**1.Introduction:**

In an era defined by the convergence of technology and sports, the project "Cricket Sports Video to Text Summarization" represents a monumental leap in how cricket enthusiasts engage with their favourite sport. This innovative endeavour seamlessly translates dynamic cricket video content into concise, accessible textual summaries. Traditionally, summarizing cricket events has been a labour-intensive task, reliant on manual interpretation and subjective judgments. However, propelled by cutting-edge neural network methodologies, a new era unfolds. This project harnesses the power of advanced technologies, integrating computer vision and natural language processing to automate the generation of multi-language text summaries exclusively tailored for cricket. Beyond its technical prowess, this innovation holds the promise of transcending linguistic barriers. By enabling real-time translation of summaries into multiple languages, the project democratizes access to cricket content on a global scale. This inclusive approach aims to make the exhilarating world of cricket universally accessible, opening doors to fans and enthusiasts regardless of their native tongue. This project's evolution into a specialized cricket summarization tool signifies its commitment to providing cricket enthusiasts with an unparalleled sports content experience. As we delve deeper into the architecture, methodologies, and potential applications, it becomes evident that this endeavour holds the potential to redefine how we perceive, analyse, and appreciate cricket in the digital age. It's a journey that empowers cricket lovers, offering a refined lens through which to dissect and savor the nuances of cricket events. This introduction sets the stage for an exploration of an innovative project, where technology, cricket, and language converge to create a transformative sports content experience.

**2.Defining the Scope:**

In the world of sports and digital media, Cricket Sports Video Summarization using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) units promises to be a groundbreaking technological innovation.

This project scope outlines the ambitious endeavour to develop a system capable of intelligently summarizing cricket matches, extracting key moments, and presenting them in a concise textual format and is promising, and it opens various avenues for improvement, innovation, and expansion. Here are some potential future directions and enhancements:

**2.1. Real-time Summarization:** Develop a real-time summarization system that can process and summarize live cricket matches as they happen. This would require efficient algorithms and robust streaming capabilities.

**2.2. Interactive User Interfaces:** Create user-friendly interfaces that allow viewers to interact with the summarization system, selecting specific aspects or moments of the match they want to explore in more detail.

**2.3. Customizable Summaries:** Provide users with options to customize the level of detail in their summaries, allowing them to choose between concise or detailed summaries based on their preferences.

**2.4. Multi-sport Summarization:** Extend the technology to support various sports summarization, catering to a wider audience of sports enthusiasts.

**2.5. Social Media Integration:** Integrate the summarization system with social media platforms, enabling users to share and discuss summarized cricket match highlights more easily.

**2.6. Automated Highlights Generation:** Implement a feature that automatically extracts and compiles the most exciting moments into short highlight reels, making it easier for fans to catch up on matches quickly.

**2.7. Scalability and Performance Optimization:** Optimize the system's scalability, performance to handle a larger volume of video data and accommodate various video resolutions and qualities.

**2.8. Multi-language Support:** Enable the summarization system to provide summaries in multiple languages, expanding its accessibility to a global audience.

**2.9. Collaboration with Sports Broadcasters:** Collaborate with sports broadcasters and streaming platforms to integrate summarization technology into their services, enhancing the viewer experience.

**2.10. AI-Driven Insights:** Explore the use of AI to provide deeper insights into player performance, team strategies, and match statistics, helping coaches and analysts gain valuable information.

**2.11. Adaptive Summarization:** Develop adaptive summarization models that can learn from user feedback and preferences to tailor summaries to individual tastes.

**2.12. Ethical AI:** Continue to prioritize ethical considerations, data privacy, and transparency, and work on AI systems that are responsible and trustworthy.

**2.13. Research and Collaboration:** Engage in collaborative research with academic institutions and industry partners to push the boundaries of video summarization technology and explore new applications.

The future of cricket sports video summarization is bright, with the potential to transform how fans engage with cricket matches and other sports events. As technology continues to evolve, innovators in this field can provide more immersive and personalized experiences for sports enthusiasts.

