Journal of Theoretical and Applied Information Technology

30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEEP YOLOv8-BASED HANDBALL DETECTION SYSTEM WITH TRANSFER LEARNING APPROACH

[1]R.J.POOVARAGHAN, [2] P. PRABHAVATHY

[1]Research Scholar, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

[2]Professor, School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India

[1]poovaraghan.rj2017@vitstudent.ac.in, [2]pprabhavathy@vit.ac.in

Corresponding Author: *[2] P. PRABHAVATHY

ABSTRACT

Utilizing advanced computer vision techniques such as deep learning and object tracking algorithms, AI-powered active player detection in handball videos offers the capability to automatically track players' movements within high-speed matches. This innovation not only enriches coaching insights into player performance and team dynamics but also elevates viewer engagement through real-time analysis and augmented reality enhancements. In the context of practice-based handball videos, where multiple players frequently appear together, not all participants engage in the specific exercise or adopt the recommended handball techniques. This study explores the novel approach of employing the CNN based YOLOv8 pre-trained model in conjunction with transfer learning techniques for enhanced handball recognition. The YOLOv8 architecture's advanced capabilities are harnessed to address existing gaps in player tracking, ball trajectory prediction, and complex player interactions. Through transfer learning, the model is fine-tuned using handball-specific data, enabling adaptation and specialization in identifying players, the ball, and key elements. The method leverages YOLOv8's real-time processing and multi-scale analysis to improve accuracy in dynamic game scenarios, overcoming challenges like occlusion and rapid motion. By integrating the YOLOv8 pre-trained model with transfer learning, this approach showcases a promising advancement in achieving comprehensive and efficient handball recognition, significantly enhancing insights into player dynamics, ball movement, and overall gameplay. The fusion of YOLOv8 with transfer learning involves leveraging YOLOv8's pre-trained features for extracting object characteristics, followed by fine-tuning the model on handball-specific data to enhance its ability to recognize players, the ball, and other essential elements in the context of handball recognition. We systematically evaluated the proposed approach using a custom dataset of 751 handball scene videos captured during training sessions for young cadets and handball schools for both girls and boys [22]. Testing encompassed nearly 60,000 frames and incorporated metrics such as sensitivity, specificity, and accuracy. The results demonstrated that our method surpassed state-of-the-art techniques, showcasing heightened accuracy. Notably, the proposed method exhibited enhanced efficiency, achieving a sensitivity 92.18%, specificity 91.13%, accuracy 93.57% and F-score 94.33% respectively.

Keywords: *Handball Recognition, Deep Learning, YOLOv8, Transfer Learning, Computer Vision*

1. INTRODUCTION

In the realm of handball recognition, the fusion of computer vision, artificial intelligence, and deep learning has yielded a groundbreaking approach that transcends traditional analysis. By harnessing the power of computer vision, machines can decipher intricate visual cues from handball videos, identifying players, ball trajectories, and tactical maneuvers. The integration of artificial intelligence imparts decision-making capabilities, enabling systems to comprehend player strategies and game dynamics. Meanwhile, the inclusion of deep learning, a subset of AI, empowers algorithms to autonomously learn from data, uncover complex patterns, and predict player movements. This synergy of disciplines holds transformative potential, reshaping coaching methodologies,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

offering real-time insights, and immersing viewers in a dynamic and interactive handball experience. As technology blurs the lines between observation and intelligence, handball recognition enters a new era driven by the convergence of computer vision, artificial intelligence, and deep learning. Handball is a dynamic indoor team sport that involves two teams, each comprising seven players, including a goalkeeper. The primary objective is to score goals by propelling a ball into the opposing team's goalpost. Players use techniques like passing, dribbling, and shooting to create scoring opportunities while adhering to regulations that govern physical contact which shown in Figure 1. The game consists of two halves, typically lasting around 30 minutes each, with the team accumulating the most goals emerging victorious. Handball blends elements from basketball, soccer, and hockey, incorporating rapid ball movement, strategic positioning, and interactive player engagements. Matches are played on a court with specific dimensions and markings, showcasing the sport's agility, teamwork, and tactical intricacies [1].

Handball recognition encounters a set of intricate challenges rooted in the dynamic nature of the sport and the complex environment of the playing field. Fast-paced player interactions often lead to

occlusion, where players obstruct each other's visibility, making accurate tracking a formidable task. The sport's rapid motion further compounds the issue, demanding recognition systems capable of capturing and interpreting swift player movements. The dynamic background of the handball court adds another layer of complexity, as the changing scene can confuse recognition algorithms attempting to differentiate players from their surroundings. Additionally, the variability in player appearance, the small and agile ball, and the low lighting conditions within indoor courts pose further obstacles to accurate player and ball tracking. Addressing these challenges necessitates the development of sophisticated computer vision techniques, machine learning models, and innovative algorithms, each tailored to handle the intricacies of handball's diverse game situations and interactions.

Figure 1. Different Actions in Handball Recognition

In the domain of computer science, handball recognition relies on a diverse range of techniques to derive insightful information from game videos. Central to this is computer vision, where sophisticated object detection algorithms like YOLO and Faster R-CNN [2-8] come into play to pinpoint players, the ball, and critical elements within the handball footage. Motion tracking algorithms work in tandem, allowing the monitoring of player movements through the identification of distinctive markers across consecutive frames, enabling the reconstruction of their trajectories and positional data. Machine learning and deep learning, encompassing CNNs and RNNs, are harnessed to not only classify player actions and predict ball trajectories but also decipher intricate patterns within gameplay [9-12]. The extraction of essential features, such as speed and acceleration, contributes to nuanced player analysis and strategic comprehension. Further enhancing this field are pose estimation techniques, which unravel players' body postures and movements, providing insights into their dynamics during different in-game scenarios. As technology evolves, the fusion of multi-sensor data, real-time processing, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

augmented reality applications amplify the dimensions of handball recognition in computer science, significantly augmenting coaching methodologies, player performance evaluation, and viewer engagement [13-15].

