**Journal of Intelligent Systems and Internet of Things**
**Vol. 09, No. 02, PP. 93-107, 2023**
**DOI:** https://doi.org/10.54216/JISIoT.090207

# Modeling Sports Event Tasks in Augmentative and Alternative Communication Using Deep Learning

**Noora Hani Sherif[1,*], Eay Fahidhil[2], Najlaa Nsrulaah Faris[3], Hussein Alaa Diame[4], Raaid Alubady[5], Seifedine Kadry[6, 7, 8]**

[1]Computer Technologies Engineering, Al-Turath University College, Baghdad, Iraq
[2]Medical instruments engineering techniques, Al-farahidi University, Baghdad, Iraq
[3]Department of Medical Devices Engineering Technologies, National University of Science and Technology, Dhi Qar, Nasiriyah, Iraq
[4]Technical Computer Engineering Department, Al-Kunooze University College, Basrah, Iraq
[5]Technical Engineering College, Al-Ayen University, Thi-Qar, Iraq
[6]Department of Applied Data Science, Noroff University College, Kristiansand, Norway
[7]Artificial Intelligence Research Center (AIRC), Ajman University, Ajman, 346, United Arab Emirates
[8]Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon
Emails: noura.hani@turath.edu.iq; EayFahidhil@uoalfarahidi.edu.iq; najlaa.faris@nust.edu.iq; Hussein.Alaa@kunoozu.edu.Iq; alubadyraaid@alayen.edu.iq; skadry@gmail.com
[*]Corresponding Author: noura.hani@turath.edu.iq

**Abstract**

Rapid changes in modern technology and sports have impacted society and lifestyle. Augmentative and Alternative Communication (AAC) technology helps to speak and play videos in various sports applications. In the current sports event, AAC's utilization to validate the players' complex moves exclusively has been considered a significant challenge that includes athlete moves in athletics and penalty shots in Soccer. Deep Learning-based Video Segmentation and Video mining (DL-VSVM) with eyeball tracking assistance are proposed to validate the task modeling of sports event video streaming in AAC. The user could select the specific event in the sport and sub-event using eyeball tracking assistance. The AAC is installed with unique icons to identify circumstances. A deep learning-based Sports Task model is created to recognize the required data to be streamed, and the model will help them view the specific sports event they need to watch. The numerical outcomes demonstrate that the suggested DL-VSVM model enhances the segmentation accuracy ratio of 95.3%, tracking ratio of 97.6%, prediction ratio of 98.7%, and reduces the cost function of 5.6% and the error rate of 20.1% compared to other existing models.

**Received: February 21, 2023 Revised: May 13, 2023 Accepted: September 04, 2023**

**Keywords:** Segmentation; deep learning; video streaming Task model and eyeball tracking.

## 1. Introduction

The internet has slowly changed all aspects of our day-to-day lives, from work to entertainment to learning and beyond [1]. With the rapid advancement of multimedia systems, the intense spread of sports video is made available on the Internet and broadcast [2]. The extensive diversity of services through modern media channels like mobile network devices and TV has shown the remarkable commercial possible of sports video and brought enormous, personalized sports videos consistent with customers' preferences [3]. The conventional one-to-numerous broadcast mode cannot encounter various viewers' demands. It is vibrant that when retrieving long, capacious video events, the capability to access high points and skip the fewer stimulating parts of the video will save the audience's time and data airtime/downloading costs receive video wirelessly from a remote server [4]. Furthermore, it will be eye-catching to access and view content based on their favorites [5]. The source video must be marked with the semantics label to comprehend the above requirements. These labels' obligation is broad to conceal the video's overall events, such as fouls in soccer, near-misses, and goals scored. It has to be adequately profound to cover the events' semantics; for example, the names of the performers are convoluted [6]. This is a perplexing task and could need to employ multi-modal and multi-context methods. Consequently, sports media content analysis information has gathered consideration from different studies in recent years. Sports data analysis fetches a high scale, differentiated, and shared [7]. The most persistent issue presently is how to access the most significant data in a small period. Due to the huge demand for sports video dissemination, several enterprises like Systems Applications and Products (SAP), Vizart and Bloomberg service sports content analytics [8]. Content analysis with big data has become the main developing industry. In online service, exposed latent knowledge can be used for real-time tactic suggestions. The record can be utilized in offline services to examine video content via deep learning [9]. Most sports games are naturally prearranged into successive and interchanging plays of defense and offence, cumulating at events like attack or goal [10]. If a sports video can be segmented according to these semantically meaningful events, several applications can improve their values and augment the user's observing experiences [11].

Communication in sports training grants numerous difficulties and interferences, resulting from the noise dispensed by the sportsperson and objects scrolling and hitting, the concurrent verbal communication of a high number of personalities, and the lack of acoustic appropriateness in the spaces utilized in outdoor facilities, teach lessons, with reverberation issues or external noises [12]. Similarly, the emotional side consequences are rather significant in Sports Education communication because the instructor must transmit a message and deliver sportsperson self-assurance to carry out the various actions positively and evade risky and insecure situations [13]. The messages transmitted in sports environments do not invent to reach identical integration levels in all sportspeople since they are habitually subject to diverse clarifications of the rules and morals that encourage originality and novelty of motor reactions [14]. The wide variety of sports videos offered by new media chains such as mobile network appliances and television have demonstrated outstanding commercial potential and have introduced huge, tailored sports films in keeping with customers' wishes.

Deep learning and Augmentative and Alternative Communication for video analysis is an emergent and vibrant field [15]. The relationship between video processing and video mining is very subjective. The video data mining objective is to extract a video series pattern, whereas video processing concentrates on understanding and extracting a feature from the video databank [16]. Video data mining and pattern recognition regions share the feature extractions phases and vary in pattern specificity [17]. Pattern recognition aims to identify a particular classification pattern, analysis, and generation. Pattern identification is pampering in research on categorizing superior samples with a current model, while video mining is convoluted in determining rules and sample patterns without or with image processing [18]. Video data mining aims to produce all-important patterns without preceding information. The ever-growing quantity of video archives creates physical annotation tremendously expensive and time-consuming [19]. There is a cumulative requirement for automatic video semantics annotation since individuals use high-level semantic notions when browsing and querying video databank. Video annotation can enable semantic video analysis like personalized video summarization and retrieval. DL-VSVM is suggested to validate the work of modelling sports event video streaming in the AAC with ocular tracking aid. With eyeball-tracking help, the user may pick a special event in the sport and sub-event. The AAC has a unique collection of symbols to recognize specific conditions. The deep learning Sports Task model is developed to identify the data necessary for streaming and helps observe the sports event that needs monitoring.

The main contribution of the paper is,

- To propose a DL-VSVM model for sports event task modelling for customized video streaming information in augmentative and alternative communication.
- Evaluating the mathematical model for sports video segmentation.
- Numerical outcomes have been implemented, and the recommended model improves the accuracy, prediction, tracking and error compared to other existing methods.

The rest of the study is organized as follows: section 1 and section 2 deliberate the introduction and existing video data segmentation and mining models. Section 3 discusses the DL-VSVM model. In section 4, experimental outcomes have been implemented. Finally, section 5 concludes the research study.

## 2. Literature Survey

Emily E. Cust et al. [20] proposed the Support Vector Machine algorithm (SVMA) for sport-specific motion recognition. This analysis aimed for the inertial measures unit (IMU) and computer vision information input to systematically analyze ML and DL literature for sport-defined motion recognition. A search has been carried out for various databases. The studies included would explore a sport-related movement and evaluate the production model using machinery or deep learning techniques. The inclusion and exclusion criteria have been followed in all tests. There is a wide range of data preprocessing, analysis, model creation and methods for evaluation. Supervised classification techniques mainly developed models for the recognition of movement.

