Model for Practice Badminton Basic Skills by using Motion Posture Detection from Video Posture Embedding and One-Shot Learning Technique

Nuttachot Promrit
Center of Excellence in AI and NLP
Department of Computing
Faculty of Science
Silpakorn University
Promrit_n@silpakorn.edu

Sajjaporn Waijanya
Center of Excellence in AI and NLP
Department of Computing
Faculty of Science
Silpakorn University
Waijanya_s@silpakorn.edu

ABSTRACT
Badminton is a sport that is very popular in Asia for players of all gender and all ages. The practice of basic badminton skill such as badminton posture practice is essential because the correct posture help to avoid injury and help for improving player skill. In this article, we propose The Model for Practice Badminton Basic Skills. We have created the video posture embedding by using the Triplet-Loss technique and develop the badminton player's motion posture detection by using the One-Shot Learning technique. The motion posture detection consists 8 badminton's postures and 4 other postures instances 1) Forehand Clear 2) Backhand Clear 3) Forehand Drop Shot 4) Backhand Drop Shot 5) Forehand Smash 6) Backhand Smash 7) Forehand Serve 8) Backhand Serve 9) Raise the right hand 10) Raise the left hand 11) Stand/Walk and 12) Run. The performance of the model present the Precision is 0.856, Recall is 0.845, F-Measure is 0.847 and the Accuracy is 0.845

CCS Concepts
• Computing methodologies → Artificial intelligence → Computer vision → Computer vision tasks → Activity recognition and understanding

Keywords
Triplet Loss; One-Shot; motion posture detection; badminton posture detection; video posture embedding; posture embedding; practice badminton posture model

1. INTRODUCTION
Nowadays, digital technology is increasingly in the health care and sports industry to use as tools for analyzing exercise behavior and statistical analysis from video images. The processes used may consist of machine learning, image processing and sensors. Videos of sports that have analyzed in a variety of categories such as dancing, golf, running, basketball, tennis and badminton. For Badminton, the research about image processing is now using for badminton analytic in various terms or views from the video, creating motion analysis and the researchers also interest in badminton player tracking and detection but not much research to create a model for practice badminton by machine.

AICCC 2019, December 21–23, 2019, Kobe, Japan

ACM ISBN 978-1-4503-7263-3/19/12…$15.00
DOI: https://doi.org/10.1145/3375959.3375981

Badminton is an extremely popular sport especially in Asia because it doesn't need much equipment just only badminton racket and bird or shuttlecock. The equipment price is various, high price is for the professional player and affordable price is for newbie who just want to exercise only. To play badminton, there are two teams to hit the shuttlecock and enjoy the game. The place for playing badminton mostly play in standard court but someone who just want to exercise can play in the simulated court. In Thailand we often found some people play badminton with neighbors and their families in the front yard and alley.

The sportspersons who want to progress in badminton or someone who want to archive efficiency exercising to play badminton in standard court or having a professional coach or teaching assistant are can help. The badminton basic practice for the players composing of how holding the badminton racket properly, focusing the position of shuttlecock and badminton stance. The badminton stance is an important for all badminton players according to making your performing properly can help in the game especially the newbies or the beginners who just start playing badminton the correct stance is a general problem they suffered. Another problem about power use for hitting the shuttlecock incompletely maybe the cause of game is not fun anymore for the competitor who play better. And another effect from performing with incorrect instance is an injury so learning the correct of basic instance stroking and footwork are recommended. Practice with professional coach or professional badminton player spend a lot of money but also practice with friends in the public court maybe the cause of awestruck or unconfident to do. It is better if the beginners can practice by themselves or using assistant tool for performing analysis.

The basic performing skills of badminton consist of gripping, footwork, serve, clear, stance, smash and short. The correctness of postures are essential for improving the performance. This article would like to present the posture classification model and real-time accuracy evaluation of badminton basic instance by Triplet Loss Model technique and One-shot model. The badminton basic consists of 8 instances 1) Forehand Clear 2) Backhand Clear 3) Forehand Drop Shot 4) Backhand Drop Shot 5) Forehand Smash 6) Backhand Smash 7) Forehand Serve and 8) Backhand Serve and these are tool for basic practice.

The raise of this article is organized as follows. The section 2 will present the related researches about the Image Processing and Artificial intelligence in Badminton sport. The section 3 will present the methodology and model that used in this research. And the section 4 will present the results of each model and the prototype of a practice tool. The research summary, research conclusion and research guidelines in the future will be presented in section 5.

2. RELETED WORKS

In recent years, video analytics technology is growing rapidly, and sports videos are a group of data sources that researchers are interesting. The technology that researchers often use to analyze sports videos is the use of deep learning. For example, the classification of sports video using Deep Convolutional Neural Networks [1], sports classification in sequential frames using the Convolutional Neural Networks combine with the Recurrent Neural Network to classify 5 sport categories [2] and another work researcher use the combination of deep learning and transfer learning apply with VGG-16 to classify 15 sport categories [3]. Not only the sport type classification, but many works in artificial intelligence and computer vision are also focusing on the specific type of sport such as Football, Basketball, Volleyball, Tennis and Badminton. Most of the research uses sports video from various tournament competitions to analyze play patterns in the area or player tracking to use the information from those videos.

For badminton, most of the research proposes the methodology for player detection and player movement from the video of the tournament as well. The game statistic base on classification result is a goal that researchers use to develop models. Probabilistic Hough Transform and Gaussian mixture model have used for the spatiotemporal and strokes classification using the dataset from Badminton World.TV channel [4]. Since professional players' playing statistics are useful information, The another researchers propose a method to analyze the video from 10 Olympic matches [5]. They analysis the player's detection and identification each score point and also segment the player strokes of top player and bottom player by using TCN (Temporal Convolutional Network). The application of computer vision and vector space model for tactical movement classification in badminton [6] use the video from the highest level tournament and Olympic Game. This work proposed the automate data gathering consist of player position detection, position annotation, trajectory extraction and tactical movement extraction and classification.