**3.Search Strategy:**

Cricket Sports Video to Text Summarization (CSVTS) represents an evolving frontier in the realm of sports content consumption, poised to reshape how we engage with cricket matches. In this context, the implementation of a comprehensive and meticulous search strategy becomes pivotal to ensure the encapsulation of the latest research and advancements in this burgeoning field. Our search strategy for conducting a literature review on CSVTS is thoughtfully structured in two distinct stages: Information Gathering and Article Filtering and Reviewing. These stages are carefully orchestrated to facilitate a thorough examination of the vast body of work in the CSVTS domain.

**3.1. Information Gathering:** Navigating the Sea of Data

The initial phase of our search strategy revolves around information gathering, a critical foundation for any literature review. This stage commences with the identification and refinement of a set of pertinent search terms closely aligned with the CSVTS landscape. These search terms traverse a wide spectrum, encompassing fundamental phrases like "cricket sports video to text summarization" to more nuanced expressions such as "event-based cricket video analysis" and "real-time cricket video text summary generation." This extensive keyword selection process ensures that we cast a wide net in our search, leaving no relevant stone unturned. Having curated an exhaustive list of search terms, we then embark on the task of selecting appropriate databases to explore for research articles. Our chosen databases are recognized for their robust repositories of scientific and technical literature. They include venerable platforms such as IEEE Xplore(20), ACM Digital Library(155,832), Google Scholar(2,210), SpringerLink(328), and ResearchGate(42). To further expand our horizons and glean insights from diverse academic domains, we extend our search to encompass arXiv and JSTOR, esteemed for their extensive collections spanning fields like information technology, computer sciences, computer engineering, and communication. Once our arsenal of search terms and databases is firmly established, we meticulously craft search queries tailored to each individual database. These queries are meticulously structured to include a judicious blend of search terms and publication date filters, the latter serving as a temporal boundary. Specifically, we restrict our search to articles published between 2015 and 2023, a time frame carefully chosen to align with the surging interest and proliferation of CSVTS-related research.

**3.2. Article Filtering and Reviewing:** Navigating the Maze of Literature

The second stage of our search strategy, Article Filtering and Reviewing, represents the heart of our literature review process. The primary objective here is to sift through the troves of retrieved articles and classify them methodically based on four pivotal facets of CSVTS: design, implementation, application, and evaluation methods. We initiate this phase by subjecting our search results to rigorous title-based filtration. In doing so, we place a premium on articles that offer tangible instances of technical implementations and developmental strides within the CSVTS landscape. We also recognize the value of publications contributing to a broader understanding of CSVTS through literature reviews and surveys, as these illuminate the evolutionary trajectory of the field. Simultaneously, we exercise discernment by excluding articles predominantly focused on the marketing facets of CSVTS or those veering into social studies involving CSVTS, as these themes deviate from the core objectives of our study. Our primary goal is to explore the cutting-edge Deep Learning algorithms and Natural Language Processing techniques employed in contemporary CSVTS implementations. According to Google Scholar, there are 102 publications on the topic of "cricket" AND "sports video" AND "text summarization" from 2015 to 2023.There are 1,234 publications on the topic of "cricket" OR "sports video" OR "text summarization" from 2015 to 2023, and there are 308 publications on the topic of "cricket" OR "sports video" AND "text summarization" from 2015 to 2023.To augment our repository, we embark on a quest for additional sources. We delve into arXiv to unearth potentially relevant articles while meticulously cross-referencing to eliminate redundancy. Furthermore, we navigate the bibliographies of previously identified articles, unearthing hidden gems that may have eluded our initial search. This meticulous process culminates in the assembly of a robust collection of pertinent articles, thereby enabling a comprehensive analysis of the dynamic CSVTS landscape.

**4.Selection criteria:**

**4.1. Relevance to Innovative Sports Video Technology:**

- The selected sources must directly pertain to innovative sports video technologies, encompassing areas such as video summarization, neural network methodologies, and multilingual translation.

**4.2. Methodological Rigor and Technical Depth:**

- Emphasis will be placed on sources that demonstrate methodological rigor in the fields of computer vision, natural language processing, and neural networks.

- The literature should exhibit technical depth in the development and application of neural network architectures, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based models.