Ali Javed et al. [21] suggested the thresholding-based approach, Gaussian mixture model and Extreme learning machine (TA-GMM-ELM) for Replay and key-events detection for sports video summarization. To sum up, a novel framework is introduced in this paper: sports videos. The replays inside a sporting video are considered the main incidents that can be used to summarize the video. These events. Replays were mostly sandwiched from beginning to end of graduation transfers. A threshold-based approach is used to identify the gradual effect on sports video transformation (i.e. fade-in, fadeout). The Gaussian Mixture Model (GMM) was then used to remove silhouettes and create a Motion History (MHI) picture for each main event. For each major event. The MHIs were processed for role extraction via CELP (CE-LTP) confined elliptical local ternary patterns. The ELM classifier system was used to learn the underlying events model. For main event identification, a qualified ELM-based classifier was used.

Muhammad Rafiq et al. [22] introduced the pre-trained AlexNet Convolutional Neural Network (PT-AlexNe-CNN) for Scene Classification for Sports Video Summarization. They reflect cricket as a case study and categorize five scene groups, i.e., bowling, boundary, batting, crowd, and close-up. The suggested approach employs a new, full connection layer in an encoder fusion. They use data augmentation to attain a high accuracy of 99.2% over lesser datasets. They conduct a performance comparison in contradiction to standard methods to demonstrate the advantage of the system and state-of-the-art methods. Their experiments show that the proposed system with AlexNet CNN generates improved outcomes than prevailing models.

Khan Muhammad et al. [23] initialized the Deep Convolutional Neural Network framework (DCNN) for Cost-efficient Video Summarization with Hierarchical Weighted Fusion. The first phase of their outline designs rich discriminatory features extracted for shot segmentation from the deep convolutional neural network. Then, along with aesthetic and entropic features, they use image memorability predicted by a sophisticated convolutional neural network model to make the summary exciting and varied. Thirdly, it is proposed to create a hierarchical, weighted fusion mechanism for an accumulation to calculate the features extracted efficiently. Finally, the focus curve for determining outstanding images for the video's final summary is defined using the aggregated score. Experiments were carried out with benchmark datasets to verify their Framework's value and efficacy, which outperforms other state-of-the-art frameworks. Different viewers cannot encounter a typical, one-to-numerous transmission manner. It was great that finding large and powerful videos saved viewers the time and data expenses of airtime/downloading movies from a remote server without having to access high points and skip the least exciting sections of the film. In addition, access to and see items depending on their advantages will be attractive. The source video must be labelled with the semantic label to understand the conditions above.

There are numerous challenges in the existing model, to overcome these problems, in this research, DL-VSVM has been proposed. The subsequent section deliberates the recommended methods briefly.

## 3. Deep Learning-based Video Segmentation and Video mining (DL-VSVM)

The present study proposes the DL-VSVM model for sports event task modelling in additional and alternative communication for personalized video streaming data. Assessment of the sports video segmentation mathematical model. Experimental findings have been carried out and compared with other current models, and the model presented boosts the accuracy, prediction, tracking and error rate. Large-scale sports video analysis with augmentative and alternative communication systems and DL is an emergent research area in several application fields. One of the most predominant trends is learning features utilizing a deep differentiative trained neural network for activity recognition. AAC's usage to verify the players' complex moves is considered a major challenge at a modern sporting event, such as competitors' moves and penalties in soccer. Use eyeball tracking to pick the specific event in the sport and sub-event. The AAC has a unique set of icons to identify specific circumstances. The Sports Task model is developed to determine the information needed for the webcast. The model allows them to interpret the sports event they need to watch. In this study, the DL-VSVM model has been suggested for task-modelling sports event video streaming in AAC. A method of video analysis may be considered a content reasoning system. The object is defined as object tracking to extract useful information from the video scene and track the movement, orientation, occlusion, etc. Tracking objects is a difficult problem, in general. There are many problems in detecting objects due to sudden motions, shifting presence of the scene and object, non-rigid object arrangements, object-to-scenes or camera motion and object-to-object occlusions. Video segmentation refers to the decay of video data into significant elementary parts that correlate strongly with the video data's real world. Video segmentation consists of a series of segments collectively covering the whole video data. The key distinctions between the image signal and the video signal are that a video signal incorporates temporal details that involve camera motion and introduce the idea of object movement. The video segment can be temporal, spatial and Spatio-temporal. The spatial domain segmenting of a video is as a static image. A time segmentation or shot detection is the name of a video frame within the temporal field. For video data, temporal segmentation is utilized for identifying the shots boundary may function both in the uncompressed and compressed field. Applications of segmentation contain video description, region-based image and retrieval and indexing, video summarization, detection of objects, interactive region-based annotation schemes, region-based coding that can serve as prompts for events recognition, etc. predominantly video and image description, retrieval and indexing have been on the emphasis of attention of several examiners functioning on video segmentation. It is a perplexing task that several multi-context approaches may need to be employed. As a result, several researchers acquired information on sports media content analysis last year. The examination of sports data is high-level, distinctive, and shared.

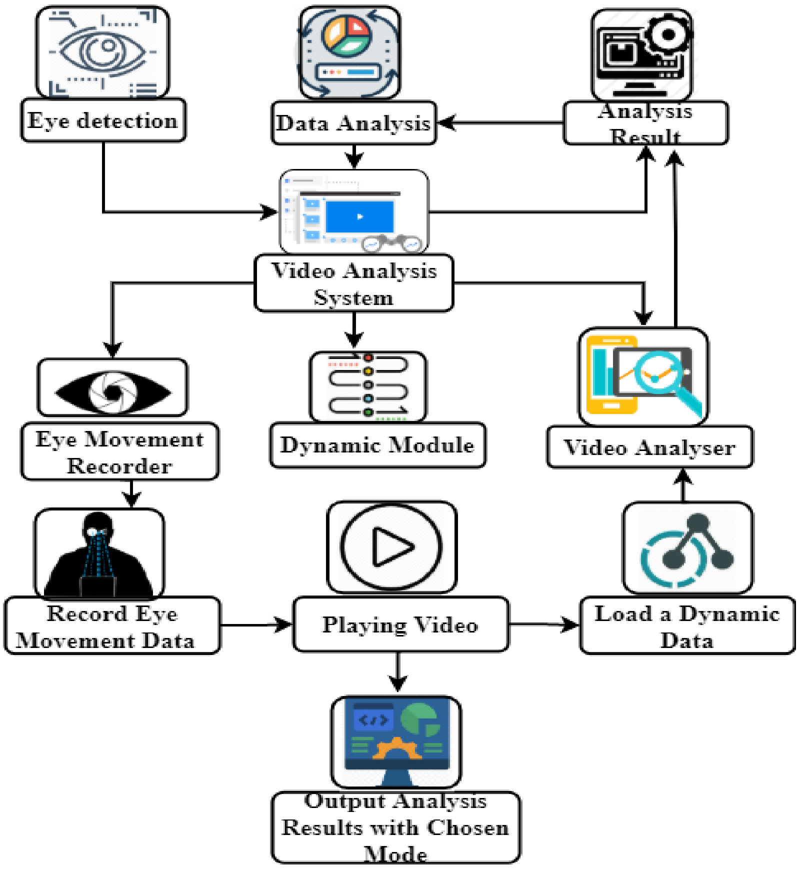


Figure 1: Eye Tracking in Video Analysis

Amplified, alternative and communication systems for large-scale sports video analysis and DL are emerging research fields in numerous application sectors. One of the most prevalent trends is using a deep differentiation network for brain activity identification. The use of AAC to verify the complicated movements of the participants in current athletic events, such as competitive movements and penalties in soccer, is considered a serious issue. To choose the exact competition in the sport and sub-event with the help of eyeball tracking.