The application of computer vision techniques in handball recognition reveals notable gaps and challenges that offer avenues for further development. The issue of player occlusion persists, impacting the accurate tracking of players' movements and the analysis of their trajectories when obstructed by others. Equally, achieving precise ball tracking amidst fast-paced exchanges remains a challenge, affecting the overall understanding of the game's dynamics. The intricacies of complex player interactions, such as coordinated passes and strategic formations, continue to present difficulties for accurate recognition. Striking a balance between real-time processing and recognition accuracy, especially during live matches characterized by swift movements, remains a complex endeavor. Additionally, adapting to varying lighting conditions, ensuring reliable player identification across changing appearances, and addressing data annotation scarcity all contribute to the existing gaps. Bridging these challenges necessitates interdisciplinary collaborations, dataset availability, the refinement of computer vision models, and a consideration of ethical concerns surrounding data privacy. By addressing these gaps, computer vision techniques can further enrich handball recognition, advancing the sport's analysis and engagement. Existing handball recognition models may exhibit gaps such as challenges in accurate player tracking, ball trajectory prediction, and handling complex player interactions. These gaps can lead to compromised performance in scenarios involving occlusion, fast motion, and varied lighting conditions. Furthermore, real-time processing and scalability might be limiting factors in some models, hindering their applicability to live matches and large datasets. Figure 2. Shows Handball Training Court.

Figure 2. Handball Training Court

The novelty introduced by using the YOLOv8 pre-trained model, along with transfer learning techniques, lies in its potential to bridge these gaps. YOLOv8's advanced architecture enhances accuracy and speed, making it adept at precisely identifying players and objects in real-time, even within dynamic handball game scenarios. Through transfer learning, the model's existing knowledge is leveraged and fine-tuned on handball-specific data, enabling it to adapt and specialize in detecting players, the ball, and relevant elements in the sport. This approach capitalizes on YOLOv8's multi-scale analysis to address challenges like occlusion and rapid motion, enhancing player tracking and trajectory prediction accuracy. Transfer learning also allows the model to learn from diverse handball scenarios, reducing the impact of variations in lighting conditions and player appearances. By fine-tuning the YOLOv8 pre-trained model, the new approach maximizes recognition accuracy while benefiting from the model's inherent efficiency. Overall, the integration of YOLOv8 using transfer learning techniques introduces a novel solution that bridges the gaps in existing models, offering a comprehensive and real-time handball recognition system that can accurately analyze player movements, ball trajectories, and game dynamics.

The primary contribution of this work lies in the novel utilization of the YOLOv8 pre-trained model combined with transfer learning techniques to significantly advance handball recognition,

- By harnessing the YOLOv8 architecture's state-of-the-art object detection capabilities, this approach addresses critical gaps in existing models, notably in accurate player tracking, precise ball trajectory prediction, and the handling of intricate player interactions.
- The application of transfer learning further refines the model's performance by adapting it to handball-specific data, enabling it to specialize in detecting players, the ball, and pertinent elements within the game. Leveraging YOLOv8's real-time processing and multi-scale analysis further improves recognition accuracy, particularly in scenarios characterized by occlusion, rapid motion, and varying player appearances.
- As a result, this integrated method yields a comprehensive handball recognition system that offers real-time insights into player movements, ball dynamics, and game strategies, ultimately enriching

coaching strategies, player performance assessment, and viewer engagement in the realm of handball.

The fusion of YOLOv8 with transfer learning involves leveraging YOLOv8's pre-trained features for extracting object characteristics, followed by fine-tuning the model on handball-specific data to enhance its ability to recognize players, the ball, and other essential elements in the context of handball recognition. We systematically evaluated the proposed approach using a custom dataset of 751 handball scene videos captured during training sessions for young cadets and handball schools for both girls and boys [16]. Testing encompassed nearly 60,000 frames and incorporated metrics such as sensitivity, specificity, and accuracy. The results demonstrated that our method surpassed state-of-the-art techniques, showcasing heightened accuracy. Notably, the proposed method exhibited enhanced efficiency, achieving a sensitivity 92.18%, specificity 91.13%, accuracy 93.57% and F-score 94.33% respectively.

The rest of this paper is organized as follows. Section 2 gives Related Work on using various AI techniques for Active Player detection in Handball Recognition System. Section 3 gives proposed method for Active Player detection. Section 4 Discuss with Performance evaluation for Active Player detection in Handball Recognition System. Section 5 our conclusion is specified.

2. RELATED WORKS

Ayako Abe et al. [17] utilized a temporal and geometrical calibration approach to achieve precise 3D ball position estimation. Despite an 86.7% success rate in elimination, this method is limited when the ball is thrown very close to the goal, causing discrepancies between the estimated line and the ball trajectory. Proposed by T. Hlupic et al. [18], the system relies on a web application with a relational database, complemented by a mobile application using CouchDB. This approach yields valuable experiential insights. Continued exploration in this area could pave the way for a self-learning system leveraging machine learning techniques. Such a system would identify and display diverse anomalies, proving beneficial for optimizing handball teams. M. Pobar et al. [19] utilized the Mask R-CNN and Optical flow approach to identify active players executing specific actions from the entire player ensemble in a scene. This method attains an 85% efficiency rate for active player detection; however, it doesn't automatically address shifts in player activity across extended video sequences. Marina Ivasic-Kos et al. [20] employed spatio-temporal interest points (STIPs) and optical flow (OF) techniques to detect active players. The STIPs-based approach achieved a 69% true positive rate, while the OF-based approach achieved 56%. Notably, the STIPs-based method demonstrated better performance but requires more computational resources and time for feature calculation.

