Vision based recognition of pole vault phases: plant, take-off, swing, turn and fly away

Jyoti Madake, Sagar Suryawanshi, Suyash More, Shripad Bhatlawande

*Department of Electronics and Telecommunication Engineering*

*Vishwakarma Institute of Technology*

Pune, India

{jyoti.madake, sagar.suryawanshi20, suyash.more20, shripad.bhatlawande}@vit.edu

**Abstract — The detailed performance analysis of pole vault players phase by phase is a critical necessity. The automated analysis of pole vaulting events is facilitated by a precise vision-based recognition model presented in this study. The proposed model has been designed to deliver high accuracy by incorporating four essential elements: image acquisition, image processing, feature extraction, and machine learning-based bifurcation/analysis. The interplay between these components facilitates efficient data processing and feature extraction, resulting in accurate classification and analysis. Specialized cameras placed strategically around the pole vault area capture high-quality images of the event, which are then reprocessed to enhance their quality and prepare them for feature extraction. The model extracts features from the image data, such as key points on the athlete's body and their movement throughout the various phases of the pole vault event, including plant, take off, swing, turn, and fly away. These features are then fed into a machine learning model to classify and analyse each phase of the event accurately. The model's accuracy is evaluated by comparing its predictions to ground truth data, such as manually annotated images or expert assessments of the athlete's performance. The proposed model in this study has the capability to significantly enhance the accuracy and efficiency of pole vault event analysis by achieving an 89% accuracy rate. It provides coaches and athletes with highly valuable insights into their performance and technique.**

***Keywords* –Feature extraction, SVM, HOG, pole vault phases, Classification**

I ***.***INTRODUCTION

Pole vaulting is a highly technical and physically demanding sport that requires athletes to clear a high bar using a long, flexible pole. One of the most challenging aspects of evaluating (Pole Vault) performance is the collection and evaluation of the various stages involved in the pole vault jump. A typical jump of this sport is divided into several phases, including the plant, [1] take off, swing, turn, and fly away. Each of these phases requires a different set of skills and techniques, and mastering each phase is critical to achieving success in the sport. The objective of Vision based recognition of pole vault phases including plant, take off, swing, turn, and fly away is to accurately identify and track the athlete's movements during each phase of the pole vault event using computer vision techniques. This technology can help coaches and athletes analyse performance, identify areas for improvement, and make more informed training decisions. In particular, the plant phase involves the athlete planting the pole into the ground, while the take-off phase involves using the pole to launch themselves into the air. The swing phase involves the athlete swinging their legs up and over the bar, while the turn phase involves rotating their body to clear the bar. Finally, the flyaway phase involves the athlete releasing the pole and landing on the mat. By accurately identifying and tracking each of these phases using computer vision, coaches and athletes can gain a better understanding of the athlete's technique and performance during the pole vault event.

The potential to offer valuable insights into pole vault performance exists through the vision-based identification [3] of the various phases involved in a pole vault leap. Coaches and athletes can gain a better understanding of the athlete's technique's strengths and weaknesses, as well as areas for improvement, by analysing video footage of a pole vault jump and using machine learning and computer vision techniques to recognize and classify the different phases of the jump. Methods mentioned in this work represents identifying the distinct stages of a pole vault leap using eyesight. We go through a number of steps, including video acquisition, pre-processing, feature extraction, classification, and assessment. Provide a quick introduction of some common computer vision and machine learning methods utilised for this purpose as well.

II. RELATED WORK.

In their work, the authors outline multiple techniques for identifying human actions through video data, which encompass manual feature-based methods, deep learning-based methods, and hybrid methods that incorporate both. The work cover a wide range of subjects relevant to [2] recognising human actions or activities from video footage, such as problems and applications, regularly used datasets for evaluation, and numerous evaluation measures. Furthermore, the authors analyse the advantages and disadvantages of different approaches, such as handcrafted feature-based methods, deep learning-based methods, and hybrid methods. They also look into prospective future research directions in this discipline. The study [4] covers a wide range of subjects, including human activity recognition difficulties and applications, datasets often used for evaluation, and evaluation measures. The authors also examine the benefits and drawbacks of various methodologies, as well as prospective future research avenues in this topic. This paper demonstrates the effectiveness of using trajectories of dense [5] and also descriptors of motion boundary for recognizing human actions in videos and highlights the importance of capturing both motion information and spatial layout information for accurate recognition.