**4.3. Cutting-Edge Innovations:**

- Priority will be given to studies that introduce cutting-edge innovations in the realm of sports video analysis, including novel techniques for video feature extraction, temporal relationship modelling, and neural machine translation.

**4.4. Real-World Applications and Practicality:**

- Sources that showcase real-world applications and practical implementations of automated sports video summarization and multilingual translation will be favoured.

**4.5. Interdisciplinary Insights:**

- The criteria will consider sources that offer interdisciplinary insights, bridging the domains of sports analytics, computer vision, and natural language processing, to provide a holistic perspective.

**4.6. Credibility and Publication Venue:**

- Reputable publication venues, including renowned conferences and journals in the fields of machine learning, computer vision, and natural

language processing, will be prioritized.

- Authors with expertise and credibility in the areas of sports video analysis and neural network methodologies will also be considered.

**4.7. Temporal Relevance and Up-to-Date Information:**

- Preference will be given to recent studies and publications that reflect the latest advancements and trends in the application of neural networks for sports content summarization and multilingual translation.

**4.8. Accessibility and Availability:**

- Efforts will be made to include literature that is readily accessible through open-access platforms, institutional subscriptions, or reputable online repositories.

**4.9. Contribution to the Field:**

- The criteria will identify and prioritize sources that have made significant contributions to the field of sports video analysis and multilingual summarization through innovative methods, novel techniques, or practical applications.

**4.10. Practical Utility and Scalability:**

- Sources highlighting the practical utility and scalability of the proposed methodologies for real-time sports content processing and translation will be considered valuable.

**4.11. Alignment with Project Objectives:**

- The selection criteria will ensure that the chosen literature aligns with the project's objectives of automated sports video summarization, multi-language text summarization, and real-time multilingual translation.

These refined selection criteria provide a focused and targeted approach for sourcing relevant literature for your project.

**5.Data Extraction**

Certainly, here's a step-by-step explanation of the process for "Cricket Sports Video Text Summarization" using Data Extraction with CNN and RNN (LSTM) and Natural Language Generation (NLG):

**5.1. Data Collection**- Gather a dataset of cricket sports videos along with associated textual descriptions or transcripts. This dataset should contain a variety of videos, including different matches, events, and commentary.

**5.2. Preprocessing**- Preprocess the textual data by cleaning and tokenizing it. This involves removing special characters, lowercasing, and splitting the text into individual words or tokens.

**5.3. Feature Extraction with CNN**- Utilize Convolutional Neural Networks (CNN) to extract relevant features from the video frames. This process involves analysing the visual content of the video to identify key moments, actions, or highlights. CNNs are well-suited for image and video analysis.

Here we use VGG 16 CNN neural network model,

VGG16 is a convolutional neural network (CNN) architecture used for image classification tasks. It consists of 16 layers, primarily composed of small 3x3 convolutional filters and 2x2 max-pooling layers. VGG16 is known for its simplicity and effectiveness, with the ability to learn hierarchical features from input images. It's often used as a feature extractor, and pretrained versions are available for transfer learning. However, it has many parameters, making it computationally intensive compared to more recent CNN architectures like ResNet and Inception.

**Architecture Overview**- VGG16, short for "Visual Geometry Group 16," is a deep convolutional neural network designed for image classification tasks. It was developed by the Visual Geometry Group at the University of Oxford. The "16" in VGG16 signifies the network's depth, which includes 13 convolutional layers and 3 fully connected layers.

**Convolutional Layers-** VGG16 uses a series of convolutional layers to learn hierarchical features from input images. Convolutional layers apply small 3x3 filters to the input with a stride of 1 and zero-padding to maintain spatial resolution. The number of filters increases as you go deeper into the network, allowing it to capture increasingly complex and abstract features.

**Pooling Layers-** After every two convolutional layers, there is a max-pooling layer with a 2x2 window and a stride of 2. Max-pooling reduces the spatial dimensions of the feature maps while retaining the most important information. This helps in reducing computational complexity.

**Fully Connected Layers-** VGG16 concludes with three fully connected layers. The first two fully connected layers have 4,096 neurons each, making the network deep and capable of learning intricate patterns. The final fully connected layer typically has as many neurons as there are classes in the classification task. It produces class probabilities using the SoftMax activation function.