Figure 1 shows the eye-tracking in video analysis. This study proposed DL-VSVM for the Eye Motion Data Analysis system to recover eye-tracking data, complex settings, and video readings. The system has three major components: the Eye Movement Monitor, Dynamic Module and Video Analyzer. Eye Movement Recorder connects with the eye-tracking software and tracks the participants' movement information while viewing the video. Dynamic Module produces and preserves video data independent of the monitoring and study of eye expression. Video Analyzer analyses the video's eye movement details by loading the registered eye motion data and saved data. Results of performance analyses in the Video Analyzer mode build trends of analysis results and save the analysis results. The Eye Moving Module Recorder monitors the subject's eye movement details with interactive video simulation. The Dynamic Module offers two types of settings: Manual Setting Mode and Automatic Setting Mode, allowing users to easily identify dynamic objects. The manual mode helps the user to describe the video dynamic with basic mouse operations. The Module of Video Analyzer is designed to measure, store, and display the individual subjects' view status in the video. In this module, the eye movement data will be collected, and the specified dynamic data will be combined with the fixation for every subject. The ongoing question now is how the most important data can be accessed in a short time. Due to the great demand for video sports distribution, several companies, such as SAP, Vizart and Bloomberg, provide sports content analysis services. Big data content analysis has become the key industry in development. The web service can use exposed latent knowledge for tactical suggestions in real-time. The record may be used to examine video material by thorough study in the off-line service.

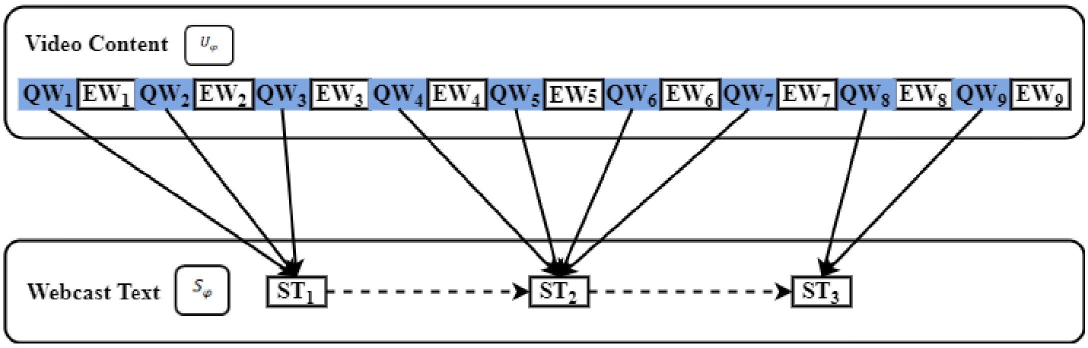


Figure 2: Webcast Text

Figure 2 illustrates the Webcast Text. This paper discusses the problem concerning sports videos' annotation. Conventional video annotation systems are focused mainly on interpreting video content, which demands many human inputs while gathering and organizing training details. Those structures can become unreliable if the audiovisual pattern is not incorporated in the training details. The DL-VSVM annotation is proposed for short-level video content and aligns great-level webcast texts. Video quality includes low-level audiovisual metrics to classify incidents in sports game videos. At the same time, data are parsed to provide detailed game information for high-level sports keywords. The two-tier characteristics complement each other. Require video analysis for sports. In addition, the connection between the video material and the source information can be discussed to integrate textual annotations with video fragments. The use of AAC to verify the complicated movements of the participants in current athletic events, such as competitive movements and penalties in soccer, is considered a serious issue. To select a certain event in the sport and the sub-event with the help of eyeball tracking. The AAC includes a unique collection of symbols for identifying certain conditions. The Sports Task Model is designed to identify the information required for the webcast, and the model can comprehend the sporting event to be seen.

Let's consider the video content $U_\varphi = \{(QW_1, EW_1), (QW_2, EW_2), \dots, (QW_M, EW_M)\}$ with $M$ pairs of pitch segment ($QW$) and event segment ($EW$) and $S_\varphi = \{ST_1, ST_2, \dots ST_N\}$ denotes the webcast text of $\varphi$ with $N$ webcast text objects ($ST$). Our goal is to map every $M$ video segment pairs $\{(QW_j, EW_j)\}$ to one of the text items $\{ST_n\}$ properly, indicated as $QW_j \to ST_n$; it refers to the $nth$ event $ST_n$ can be tagged to the $jth$ video series $(QW_j, EW_j)$. This study calls this annotation task the alignment issue among webcast text and video content of sports videos. The arrangement issue can be articulated as graphs that model the mapping relationship among $U_\varphi$ and $S_\varphi$. The map is stated by an onto functions $f_\varphi: U_\varphi \to S_\varphi$, where every component $S_\varphi$ remains mapped to at least one component in $U_\varphi$. Communications in sports training give rise to several problems and interference caused by the noise from the athletes, scrolling and striking subjects, the competitor's vocal communication of many personalities and the lack of acoustic suitability in the outdoor places, teachings, reversal problems or external noise.

$$\forall j, i \in [1, M], f_\varphi(QW_j, EW_j) = ST_n \text{ and } f_\varphi(QW_i, EW_i) = ST_m,\ n \le m \text{ if } j < i \quad (1)$$

Hidden Markov Model denotes the sub shots as $WW_1, WW_2, \dots, WW_R$ and the clusters as $C_1, C_2, \dots, C_W$, where $R$ and $W$ is the numbers of sub-shots and groups, correspondingly. The groups can be observed as a Markov chain state, and state transitions graphs can be designed by negotiating the series of sub shots. This study formulates the graph as a Hidden Markov Model issue to determine the steady-state likelihoods, signifying clusters' visit frequency (states). Meanwhile, pitch segments are the first state of each sports event, and this study assumes that the group comprising pitch segments must stay in the most frequently. The group with the maximum visit frequencies are recognized as pitch clusters. Supposing $Q$ exists a Markov model categorized by an $W \times W$ matrices. $Q$ is built from three transition likelihood matrices $B, A\ and\ D$. Firstly, this study describes the matrices of inner transition likelihoods $B = [B_{ji}]$:

$$B_{ji} = \begin{cases} \frac{b_{c_j c_i}}{\sum_{i=1}^{W} b_{c_j c_i}}, & if\ j \neq 1 \\ 0, & Otherwise \end{cases} \quad (2)$$

As inferred from equation (2) where $b_{c_j c_i}$ is the number of conversions between state $C_j$ and $C_i$. Succeedingly, this paper defines the matrices of intra-transition likelihoods $A = [Bij]$ as follows:

$$A_{ji} = \begin{cases} \frac{size\,(C_j)}{\sum_{i=1}^{W} size_{C_j}}, & if\ j = i \\ \frac{(1-A_{jj})}{(W-1)}, & otherwise \end{cases} \quad (3)$$

As shown in equation (3), where size $(C_j)$ indicates the number of video frames in $C_j$. Besides, a matrix of uniform transition likelihoods $D = [D_{ji}]$ is derived for each pair of $C_j$ and $C_i$.

$$D_{ji} = 1/W \quad (4)$$

As derived in equation (4), where $D$ is utilized to guarantee that the conversion likelihood among two states is higher than 0 after the convinced transition stages. Let $Q$ exist the linear combination of $B, A, D$.

$$Q = \beta \cdot B + \alpha \cdot A + \delta \cdot D \quad (5)$$

As discussed in equation (5), where $0 \leq \beta, \alpha\ \delta \leq 1$ and $\beta + \alpha + \delta = 1$. Therefore, this study can calculate $Q's$ primary left eigenvectors, $\pi$, with the eigenvalues 1

$$\pi \cdot Q = 1 \cdot \pi \quad (6)$$

As derived in equation (6), where $\pi$ could be construed as a $1 \times M$ steady-state likelihoods vector, and the $jth$ entry $\pi_j$ is the state's visit frequency $C_j$. This study aims to connect sports video text and video content with this annotation assignment. The arrangement problem can be stated as graphs validating the mapping relationship.