Because players position are an essential in badminton sports, the badminton player detection and tracking by machine learning are proposed by other researchers. For the example, The badminton player tracking based on Tracking-Learning-Detection (TLD) improving by image pixel [7].The faster region convolution neural network has propose to use for player detection by track the position of player from broadcast video[8]. To use the information for tactical analysis, another researcher proposed the position detection based on multi-person pose estimation by use MPII Human Pose database and Fast R-CNN as the methodology [9] .

However, creating a model that aims to be used as a tool for practicing basic postures for beginners may not be suitable to be used the Athletes' posture in high level tournament videos. Creating tools for beginners may require other methods or specific data preparation for use. We found the research that has the other methodology to analyze the strokes and movement by using sensor stick to the bottom of the handle of the badminton racket. They analyze the stroke by sound [10]. And another work [11], they separate their work to many sub-project. They use IoT in part of equipment to create smart-racket and use deep learning to analyze the video.

The main objective of this article is creating motion posture detection model to be the model for practice badminton basic skill. By this reason, we have to predict the posture as real time or near real time. PoseNet Library [12,13] was used to get a key point of human from video and real time process. The researcher adopts it to detect the posture such as the abnormal human behavior using CNN and LSTM [14]. To predict the posture compare with the reference posture we will create the video posture embedding by using Triplet-loss and compare by One-short technique. The Triplet-loss is a well-known approach to learning embedding to compare images [15] and some researcher applied to use with voice [16,17]. And in the video posture embedding process, The Residual Block [18] will use to be part of Deep Network in this work.

3. METHODOLOGY AND MODEL

3.1 Process Overview

In this article, we propose The Model for Practice Badminton Basic Skills by using Motion Posture Detection from Video Posture Embedding and One-Shot Learning Technique. Since the objective of this work is creating the model for practice badminton basic skill by using the real-time camera via the web to predict player posture. Therefore, the main process of the work consists of 3 processes as follows 1) Video preparation 2) Model Training and Evaluation and 3) Real-time Detection Process. Figure 1 shown 3 main process of work.

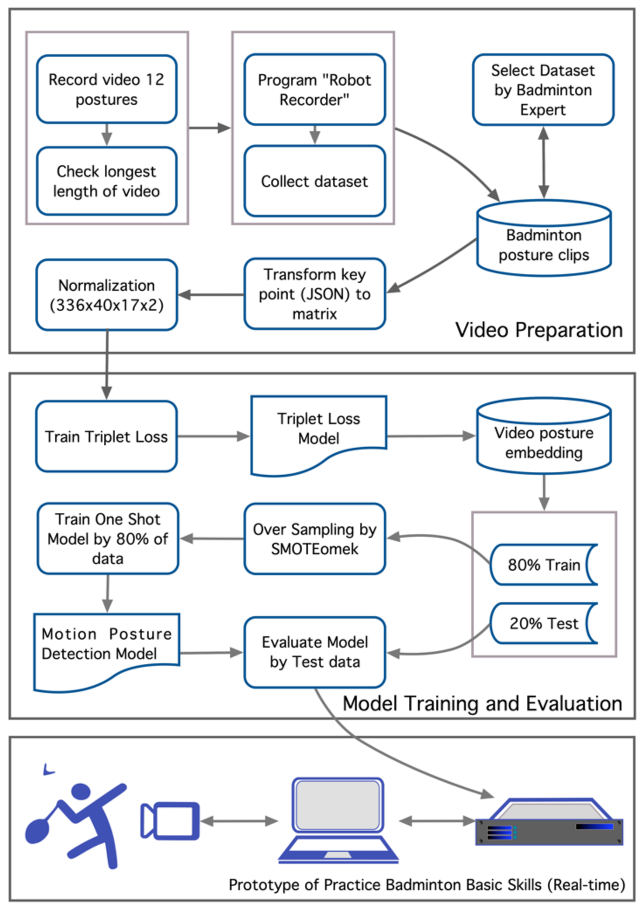

Figure 1. Process Overview

From 3 main process in Figure 1, we will describe sub-processes as following subsection 3.2 the video preparation, subsection3.3 the methodology of video posture embedding, subsection 3.4 one-shot learning and subsection 3.5 the process of real-time prediction of Motion Posture Detection Model.

3.2 Video Preparation

To prepare the dataset, we recording 12 clips video of 1 player who is Chair of Silpakorn University badminton club and another member of badminton club to be assistant. 12 video clips consisting 8 badminton posture and 4 other posture which are 1) Forehand Clear 2) Backhand Clear 3) Forehand Drop Shot 4) Backhand Drop Shot 5) Forehand Smash 6) Backhand Smash 7) Forehand Serve 8) Backhand Serve 9) Raise the right hand 10) Raise the left hand 11) Stand/Walk and 12) Run. Next, we gain the longest time from 12 postures by finding the maximum length of the video clip which long 1 second.

In the posture recording, we set the camera on the right side of the player-side court, as in Figure 2 shows the preparation of the data collection environment. The example of posture in preparation process shows in Figure 3.

