Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 2374–2381

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Robust Sports Image Classification Using InceptionV3 and Neural Networks

Ketan Joshi*, Vikas Tripathi, Chitransh Bose, Chaitanya Bhardwaj

Graphic Era University, Clement Town, Dehradun,248002, India

Abstract

In today's world of internet, a massive amount of data is getting generated every day and content-based classification of images is becoming an essential aspect for efficient retrieval of images and have attracted application in several fields and one of such field is sports. Sport is an integral part of everybody's daily life and it is very important to play the sport with the right posture and environment otherwise it can lead to medical issues. This paper presents a robust framework for classifying the sport images based on the environment and related surroundings. In this paper, our approach is based on the use of the Inception V3 for the extraction of features and Neural Networks for the classification of various sport categories. Six categories rugby, tennis, cricket, basketball, volleyball, and badminton have been used for analysis and classification. To validate the effectiveness of the framework and Neural Networks, comparisons have been done with other classifiers like Random Forest, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). Our framework has successfully achieved an average accuracy of 96.64 % over six categories which demonstrate the effectiveness of the framework and can be used for the detection and classification of various sport activities in an efficient manner.

Keywords: Sport images; Neural Networks; InceptionV3; Content-based classification.

1. Introduction

Computer vision is playing a vital role in the field of image or video identification and analysis. It deals with the understanding of machines about images and videos. Machines these days are capable enough to detect objects, videos, particular activities, etc. Humans use their eyes and brain to identify things, to sense their surroundings. Computer vision is a field that provides a similar capability to a machine. There are a lot of applications of computer vision like biometrics, gesture analysis, security and surveillance, pollution monitoring and many more [1], [2], [3], [4], [5]. The

10.1016/j.procs.2020.03.290

utmost role of computer vision is played in Human Activity Recognition (HAR). In this, the machine is trained in such a way that it becomes capable enough to identify the human motions [6].

The motivation behind the computer vision lies at the core of imitating HAR. It aims to tell apart various human actions like throwing a ball, running, hitting the ball, playing games and many more by certain kind of observations in a specific environment. Besides its tremendous growth, it has some challenges also like recognizing two or more similar looking human motions together, for example, a person kicking a ball in a ground. At this, the machine can be confused that whether to detect it as a football event or a rugby event. It was just an example of the challenges in HAR. In this paper, computer vision applications are used to identify the different types of sports activities and then classified them. There are different types of classifiers that are used in the classification process, for example, Neural Networks, Support Vector Machine (SVM), Random Forest, k-nearest neighbors (KNN), and some more. These all work differently and are suitable under different circumstances like SVM works well in smaller and complex data and can be much more efficient in model formation. In layman terms, it can be used to form clusters of multiple data and each cluster comprises of a similar kind of data. On the other hand, Neural Networks can work on larger as well as small or complex data, but the approach is different in both the models that are, the Neural Networks work somewhat as a human brain does. It has millions of neurons that process and transmits the data. KNN is a non-parametric technique that stores all the available cases and the new cases are then classified on the basis of the stored ones. Random Forest is a supervised algorithm that builds multiple decision trees and then combines them to get a stable prediction. These above-mentioned classifiers are used for our classification. For the analysis, different sports videos are used and frames are extracted from them for deliberate analysis. Now here comes the role of feature descriptors. In layman term, feature descriptors are those entities that provide the description (i.e. Coordinates of pixels) of the data to be analyzed. For our analysis, we used it to get the description of our dataset and to train the model according to that information and then tested the competence of our training. There is comparative testing in which the results provided by each classifier are compared with the results provided by the Neural Networks.

Sports play an important role in everyone's life. Everyone loves to play or watch games and sports videos. So, in this work, our aim is to classify different sports activities into different categories so that this analysis can be used to explore the required sports-related images [7] for example, nowadays, many types of research have been going on sports and dataset formation is a very tedious task. So to make it easier, this analysis seems to be helpful by providing the desired sports images. The rest of the paper is defined as Section 2 consists of the literature review that comprises of the work done in HAR under the vast field of computer vision. Section 3 consists of a methodology that has been used for the classification like classifiers used and their configurations. Section 4 consists of our results and discussions, and then in Section 5, our work is concluded.