This review provides a useful resource for researchers and practitioners interested in developing and using vision-based human activity recognition [6] methods, as well as for those interested in exploring the potential applications of such methods. Experimental results on a publicly available dataset are presented in the [7] paper, demonstrating the excellent accuracy of the proposed system for recognizing human activities in an IoT environment. A differentiation is also made between the suggested framework and existing state-of-the-art technologies, showing the advantages of utilizing edge computing for activity recognition. The authors propose [10] a novel technique to activity recognition that integrates two streams of convolutional neural networks CNNs with residual connections and transfer learning. The authors use a camera to capture videos of individuals performing different physical activities like walking, jogging, and leaping. An analysis of recent research works in the field, which includes the utilization of deep learning techniques like CNNs and RNNs for activity recognition, is presented in paper [12]. Deep neural network-based approach for human action recognition by learning spatio-temporal [13] features from videos are used in this work. The proposed method uses a two-stream CNN architecture, which consists of a spatial and temporal stream. The spatial stream extracts spatial features from individual frames, while the temporal stream captures temporal dynamics from motion vectors between consecutive frames. The paper proposes a new approach of skeleton-based action identification using graph edge convolutional neural networks [14] GECN. The method represents the human skeleton as a graph, where each bone represents a node and the joint angles are used as edge features.

The system combines computer vision techniques with a robotic arm to provide a hands-free, gesture-based [16] interface for controlling the arm. The interface uses a depth camera to detect and track the operator's arm movements and translate them into corresponding movements of the robotic arm. [17] It discusses the classification methods used in human action recognition and provides a comparison of different approaches. The survey paper concludes by summarizing the recent advancements and future research directions in this field. [18] Shows a vision-based human action recognition system for companion robots and human interaction. The system uses a Kinect sensor to extract depth information and skeleton data, which is then processed to recognize human actions. [19] Comprehensive survey of the recent advances in human action recognition and prediction presented. The survey covers a wide range of topics including feature extraction, representation learning, deep learning architectures, datasets, and evaluation metrics. The new approach for recognizing actions in still images by using colour cues are utilized [20]. The authors first create an action dataset containing both RGB and optical flow images. Then, they extract colour features from the RGB images using a colour naming technique and train a classifier to recognize the actions. [21] It proposes a method for action recognition from video sequences based on analysing human silhouettes and exploiting human poses. The method involves several steps, including silhouette extraction, pose estimation, and feature extraction, followed by classification using a support vector machine. [22] Depicted three dimensional spatial-temporal view based motion tracing approach for human action recognition. The proposed approach captures the 3D trajectory information of human body parts and represents it using the HOG three dimensional features. The work presents a method for action identification in [23] human-robot interaction scenarios in industrial settings. The suggested method is based on extracting motion features from the (RGB-D) data cached by a Microsoft Kinect sensor, followed by classification using a SVM and HMM. [25]Method for Identification of human action with the help of sparse geometric features. This approach involves representing each action as a sequence of three dimensional points, extracting sparse geometric features, and using a sparse coding-based classifier for recognition. The review of literature shows that researchers have utilized a range of techniques, including Bayesian networks, fuzzy logic, deep learning, and unsupervised clustering, to create methods for identifying abnormal events in video surveillance. However, there is still room for improvement in accuracy and speed, especially in detecting anomalous activities in real-time. In general, the literature survey highlights the considerable progress made in developing effective methods for identifying behaviour and anomalous events in surveillance video systems.