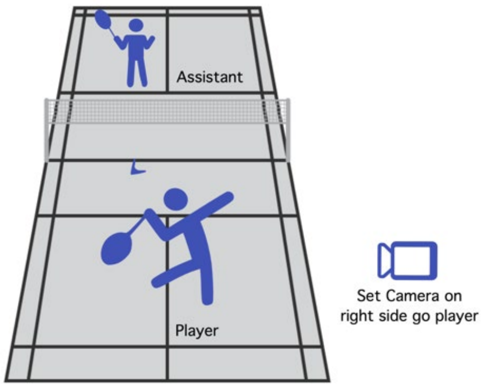

Figure 2. The preparation of the data collection environment

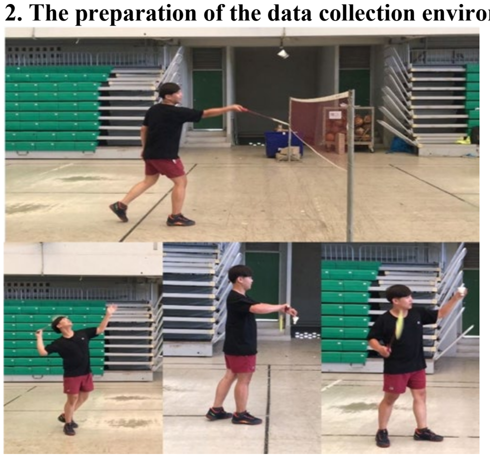

Figure 3. The example of postures in preparation process

We program the Robot Recorder script by using PoseNet library to open the camera via the web browser to get key points of 1 frame and sent via RabbitMQ. The Python script run as backend and receive key points from RabbitMQ. Because the laptop used in the development of the model in this research can render the key points from the video at 40 frames per second. Therefore, 1 badminton posture has a length of 40 frames per second. The key point will show in the monitor as mirror. The example of badminton key point detection is show in Figure 4.

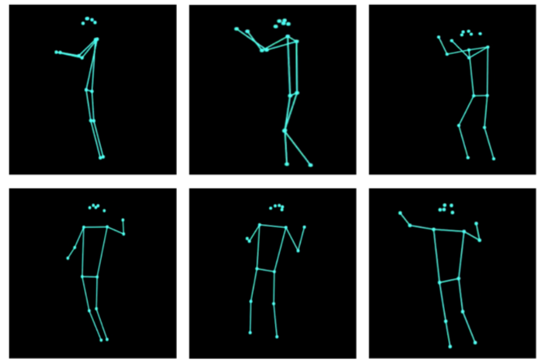

Figure 4. The example of badminton key point detection (mirror display)

We collect 40 frames of 1 posture and store in JSON format. The total number of video posture collection is 720 clips. Then we sent all clips to badminton expert for selecting the valid posture videos 336 clips. To prepare date for training model, we transform key points to matrix by [number of clips X number of frame X number of key point X dimension] therefore matrix size will be [366 X 40 X 17 X 2]. Then normalize the data to be between 0 and 1.

3.3 Video Posture Embedding by Triplet Loss

Triplet loss is a well-known approach to learning embedding to compare images especially in research about face recognition and authentication. The Motion pictures or videos are made up of multiple frames of overlapping images. Therefore, Triplet Los suitable to be built Video embedding as well. Triplet-loss function instants of 3 inputs which are baseline (Anchor: a), right value (Positive: p) and wrong value (Negative: n).

The video posture embedding in this article use input in term of the matrix of clip video size 40X17X2. The anchor and positive is clips video of same posture whereas negative is the image of different posture. In order to train the Triplet loss to create the video posture embedding, there is a training process as shown in Figure 5 and the details of Deep Network shown in Figure 6.

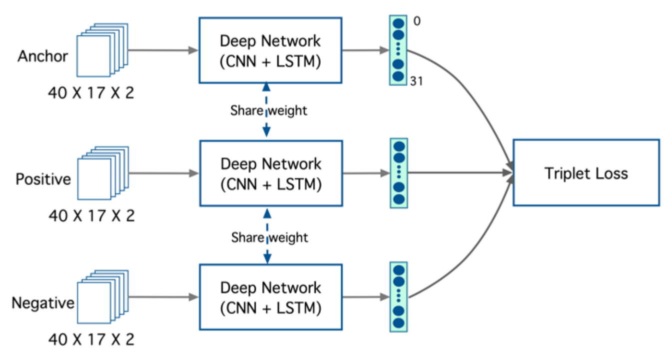

Figure 5. Triplet Loss Model Training

The Figure 6 shown Deep Network by The Convolutional Neural Network and LSTM. The input of Residual block is a frame of video in matrix form size 17X2 (17 key points). The output of Residual is feature map. Then we flatten feature map to 1 dimension vector and forward to LSTM.

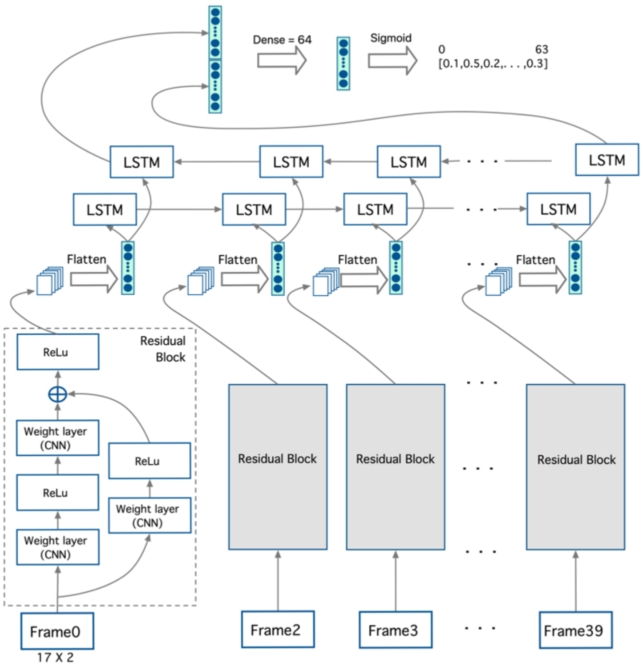

Figure 6. Deep Network Detail

In part of LSTM layer we use Bidirectional LSTM to create output as vector 1 dimension with size is 128 and forward to dense layer that has size is 64. The sigmoid is activate function use for adjust the output of Deep Network. After this process, the vector of each image will be calculated the loss by the loss function as in equation (1) and train the Video Posture Embedding 1,000 epochs with batch size is 32, then select weight from epoch with minutest loss value

$$Loss = \sum_{i=1}^{N} \left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha \right]_+ \quad (1)$$