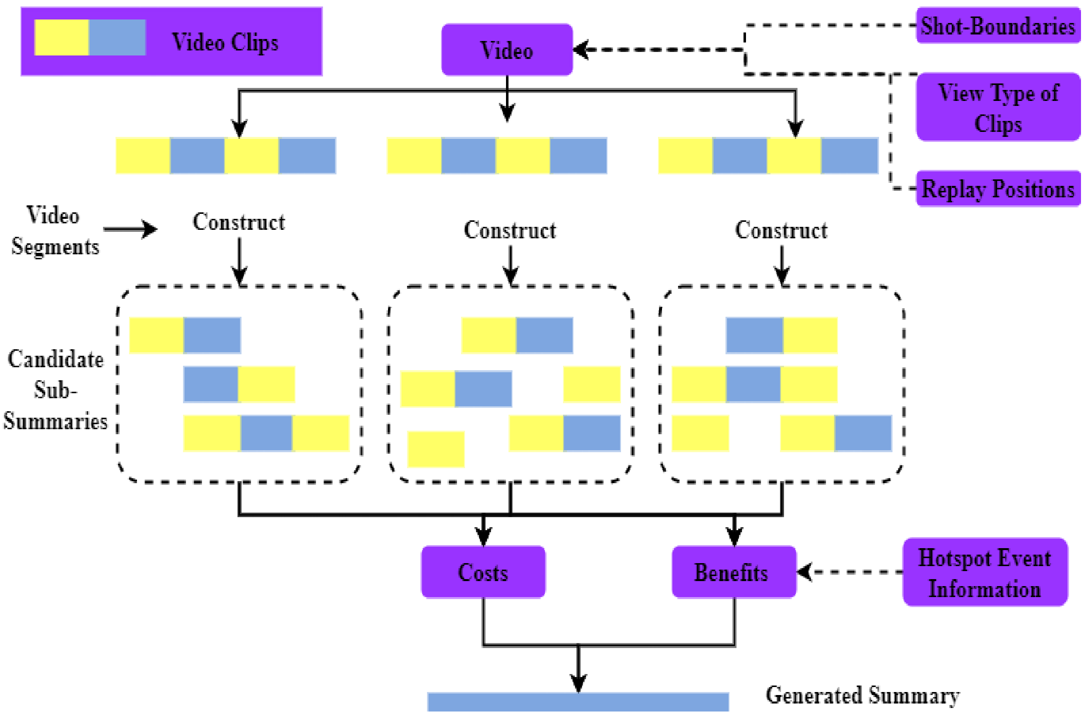


Figure 3: Summarization of Sports Video.

Figure 3 demonstrates the Summarization of Sports Video. Our overview process's general structure, where the video summary can be viewed as an optimization problem for finding a clip collection technique that gains the most under the restricted summary duration. The material the control room commonly cast is included in this text. Figure 3 shows how personal resumes are built-in non-overlapping and semantically significant segments based on this fragmentation content. The object is set as object tracking for the extraction and monitoring of the video scene's movement, orientation, occlusion, etc of valuable information. Tracking things is often a challenging issue. To recognize objects, there are several challenges owing to rapid motions, the shifting scene and item presence, non-rigid object arrangements, camera movement and object-to-object occlusions. Especially split a video into a series of clips with various view styles with all the shot boundaries known. Both unreplayed clips of a sports film are momentarily unjoined, and there is no drastic camera moving during crucial acts, contrary to the rules of creating sports videos. Therefore, DL-VSVM plans to customize the description in the paradigm. The proposed arrange these clips in non-overlapping segments by examining patterns of camera switching. The objective of model recognition is for identification of specific categorization patterns, model analysis, and creation. The pattern identification is relaxed in a study into the categorization of topics using an actual model. At the same time, video mining is involved in developing rules and sample patterns with or without the processing of images. Video data mining aims to create all-important patterns without prior information. The expanding number of video archives produces extremely costly and time-consuming physical annotations.

Each section is a short sub-story comprising sequentially linked consecutive clips. When a sub summary, called a narrative alternative or local storey, is defined as one way to pick videos from a chapter, see the final resumé as a non-compilation set. Their advantages and costs judge both optimum combinations of clips within each section, as seen in clips and incidents in this film. The suggested structure is general. Based on the application setting, the information accessible to or assumed on the scene at hand, and the user's narrative interests, each summary's advantages can be described in several ways. Although only productivity is explored using soccer, basketball and volleyball footage, the approach suggested can be used for other team sports by correctly changing the segmentation and advantage meanings guidelines.

Video segmentation refers to the degradation of video data into important primary sections that significantly correlate with the real world of video data. Video segmentation comprises several segments which jointly cover the entire video data. The main distinguishing feature between the image and the visual signals is that video signals include time elements, including camera movement and the notion of the motion of objects. The segment of videos might be time, space, and space-time.

This paper explains the building of a sub-summary by clips collection within video segments. The aim of this subdivision is twofold. First, this article describes a separate sub-summary (local task) based on scene types and view types for every segment clip. Second, this paper derives an advantage and cost metrics for every sub-summary utilized throughout resource distribution.