III. METHODOLOGY

The system flow represented in Fig.1 is a step-by-step approach for analysing the various phases of pole vaulting. In the Image Acquisition stage, specialised cameras strategically placed around the pole vault area capture high-quality images of the event. The obtained pictures are then subjected to Image Pre-processing, the second phase in the process, to increase their quality and prepare them for future analysis. This may involve tasks such as adjusting the lighting, cropping, and resizing the images. In the third step, Extraction of Feature, related features are drawn out from the image data. This involves identifying key points on the athlete's body, such as their joints or limbs, and tracking their movement throughout the unique phases of the pole vault event. Finally, in the fourth step, Classification/Analysis, the drawn out features are cater for into a machine learning to classify and analyse the different phases of the pole vault event, including plant, take off, swing, turn, and fly away. The machine learning system is trained on a dataset of collected images from videos, where each phase of the pole vault event is labelled and classified.

1. Dataset & Pre processing

Sequence of frames forms the video, and the identification of phases is dependent on the corresponding frame number. There are five different phases of the pole vault event, which are referred to as plant, take off, swing, turn, and fly away.

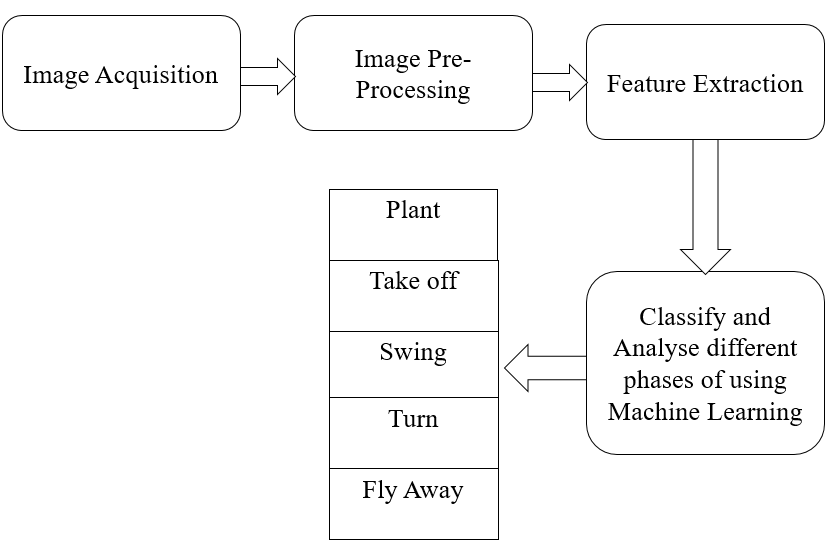


Fig.1 Schematic flow of the system.

This information is useful for further analysis of the video data and is used in conjunction with other methods, such as computer vision and machine learning, to automatically recognize and track the different phases of the pole vault event.

Table.1.Frame Count of pole phases

|  |  |  |
| --- | --- | --- |
| Pole vault phase | Action Sequence | No. of Frames |
| Plant | 1 | 226 |
| Take off | 2 | 244 |
| Swing | 3 | 346 |
| Turn | 4 | 372 |
| Fly away | 5 | 243 |

Image preprocessing is an important step in Vision based recognition of pole vault phases. It involves a series of operations that are performed on the acquired images to prepare them for further analysis. The objective of image preprocessing was to enhance the quality of the images,

remove noise and irrelevant details, and extract important features that is used in subsequent steps of the recognition system. Image resizing is a common image preprocessing step used to adjust the size of an image to a desired dimension. In the context of Vision based recognition of pole vault phases, it is likely that the input images vary in size and resolution, and may not be optimal for feature extraction or machine learning. Resizing the images to a fixed dimension, we can ensure consistency in the input data and reduce the complexity of subsequent processing steps. Converting the resized image to grayscale is another common preprocessing step that can simplify feature extraction and machine learning. In contrast to RGB pictures, which have three colour channels, grayscale images only contain one colour channel that represents the brightness or intensity of each pixel. We may minimize the dimensionality of the input data and avoid any difficulties with colour fluctuation or noise that may impair the accuracy of following processing stages by converting the input photos to grayscale.