**Activation Function-** Throughout the network, VGG16 uses the rectified linear unit (ReLU) activation function after each convolutional and fully connected layer. ReLU introduces non-linearity, enabling the model to learn complex relationships in the data.

**Classification-** VGG16 is often used for image classification tasks, where it assigns one or more labels to an input image. The final fully connected layer is adapted to the specific classification problem by adjusting the number of output neurons and using SoftMax activation to produce class probabilities.

**Pretrained Models-** Pretrained versions of VGG16, trained on large image datasets like ImageNet, are available. These pretrained models can be fine-tuned on smaller, task-specific datasets to achieve good performance with limited labelled data, a technique known as transfer learning.

In summary, VGG16 is a deep CNN architecture known for its simplicity and effectiveness in image classification tasks. Its use of small convolutional filters and max-pooling layers allows it to learn intricate features from images. While it's computationally intensive, pretrained versions make it valuable for transfer learning applications.

**5.4. Temporal Context with RNN (LSTM)-** Employ Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) cells, to capture the temporal context of the video. LSTM networks can model sequences effectively, making them suitable for understanding the chronological flow of events in a cricket match.

**Data Preparation**- Collect the cricket sports videos and their corresponding textual descriptions or transcripts. Ensure that the videos are divided into small, manageable segments or frames. Each frame should be associated with a timestamp.

**Temporal Sequence Creation**- Organize the frames from the video into a temporal sequence. This sequence represents the chronological order of frames as they occur in the video.

**Feature Extraction-** Extract relevant features from each frame of the video. These features can include visual information like color histograms, motion vectors, or even deep learning-based features from convolutional neural networks (CNNs).

**Sequence Encoding with LSTM-** Utilize Long Short-Term Memory (LSTM) cells to encode the temporal sequence of features. LSTM is a type of recurrent neural network (RNN) that is particularly well-suited for modeling sequences due to its ability to capture long-range dependencies and handle vanishing gradient problems.

**Input Sequence to LSTM**- Feed the sequence of extracted features into the LSTM network one frame at a time. At each time step, the LSTM processes the current frame's features and updates its hidden state, which acts as a memory of past information.

**Capturing Temporal Context-** As the LSTM processes the frames sequentially, it captures the temporal context of the video. It remembers the information from previous frames and incorporates it into its understanding of the current frame. This enables the network to understand the chronological flow of events in the cricket match.

**Hidden State and Output**- The LSTM network's hidden state at the final time step contains a summary of the temporal context and information from all previous frames. This hidden state can be used to make predictions or generate textual summaries of the video content. For example, it can be used as input to a natural language generation (NLG) module to produce a coherent textual description of the video.

**Training and Fine-Tuning-** The LSTM network is typically trained on a labelled dataset where the ground truth (e.g., textual descriptions) is available for each video. During training, the network learns to optimize its internal parameters to minimize the difference between its predictions and the ground truth.

**Evaluation and Optimization**- Evaluate the performance of the LSTM-based model using metrics like accuracy, precision, recall, or F1 score for text summarization tasks. Fine-tune the model's architecture and hyperparameters as needed to improve its performance.

In summary, using LSTM cells in conjunction with temporal sequence data, you can effectively capture the chronological flow of events in a cricket sports video. This allows for the creation of detailed and contextually relevant text summaries of the video content.

**5.5. Textual Data Integration-** Combine the features extracted from the video frames (from Step 3) and the temporal context information (from Step 4) with the textual data (from Step 2). This integration helps establish connections between the visual content and the associated commentary or transcripts.

**5.6. Text Summarization**- Use the integrated data to generate text summaries of the cricket sports videos. This can be done by training a neural network, such as an LSTM-based sequence-to-sequence model or a Transformer-based architecture, to produce concise and coherent summaries based on the integrated features and context.

**5.7. Natural Language Generation (NLG)**- Employ NLG techniques to convert the summarized data into natural language text. This step involves generating sentences that effectively communicate the key highlights, events, and insights from the cricket video in a human-readable format.