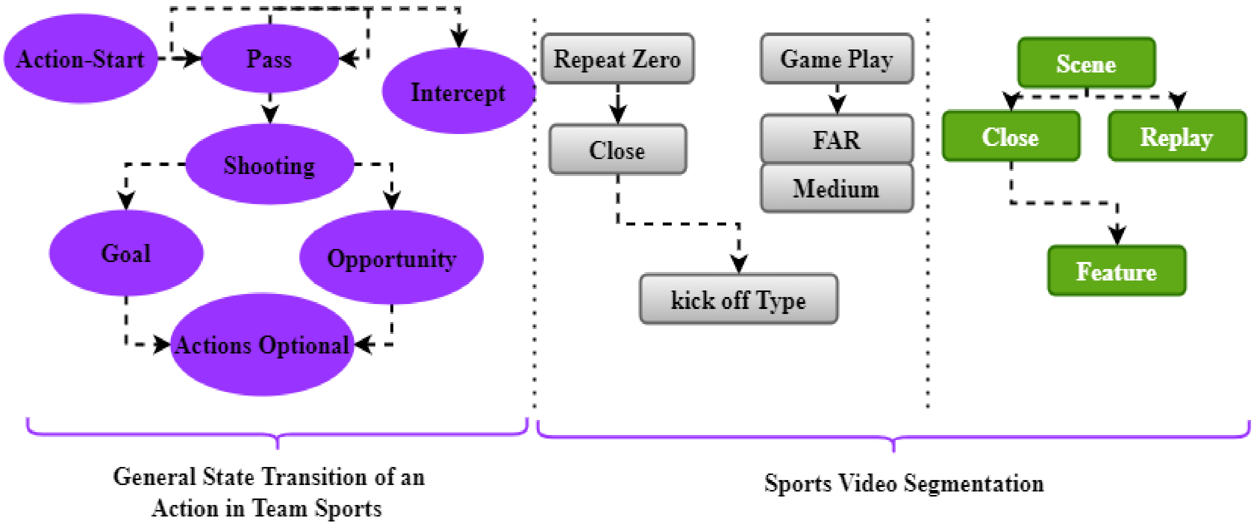


Figure 4: Team sports video Segmentation

Figure 4 shows the Team sports video Segmentation. Video segment as a duration that involves numerous successive clips narrowly linked in terms of semantic plot, clips for a football attack including the free-kick and its results. When considering the independent development of sub-summaries for each section, they sacrifice the summary between measurement efficiency and the tale organization's controllability. In the photo. 4, clarify the sports video segmentation law that imagines. It should be remembered that only information immediately accessible via the production process is needed for the proposed segmentation process. In other words, presume that the video sequence to section is not complicated by hand annotation or complex automated processing. Instead, take advantage of the game state changes that inspire the scene change to segment the video based on tracking creation behaviour rather than (complex) semantical scene analysis techniques mirrored in production actions. This means that general values of sports development are used to assess the game status by analyzing production behaviour as reflected by transitions in the clip view. In one round of offence/defence, a general diagram of state transformation is provided on the left of the Figure. 4. Starting at a kick-off style play, the offensive side checks the score after some passing acts. This test results in one of three potential outcomes: a ranking, an interception or a chance. Unusual acts, including foul stuff, medical attention and player trade, could happen before a new round. The graph to the right of the diagram is parallel. Figure 4 presents the sequential arrangement of video clips from implementing the concepts of sports development of teams' acts to deal with the change in the game's state. It demonstrates the traditional visual structure and highlights how the scene's textual substance relates. One section usually begins with a narrow view to reveal the player that starts. The key aspect of saying the past of a section is a series of long and medium views. Emotional degree may be administered to boost any close-up video after a significant incident. As seen in Figure. 4, proper video segmentation involves knowing the near views a player may improve before a corner or kick. A still action marks the kick-off's imminence from the production angle, resulting in a camera absence. Therefore, based on the calculation of dislocation of field lines, the kick-off detector is applied and efficiently relative to the camera image. For video data, temporal segmentation may function both in the uncompressed and compressed field to determine the shot boundaries. Segmentation applications include videos, region-based images, recovery and indexing, video summary, object identification, interactive region-based annotation systems, region-based coding, etc. Most of the focus of the examiners on video segmentation was video and picture description, recovery, and indexing.

Let's deliberate the $nth$ segments and elucidate how its sub-summary and related cost and advantages are calculated. For representation easiness, this paper omits the indices $n$ of segments under investigation. Statistically, undertake that the segment is collected of $M$ consecutive clip. For the $jth$ clips, this paper determines its scene types $w_j$, view types $u_j$, and replay condition $r_j$. This study set $w_j = 0$ for the public scene and $w_j = 0$ for the sports scenes and set $u_j$ to $0, 1$ for the medium, far and views, correspondingly. $r_j = 0$ for $jth$ clips in replays modes and $r_j = 1$ for standard sports. Let's preassume that an interest level $J_j$ calculated for every clip $j$, depending on the transmission of audios connotation to distinct clips. Then use $b_{lj} = 1$ to signify the adoption condition of picking the $jth$ clips into the $lth$ sub-summaries of segments, and $b_{lj} = 0$ for not choosing the clips. The cost of $b_l$ denotes its lengths, which is stated as $|b_l| = \sum_{j|b_{lj}=1} |\tau_j^E - \tau_j^W|$. This paper uses $\tau_j^W$ and $\tau_j^E$ to represent the beginning time and finish period of the $jth$ clips in the webcast videos. This article then proposes to describe the advantage expanded utilizing $b_l$ to express the $lth$ sub-summaries of the segments:

$$\mathbb{C}(b_l) = \sum_{j|b_{lj}=1} J_j \left(1 + \Phi\left(b_{l(j-1)} + b_{l(j+1)}\right)\right) \Gamma(b_l) \quad (7)$$

The term $b_{l(j-1)} + b_{l(j+1)}$ is comprised to augment extra advantage from the steadiness of task modelling by choosing sets of consecutive clips. Variable $\Phi$ controls the significance of task continuousness, where greater $\Phi$ hints at a more uninterrupted summary. $\Gamma(b_l)$ denotes the penalty carried by idleness in the sub summaries and other prohibited cases; for example, replays are designated while no regular play slice is chosen,

$$\Gamma(b_l) = \left[ \frac{\sum_{\substack{j|b_{lj}=1 \\ u_j>0, r_j=1}} |\tau_j^E - \tau_j^W|}{\sum_{j|b_{lj}=1} |\tau_j^E - \tau_j^W|} \right] \Gamma_2(b_l) \quad (8)$$

As shown in equation (8), the bracket term shows replays' idleness rate and close view. Note here that the $jth$ clips are not replays if $r_j = 1$, and is not close views if $u_j > 0$. Henceforth, greater $\delta$ endures a smaller amount of idleness, producing summaries with more and smaller sub-summary with fewer replays. At the same time, lesser $\gamma$ generates summaries containing fewer and lengthier sub-summary with comprehensive replay. $\Gamma_2(b_l)$ changes its value between 0.1 and 1, consistent with whether $b_l$ is of a prohibited form or not. In our subsequent employment for sports video, this study has described the following forbidden cases: 1) kick-off activity without the initial medium/far view clips; 2) a sub-summary with only replay, and 3) from the finishing period of the last medium/far view in segments, the incessant time in summaries for elucidation the result of the activity is smaller than a provided length.

This study explains how the importance of the highlighted instants perceived in segments can be interpreted into an attention $J_j$ for the segment's $jth$ clip. Good-sense and heuristic policies are suggested. To interpretation for assorted user favourites concerning the enclosure of extremely emotional clips, i.e., replays and close views, this paper derives both an emotional level $J_j^E$ and game relevance $J_j^G$ for the $jth$ clips. The interest levels $J_j$ is calculated as

$$J_j = \gamma J_j^E + (1-\gamma) J_j^H \quad (9)$$

As inferred from equation (9), where $\beta$ is a hyper-parameter of consumer favourite to control the comparative significance of game relevance and emotional levels. Emotional level $J_j^E$ and game significance $J_j^H$ of the $jth$ clips are calculated by adding all associated events for the $nth$ segments in line with its view-type construction,

$$J_j^E = T_n^E \sum_k O_{kj}^E R_k^E \quad J_j^H = T_n^H \sum_k O_{kj}^H R_k^H \quad (10)$$

As shown in equation (10) where $R_k^E$ , $R_k^H$ denote the emotional levels and game relevance allocated to the $kth$ emphasized events, correspondingly. $O_{kj}^E, O_{kj}^H$ indicate the proportion of the game/emotional interest in the $kth$ events allocated to the $jth$ clips.

The universal summary length resource is assigned between the accessible local sub-summary to exploit the combined advantage. This study refers openly to the segment indices $n$, and let $b_{nl}$ signify the $lth$ sub-summaries of the $nth$ segments. The total advantage of the entire summaries are stated as added advantages of every chosen sub-summary, that is

$$\mathbb{C}(\{b_l\}) = \sum_n \mathbb{C}(b_{nl}) \quad (11)$$

With $\mathbb{C}(b_{nl})$ stated in (7) are user function inclinations and the emphasized instants. Our main tasks are to examine for the sub-summary index set $\{l^*\}$ that enhances the overall payoff $\mathbb{C}(\{b_{nl}\})$ under the length constraints $\sum_n |b_{nl}| \leq v^{LEN}$, with $v^{LEN}$ being the user-preferred lengths of the summaries.