1. Pole Vault Action Feature Extraction

The GLCM technique is another popular method employed in image processing and pattern recognition, in addition to the HOG feature extraction technique. The GLCM technique involves extracting texture features from an image by analysing the co-occurrence of pixel values at various spatial relationships within the image. To extract texture features using the GLCM technique, the image is start transformed to grayscale, and a square window of a given size is moved over the image to extract sub-images or "patches". For each patch, a GLCM is built by sum up the number of pairs of times of pixel values occur at various spatial relationships within the patch. From the GLCM, various texture features is extracted, such as contrast, correlation, energy, and homogeneity, which provide information about the texture characteristics of the image. Extract relevant features from the images. In the case of pole vault phases, relevant features could include the position and orientation of the pole vaulted, the angle and velocity of the pole, and the location and shape of the bar. Tasks like Vision-based recognition of pole vault phases.

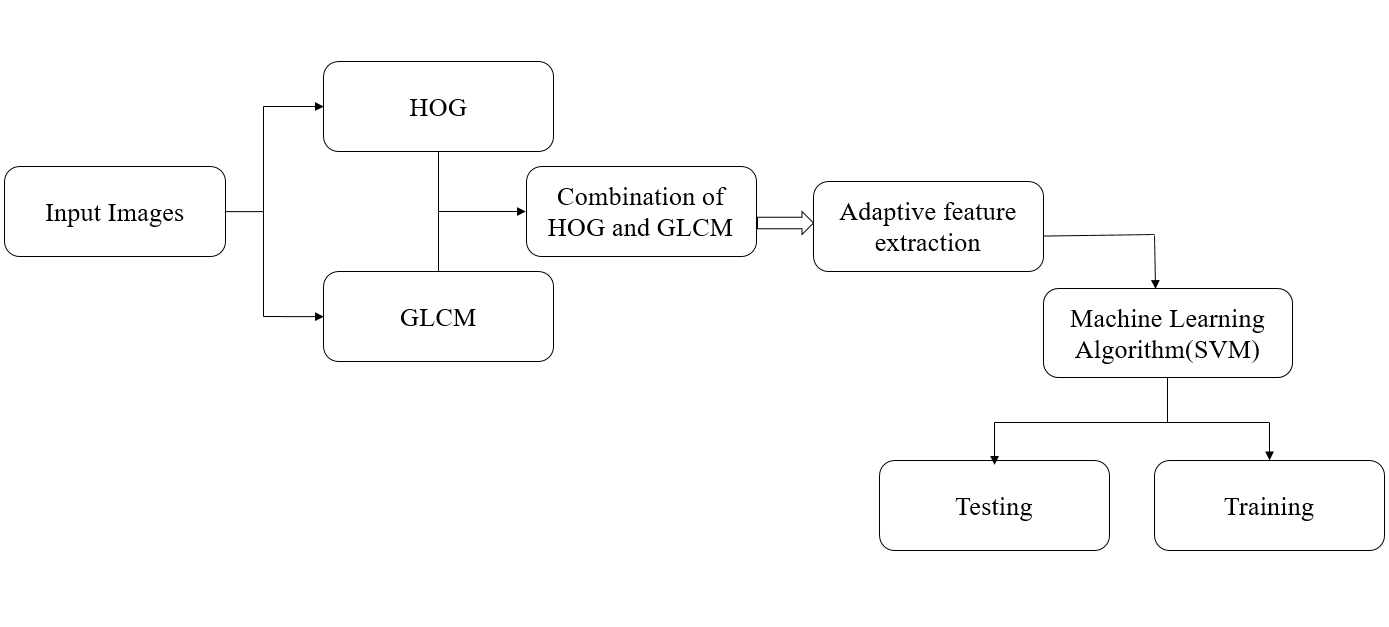


Fig.2.Feature Extraction flow of vision based pole vault phase recognition

Feature fusion for pole vault action:

