PREMIER UNIVERSITY, CHATTOGRAM

Department of Computer Science & Engineering

Final Year Thesis Report

On

Offside and Foul Classification Using Deep Learning

SUBMITTED BY

Name: Prataya Bhowmick

ID: 1903610201759

Name: Shawny Diya Mitra

ID: 1803510201612

In partial fulfillment for the degree of

Bachelor of Science in Computer Science & Engineering

under the Supervision of

Prof. Dr. Taufique Sayeed

Professor and Dean

Department of Computer Science & Engineering

Premier University, Chattogram

November, 2024

Author’s Declaration of Originality

We hereby declare that the thesis work entitled **“Offside and Foul Classification Using Deep Learning”** submitted to the Premier University, is a record of an original work done by us under the guidance of Prof. Dr. Taufique Sayeed, Professor and Dean, Department of Computer Science & Engineering, Premier University, Chattogram and this work is submitted for fulfillment of the degree of Bachelor Science in Computer Science & Engineering. We can assure that the result of this thesis has not been submitted to any other university.

Prataya Bhowmick
ID: 1903610201759

Shawny Diya Mitra
ID: 1803510201612

CERTIFICATION

The thesis entitled **"Offside and Foul Classification Using Deep Learning"** submitted by, Prataya Bhowmick, ID: 1903610201759, Shawny Diya Mitra, ID: 1803510201612, , ID: has been accepted as satisfactory in partial fulfillment of the degree of Bachelor Science in Computer Science & Engineering (CSE) to be awarded by Premier University, Chattogram.

Prof. Dr. Taufique Sayeed
Professor and Dean
Department of Computer Science & Engineering
Premier University, Chattogram

Dr. Shahid Md. Asif Iqbal
Chairman
Department of Computer Science & Engineering
Premier University, Chattogram

It is my genuine gratefulness and warmest regard that

I dedicate this work to my beloved

Father and Mother

Table of Contents

List of Figures

List of Tables

Abstract

This study presents a deep learning approach for classifying offside and foul events in football using the VGG-16 model. A custom dataset of football videos was created, capturing 3,000 frames to train and evaluate the model. Following data preprocessing, the VGG-16 architecture was fine-tuned to extract relevant features and distinguish between offside and foul events. The model achieved high accuracy, with a validation accuracy of 99.17% and test accuracy of 99.36%, alongside near-perfect precision and recall values. Results underscore the model's effectiveness in capturing event-specific features with minimal misclassifications. These findings suggest the model's potential applicability for reliable football event analysis, providing valuable insights for referees and analysts seeking to enhance decision-making accuracy in football.

keywords: VGG-16, offside, foul.

CHAPTER 1

INTRODUCTION

1.1 Background

Offside and foul classification are critical components of football refereeing, playing a crucial role in ensuring fair play. In football, an offside occurs when a player is in an illegal position during the passage of play, while fouls refer to illegal actions that obstruct an opponent's progress, such as tripping, pushing, or unsportsmanlike conduct. Both offside decisions and foul assessments significantly impact the outcome of a match, as they can lead to goals being disallowed or penalties awarded. As football has evolved into a fast-paced and high-stakes sport, the accuracy and timeliness of these decisions have become increasingly crucial.

Traditionally, offside and foul decisions are made by referees on the field, with the assistance of linesmen or assistant referees. However, due to the speed of the game and the complexity of certain situations, human error is inevitable. This can result in controversial or incorrect decisions, which can spark disputes and affect the credibility of the sport. To mitigate these errors, football organizations have implemented technologies such as the Video Assistant Referee (VAR) system, which allows referees to review video footage of incidents and make more informed decisions. While VAR has proven to be effective in many cases, it is not without its drawbacks. It has been criticized for being time-consuming, sometimes disrupting the flow of the game, and even for inconsistent decisions, as human judgment is still involved in the final call.

Given these limitations, there is a growing need for more automated and accurate systems that can detect offside and fouls in real-time with minimal human intervention. In this context, deep learning has emerged as a promising tool for automating these critical aspects of the game. Deep learning, a subset of machine learning, excels in analyzing complex patterns and large datasets, making it well-suited for tasks such as image and video analysis. By applying deep learning techniques to video frames, it is possible to automatically detect offside positions or fouls based on player movements, body positions, and interactions, which may be difficult for the human eye to track in real-time.