More specifically, this study resolves this resource allocation issue utilizing the Lagrangian slackening, whose primary theorem declaims that if $\lambda$ is a non-negative Lagrangian multiplier and $\{l^*\}$ denotes the optimum sets that increase,

$$K(\{L\}) = \sum_n \mathbb{C}(b_{nl}) - \lambda \sum_n |b_{nl}| \quad (12)$$

Total possible $\{l\}$, then $\{{b_{nl}}^*\}$ maximizes $\sum_n \mathbb{C}(b_{nl})$ overall $\{b_{nl}\}$ such that $\sum_n |b_{nl}| \leq \sum_n |{b_{nl}}^*|$. If $\{k*\}$ resolves the unimpeded issue in (12), then it offers the optimum resolution to the reserved issue in (11), with $v^{LEN} = \sum_n |{b_{nl}}^*|$. Since the assistances to the advantage and cost of every segment are independent and preservative,

$$\sum_n \mathbb{C}(b_{nl}) - \lambda \sum_n |b_{nl}| = \sum_n (\mathbb{C}(b_{nl}) - \lambda |b_{nl}|) \quad (13)$$

From the curves of $\mathbb{C}(b_{nl})$ concerning their respective summary length $|b_{nl}|$, the gathering of points increasing $\mathbb{C}(b_{nl}) - \lambda |b_{nl}|$ with similar slopes, $\lambda$ creates one unconstrained optimal. Diverse selections of $\lambda$ lead to various summary lengths. If it builds a series of curved exteriors from $\mathbb{C}(b_{nl})$ curvatures concerning $|b_{nl}|$, this study can utilize a search for the finest under the provided constraints $v^{LEN}$.

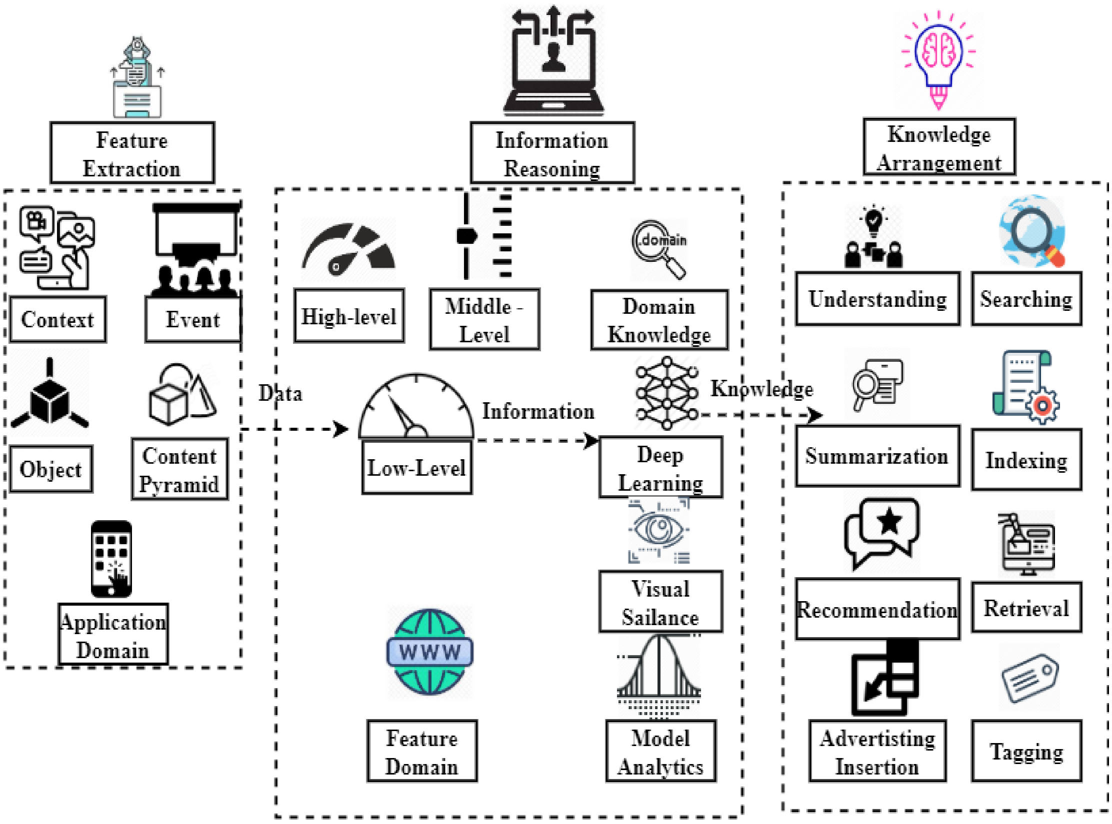


Figure 5: Content-Aware Video Mining Framework

Figure 5 shows the Content-Aware Video Mining Framework. The recovery process must be planned according to the user's purpose in designing a content-aware recovery method. Various sports video analytics tools have been developed to handle spatial-visual elements. A general system for evaluating short-net sports videos from pixel, object, and scenario perspectives. This system allows for detecting a set of predefined incidents at various stages. The video material is divided into four layers according to the semantical importance: video, object, behaviour and conclusion. The layer volume represents the number of the definitions inferred. The pyramid concept is used to examine video entities' contextual issues. The numerous core components to define video quality in every layer of the content grading are. The video layer comprises video clip frames, consisting of raw video content and video clip tag. The class object layer comprises an object frame containing the video's main object. An object tag and a pointer connect any main object to each object frame's respective video clips. An object with or interacts with a specific location or activity offers activities or interaction tags. The layer of the event class contains an event frame reflecting the main object's behaviour. Acts, in combination with scene details, create an incident identifier. Every event frame includes event and object tags that describe the behaviour or connection with several objects. The top layer is the final layer and contains the final frame demonstrating the video sequence's semantical description. The case tags and subsequent outcomes comprise each concluding frame. A game description is drawn up in conjunction with transcripts and event results. The suggested DL-VSVM model achieves a high segmentation accuracy ratio, tracking ratio, prediction ratio, reduced cost function, and error rate compared to other existing models.

## 4. Results and Discussion

The suggested DL-VSVM model's trial outcomes have been implemented based on the performance metrics like segmentation accuracy ratio, tracking ratio, prediction ratio, cost function, and error rate. This research experimented with video/text alignment to perceive event boundaries in webcast soccer videos. This study analyzes the 100 sports video events from online sources.

### (i) Segmentation Accuracy Ratio

Analysis of eye monitoring material is one of the most active fields of inquiry. This method is successfully extended for behaviour identification by evaluating and assessing intra-object pose. An accurate body position and precise motions are important to achieve the anticipated output in sports trial and field games. The required output in sports is accomplished by inter-object interactions for two-object applications, whereas inter-group interactions are used in sports. The goal is to decide how many cameras can be combined to produce a more precise tracking outcome. An accurate human body location with expressed components and movement variables is required in the person class to assess local and global gesture performance. Figure 6 shows the Segmentation Accuracy Ratio.