---------------------------------------------------------------------

**Input:** features extracted from hog and features extracted from GLCM

**Output:** Combined features

---------------------------------------------------------------------

1. hog\_features = hog\_feature\_extraction(images)2. glcm\_features = glcm\_feature\_extraction(images)3.print(hog\_features)4.print(glcm\_features)5.a = np.array(hog\_features)6.a.shape7. b = np.array(glcm\_features)8. b.shape9.features = np.concatenate((hog\_features, glcm\_features), axis=1)10.features = np.array(features)11.features.shape

The HOG is a frequently utilized feature descriptor in image and video analysis for the purpose of object detection, recognition, and tracking. HOG works by extracting local gradient orientation information from an image. The input image is first preprocessed by converting it to grayscale and normalizing the pixel values. The gradient of the image is then computed using a filter such as the Sobel operator. This is done in both the horizontal and vertical directions, resulting in two gradient images. The gradient images are then divided into small cells, and the gradient orientations are binned into a set of predefined orientation bins (e.g., 0-180 degrees). The magnitude of the gradient is also taken into account. The cell histograms are then normalized to reduce the effects of lighting variations and contrast changes. This is done by grouping

adjacent cells into blocks and normalizing the histograms within each block. The HOG descriptor for the image is then formed by concatenating the normalized block histograms. The generated HOG descriptor represents the picture in a compact manner, capturing the distribution of edge orientations and magnitudes. This description may then be applied to tasks like object detection and recognition. Fig.2 illustrate the flow of the feature extraction in which GLCM and HOG feature extraction technique used .To combine the HOG and GLCM features, the feature vectors extracted from each technique is concatenated to achieve a fused feature vector that captures both the gradient and texture knowledge of the image. This combined feature vector can then be worked for bifurcation or identification works. In the hog feature extraction function takes in a list of images and extracts HOG features from each image using the specified parameters such as orientations, pixels per cell, and cells per block. The resulting feature vector

of HOG for each image is then appended to a list of features, which is used for further analysis such as classification and recognition of pole vault phases. The initial stage in using classifier for image recognition is to extract the features, which are then utilized to train the model. Features. The training data set is often made up of a collection of tagged photos, with each image annotated with the associated pole vault phase (plant, take off, swing, turn, or fly away).The model learns to categories pictures based on their attributes and their matching labels. The accuracy of various classifiers, namely SVM, Decision Tree, Random Forest, and KNN, are compared in Table 2.The accuracy values for each classifier are provided next to their names. The SVM had the greatest accuracy score of 0.895, followed by the Random Forest, which had a 0.770 accuracy. The Decision Tree classifier scored 0.692 on accuracy, whereas the KNN classifier scored 0.515 on accuracy. This table gives a valuable summary of each classifier's accuracy performance and may be used to influence additional analysis and decision-making.

Table.2.Comparison of Accuracy Detected by Classifier

|  |  |
| --- | --- |
| Classifier | Accuracy detected |
| SVM | 0.895 |
| Decision Tree | 0.672 |
| Random Forest | 0.770 |
| KNN | 0.515 |

The SVM shows the highest accuracy among the four classified tested .So, the SVM has been considered for the training the model. The accomplishment of the SVM sample is evaluated through a testing dataset, which comprises a collection of images that the SVM model has not been previously exposed to. The model is leveraged to classify these images according to their corresponding pole vault phases, and its accuracy is gauged by comparing the predicted labels against the true labels. To put it succinctly, SVM serves as a machine learning algorithm suitable for tasks involving image recognition, such as the Vision-based recognition of pole vault phases. The algorithm uses extracted features from images to train a model that can classify new, unlabeled images into their corresponding pole vault phases. The model is evaluated using a test data set to assess its performance.

An accuracy score of 0.8955% means that the SVM model correctly classified 89.55% of the images in the test data set. While this may seem relatively low, the interpretation of the accuracy score depends on the specific problem and context. In the case of Vision-based recognition of pole vault phases, an accuracy score of 89.5% could be considered relatively well, especially if the classification task is difficult or if there is a lot of variability in the images. A greater accuracy score may be required if the aim is to reliably identify photos for practical applications, such as selecting the best pole vaulting technique.