2. Literature Review

Since the last decade, the work in computer vision has grown exponentially high. A lot of researches have been done on its applications. The classifiers like Neural Networks, SVM, KNN, and Random Forest are used in the HAR because of their good analysis and results. In [8] Amin Ullah et al. have proposed their framework in action recognition in video sequencing using deep bidirectional LSTM with CNN features. In this framework, a Convolutional Neural Networks (CNN) and deep bidirectional long short term memory (DB-LSTM) was used to process the data, and then the deep features were extracted from the video data to help reduce the complexity and redundancy. This framework was capable to process long videos by analyzing features for a short period of time. Similarly, in [9] Pavel Zhdanov et al. have proposed a model for increasing human action recognition through hierarchical Neural Networks. In this paper, twenty most challenging and difficult to diagnose actions from the kinetic dataset are recognized using CNN, which as a matter, of course, improves the quality of recognition for those activities. In HAR and specifically in the field of sports, Cem Direkoglu and Noel E. O'Connor in [10] introduced an approach for team activity recognition in which SVM classification and proposed motion descriptors are used to evaluate the European handball dataset that can classify six different sport categories. In this approach, the positions of team players in the ground were known and with the help of Poisson equation, a smooth distribution defined on the whole playground is generated and termed smooth distribution

as position distribution. After computing the position distribution for each frame the sequence of distributions was provided that were processed to extract the motion features for team activity recognition. In [11] David A. Sadlier and Noel E.O' Connor proposed a framework based on audio and visual features to detect events in the videos of multiple different field sports. In this model, some particular features were extracted and were rooted in the category of similar characteristics. Here, SVM is used for the combination of features gathered, which assumes the repetition of an event, based on the trained model. The model was tested across multiple categories, including six different sports videos and concluded that high event retrieval and content brush-off is achievable. In [12] LingYu Duan, Min Xu, Qi Tian, and two others proposed a model for semantic classification in sports videos. To achieve this, they implemented supervised learning to perform a top-down video shot classification. The work was divided mainly into three tiers: 1) for each sport, spotting the video shot; 2) establishing the group of color, shot length-related mid-level representations like player motion or the pattern of the motion of the camera; 3) group of common motions. The nonparametric feature analysis has been done to map low-level features to the video shot attributes like the motion of the player while playing a particular type of sport, the attributes of the playing ground or court such as its shape. That is, the low-level features like the texture of the shots, color, motion vector field, shot length were mapped with the mid-level features such as motion entropy, active regions shot pace and some more for better analysis of the model. The model is working smoothly with decent results. Even the work has been done activity detection and recognition for sports [13], where single player activity recognition is done on a couple of volleyball videos. The investigation is done frame-by-frame by using Histogram of Gradients (HOG) features [14] and Histogram of Oriented Flow (HOF) [15] features, these both methods work well if and only if the object regions are known. The classification here is done by SVM and the results as context information are embedded by using Activity Context (AC) descriptor that is used for describing the activity probability on the court. In [16] Billur Barshan and Murat Cihan Yuksek worked on recognizing sports activities using body-worn sensor units in two open source machine learning environments and provided a comparative appraisal on various techniques to classify human activities by wearing body sensors on various parts of the body that are mostly used in any kind of athletic activity like arms and legs. The comparison was shown between classifiers like Naïve Bayesian, Artificial Neural Networks (ANN), Gaussian Mixture Models (GMMs) and SVMs. According to their study, the sensors that were on legs were the most informative than the sensors on other positions like arms or chest. The comparison between the classifiers was shown on the basis of confusion matrices, computational costs, and the correct differential rates. The relation between the commonly used machine learning tools like WEKA and PR Tools in terms of their functionality and performance of their classifiers were also compared.

A lot of work has been done in the activity recognition domain by taking Neural Networks, SVM, and Random Forest, etc. From the literature review, it has been identified that Neural Networks can be used for sport image segmentation and analysis. In this paper, a framework based on Neural Networks for efficient classification of images in different sports categories is presented and comparative analysis of various sports activities using different classifiers like KNN, SVM, Random Forest, and Neural Networks has been performed to validate our framework.