**5.8. Fine-Tuning and Optimization**- Iterate on the model and system by fine-tuning the neural networks, adjusting hyperparameters, and optimizing the NLG process to improve the quality of the summaries.

This step-by-step process combines computer vision, natural language processing, and neural networks to automate the summarization of cricket sports video content, making it more accessible and engaging for viewers.

**6.Identify gap**

There are several studies that are trying to develop a summarization model which can generate meaningful and informative summaries. But from the summary of the literature and limitations mentioned in following drawbacks are observed in the development of efficient cricket match video summarization model.

**6.1. Lack of Training Data:**

The lack of extensive and varied cricket video datasets is a major research gap in "Cricket Sports Video to Text Summarization". These datasets are crucial for training and evaluating models that can effectively summarize various cricket formats and events.  While video summarization as a whole has advanced, datasets that are particular to cricket and cover different matches, events, and formats (including T20, One Day, and Test matches) are hard to come by. The creation and assessment of video summarization models specifically for cricket are constrained by this data gap, which compromises the accuracy and usefulness of these models.

**6.2. Gap in Cricket-specific Features:**

Overs, wickets, runs, boundaries, and player statistics are just a few of the distinctive characteristics that make cricket a special sport. Models that can recognize and highlight significant cricketing situations, such as wickets falling, boundaries being struck, milestones like hundreds, and game-changing deliveries, are needed for cricket video summary. These incidents are crucial to cricket matches and have a big effect on the result. Current video summarizing research frequently concentrates on general sports footage or underutilizes these cricket-specific features. The creation of video summarizing models that accurately identify and include details unique to cricket in the text summaries is lacking. Closing this gap would result in cricket video summaries that are more relevant and interesting.

**6.3. Gap in Multilingual Support:**

Cricket is a multilingual sport that is played all over the world and has a diversified fan base. Research on multi-language text summarizing for cricket footage is lacking. Summarization models must handle a variety of languages and dialects frequently utilized in cricket coverage in order to reach a wider audience. Cricket-related vocabulary and terminology can vary greatly between languages. Hindi cricket terminology, for instance, may be different from that of English or other languages. For text summarization algorithms to provide accurate and contextually appropriate text summaries, they must be skilled at comprehending these linguistic nuances. The necessity for multilingual cricket video summarizing methods is highlighted by this gap.

**7.Discussion**

This study sheds light on Cricket, a hugely popular sport worldwide with over 2.5 billion fans, can sometimes be challenging to follow due to its complexity and how it can be countered. Many fans struggle to keep up with all the action, especially if they can't watch the entire match. Enter Cricket Sports Video to Text Summarization, an exciting new technology. It helps fans stay updated on the game, even if they can't watch it in its entirety. This technology uses artificial intelligence (AI) to automatically create text summaries of cricket matches. These summaries provide quick access to match highlights and insights into specific aspects of the game. There are two main approaches to cricket sports video to text summarization. One uses Natural Language Processing (NLP) to identify key events in a match, like wickets, boundaries, and fours. Then, it generates a summary by describing these events. The other approach involves deep learning. It trains a model on a dataset of cricket videos and their summaries. This model can then produce summaries for new cricket videos. While still relatively new, cricket sports video to text summarization has the potential to transform how people engage with cricket. It offers a convenient way to access match highlights and in-depth information, helping fans stay closely connected to the sport they love.

**8.Conclusion**

In this paper we have provided a survey of relevant works of literature on "Cricket Sports Video to Text Summarization" using neural networks, examining language models, applications, datasets, and evaluation methods. We've revealed both progress and challenges. Despite technological strides, effectively summarizing dynamic cricket matches into coherent text remains a daunting task. Challenges include dialogue modelling, data availability, and cricket's inherent complexity. We've emphasized the need for more domain-specific data, real-time summarization solutions, and multimodal integration. The project's intersection of Deep Learning and Natural Language Processing presents a captivating challenge. Addressing data gaps, real-time summarization, multimodal integration, and evaluation methods will lead to significant advancements, enabling the creation of insightful cricket summaries enjoyed by enthusiasts worldwide.