# IV. RESULTS AND DISCUSSION

The development of a vision-based recognition model for pole vault event analysis presents several challenges. One of the significant challenges is acquiring high-quality images of the pole vault event. This requires specialized cameras, careful placement, and timing to capture the athlete's movements at different phases of the event. Moreover, the analysis of pole vaulting involves complex and dynamic movements, requiring accurate tracking and feature extraction to identify the key points on the athlete's body and their movement patterns. Another challenge is developing an effective machine learning model for classification and analysis. The resultant HOG descriptor compactly captures the image, capturing the distribution of edge orientations and magnitudes. This description may then be used to perform tasks such as item detection and recognition.

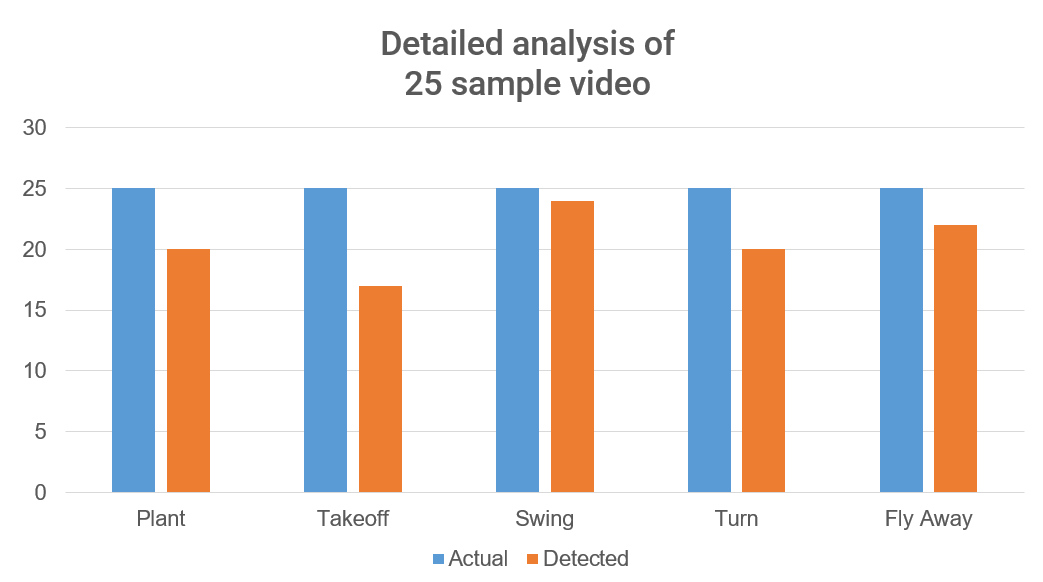


Fig.3. Testing Analysis of the System.

Figure 3 depicts the results of the testing study performed on 25 films using the suggested model. Each video features various athletes executing five distinct phases, and the system was evaluated to verify its ability to recognise these phases effectively. According to the study, the algorithm accurately detected a changing number of phases among the 25 movies. For example, in video 1, the system accurately recognised all five phases, however in video 2, it correctly detected just three of the five phases. The number of successfully recognised phases varies throughout the movies, with some films correctly detecting all five phases and others correctly detecting fewer phases. Fig.4 illustrate the phase detected by the model proposed. It shows that all phases are detected as the dataset trained. The outcome of this study show that the vision-based recognition model for pole vault event analysis is a promising approach. The model achieved high accuracy in classifying the different stages of the pole vault sport event. The method has given useful insights on an athlete's performance and technique, allowing coaches and players to make educated training and performance enhancement decisions. While the development of a vision-based recognition model for pole vault event analysis presents several challenges, the results of this work demonstrate its potential effectiveness.