Xiong, J et al. [21] tackled the fine-grained action recognition challenge, focusing on differentiating successful and failed ball-stopping actions in soccer training. Their model, based on YOLOv3, excels in detecting players and the ball. Leveraging a custom dataset comprising 2543 annotated ball-stopping action videos from training sessions, the model achieved an accuracy of 93.24%. Additionally, it incorporates object-level trajectories to depict human-object interactions and utilizes human postures estimated as key-point sequences for enhanced representation of middle-level semantics, particularly valuable for finer-grained actions. M. Oytun et al. [22] implemented various machine learning models to predict specific types of athletic performance in female handball players. The results ranged between 0.86 and 0.97 in accuracy. Further research is needed, particularly for specific types of athletic performance that involve a larger number of participants and parameters. Additionally, exploring the potential of other artificial intelligence models may be beneficial in advancing this field of study. Host, Kristina et al. [23] implemented the DeepSort algorithm for player tracking after detecting the players with the YOLOv3 object detector. The results showed high accuracy and precision, reaching 99%. However, these results may not be relevant since the same detector was used to generate both the ground truth detections and the detections for the tracker.

Wu, L et al. [24] employed a two-stream 3D-CNN model integrating motion patterns and visual key information to recognize group activities in basketball and predict scoring occurrences. They achieved an accuracy of 50.6% and a mean average precision (MAP) of 60.9% on the NBA&CBA dataset. However, these methods did not perform well in predicting free throw activities that involve mixed motion. Elaoud A et al. [25] conducted a comparison and evaluation of handball players' performance during the "Throw" action. They employed dynamic time warping to analyze skeleton data extracted from Red Green Blue-Depth (RGB-D) images, with a specific focus on central angles relevant to the action. Their analysis involved assessing handball players' performance using angles derived from the Kinect skeleton. Notably, this sensor can capture point cloud data, making it a cost-effective solution that doesn't necessitate attached markers or calibration.

Pobar, M et al. [26] proposed an active player detection method that integrates the Yolo object detector, activity metrics, and tracking techniques to detect and track active players in real-time. They achieved a correct sequence rate of 67%. However, this method is less effective for handling long-term video clips in which the leading player changes at varying intervals. Nevertheless, it shows improvement in analyzing multi-player actions like crossing or defense. J. Xu et al. [27] employed the YOLO method for detecting objects in handball games. They achieved notable accuracy with a recall of 0.840, precision of 0.913, and a detection speed of 8.2 frames per second. Julio Alberto Lopez-Gomez et al. [28] introduced six metaheuristic algorithms, including Particle Swarm Optimization, Genetic, Bat, Artificial Bee Colony, Gravitational Search, and Chaotic Gravitational Search, as well as two memetic algorithms, Gravitational Search and Memetic Chaotic Gravitational Search Algorithms. These algorithms were employed to assess the performance of a handball goalkeeper, treated as a feature weighting optimization problem. The results obtained contribute to identifying the best goalkeeper in a tournament and enhancing optimization methods for recognizing the top five goalkeepers.

Rikuya Kawamura et al. [29] utilized this study to categorize shooting by collecting joint coordinates and then classifying them using publicly available footage of shooting scenes in handball games. IT tools were employed to enhance player evaluation and training, with the aim of fostering improvement regardless of the setting or the coach's skill level. Host, K et al. [30] employed a CNN model that categorizes each frame into an action class, augmenting it with LSTM and MLP-based models to incorporate temporal information from the input video. These models underwent training and testing with varying lengths of input sequences, ranging from 20 to 80 frames. The study revealed that increasing the number of frames in the input sequence led to improved results for the MLP-based model, while it did not have the same impact on the performance of the LSTM model. Host K et al. [31] employed automatic player detection and tracking, along with models for handball action recognition and localization based on Inflated 3D Networks (I3D). The action recognition models achieved strong performance on the test set, consisting of nine handball action classes, with average F1 measures of 0.69 and 0.75 for ensemble and multi-class classifiers, respectively.

Traditionally, the process of feature extraction has been a manual endeavor, characterized by its time-consuming nature, subjectivity, and heavy dependence on technical expertise. Recent research highlights the success of deep learning techniques in effectively identifying objects within handball video scenes. To address these challenges, convolutional neural networks (CNNs) emerge as a promising solution. CNNs exhibit remarkable performance in the realm of active player detection within handball video sequences, owing to their innate ability to automatically extract pertinent features. Despite the array of machine learning and deep learning approaches available, there persists a demand for a robust framework tailored to the specific task of detecting active players and the ball in handball videos. To meet this need, our proposed method leverages pre-trained models like YOLOv8, which have been fine-tuned through transfer learning techniques to enhance their proficiency in handball recognition.

3. PROPOSED METHOD

Based on the insights gleaned from the literature review, it is imperative to cultivate novel methodologies for the detection of active players in handball recognition, as this is essential for addressing a diverse array of variances. The objective of the proposed approach is to identify, within video sequences, the players who are actively participating during specific game moments or training sessions and discern which of them are executing the desired actions at any given time. The

Journal of Theoretical and Applied Information Technology
30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proposed method is employed the CNN based YOLOv8 pre-trained model in conjunction with transfer learning techniques for enhanced handball recognition (CNN-Y8+TL). The YOLOv8 architecture's advanced capabilities are harnessed to address existing gaps in player tracking, ball trajectory prediction, and complex player interactions. Through transfer learning, the model is fine-tuned using handball-specific data, enabling adaptation and specialization in identifying players, the ball, and key elements. The method leverages YOLOv8's real-time processing and multi-scale analysis to improve accuracy in dynamic game scenarios, overcoming challenges like occlusion and rapid motion. By integrating the YOLOv8 pre-trained model with transfer learning, this approach showcases a promising advancement in achieving comprehensive and efficient handball recognition, significantly enhancing insights into player dynamics, ball movement, and overall gameplay. The fusion of YOLOv8 with transfer learning involves leveraging YOLOv8's pre-trained features for extracting object characteristics, followed by fine-tuning the model on handball-specific data to enhance its ability to recognize players, the ball, and other essential elements in the context of handball recognition.

Within a training scenario, the field typically accommodates multiple players. Nevertheless, the primary aim is to precisely monitor individuals who are actively involved in executing specific handball techniques or actions, such as shooting, dribbling, or passing the ball. It is of utmost importance to distinguish these engaged participants from those who are retrieving balls, waiting in a queue, seated on the bench, or simply observing the proceedings—these collectively referred to as inactive players in this context. Examining segments of the training sessions that specifically showcase players demonstrating handball techniques holds significance. This analysis serves the dual purpose of offering constructive feedback to enhance their performance and refine their individual style, while also enabling the tracking of their ongoing progress. The primary objective of the proposed method is to autonomously identify and distinguish the actively participating players within the scene.