3.4 One-Shot Learning for Video Posture

One-shot learning is widely used in an image classification domain and appropriate for limit training data. In this article, we recorded each video posture instead of using the pose from the video from the badminton tournament as other research uses. Since we want to use the video posture as a reference for practice and the pose from the badminton tournament players must move their posture by the situation, it is not suitable for our objective. After the experts select quality videos, only 336 video content is left. From the above reasons, learning with a One-Shot is an appropriate method of comparing badminton video postures.

From 336 video posture, we separate 296 video posture to train One-Shot. Because of the dataset is imbalance class, therefore we balance the data by over-sampling using SMOTEomek [19] (Synthetic Minority Over Sampling Technique). We separate data after over-sampling to be 80% for training and 20% to validating the One-Shot learning. We pair 2 video posture and label 1 for same posture and label 0 for different posture then transform to video posture encoding by triplet loss. Then vector of video posture will send to the Two Fully Connected Network + Sigmoid then find the similarity score of 2 video posture by Similarity Function and Sigmoid Function. One-Shot Learning shows in Figure 7.

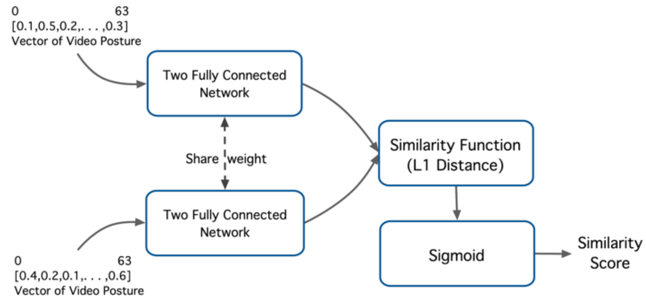

Figure 7. One-Shot Learning

The output of One-Shot learning is Similarity score that have value between 0 to 1. The One-Shot learning model will use 12 posture as a reference in real-time detection process that we will describe in the next subsection

3.5 Real Time Detection Process

Reference to the objective of this article, We want to create the Model for Practice Badminton Basic Skills by using Motion Posture Detection that mean the model should be detect the player posture as real-time or near real-time. Therefore we design the real-time detection process as shown in Figure 8.

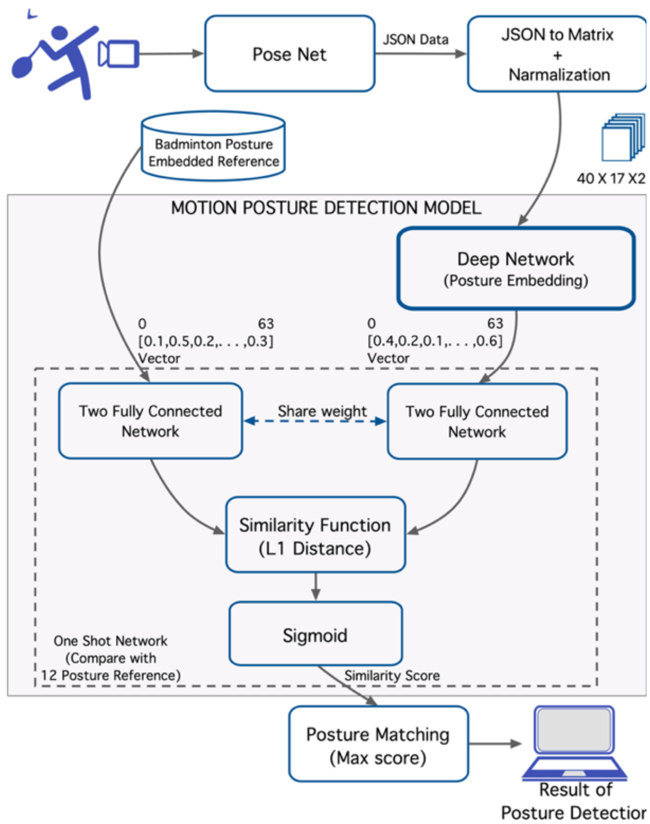

Figure 8. Real-time Detection Process

The process of Real-time detection has the details are as follows.

(1) The player opens the web application, and web application use PoseNet library to open the camera via the web browser and get key points of posture.

(2) The keypoint of posture has normalized to 40 frames and send to the Motion posture detection model.

(3) Sub-model Deep Network will transform the video of posture to the video posture embedding and compare with the posture embedding from the badminton posture

embedding reference (data from triplet loss training) by using One-Shot model.

(4) The similarity score will send to posture matching process to find max score and report to player.

From the process designed in Figure 8, we use the prototype of the web application for practice badminton basic skills to evaluate our model as real-time uses.

4. RESULTS AND DISCUSSION

In the experiment, we will report the results of The Motion Posture Detection Model according to the working process of the model, beginning with 4.1) video embedding results 4.2) One-Shot Learning Evaluation 4.3) Measurement by Confusion Matrix and 4.4) The discussion of results, respectively as follows.