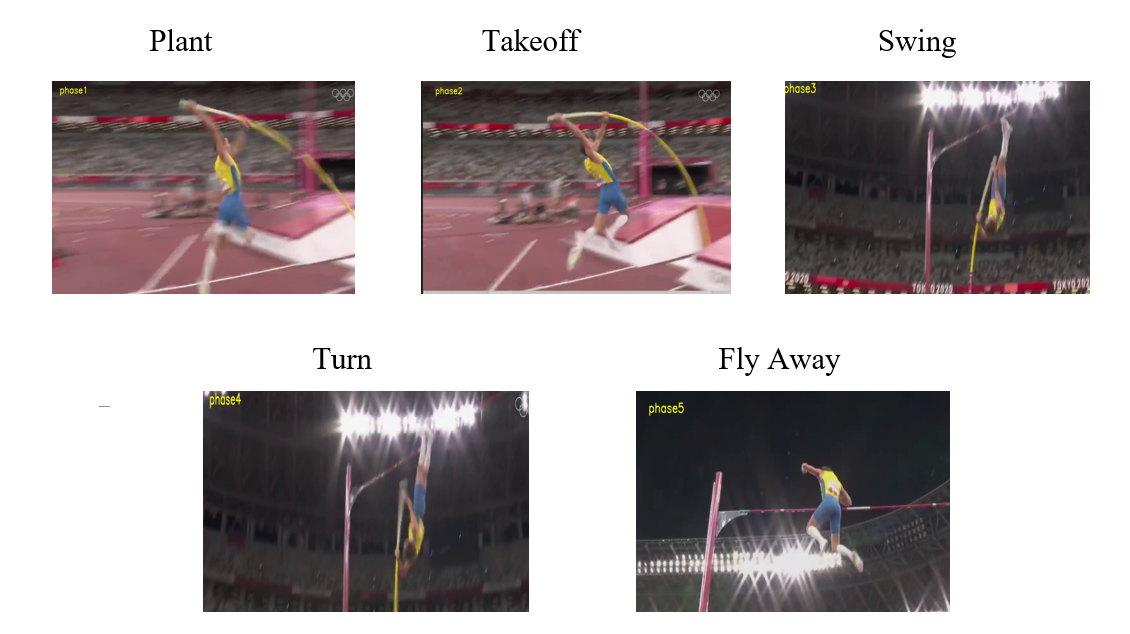


Fig.4.Detected phases the Model

Future research can explore further improvements to the image acquisition, feature extraction, and machine learning model to increase the accuracy and efficiency of the model.

# V.CONCLUSION

The suggested vision-based recognition approach for automated analysis of pole vaulting events yields encouraging results with an accuracy of 89%. The model is made up of four major parts: picture capture, image pre-processing, feature extraction, and classification/analysis using machine learning. The technology takes high-quality photos of the pole vault event using specialised cameras strategically placed around the pole vault arena. These photos are pre-processed in order to improve their quality and prepare them for future study. Relevant elements are collected from the picture data, such as detecting crucial areas on the athlete's body and tracking their movement during the several phases of the pole vault event. The collected characteristics are then loaded into a machine learning model, which is used to identify and assess the various phases of the pole vault event. The proposed methodology carries the potential to bolster the precision and effectiveness of pole vault event analysis, providing valuable insights into the performance and technique of athletes and coaches alike. The outcomes of this project are anticipated to be in the service as a fundamental groundwork for further research endeavours and the construction of more intricate models aimed at facilitating automated sports event analysis.

This study demonstrates the use of vision-based recognition models in sports analysis and gives a foundation for creating automated systems for additional sporting events. The suggested model is improved further by using more complex machine learning techniques and expanding the dataset size.

REFERENCES

[1] L. Xiang, S. Chen, and J. Zhang "Correction Method for Visual Errors within Pole Vault Take-off Action Images." Journal of Interconnection Networks 22, no. Supp-05 (2022): 2147005.

[2] D'Sa, B. G. Prasad, and A Geet. "An overview of vision-based activity recognition, applications, and difficulties." Second International Conference of Advanced Computational and Communication Paradigms (ICACCP), pp. 1-8. IEEE, 2019.