3. Methodology

Image categorization involves two major steps feature descriptor calculation and classification. The visual representation of our whole framework is shown in Fig. 1. Firstly, several sports videos of each category are collected that is badminton, rugby, basketball, tennis, cricket, and volleyball from YouTube and then extracted frames from them and made our own dataset. This data set is fed into image embedder for extraction of feature descriptor values. For feature extraction, the Inception V3 model is used. It is Google's pre-trained model which has been trained over 1000 classes and over 1.4 million images. The Inception V3 model is an image recognition model for feature extraction with the help of the Convolutional Neural Networks. Further classification is performed with fully- connected and softmax layers. After that, a data sampler is used to sample the data as training and testing. In our case, 70% of our data is kept as training data and rest 30% as testing data and then sent the training data to our classifier i.e., Neural Networks for training. After that, the rest 30% of our data were sent to the prediction section and at last the confusion matrix was built to check the competence of our training.

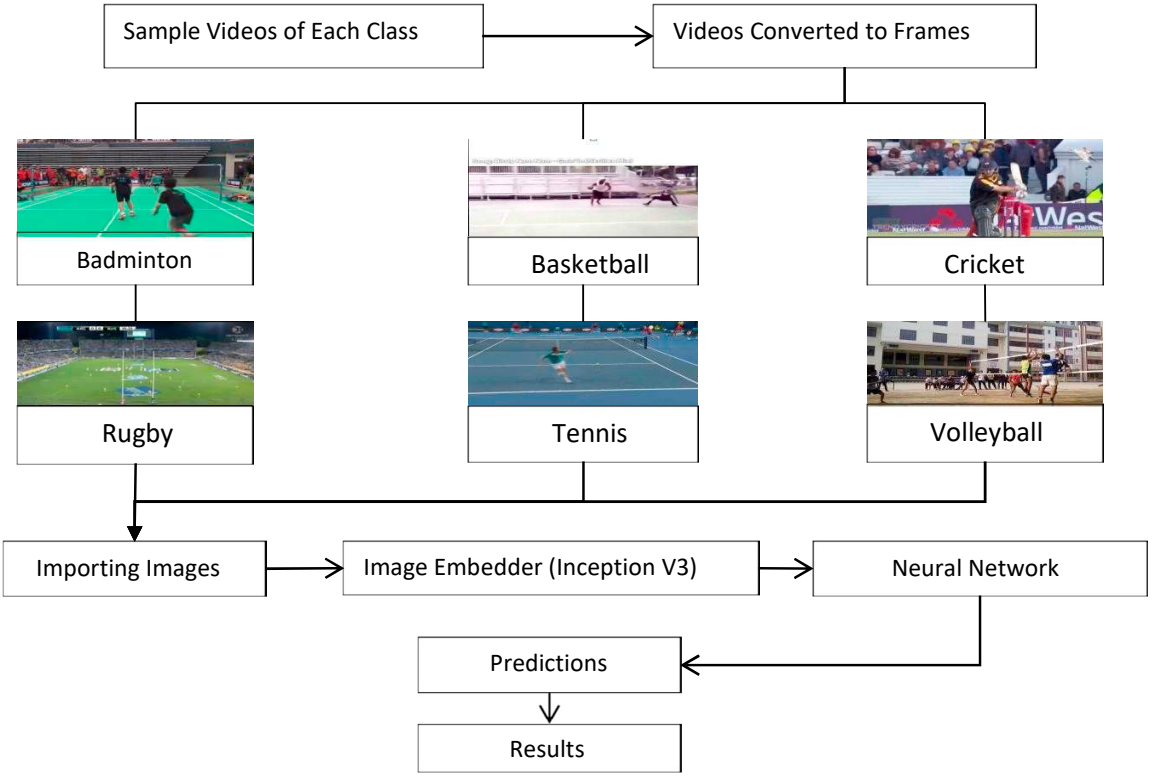

Fig. 1. Framework for sports Image classification

The Inception V3 model extracts useful features from given input images in the training part and further deploy classification based on the extracted features in the second part. The diagrammatic representation of the working of Inception V3 is shown in Fig. 2.