Identifying the player executing a specific action of interest becomes notably challenging in the context of training sessions. This complexity arises due to the presence of more players on the field than allowed in a typical game, the simultaneous existence of multiple balls, each player having their own ball, and the occurrence of various actions occurring concurrently to optimize skill development. The inherent complexity of this task exceeds the capabilities of straightforward methods like background subtraction or Chroma keying for people segmentation. To address this, we propose an active player detection approach that integrates player detection and tracking while employing an activity measure to pinpoint the key players in these scenarios. To perform player detection, we recommend the utilization of deep convolutional neural networks (CNNs), which have demonstrated their efficacy in classifying and detecting objects within real-world handball videos. In this study, we employ CNN based YOLOv8 with Transfer Learning (CNN-Y8+TL), a model known for its exceptional accuracy in object detection tasks within handball videos.

An overview of the proposed active player identification technique is shown in Figure 3. Person appearances are recognized and marked with bounding boxes in each frame. In addition, features that are indicative of moving objects are extracted, such as interest points that exhibit significant variations in speed and appearance between consecutive frames or optical flows resulting from the time-varying intensities of all surface points in the 2D motion field of the images.

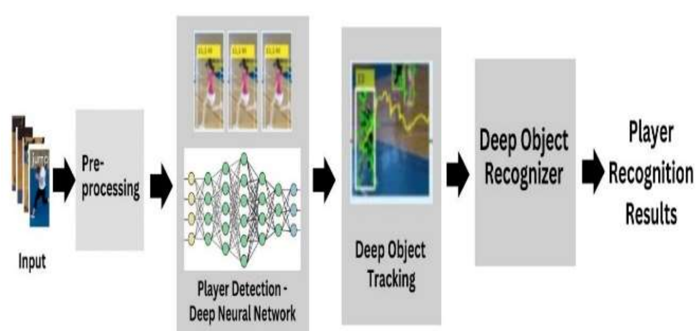

Figure 3. Overview of the proposed method

3.1 Player Detection using CNN based YOLOv8 with Transfer Learning (CNN-Y8+TL)

YOLOv8 is a one-stage object detection model that predicts bounding boxes and class probabilities directly from the input image of handball video. It can be divided into two main parts: the backbone network and the head network.

a) Backbone network:

The backbone network is a convolutional neural network (CNN) that extracts features from the input image from handball video. YOLOv8 uses the Cross Stage Partial Networks (CSPNet) architecture as its backbone. CSPNet is a lightweight and efficient CNN architecture that is well-suited for object detection. Cross Stage Partial Networks (CSPNet) are a type of convolutional neural network (CNN) architecture that are designed to be lightweight and efficient, while still maintaining high accuracy. CSPNets are based on the idea of splitting the feature map of each layer into two parts, and then processing each part separately. This allows CSPNets to reduce the number of computations required, without sacrificing accuracy. CSPNets are typically used for object detection and image classification tasks. They have been shown to achieve state-of-the-art results on a variety of benchmarks, while also being significantly faster than other CNN architectures.

The input image is passed through a series of convolutional layers. At each layer, the feature map is split into two parts. One part of the feature map is passed through a regular convolutional layer. The other part of the feature map is passed through a dense block. The outputs of the two convolutional layers are then concatenated together. The concatenated feature map is then passed on to the next layer. The dense block is a key component of the CSPNet architecture. A dense block is a type of CNN architecture that connects all of the layers in the block to each other. This allows dense blocks to learn more complex features than regular convolutional layers.

The dense block can be mathematically represented as follows:

$$X\_I = H\_I(X\_\{I-1\}) + X\_\{I-1\}$$
(1)

Where, X_I is the output of the Ith layer in the dense block and H_I is the convolutional layer in the Ith layer of the dense block. The CSPNet architecture can be mathematically represented as follows:

$$F\_I = C\_I(X\_\{I-1\}) + D\_I(X\_\{I-1\})$$
(2)

Where, F_I is the output of the Ith layer in the CSPNet. C_I is the convolutional layer in the Ith layer in the CSPNet. D_I is the dense block in the Ith layer in the CSPNet.

b) Head network:

The head network takes the output features from the backbone network and predicts the bounding boxes and class probabilities for each object in the image. The head network of YOLOv8 is divided into three branches: (1) Bounding box branch, (2) Objectness branch, and (3) Class probability branch respectively. Bounding box branch predicts the coordinates of the bounding box for each object. Objectness branch predicts the probability that a bounding box contains an object. Class probability branch predicts the probability that an object belongs to a particular class.

The output of the head network is a tensor of shape,

$$[B, S, S, (C + 5)]$$
(3)

Where, B is the batch size. S is the size of the output grid. $C$is the number of object classes. The 5 additional channels contain the bounding box coordinates and objectness probability for each cell in the output grid.

The general formulation of YOLOv8 can be summarized as follows:

$$y = f(x)$$
(4)

Where, x is the input image of handball scene, y is the output tensor of the head network, and $f$is the YOLOv8 model. The function f takes the input image x and predicts the bounding boxes and class probabilities for each object in the image.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

YOLOv8 is trained using a supervised learning approach. The training data consists of images with labeled objects. The model is trained to minimize the loss between the predicted bounding boxes and class probabilities and the ground truth labels. At inference time, YOLOv8 takes an input image and predicts the bounding boxes and class probabilities for each object in the image. The model uses a non-maxima suppression (NMS) algorithm to remove duplicate bounding boxes and generate the final output.