4.1 The Video Embedding Results

To create the video embedding, we had defined the parameter of Triplet Loss Model by used dimension is 64. We had defined 1,000 epochs and Batch size id 32 for training the video posture embedding. The result of video posture embedding shown in Figure 9. which can be seen that there are clusters of each posture. We can be seen that each position held separately quite clearly And good for prediction in the One-Shot Model.

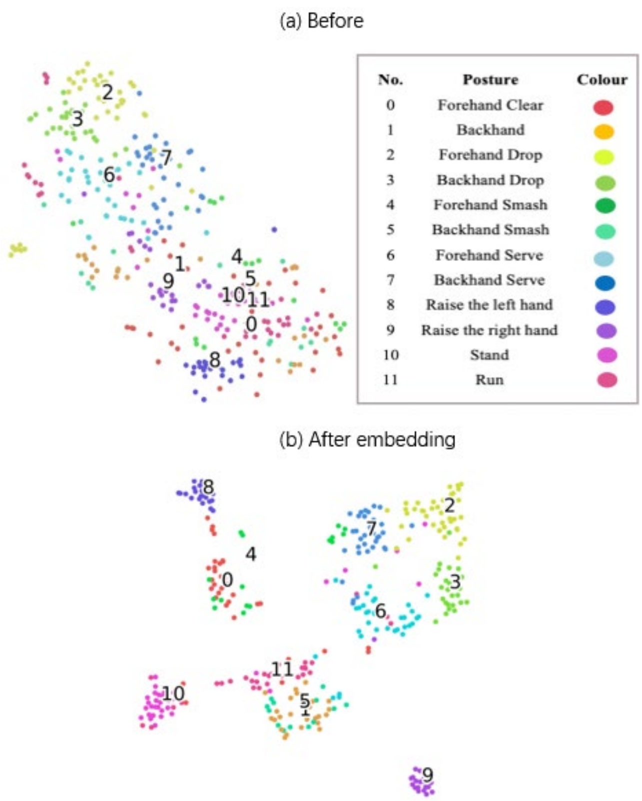

Figure 9. Video Posture Embedding.
(a) The posture spots distribution before embedding.
(b) The posture spots distribution after embedding.

4.2 One-Shot Learning Evaluation

For One-Shot Model, we defined dimension is 64 batch size is 32. The dataset 336 postures consist of 12 classes. We separate those to 80% (256 postures) to use in One-Short learning and 20% (84 postures) use for test the model. But the dataset in each class in our experiment is an imbalance, then we balancing by over-sampling with SOMTEomek. Finally, we got the balancing data is 356 postures. We have the assumption that imbalance data will affect the accuracy of the model. Therefore we experiment with the imbalance and balance data then compare the result of both. Table 1 shows the number of postures in each class on imbalance and balance data for One-Short learning Model.

Table 1. Number of Imbalance and Balance in each class

Postures	Imbalance (Im)	Balance (B)
Forehand Clear	25	30
Backhand Clear	18	29
Forehand Drop Shot	30	29
Backhand Drop Shot	20	30
Forehand Smash	12	30
Backhand Smash	11	29
Forehand Serve	26	30
Backhand Serve	25	29
Raise the right hand	18	30
Raise the left hand	16	30
Stand/Walk	26	30
Run	25	30

The experiment with imbalanced data, we use 256 video postures and expand to 3024 pairs then separate data 80% (2419 pairs) for training and 20% (605 pairs) for validating. And the experiments with balanced data, we use 356 video postures and expand to 4272 pairs then separate data 80% (3417 pairs) for training and 20% (855 pairs) for validating. We define the epochs is 30,000 to train and validate the model. The validation loss (minimum) of model with imbalanced data is 0.06486 and the model with balanced data is 0.00028. The result comparison of both experiment shown in Figure 10.

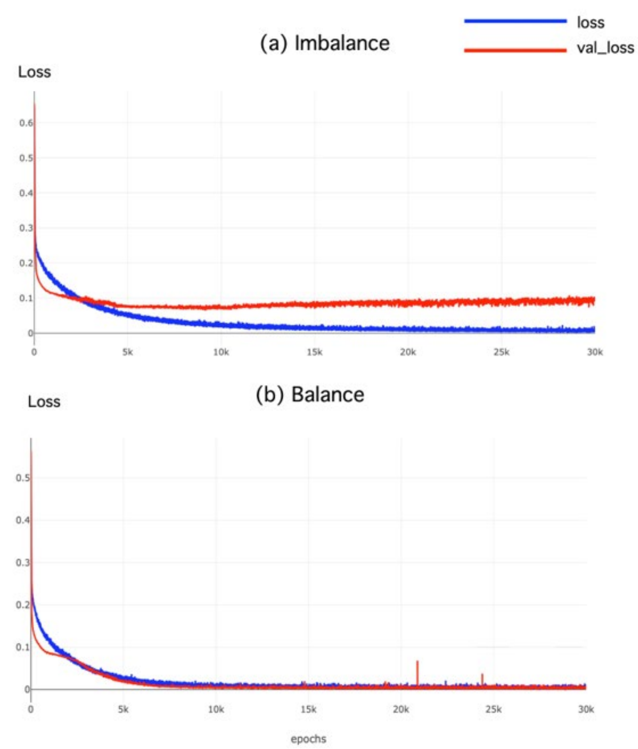

Figure 10. Loss of the video posture One-Shot model.
(a) Loss of the model with imbalanced data
(b) Loss of the model with balanced data

The validation accuracy of model with imbalanced data is 0.988 and the validation accuracy of model with balanced data is 1. The accuracy result comparison of both experiment shown in Figure 11.