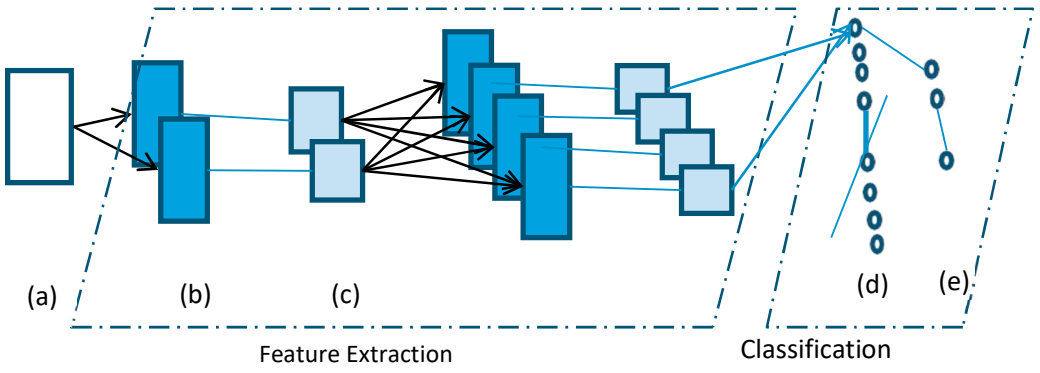

Fig. 2. Diagrammatic representation of Inception V3 where (a) is the input image, (b) is convolution layer, (c) is subsampling layer, (d) and (e) are fully connected output layers

The activation function used is a Rectified linear unit (Relu). It is linear for all positive values and is zero for all negative values. It is easy to compute and hence the model gets less time to train. This function is used because it does not have vanishing gradient problem which is present in other activation functions like sigmoid and tanh. Mathematically, it can be expressed as shown in eq (1):

$$X(x) = \max(0, x) \tag{1}$$

Here, X () is the function, 0 is the initial value, x is the input and a maximum of 0 and the input value is taken. As the *Relu* function is 0 for all negative values, hence, the initial value is set to be 0.

4. Results and discussions

The proposed framework is deployed on a system with configuration as Intel(R) Core(TM) i7-7700 CPU @ 3.60x8 GHz with 8 GB RAM. Images used for analysis are having a size of 320x240. The images for all six categories are chosen in such a way that analysis becomes complex, for example, some images are blurred, some are dark and some are clear, apart from that, images are taken from various resources so that background and camera angle are not uniform. The non-uniformity in the background, angle of the camera, image blurring makes our dataset challenging as well as suitable for analysis. Sample frames for all six categories of images are shown in Fig. 3.

Fig. 3. Sample frames of different sports activities.

The dataset is made by taking videos of each sports category [17], [18], [19], [20], [21], [22] from YouTube and then frames are extracted from each video. A total of 10,000 frames is in total with 1,650 (approx.) frames in each class. As per standard used for classifying the dataset, it is divided into the testing (30%) and the training (70%) parts. Table 1shows detailed statistics about images used for different sports activities for training and testing.

Table 1. Dataset divided into training and testing data.

Sports Category	Total Images	Training Data	Testing Data
Badminton	1645	1151	494
Basketball	1700	1190	510
Cricket	1635	1144	491
Rugby	1650	1155	495
Tennis	1670	1169	501
Volleyball	1650	1155	495
Total	**9950**	**6964**	**2986**

The classification is done with the help of Neural Networks with an accuracy of 96.64 %. To check the competence of our training the dataset is trained with the help of other classifiers like KNN, SVM, and Random Forest The confusion matrix of Neural Networks is shown in Table 2.

Table 2. Confusion matrix achieved by Neural Networks over the testing dataset.

	Badminton	Basketball	Cricket	Rugby	Tennis	Volleyball	∑
Badminton	**494**	0	0	0	0	0	**494**
Basketball	0	**510**	0	0	0	0	**510**
Cricket	0	0	**464**	27	0	0	**491**
Rugby	0	0	30	**465**	0	0	**495**
Tennis	6	0	12	25	**458**	0	**501**
Volleyball	0	0	0	0	0	**495**	**495**
∑	**500**	**510**	**506**	**517**	**458**	**495**	**2986**

Since our dataset is new and no previous works have been done on this dataset before therefore, we compared our results obtained from Neural Networks with the results of other classifiers. The comparison between the results can be seen in the Table 3., for the class Badminton the accuracy obtained is very less in case of KNN thus decreasing the overall accuracy of the class, in the case of Random Forest the accuracy procured in Cricket and Rugby are comparatively less, thus affecting its overall accuracy. In the case of SVM, all other classes give good results except the Cricket which gives less accuracy as compared to other classes. It can be seen that Neural Networks is providing very stable results for all classes as compared to other classifiers.

