rev.)*, vol. 42, no. 6, pp. 790–808, 2012.

[3] H. Leutheuser, D. Schuldhaus, and B. M. Eskofier, "Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset," *PloS one*, vol. 8, no. 10, p. e75196, 2013.

[4] M. Zok, "Inertial sensors are changing the games," in *Inertial Sensors and Systems (ISISS), 2014 Int. Symp. on*. IEEE, 2014, pp. 1–3.

[5] R. Srivastava and P. Sinha, "Hand movements and gestures characterization using quaternion dynamic time warping technique," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1333–1341, 2016.

[6] R. Srivastava, A. Patwari, S. Kumar, G. Mishra, L. Kaligounder, and P. Sinha, "Efficient characterization of tennis shots and game analysis using wearable sensors data," in *Sensors, 2015 IEEE*. IEEE, 2015, pp. 1–4.

[7] D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, M. Walsh, and C. O'Mathuna, "Multi-sensor classification of tennis strokes," in *Sensors, 2011 IEEE*. IEEE, 2011, pp. 1437–1440.

[8] C. Ó Conaire, D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, and J. Buckley, "Combining inertial and visual sensing for human action recognition in tennis," in *Proc. of the 1st ACM int. workshop on Analysis and retrieval of tracked events and motion in imagery streams*. ACM, 2010, pp. 51–56.

[9] P. Blank, J. Hoßbach, D. Schuldhaus, and B. M. Eskofier, "Sensor-based stroke detection and stroke type classification in table tennis," in *Proc. of the 2015 ACM Int. Symp. on Wearable Computers*. ACM, 2015, pp. 93–100.

[10] Z. Wang, M. Guo, and C. Zhao, "Badminton stroke recognition based on body sensor networks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 769–775, Oct 2016.

[11] R. Bartlett, *Introduction to sports biomechanics: Analysing human movement patterns*. Routledge, 2007.

[12] "Ergonomics," https://www.oshatrain.org/courses/mods/810m1.html.

[13] "Better badminton: The power of 4," http://www.labadmintonclub.com/index.php?option=com_content&view=article&id=979&catid=1.

[14] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.

[15] M. A. Hall, "Correlation-based feature selection of discrete and numeric class machine learning," 2000.

[16] J. B. et al., "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.

[17] M. Aly, "Survey on multiclass classification methods," *Neural Netw*, vol. 19, 2005.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[19] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2015, pp. 3128–3137.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks." in *Aistats*, vol. 9, 2010, pp. 249–256.

[24] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5, pp. 602–610, 2005.