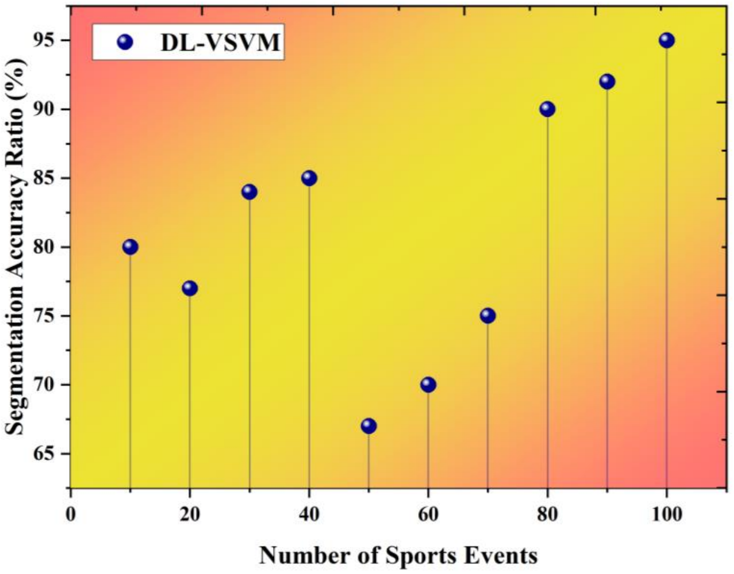


Figure 6: Segmentation Accuracy Ratio

**(ii) Tracking Ratio**

Object monitoring is an important technology in the study of sports content. The aim is to identify the objects inside a video sequence and to include video surveillance, AAC, and visual indexing for various applications. The monitoring of objects is used in sports to observe the ongoing movements of objects in the playfield. The camera motion is nevertheless a significant restriction in object detection and tracking. A genetic algorithm for fast camera movement effect removal in target scene and player tracking. A region-based algorithm is utilized to monitor athletes, and an occlusion statement between players is added to either the prototype matching or the split-and-fusion solution. A recording and monitoring subsystem is used to assess the spatiotemporal trajectory, which can be used for incident prediction and sports expert analysis for the input video series. However, monitoring success is enough as athletes move in clusters during lengthy and continued full occlusions. Figure 7 shows the Tracking Ratio.

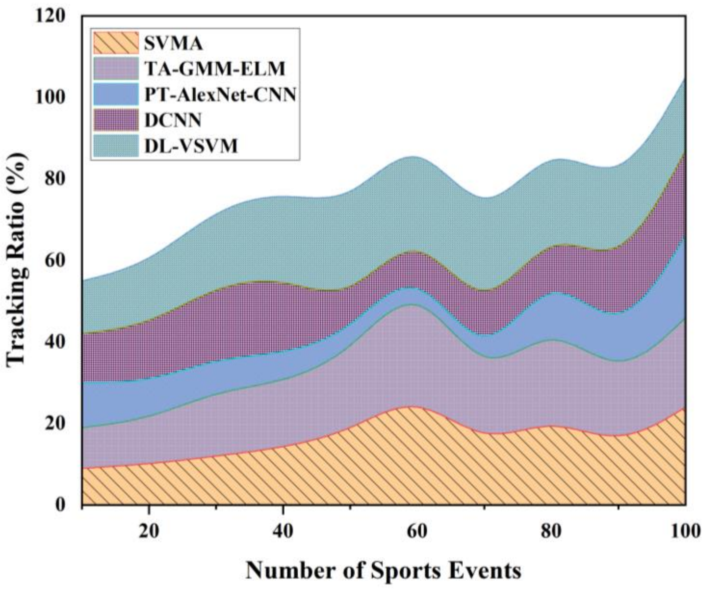


Figure 7: Tracking Ratio

#### (iii) Prediction Ratio

In sports person-object interaction events, a reciprocal meaning between objects and sports person poses will make the other more identifiable. Specifically, the shared meaning model takes care of two pieces of contextual knowledge. Deep learning has been used in sports to examine pathways and predict when a three-point shot is successful. These context models the co-incidence statistics among target and particular forms of athletes pose in any activity. According to the video frame, the loaded image is obscured by a black block to deter participants from anticipating stimuli according to the video frame. The resizing of the window facilitates the precise monitoring of the video stimulus position of the gaze. Figure 8 shows the Prediction Ratio.

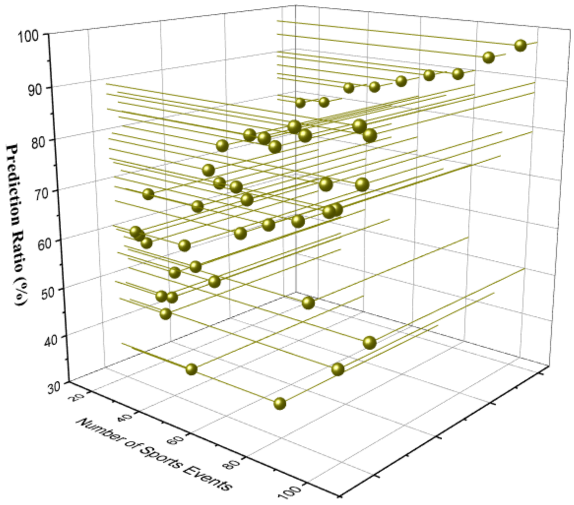


Figure 8: Prediction Ratio (%)

#### (iv) Error Rate

A classification system is used to classify the types of behavior as function vectors are extracted. For example, the proposal is based on a newly proposed over-complete independent variable DL-VSVM model for an efficient action recognition technique. The reaction characteristics of over-full DL-VSVM are taken instead of a pretrained classification. They are graded. In this method, 3D patches from a training video for every operation practice a series of over-completed functions for DL-VSVM. The videos can be classified as an activity class with the slightest errors as the base functions for restoring the video. By handling the clips via four stages: field extraction, cameras' matrix, matrix estimation, refinement and clip refinement, the genetic algorithm provides detailed camera matrices for each frame copied. Figure 9 shows the Error Rate.

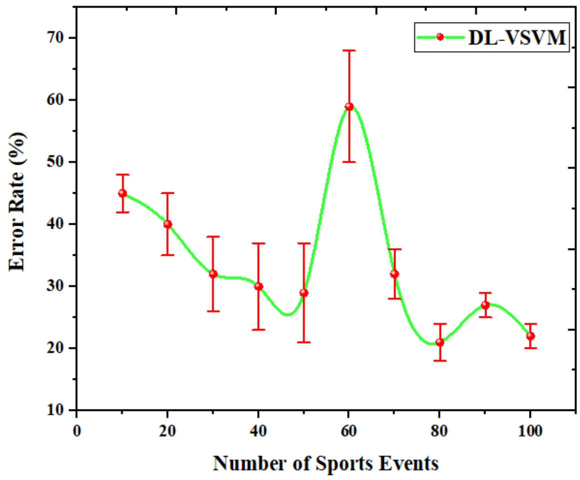


Figure 9: Error Rate.

**(v) Cost Function**

This paper can be achieved by evaluating the graphic nature of the video being transmitted, taking account of generic concepts of sports video creation, and by creating cost-benefit functions to determine how the diverse sections of the video add to the viewer's gratification, for of section clip, knowledge of the scenario form and scene type. Second, obtain a gain and cost metric for the distribution of resources for each sub-summary. Recordable conveniently from the manufacturing process, the proposed DL-VSVSM system allows content suppliers to deploy additional resources cost-effectively to generate more benefits. The optimal combinations in each field are judged by their advantages and costs in compliance with user preferences and understanding of the types of clips and incidents in the video. Figure 10 shows the Cost Function.

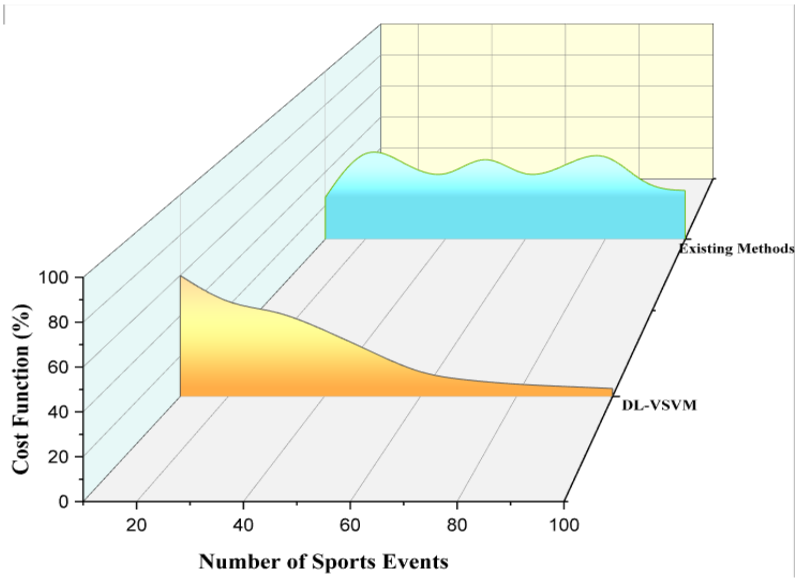


Figure 10: Cost Function.