Table 3. Detailed category wise comparison between accuracies of various classifiers.

Sports Category	Random Forest	KNN	SVM	Neural Networks
Badminton	95%	67.17%	100%	100%
Basketball	91.22%	96.42%	91%	100%
Cricket	82%	96.96%	84.76%	94.50%
Rugby	75.65%	97.16%	93.27%	93.93%
Tennis	99.55%	100%	100%	91.41%
Volleyball	97.2%	97.78%	99.59%	100%
Total %	**90.10%**	**92.58%**	**94.77%**	**96.64%**

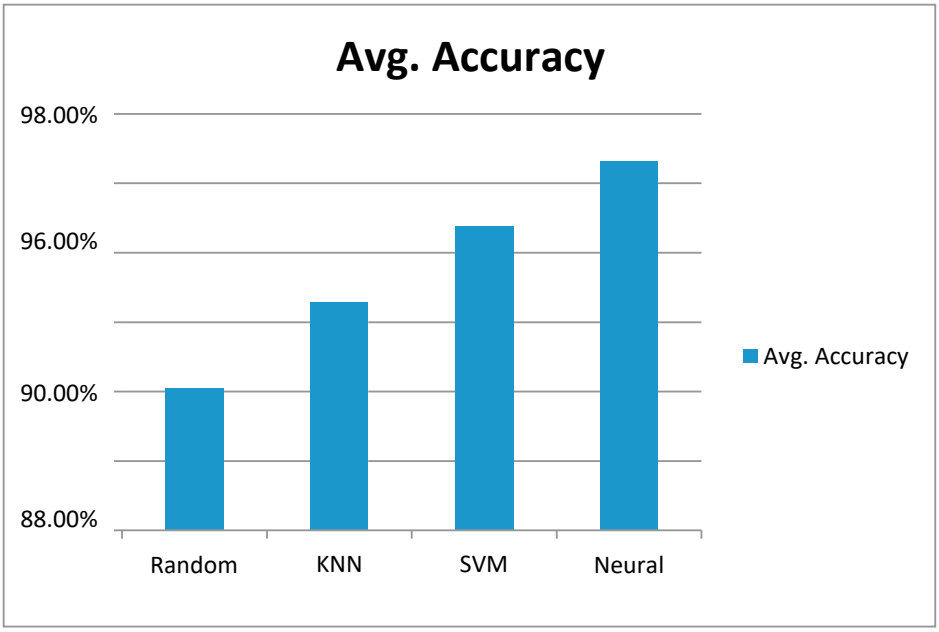

Fig. 4. Comparative analysis between different classifiers over the dataset

The comparative analysis shows that the results provided by the Neural Networks are most efficient with an average accuracy of 96.64% as shown in Fig. 4. The reason behind providing the best results is that Neural Networks work more efficiently when more data is provided to it, unlike other classifiers. The efficiency and computation of the Neural Networks increase as the amount of data increase and in our case, our dataset is quite large. But, in other classifiers, the performance becomes constant after a certain limit of data.

5. Conclusion

This paper presents a robust framework for the classification of various categories of sports using Neural Networks and InceptionV3. Sports images are divided into six classes, namely, rugby, basketball, tennis, badminton, cricket, and volleyball. As shown in the results, our framework has achieved an average accuracy of 96.64 %. To validate the effectiveness of Neural Networks, the results of other classifiers have been calculated. As the results show that the Neural Networks have achieved best accuracy as compared to other classifiers due to its ability to handle large datasets efficiently. These accuracies can also be increased as there is scope for the rectification of images as some images are blurred and dark which can be preprocessed. Further, the future scope of this paper is wide open. More categories can be incorporated for analysis. Another feature descriptor can be utilized for more effective results. Further, it can be utilized for the development of dataset generator for sports as through this framework, it can already identify sport activities efficiently. By embedding temporal information, the framework can be used for activity analysis and video searching.