Algorithm: CNN based YOLOv8 with TL (CNN-Y8+TL)
Input: *-Image I as a matrix* *-Pre-trained YOLOv8 model Θ* *-Confidence Threshold* $T_{confidence}$ *-Intersection over Union (IoU) Threshold* T_{IoU} *-Anchor Boxes A* ***Output:*** *-List of Detected Objects D*

$\boldsymbol{Step\ 1: Pre-processing}$
$I' = preprocess(I, target\_dimensions)$
$Step\ 2: Forward\ Pass$
$P = \Theta(I')$
$Step\ 3: Decode\ Predictions$
$D = \{G, B\}$
$for\ each\ grid\ cell\ G\ in\ P$
$for\ each\ anchor\ box\ B\ in\ A$
$//Decode\ bounding\ box\ coordinates$
$(x, y, w, h) = decode\_bbox(G, B)$
$next\ j$
$next\ i$
$\boldsymbol{//Extract\ class\ probabilities}$
$for\ each\ class\ C$
$p(C) = extract\_class\_prob(G, C)$
$next\ i$
$\boldsymbol{//Find\ the\ class\ with\ maximum\ probabili}$
$C_{max} = \arg\max(p(C))$
$Confidence\ Score = \max(p(C))$
$if\ \max(p(C)) > T_{confidence}$
$\boldsymbol{//Add\ the\ detected\ object\ to\ D}$
$D = D \cup \{(class_{id} = C_{max}, confidence\_score, bounding\_box = (x, y, w, h))\}$
$end\ if$
$\boldsymbol{Step\ 4: Non-Maximum\ Suppression}$
$D = non\_max\_suppression(D, T_{IoU})$
$\boldsymbol{Step\ 5: Output}$
$return\ D$

4. EXPERIMENTS

In the experiment part, [22] the customized dataset comprises 751 videos, each showcasing one of seven handball actions, including passing, shooting, jump-shot, dribbling, running, crossing, and defense. These videos were meticulously curated by manually extracting specific scenes from longer recordings of handball practice sessions. Stationary GoPro cameras, strategically positioned on the left or right sides of the playing field and capturing footage from various angles, were employed for this purpose. The videos were consistently recorded in high-quality, meeting or exceeding full HD (1920 × 1080) resolution, and maintaining a frame rate of at least 30 frames per second. On average, each scene features approximately 12 players, although the primary focus remains on one or two players executing the target action. The experimental segment evaluates the performance of the proposed technique through four metrics: sensitivity, specificity, accuracy, F-Score.

The training process encompassed the analysis of frames extracted from a 60 videos out of 751 videos from custom the dataset in our training set. Each frame was meticulously annotated, categorized as either depicting an active or inactive player. The training parameters included a learning rate of 0.001, a momentum of 0.9, and a decay rate of 0.0005. For image input, a fixed size of 640 × 640 pixels was employed, without employing YOLOv8's multi-scale training feature. Data augmentation techniques were employed, involving random adjustments of image saturation and exposure, with maximum factors of 1.5, along with random hue shifts of up to 0.1.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. Experimental Setup for the proposed method

Dataset	Various Handball Actions	*.avi format (videos)	*.CSV file	*.mp4 format (videos)
Custom dataset - 751 videos	Crossing	250	129	129
	Defense	35	18	16
	Dribbling	26	24	24
	Jump-shot	248	387	370
	Passing	141	104	104
	Running	13	09	09
	Shot	98	102	102

4.1 Results and Comparison with Other Existing Methods

The proposed handball recognition system has been experimented with the benchmark dataset mentioned in the experiment setup column. The CNN-Y8+TL (CNN based YOLOv8 with Transfer Learning) approach for active player detection in handball videos exhibits strong performance across various handball actions, achieving improved sensitivity, specificity, accuracy and F-score values. Table 2 illustrates the notable precision achieved in detecting active players in handball videos. Furthermore, Table 2 presents the average performance metrics for a range of handball action types. These results, as shown in Figure 4, reflect the promising outcomes produced by the algorithm for active player detection using the CNN-Y8+TL approach.

Table 2. The comprehensive effectiveness of the proposed method

Different Actions	Sensitivity (%)	Specificity (%)	Accuracy (%)	F-Score (%)
Crossing	90.08	90.45	90.05	90.01
Defense	90.06	88.94	89.68	90.07
Dribbling	89.99	90.20	91.13	90.22
Jump-shot	90.1	90.44	91.66	91.26
Passing	87.14	85.67	87.76	88.87
Running	91.03	90.43	89.33	90.52
Shot	92.18	91.13	93.57	94.33

The proposed system shows a better performance for different actions in handball recognition system with the following measures such as sensitivity, specificity, accuracy, and F-score respectively. The crossing action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 90.08%, 90.45%, 90.05% and 90.01% respectively. The defense action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 90.06%, 88.94%, 89.68% and 90.07% respectively. The dribbling action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 89.99%, 90.20%, 91.13% and 90.22% respectively. The jump-shot action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 90.1%, 90.44%, 91.66% and 91.26% respectively. The passing action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 87.14%, 85.67%, 87.76% and 88.87% respectively. The running action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 91.03%, 90.43%, 89.33% and 90.52% respectively. The crossing action class, the average measures of sensitivity, specificity, accuracy, and F-score rates are 92.18%, 91.13%, 93.57% and 94.33% respectively.

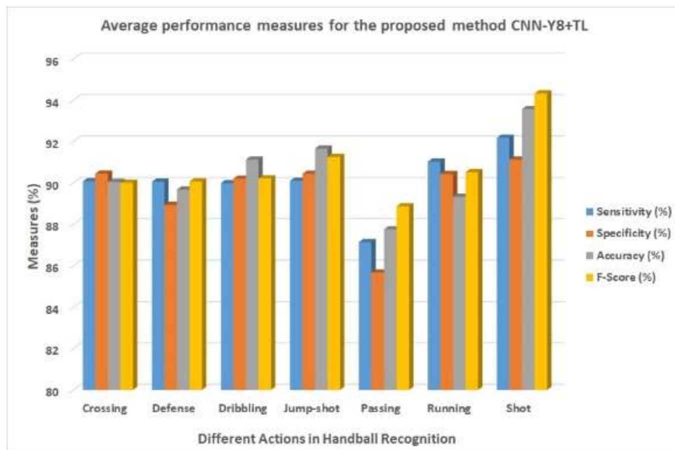

Figure 4. Average performance measures for the proposed method CNN-Y8+TL

Figure 5 shows the accuracy of the proposed model. High training accuracy typically indicates that the model has effectively learned and adapted to the training data, which is a positive sign during the training process.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