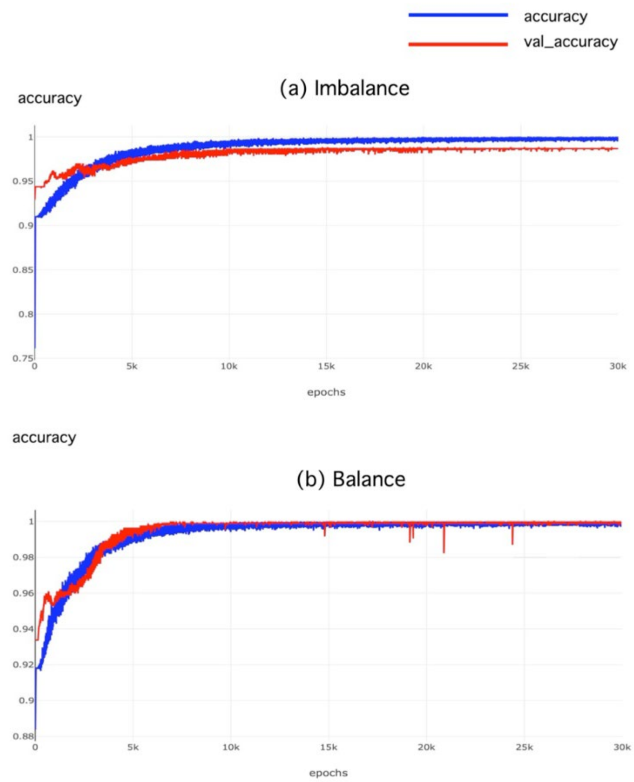

Figure 11. Accuracy of the video posture One-Shot model. (a) Accuracy of the model with imbalanced data. (b) Accuracy of the model with balanced data

From images 10 and 11, which show graphs of training and validate The Motion Posture Detection Model, We found that when balancing data in each class the loss value of with balanced data is lower and the accuracy is higher than the model with imbalance.

4.3 The Measurement by Confusion Matrix

To measure and evaluate the model, we use test data 84 postures and use the confusion matrix to calculate the precision, recall and F1-score of validate data. The number of test data in each class shown in Table 2.

Table 2. Number of posture in each class

Postures	Num	Postures	Num
A: Forehand Clear	9	G: Forehand Serve	8
B: Backhand Clear	6	H: Backhand Serve	8
C: Forehand Drop Shot	10	I: Raise the right hand	6
D: Backhand Drop Shot	7	J: Raise the left hand	6
E: Forehand Smash	4	K: Stand/Walk	8
F: Backhand Smash	3	L: Run	9

The confusion matrix of model with imbalanced data shown in Figure 12 (a) and with balanced data shown in Figure 12 (b)

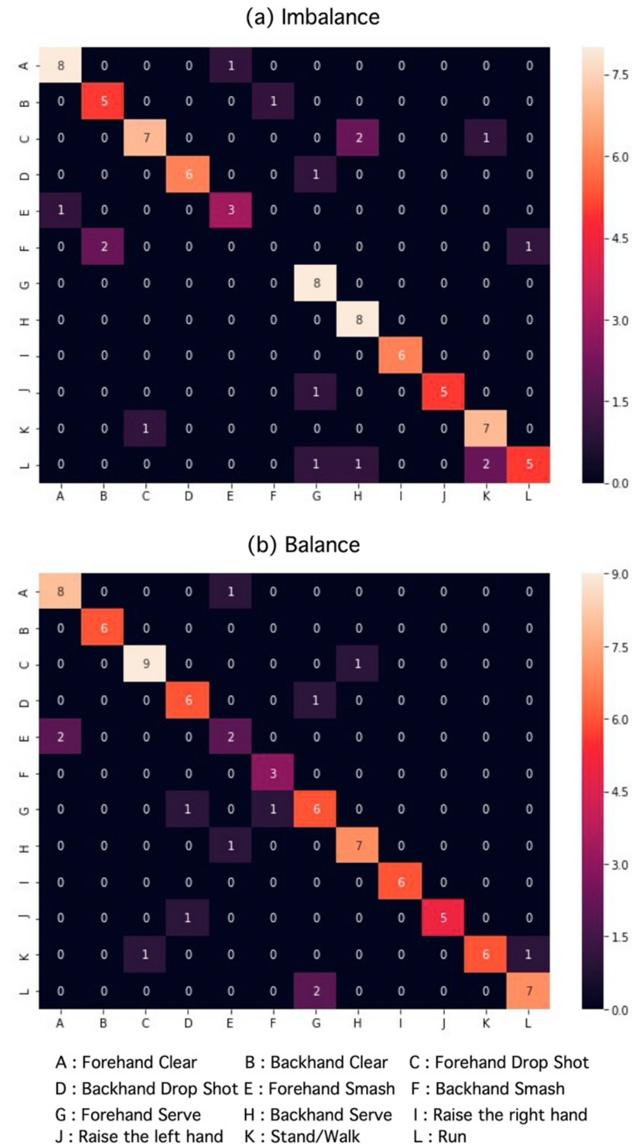

Figure 12. Confusion Matrix of the video posture detection model (a) Confusion Matrix of the model with imbalanced data. (b) Confusion Matrix of the model with balanced data

In Figure 12, when we compare the confusion matrix of model with imbalance data and model with balance data the matrix in Figure 12 (b) shows the result of the model with balance data is improved overall. In class B: Backhand Clear improve to 100% and F: Backhand Smash, the model with balance data can improve from 0 to 100%. The number of each class in the confusion matrix will use to calculate for measure the performance. The Precision, Recall and F1-score were calculated and shown in the Table 3. And the Table 4. show the result of macro average and weight average and accuracy of 2 experiments. In both of table, The column “Im” means the result of the model with imbalanced data and the column “B” means the model with balanced data.