6. References

[1] Tripathi, Vikas, Mittal Ankush, Gangodkar Durgaprasad, and Kanth Vishnu. (2019) “Real-time security framework for detecting abnormal event at ATM installations.” *Journal of Real-time image processing* 16(2): 535-545.

[2] Sanchez, Reillo, Sanchez Avila, and Gonzalez Marcos. (2000) “Biometric identification through hand geometry measurements.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10): 1168-1171.

[3] Thomas, Huang, and Vladimir Pavlovic. (1995) "Hand gesture modeling, analysis, and synthesis." *Proc. of IEEE International Workshop on Automatic Face and Gesture Recognition: 73-79*.

[4] Sharma, Abhishek, Tripathi Vikas, and Gangodkar Durgaprasad. (2019) "An Effective Video Surveillance Framework for Ragging/Violence Recognition." In: Kulkarni A., Satapathy S., Kang T., Kashan A. (eds) Proceedings of the 2nd International Conference on Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing, 828. Springer, Singapore.

[5] Guanochanga, Byron. (2018) "Towards a Real-Time Air Pollution Monitoring Systems Implemented using Wireless Sensor Networks: Preliminary Results." *IEEE Colombian Conference on Communications and Computing (COLCOM)*.

[6] Tripathi, Vikas, Durgaprasad Gangodkar, Ankush Mittal, and Vishnu Kanth .(2017) "Robust Action Recognition framework using Segmented Block and Distance Mean Histogram of Gradients Approach." Procedia computer science 115: 493-500.

[7] Jinjun, Wang, Changsheng Xu, and Chng. (2006) "Automatic Sports Video Genre Classification using Pseudo-2D-HMM." 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong:778-781.

[8] Ahmad, Ullah, Jamal Ahmad, Kamran Muhammad, and Mohammad Sajjad. (2018) "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features." *IEEE Access,* 6:1155-1166.

[9] Zhdanov, Pavel, and Adil Khan. (2018) "Improving Human Action Recognition through Hierarchical Neural Networks Classifiers." *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro: 1-7.

[10] Direkoğlu, Cem, and Noel E. O'Connor. (2012) "Team Activity Recognition in Sports. In: Fitzgibbon." *Computer Vision-ECCV 2012. ECCV2012. Lecture Notes in Computer Science*, 7578. Springer, Berlin, Heidelberg

[11] Sadlier, D.A, and Noel E. O'Connor. (2005) "Event detection in field sports video using audio-visual features and a support vector Machine." *IEEE Transactions on Circuits and Systems for Video Technology*, 15(10):1225-1233.

[12] Duan, Ling Yu, Min Xu, Qi Tian, Chang Sheng Xu, and J.S. Jin. (2005) "A unified framework for semantic shot classification in sports video." *IEEE Transactions on Multimedia*, 7(6):1066-1083.

[13] Waltner, Georg, Thomas Mauthner, and Horst Bischof. (2014) "Indoor activity detection and recognition for sports games analysis." *arXiv preprint arXiv:1404.6413*.

[14] Laptev, Marszalek, Schmid, and Rozenfeld. (2008) "Learning realistic human actions from movies." *Proceedings CVPR. IEEE: 1-8*.

[15] Ivan, Laptev, and Tony Lindeberg. (2003) "Interest point detection and scale selection in spacetime." *Proceedings of the International Conference on Scale Space Methods in Computer Vision*: 372–387. Springer-Verlag.

[16] Barshan, Billur, and Murat Cihan Yüksek. (2014) "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units." *The Computer Journal*, 1649-1667.

[17] Rugby March (2019): https://www.youtube.com/watch?v=so_BsA7THUY

[18] Tennis March (2019): https://www.youtube.com/watch?v=oyxhHkOel2I

[19] Cricket March (2019): https://www.youtube.com/watch?v=MPoasv2-hzY

[20] Basketball March (2019): https://www.youtube.com/watch?v=nAihXfsxDww

[21] Volleyball March (2019): https://www.youtube.com/watch?v=qLUoU7a5tY8

[22] Badminton March (2019): https://www.youtube.com/watch?v=4kvpgiDyElI