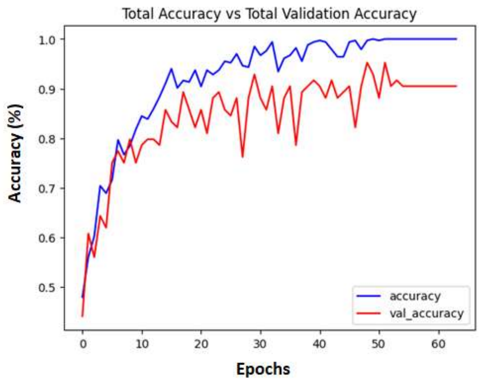

Figure 5. Visualize the training and validation accuracy metrics

Figure 6 depicts the average loss values for both the training and validation datasets throughout the learning process. Notably, the training loss consistently remains lower than the validation loss. This observation highlights the ability of the proposed method, CNN-Y8+TL, to acquire a precise representation of normality, as evidenced by the consistently low loss on the training data

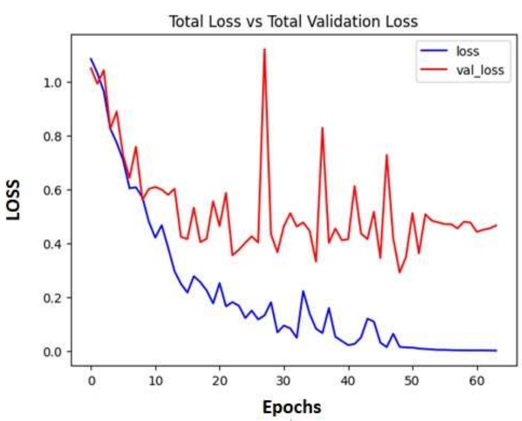

Figure 6. Visualize the training and validation loss metrics

Figure 7. Player's Detection - Background players, partially occluded or hidden, remain undetected.

Figure 7. Illustrates a player's detection challenge in a 24-player scene. Most players in the row are detected, but the player shooting the ball towards the goal goes unnoticed possibly due to atypical body positioning and a t-shirt color resembling the playground. Additionally, background players, partially occluded or hidden, remain undetected.

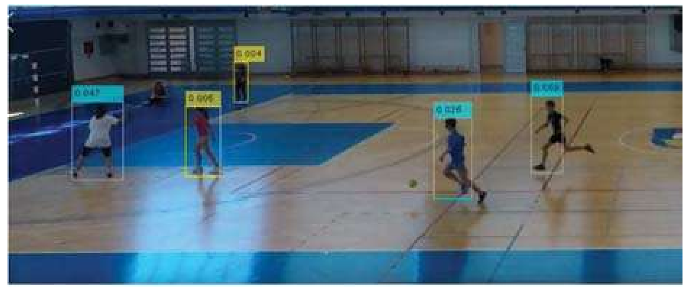

Figure 8. Active Players detection - running, ball leading, and goal-shooting

Refer to Figure 8, where active players are highlighted by blue bounding boxes. Within the image featuring six players, five are successfully detected, with three of them identified as active. These active players are engaged in activities such as running, ball leading, and goal-shooting. Active players are indicated by blue bounding boxes in Figure 9. In the image containing six players, three of them are identified as active. These active players engage in running and ball throwing.

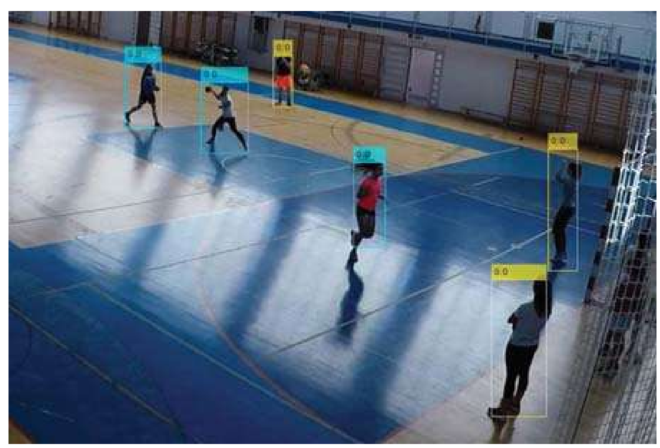

Figure 9. Active player's detection running and ball throwing

Journal of Theoretical and Applied Information Technology
30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3. Comparison of average performance measures of proposed method CNN-Y8+TL and other methods

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)	F-Score (%)
Mask R-CNN [21]	76	98	76	85
CNN-DT+Y [28]	63	87	74	73
I3D_40D 2023 [33]	77	80	79	78
CNN-Y8+TL **(proposed method)**	92.18	91.13	93.57	94.33

The results of the proposed system show a clear improvement over the I3D_40D [33], CNN-DT+Y [28], Mask R-CNN[21]. The proposed system shows a better performance with sensitivity 92.18%, specificity 91.13%, accuracy 93.57% and F-score 94.33% respectively. The proposed method CNN-Y8+TL is compared with I3D_40D method, the average measures of sensitivity, specificity, accuracy, and F-score rates are 77%, 80%, 79% and 78% respectively. The CNN-DT+Y method, the average measures of sensitivity, specificity, accuracy, and F-score rates are 63%, 87%, 74%, and 73% respectively. In the Mask R-CNN method, the average measures of sensitivity, specificity, accuracy, and F-score rates are 76%, 98%, 76% and 85% respectively. Comparison analysis of average performance measures of the proposed method CNN-Y8+TL and other existing methods as shown in Table 8 and Figure 10.