The suggested DL-VSVM model enhance the segmentation accuracy, prediction, and tracking ratio and reduces the cost function and error rate when compared to other exiting Support Vector Machine algorithm (SVMA), thresholding-based approach, Gaussian mixture model and Extreme learning machine (TA-GMM-ELM), pre-trained AlexNet Convolutional Neural Network (PT-AlexNet-CNN), deep CNN framework (DCNN) models.

## 5. Conclusion

This paper suggested a DL-VSVM model with AAC for generating personalized summarization of sports videos. The video is separated into many clips. This study dealt with short clips within every local period segment to establish storytelling in the resource allocation model. Lagrangian relaxation has been utilized to determine the optimal assortment of clips to form the concluding summary. Our system is malleable because it helps the dissimilar definition of advantage to adapt to the summarization procedure. Video annotation depended on rearranging webcast text and presented video content. The proposed genetic algorithm optimization method can resolve the alignment problem, integrating multi-modal cues from great-level webcast text and low-level video contents. The Hidden Markov Model for recognition of pitch segments decreases the necessity for human interferences. Experimentations show a vigorous outcome against a diversity of high automaticity and video content in sports video annotation. The numerical outcomes show that the suggested DL-VSVM model enhances the segmentation accuracy ratio of 95.3%, tracking ratio of 97.6%, prediction ratio of 98.7%, and reduces the cost function of 5.6% and the error by 20.1% compared to other existing models.

## References

[1] Fakhar, B., Kanan, H. R., & Behrad, A. (2019). Event detection in soccer videos using unsupervised learning of Spatio-temporal features based on pooled spatial pyramid model. *Multimedia Tools and Applications*, *78*(12), 16995-17025.

[2] Yassine, S., Kadry, S., & Sicilia, M. A. (2020). Statistical Profiles of Users' Interactions with Videos in Large Repositories: Mining of Khan Academy Repository. KSII Transactions on Internet and Information Systems (TIIS), 14(5), 2101-2121.

[3] Gao, Y., Wei, X., & Zhou, L. (2020). Personalized QoE improvement for networking video service. *IEEE Journal on Selected Areas in Communications*, *38*(10), 2311-2323.

[4] Kumar, N., Lee, J. H., & Rodrigues, J. J. (2014). Intelligent mobile video surveillance system as a Bayesian coalition game in vehicular sensor networks: Learning automata approach. IEEE Transactions on Intelligent Transportation Systems, 16(3), 1148-1161.

[5] Cuevas, C., Quilon, D., & Garcia, N. (2020). Techniques and applications for soccer video analysis: A survey. *Multimedia Tools and Applications*, *79*(39), 29685-29721.

[6] D. M. Ramírez Guerra, Y. M. Gordo Gómez,L. J. Cevallos Torres, F. G. Palacios Ortiz, Social sports Competition Scoring System Design Using Single Value Neutrosophic Environment, International Journal of Neutrosophic Science, Vol. 19 , No. 1 , (2022) : 389-402 (Doi : https://doi.org/10.54216/IJNS.190135)

[7] Wang, A., Gao, X., & Tang, M. (2020). Computer Supported Data-driven Decisions for Service Personalization: A Variable-Scale Clustering Method. *Studies in Informatics and Control*, *29*(1), 55-65.

[8] Su, H., Chang, Y. K., Lin, Y. J., & Chu, I. H. (2015). Effects of training using an active video game on agility and balance. The Journal of sports medicine and physical fitness, 55(9), 914-921.

[9] Huifeng, W., Kadry, S. N., & Raj, E. D. (2020). Continuous health monitoring of sportsperson using IoT devices based wearable technology. Computer Communications, 160, 588-595.

[10] Huifeng, W., Shankar, A., & Vivekananda, G. N. (2020). Modelling and simulation of sprinters' health promotion strategy based on sports biomechanics. Connection Science, 1-19.

[11] Anjum, M. A., Amin, J., Sharif, M., Khan, H. U., Malik, M. S. A., & Kadry, S. (2020). Deep Semantic Segmentation and Multi-Class Skin Lesion Classification Based on Convolutional Neural Network. IEEE Access, 8, 129668-129678.

[12] Zhao, J., Li, K., Xi, X., Wang, S., Saravanan, V., & Samuel, R. D. J. (2020). Analysis of complex cognitive task and pattern recognition using distributed patterns of EEG signals with cognitive functions. Neural Computing and Applications, 1-10.

[13] Sharma, D., & Kumar, N. (2017). A review on machine learning algorithms, tasks and applications. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 6(10), 2278-1323.

[14] *Ji, F., Hsu, C. H., & Montenegro-Marin, C. E. (2020). Evaluating and recognizing stressful periods and events of urban migrant children from microblog. Current Psychology, 1-9*

[15] Xu, X., Chen, Y., Zhang, J., Chen, Y., Anandhan, P., & Manickam, A. (2020). A novel approach for scene classification from remote sensing images using deep learning methods. European Journal of Remote Sensing, 1-13.

[16] Ali, M.H., Al-Azzawi, W.K., Jaber, M., Abd, S.K., Alkhayyat, A., and Rasool, Z.I., 2022. Improving coal mine safety with internet of things (IoT) based Dynamic Sensor Information Control System. *Physics and Chemistry of the Earth*, 128.

[17] Saeed Kolahi-Randji, S., Nejad Attari, M.Y. & Ala. A. (2023). Enhancement the Performance of Multi-Level and Multi-Commodity in Supply Chain: A Simulation Approach. Journal of Soft Computing and Decision Analytics, 1(1), 18-38. https://doi.org/10.31181/jscda1120232

[18] Islam, M., Mahmood, A. N., Watters, P., & Alazab, M. (2019). Forensic Detection of Child Exploitation Material Using Deep Learning. In Deep Learning Applications for Cyber Security (pp. 211-219). Springer, Cham.

[19] Shi, J., Yuan, X., Elhoseny, M., & Yuan, X. (2020). Weakly supervised deep learning for objects detection from images. In Urban Intelligence and Applications (pp. 231-242). Springer, Cham.

[20] Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *Journal of sports sciences*, *37*(5), 568-600.

[21] Javed, A., Irtaza, A., Khaliq, Y., Malik, H., & Mahmood, M. T. (2019). Replay and key-events detection for sports video summarization using confined elliptical local ternary patterns and extreme learning machine. *Applied Intelligence*, *49*(8), 2899-2917.

[22] Rafiq, M., Rafiq, G., Agyeman, R., Choi, G. S., & Jin, S. I. (2020). Scene classification for sports video summarization using transfer learning. *Sensors*, *20*(6), 1702.

[23] Mostafa, S. A., Gunasekaran, S. S., Mustapha, A., Mohammed, M. A., & Abduallah, W. M. (2020). Modelling an adjustable autonomous multi-agent internet of things system for elderly smart home. In Advances in Neuroergonomics and Cognitive Engineering: Proceedings of the AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering, and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology, July 24-28, 2019, Washington DC, USA 10 (pp. 301-311). Springer International Publishing.