Table 3. The Precision, Recall, F1-Score Test by 84 Posture

Postures	Precision		Recall		F1-score		support
	Im	B	Im	B	Im	B	
A: Forehand Clear	**0.889**	0.800	0.889	0.889	**0.889**	0.842	9
B: Backhand Clear	0.714	**1.000**	0.833	**1.000**	0.769	**1.000**	6
C: Forehand Drop Shot	0.875	**0.900**	0.700	**0.900**	0.778	**0.900**	10
D: Backhand Drop Shot	**1.000**	0.750	0.857	0.857	**0.923**	0.800	7
E: Forehand Smash	**0.750**	0.500	**0.750**	0.500	**0.750**	0.500	4
F: Backhand Smash	0.000	**0.750**	0.000	**1.000**	0.000	**0.857**	3
G: Forehand Serve	**0.727**	0.667	**1.000**	0.750	**0.842**	0.706	8
H: Backhand Serve	0.727	**0.875**	**1.000**	0.875	0.842	**0.875**	8
I: Raise the right hand	1.000	1.000	0.833	0.833	0.909	0.909	6
J: Raise the left hand	1.000	1.000	1.000	1.000	1.000	1.000	6
K: Stand/Walk	0.700	**1.000**	**0.875**	0.750	0.778	**0.857**	8
L: Run	0.833	**0.875**	0.556	**0.778**	0.667	**0.824**	9

Table 4. The Average and Accuracy

Postures	Precision		Recall		F1-score	
	Im	B	Im	B	Im	B
Macro Average	0.768	**0.843**	0.774	**0.844**	0.762	**0.839**
Weight Average	0.807	**0.856**	0.809	**0.845**	0.798	**0.847**
Accuracy	**Imbalance**			**Balance**		
	0.809			**0.845**		

4.4 The discussion of results

Since the basic badminton posture detection is the objective of this article. The model from this experimental will be applied as API to the application for practice basic badminton posture. Therefore we have to collect the dataset of basic posture from the badminton player similar as we learn from the coach. The clip video of each pose has selected by the expert; those use for training and validate the model and knowledge of the model will be the reference for practicing.

In the other works, the objective is not for practicing. They collect the dataset of posture by using the clip video from the world-class tournament to help in analytic and strategy planning. Most of them shown the accuracy of player detection and stroke classification but didn't show the accuracy of badminton posture classification. Moreover, the dataset from the world-class tournament is not appropriate for our work. It because of each pose depends on the situation of the match.

In this article, we experiment by One-Shot Learning to train the model. The problem in our experiment is the imbalance of the number of data in each class after selecting by the expert. When we train and validate the model with imbalanced data the loss value still not approaching to zero, the accuracy value is 0.988. Ee balancing the data by oversampling using SOMTEomek, it shown the validation loss value to 0.00028 (near 0) and the accuracy of the model reaches to 1.0.

We set the assumption by reference from the validation loss and validation accuracy of the model with balance data. The experimental result using testing dataset should be related to the result of the validation. We compare the performance of 2 models with the measurement tool using the confusion matrix and calculate The precision, Recall and F1-Score. The overall result of the model with a balanced dataset is accurate more than the model with imbalance dataset as shown in table 4. When analyzing the results of the precision, recall and F1-score in each posture, the result of both models shows the same performance in poses I: Raise the right hand and J: Raise the left hand. For the other 10 postures, the model with a balanced dataset show higher performance in 6 postures, especially in B: Backhand Clear improves to 1.00 and F: Backhand Smash improves from 0.00 to 0.875.

5. CONCLUSION AND FUTURE WORK

In this article, we focus on the motion posture detection model. We use the 366 videos of badminton postures to create video posture embedding by using the Triplet loss model. The posture after embedding and distribute by classes of postures, the video posture embedding will use for the One-Shot learning model. To complete the motion posture detection model we using One-shot model to predict the posture. However, the video in each class is imbalanced its effect on the result of One-shot learning and show the overfitting in loss and accuracy. Therefore we balancing the data by oversampling using SOMTEomek, it shown the accuracy of the model reaches to 1.0. Finally, we test the model with a test data 84 posture and measure by the Precision, the Recall and the F1-Score. The prediction results as wrong in some postures especially "backhand pose" on triplet loss with imbalanced data. It may impacted by the environment set up in the video collection process since we set up the camera on the right side of the badminton court when the player play backhand the body will turn to left side. When we balance the data, the accuracy is improved.

We also develop the prototype via web application and test the model by real-time predict. The example of the prototype screen (It's displayed as a mirror) shown in Figure 13 and Figure 14. The result from real-time prediction by the prototype shown in the same way with the process of testing model. In the future, we will collect more video and set up the camera in the suitable position. moreover, we will continue to develop our prototype and test with the beginner player as well.

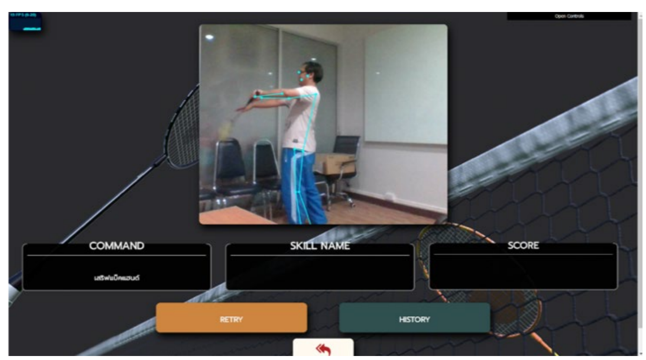

Figure 13. The prototype screen (mirror display) with real-time prediction (Backhand Serve)