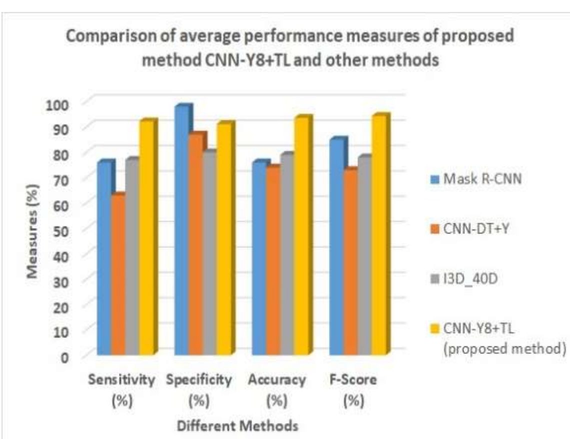

Figure 10. Comparison of average performance measures of proposed method CNN-Y8+TL and other methods

It is observed from experimentation that the CNN-Y8+TL method produces good and comparable results with sensitivity 92.18%, specificity 91.13%, accuracy 93.57% and F-score 94.33% respectively for different handball actions, including passing, shooting, jump-shot, dribbling, running, crossing, and defense as shown in Table 2. The reason for this improvement is two-fold i) transfer learning further refines the model's performance by adapting it to handball-specific data, enabling it to specialize in detecting players, the ball, and pertinent elements within the game. Leveraging YOLOv8's real-time processing and multi-scale analysis further improves recognition accuracy, particularly in scenarios characterized by occlusion, rapid motion, and varying player appearances. ii) As a result, this integrated method yields a comprehensive handball recognition system that offers real-time insights into player movements, ball dynamics, and game strategies, ultimately enriching coaching strategies, player performance assessment, and viewer engagement in the realm of handball as shown in Figure. 11.

Figure 11. Active Player Detection using CNN-Y8+TL

In this paper, the novel approach of employing the CNN based YOLOv8 pre-trained model in conjunction with transfer learning techniques for enhanced handball recognition. The YOLOv8 architecture's advanced capabilities are harnessed to address existing gaps in player tracking, ball trajectory prediction, and complex player interactions. Through transfer learning, the model is fine-tuned using handball-specific data, enabling adaptation and specialization in identifying players, the ball, and key elements. The method leverages YOLOv8's real-time processing and multi-scale analysis to improve accuracy in dynamic game

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

scenarios, overcoming challenges like occlusion and rapid motion. By integrating the YOLOv8 pre-trained model with transfer learning, this approach showcases a promising advancement in achieving comprehensive and efficient handball recognition, significantly enhancing insights into player dynamics, ball movement, and overall gameplay. The fusion of YOLOv8 with transfer learning involves leveraging YOLOv8's pre-trained features for extracting object characteristics, followed by fine-tuning the model on handball-specific data to enhance its ability to recognize players, the ball, and other essential elements in the context of handball recognition.

5. CONCLUSION

In this paper, this study explored the novel approach of employing the CNN-based YOLOv8 pre-trained model in conjunction with transfer learning techniques (CNN-Y8+TL) for enhanced handball recognition. Through transfer learning, the model has been fine-tuned using handball-specific data, enabling adaptation and specialization in identifying players, the ball, and key elements. The method leveraged YOLOv8's real-time processing and multi-scale analysis to improve accuracy in dynamic game scenarios, overcoming challenges like occlusion and rapid motion. By integrating the YOLOv8 pre-trained model with transfer learning, this approach showcased a promising advancement in achieving comprehensive and efficient handball recognition, significantly enhancing insights into player dynamics, ball movement, and overall gameplay. The fusion of YOLOv8 with transfer learning involves leveraging YOLOv8's pre-trained features to extract object characteristics, followed by fine-tuning the model on handball-specific data to enhance its ability to recognize players, the ball, and other essential elements in the context of handball recognition. The proposed method was thoroughly tested with the custom dataset using metrics such as sensitivity, specificity, and accuracy, and it was discovered that the proposed method attained better accuracy when compared to state-of-the-art methods. The results of the proposed system show a clear improvement over the I3D_40D [33], CNN-DT+Y [28], Mask R-CNN[21]. The proposed system shows a better performance with sensitivity 92.18%, specificity 91.13%, accuracy 93.57% and F-score 94.33% respectively. The proposed method CNN-Y8+TL is compared with I3D_40D method, the average measures of sensitivity, specificity, accuracy, and F-score rates are 77%, 80%, 79% and 78% respectively. The CNN-DT+Y method, the average measures of sensitivity, specificity, accuracy, and F-score rates are 63%, 87%, 74%, and 73% respectively. In the Mask R-CNN method, the average measures of sensitivity, specificity, accuracy, and F-score rates are 76%, 98%, 76% and 85% respectively.

In future work, the research will focus on refining the CNN-based YOLOv8 model through advanced transfer learning techniques to enhance its ability to recognize handball-specific actions, even in challenging conditions. Real-time recognition capabilities are a priority, allowing the system to process handball actions as they occur live. Scalability for larger datasets and versatility for multiple sports will be explored. The research also aims to develop user-friendly interfaces and advanced data augmentation techniques for improved performance and seamless integration. Collaboration with sports institutions and real-world deployment will continue to be central to its endeavors. In future directions: enhancing accuracy through dataset expansion, refining training methodologies, and exploring alternative neural network architectures; investigating the system's performance across different handball game scenarios, considering variations in lighting, player positioning, and game dynamics; and developing strategies to optimize the system for real-time deployment in dynamic, unpredictable game scenarios.

REFERENCES

[1] Buric, M.; Pobar, M.; Ivasic-Kos, M. An overview of action recognition in videos. In Proceedings of the 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 22–26 May 2017; pp. 1098–1103.

[2] R. Girshick, "Fast r-cnn," in Proc. 2015 IEEE Int. Conf. Comput. Vision (ICCV 2015), 2015.

[3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems (NIPS 2015), 2015, pp. 91–99.

[4] W. Liu et al., “Ssd: Single shot multibox detector,” in European conference on computer vision (ECCV 2016), 2016, pp. 21–37.

[5] J. Dai, Y. Li, K. He, and J. Sun, “R-fcn: Object detection via region-based fully convolutional networks,” in Advances in neural information processing systems (NIPS 2016), 2016, pp. 379–387.