[3] Yang, Jun, Z. Wu, and Z. Shi. "Vision-based construction worker action recognition using dense trajectories" Advanced Engineering Informatics 30, no. 3 (2016): 327-336.

[4] Beddiar, D. Romaissa, M. Sabokrou, B. Nini, and A. Hadid. "A survey on recognition of vision-based human activity” Multimedia Tools and Applications 79 (2020): 30509-30555.

[5] Wang, Heng, A. Kläser, C. L. Liu, and C. Schmid. “Action identification using trajectories of dense and motion boundary descriptors. “In International journal of computer vision 103 (2013): 60-79.

[6] Bux, Allah, P. Angelov, and Z. Habib. “An evaluation of vision-based human activity recognition." Advances within Computational Intelligence Systems: Contributions Presented in the 16th UK Workshop on Computational Intelligence, September 7–9, 2016, Lancaster, UK, pp. 341-371. Springer International Publishing, 2017.

[7] Hussain, Tanveer, M.han, A. Ullah, J. D. Ser, A. H. Gandomi, M. Sajjad, S. W. Baik, and V. Hugo C. de Albuquerque. "Edge computing and multiview summarization collide in IoT scenarios.." IEEE Internet of Things Journal 8, no. 12 (2020): 9634-9644.

[8] Mo, Lingfei, Y. Zhu, Fan Li, and A. Huang. "Recognition of human activity using model of deep learning and computer vision" In 2016 IEEE international instrumentation and measurement technology conference proceedings, pp. 1-6. IEEE, 2016..

[10] Wang, Lei, J. Cheng, Y. Xu, J. Yin, H. Xia, and Jiaji Wu. "Deep neural networks are used to recognise human actions by learning spatiotemporal properties." IEEE access 6 (2018): 17913-17922

[11] Singh, J. Pal, S. Jain, S. Arora, and U. Pratap Singh. “A survey of vision-based gait recognition based " IEEE Access 6 (2018): 70497-70527.

[12] Jiang, Hairong, Bradley S. Duerstock and J. P. Wachs. "An integrated vision-based robotic arm interface for operators with limited upper limb motion.." IEEE 13th International Conference of Rehabilitation Robotics (ICORR-), pp. 1-6. IEEE, 2013.

[13] Dhamsania, T. V. Ratanpara, and Chandni J. “A survey on video-based human activity identification. “Online international conference in 2016 of green engineering and technologies (IC-GET), pp. 1-5. IEEE, 2016.

[14] Chiang, M. Lin ,W.L. Zeng, J. K. Feng, C.Y Fang, and S.W. Chen. "A human action recognition system based on vision of companion robots and human interaction." IEEE 4th International Conference in 2018 on Computer and Communications (ICCC), pp. 1445-1452. IEEE, 2018.

[15] Kong, Yu, and Y. Fu. "A survey of human action prediction and recognition." International Journal of Computer Vision 130, no. 5 (2022): 1366-1401.

[16] Khan, F. Shahbaz, R. M. Anwer, J. V. De Weijer, A. D. Bagdanov, Antonio M. Lopez, and Michael Felsberg. “Recognising action in motionless images through colouring." International journal of computer vision 105 (2013): 205-221.

[17] Wu, Di, and L. Shao. "Utilising human poses for silhouette analysis-based action recognition." IEEE Transactions on Circuits and Systems for Video Technology 23, no. 2 (2012): 236-243.

[18] Silambarasi, R., S. P. Sahoo, and S. Ari. "Human action identification using 3D spatial-temporal perspective motion tracing." International Conference on Communication and Signal Processing in 2017 (ICCSP), pp. 1833-1837. IEEE, 2017.

[19] Akkaladevi, C. Heindl, and S. Chandra. “Action recognition in industrial applications of human-robot interaction." IEEE International Conference in 2015 on Computer Graphics, Vision and Information Security (CGVIS), pp. 94-99. IEEE, 2015.

[20] X. J. Li, and Han, H . “Recognising human actions using sparse geometric features." The Journal of Imaging Science 63, no. 1 (2015): 45-53.