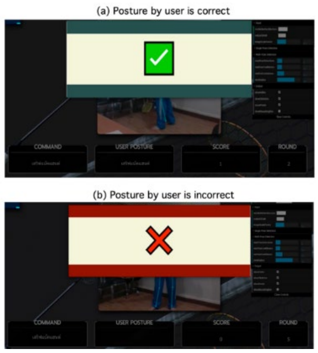

Figure 14. The prototype screen (Real-time prediction) (a) Posture by user is correct (b) Posture by user is incorrect

6. ACKNOWLEDGMENTS

Our thanks to all member of SU Badminton CLUB especially Mr. Purachet Chareonporn Chair of SU Badminton CLUB to be player for collecting data. Thanks for Mr. Parinya Bunkoed to be assistant of a player and both of them for help to develop the web application as a prototype. Thanks for Conference Secretary of AICCC 2019 who give chance to us for the paper submitted.

7. REFERENCES

[1] Nallathigal S. Rani, P. Ashok Rao, and Paul Clinton. 2018. Visual recognition and classification of videos using deep convolutional neural networks. DOI: https://doi.org/10.14419/ijet.v7i2.31.13403

[2] Mohammad Ashraf Russo, Alexander Filonenko, and Kang-Hyun Jo. 2018. Sports Classification in Sequential Frames Using CNN and RNN. In 2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT), 1–3. DOI: https://doi.org/10.1109/ICTROBOT.2018.8549884

[3] Mohammad Ashraf Russo, Laksono Kurnianggoro, and Kang-Hyun Jo. 2019. Classification of sports videos with combination of deep learning models and transfer learning. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1–5. DOI: https://doi.org/10.1109/ECACE.2019.8679371

[4] Wei-Ta Chu and Samuel Situmeang. 2017. Badminton Video Analysis Based on Spatiotemporal and Stroke Features. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR '17), 448–451. DOI: https://doi.org/10.1145/3078971.3079032

[5] Anurag Ghosh, Suriya Singh, and C. V. Jawahar. 2018. Towards Structured Analysis of Broadcast Badminton Videos. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 296–304. DOI: https://doi.org/10.1109/WACV.2018.00039

[6] Kokum Weeratunga, Anuja Dharmaratne, and Khoo Boon How. 2017. Application of Computer Vision and Vector Space Model for Tactical Movement Classification in Badminton. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 132–138. DOI: https://doi.org/10.1109/CVPRW.2017.22

[7] Takashi Kamiyama, Yoshinari Kameda, Yuichi Ohta, and Itaru Kitahara. 2017. [Paper] Improvement of Badminton-Player Tracking Applying Image Pixel Compensation. ITE Transactions on Media Technology and Applications 5, 2 (2017), 36–41. DOI: https://doi.org/10.3169/mta.5.36

[8] Ya Su and Zhe Liu. 2018. Position Detection for Badminton Tactical Analysis based on Multi-person Pose Estimation. In 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 379–383. DOI: https://doi.org/10.1109/FSKD.2018.8686917

[9] Nur Azmina Rahmad, Nur Anis Jasmin Sufri, Nurul Hamizah Muzamil, and Muhammad Amir As'ari. 2019. Badminton player detection using faster region convolutional neural network. Indonesian Journal of Electrical Engineering and Computer Science 14, 3 (June 2019), 1330–1335. DOI: https://doi.org/10.11591/ijeecs.v14.i3.pp1330-1335

[10] Juyi Lin, Chia-Wei Chang, Chih-Hao Wang, Hong-Chuan Chi, Chih-Wei Yi, Yu-Chee Tseng, and Chih-Chuan Wang. 2017. Design and implement a mobile badminton stroke classification system. In 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), 235–238. DOI: https://doi.org/10.1109/APNOMS.2017.8094141

[11] Tzu-Han Hsu, Ching-Hsuan Chen, Nyan Ping Ju, Tsi-Ui Ik, Wen-Chih Peng, Chih-Chuan Wang, Yu-Shuen Wang, Yuan-Hsiang Lin, Yu-Chee Tseng, Jiun-Long Huang, and Yu-Tai Ching. 2019. CoachAI: A Project for Microscopic Badminton Match Data Collection and Tactical Analysis. CoRR abs/1907.12888, (2019). Retrieved from http://arxiv.org/abs/1907.12888

[12] Pose estimation | TensorFlow Lite. TensorFlow. Retrieved October 12, 2019 from https://www.tensorflow.org/lite/models/pose_estimation/overview

[13] tensorflow/tfjs-models. GitHub. Retrieved October 12, 2019 from https://github.com/tensorflow/tfjs-models

[14] Thomas Gatt, Dylan Seychell, and Alexiei Dingli. 2019. Detecting human abnormal behaviour through a video generated model. In 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 264–270. DOI: https://doi.org/10.1109/ISPA.2019.8868795

[15] Shizhou Zhang, Qi Zhang, Xing Wei, Yanning Zhang, and Yong Xia. 2018. Person Re-Identification With Triplet Focal Loss. IEEE Access 6, (2018), 78092–78099. DOI: https://doi.org/10.1109/ACCESS.2018.2884743

[16] Hervé Bredin. 2017. TristouNet: Triplet loss for speaker turn embedding. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5430–5434. DOI: https://doi.org/10.1109/ICASSP.2017.7953194

[17] Nicolas Turpault, Romain Serizel, and Emmanuel Vincent. 2019. Semi-supervised Triplet Loss Based Learning of Ambient Audio Embeddings. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 760–764. DOI: https://doi.org/10.1109/ICASSP.2019.8683774

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. DOI: https://doi.org/10.1109/CVPR.2016.90

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, (June 2002), 321–357. DOI: https://doi.org/10.1613/jair.953