[6] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 2018, arXiv:1804.02767.

[7] M. Buri, M. Pobar, and M. Ivaši-Kos, “Object detection in sports videos,” in Proc. 2018 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO 2018), 2018, pp. 1034–1039.

[8] M. Buri, M. Pobar, and M. Ivaši-Kos, “Adapting YOLO network for Ball and Player Detection,” in 8th Int. Conf. Pattern Recognit. Appl. Methods (ICPRAM 2019), 2019, pp. 845–851.

[9] S. M. Smith and J. M. Brady, “ASSET-2: Real-time motion segmentation and shape tracking,” IEEE Trans. Pattern Analysis Mach. Intell., vol. 17, no. 8, pp. 814–820, 1995.

[10] D. G. Lowe, “Object recognition from local scale-invariant features,”in Proc. 7th IEEE Int. Conf. Comput. Vision (ICCV 1999), vol. 2, 1999, pp. 1150–1157.

[11] I. Laptev, “On space-time interest points,” Int. J. Comput. Vision, vol. 64, no. 2–3, pp. 107–123, 2005.

[12] A. Klaeser, M. Marszalek, C. Schmid. "A Spatio-Temporal Descriptor Based on 3D-Gradients," in Proc. 19th Brit. Mach. Vision Conf. (BMVC 2008), Leeds, United Kingdom, 2008, pp. 99.1-99.10.

[13] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, “Behavior recognition via sparse spatio-temporal features,” in 2005 IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 65–72.

[14] G. Willems, T. Tuytelaars, and L. Van Gool, “An efficient dense and scale-invariant spatio-temporal interest point detector,” in Proc. European conference on computer vision (ECCV 2008), 2008, pp. 650–663.

[15] M. Buri, M. Pobar, M. Ivaši-Kos, “Ball detection using YOLO and Mask R-CNN,” in 5th Annual Conf. on Computational Science & Computational Intelligence (CSCI’18), Las Vegas, USA, 2018.

[16] M. Ivasic-Kos and M. Pobar, "Building a labeled dataset for recognition of handball actions using mask R-CNN and STIPS," 2018 7th European Workshop on Visual Information Processing (EUVIP), Tampere, pp. 1-6, 2018.

[17] Ayako Abe, Ikuko Shimizu, “3D Shot Course Estimation System for Tactics Analysis Support of Handball Games”, Eighth International Joint IEEE Conference on Computer Science and Software Engineering (JCSSE), 2011.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[19] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.

[20] T. Hlupić, F. Jandrijević, J. Kovačev, L. Petricioli, T. Gracin and M. Baranović, "System for monitoring and advanced analysis of handball matches," 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2015, pp. 1428-1433, doi: 10.1109/MIPRO.2015.7160498.

[21] M. Pobar and M. Ivasic-Kos, "Mask R-CNN and Optical Flow Based Method for Detection and Marking of Handball Actions," 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, pp. 1-6, 2018. doi: 10.1109/CISP-BMEI.2018.8633201.

[22] Marina Ivasic-Kos, Miran Pobar, Jordi Gonzàlez, “Active Player Detection in Handball Videos Using Optical Flow and STIPs Based Measures”, 13th International Conference on Signal Processing and Communication Systems (ICSPCS), Gold Coast, QLD, Australia, 2019, pp. 1-8, doi: 10.1109/ICSPCS47537.2019.9008460.

[23] Xiong, J, Lu, L, Wang, H, Yang, J, Gui, G, “Object-Level Trajectories Based Fine-Grained Action Recognition in Visual IoT Applications”, IEEE Access, 2019, 7, 103629–103638.

[24] M. Oytun, C. Tinazci, B. Sekeroglu, C. Acikada and H. U. Yavuz, "Performance Prediction and Evaluation in Female

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Handball Players Using Machine Learning Models," in IEEE Access, vol. 8, pp. 116321-116335, 2020, doi: 10.1109/ACCESS.2020.3004182.

[25] Host, Kristina et al. "Tracking Handball Players with the DeepSORT Algorithm." International Conference on Pattern Recognition Applications and Methods, 2020.

[26] Wu. L, Yang, Z, Wang, Q, Jian, M, Zhao, B, Yan, J, Chen, C.W, "Fusing Motion Patterns and Key Visual Information for Semantic Event Recognition in Basketball Videos" Neurocomputing 2020, 413, 217–229.

[27] Elaoud A., Barhoumi W., Zagrouba E., Agrebi B, "Skeleton-Based Comparison of throwing Motion for Handball Players", J. Ambient. Intell. Humaniz. Comput. 2020;11:419–431.

[28] Pobar, M.; Ivasic-Kos, M. Active Player Detection in Handball Scenes Based on Activity Measures. Sensors 2020, 20, 1475.

J. Xu, Y. Zhang, A. Ye and F. Dai, "Real-time detection of game handball foul based on target detection and skeleton extraction," IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Fuzhou, China, 2021, pp. 41-46, doi: 10.1109/CEI52496.2021.9574504.

[30] Julio Alberto Lopez-Gomez, Francisco P. Romero, Eusebio Angulo, "A Feature-Weighting Approach Using Metaheuristic Algorithms to Evaluate the Performance of Handball Goalkeepers", IEEE Access, Volume 10, March 2022.13.

[31] Rikuya Kawamura, Yoshiro Yamamoto, "Classification of Handball Shot through Image Analysis", 20th International Conference on ICT and Knowledge Engineering (ICT&KE), 2022.

[32] Host, K., Ivasic-Kos, M., Pobar, M, "Action Recognition in Handball Scenes", In: Arai, K. (eds) Intelligent Computing. Lecture Notes in Networks and Systems, vol 283. Springer, Cham., 2023. https://doi.org/10.1007/978-3-030-80119-9_41.

[33] Host K, Pobar M, Ivasic-Kos M, "Analysis of Movement and Activities of Handball Players Using Deep Neural Networks", J Imaging, 2023 Apr 13;9(4):80. doi: 10.3390/jimaging9040080. PMID: 37103231; PMCID: PMC10144022.