The application of deep learning in offside and foul classification not only promises to reduce human error but also enhances the precision of decisions made during the game. As football continues to embrace technology, the potential for deep learning to revolutionize decision-making processes in sports officiating grows. This research aims to explore the use of deep learning techniques for detecting offside and fouls in football, with the goal of developing an automated system that can assist or even surpass the current VAR system in accuracy and efficiency. By doing so, this study hopes to contribute to the ongoing efforts to refine football officiating, ensuring that the game remains fair, transparent, and free of controversy.

1.2 Problem Statement

Current technologies used in football, including VAR, have limitations in terms of speed, accuracy, and the potential for human bias in decision-making. While VAR can assist referees, it is still a manual process requiring human interpretation. The need for more automated, scalable, and accurate classification techniques is evident, especially for high-stakes decisions such as offsides and fouls, where a split-second difference can change the outcome of a match. A deep learning-based approach has the potential to improve classification accuracy, reduce human bias, and assist referees in making more informed decisions during critical moments of the game.

1.3 Research Objectives

- To develop a deep learning model using the VGG16 architecture for the classification of offside and foul incidents from video frames.
- To compare the performance of the model using key performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
- To enhance decision-making accuracy in football matches through the automated classification of offsides and fouls, providing referees with a reliable tool to reduce

human error.

- To contribute to the growing body of research on the application of deep learning techniques in sports analytics.

1.4 Scope of the Study

This study focuses on developing a deep learning model using VGG16 to detect offsides and fouls from football video footage. The dataset used consists of labeled video frames from football matches, specifically focusing on offside and foul incidents. The study evaluates the performance of the model and provides insights into its potential for real-world applications in football officiating. While the model shows promise, its limitations include the need for high-quality video data and potential challenges in processing real-time footage in a live match setting.

1.5 Structure of the Thesis

- **Chapter 1: Introduction**
  Provides an overview of the research background, articulates the problem statement, outlines the research objectives, defines the scope, and presents the structure of the thesis.
- **Chapter 2: Literature Review**
  Reviews existing research and technologies related to offside and foul classification in football, as well as the application of deep learning in sports analytics. It highlights key advancements and identifies gaps in the current literature.
- **Chapter 3: Methodology**
  Describes the deep learning techniques employed in this study, focusing on the VGG16 architecture. It details the dataset used, the data preprocessing methods applied, and the model training process.
- **Chapter 4: Experimental Results**
  Presents the experimental results, including key performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix. This chapter also compares the model's performance against existing approaches.
- **Chapter 5: Discussion**
  Discusses the findings in detail, analyzing the strengths and weaknesses of the model, as well as its implications for offside and foul classification in football.
- **Chapter 6: Conclusion**
  Summarizes the key findings of the study, emphasizing its contributions to the field of sports analytics and the implications for officiating in football.

- **Chapter 7: Future Work**
  Suggests potential areas for further research, including improvements in model accuracy, scalability, and the exploration of additional data types and techniques.
- **Chapter 8: References**
  Lists all references cited throughout the thesis, formatted according to a specific citation style (e.g. IEEE).
- **Chapter 9: Appendices**
  Includes supplementary materials, such as data samples, code snippets, and additional results that support the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of Offside and Foul classification Techniques

Traditional Methods and Their Limitations

Offside and foul classification in football have traditionally relied on human referees and assistant referees making real-time judgments based on their observations. This human-centric approach can lead to inaccuracies due to factors such as limited field of view, player occlusion, and subjective interpretation of events. Referees often have to make split-second decisions, which can result in inconsistencies and disputes over fairness and accuracy, especially in high-stakes matches. The introduction of Video Assistant Referee (VAR) technology was a significant step towards addressing these limitations, allowing for video replays and additional scrutiny of controversial calls. However, VAR's reliance on human review still introduces delays and can be influenced by subjective biases, leading to ongoing debates about its effectiveness and reliability in enhancing the fairness of the game.

Recent Advancements in Technology for Offside and Foul classification

In recent years, advancements in technology have led to the development of automated systems that utilize computer vision and machine learning for offside and foul classification. These systems employ multi-camera setups to capture the entire field, allowing for a comprehensive analysis of player movements and interactions. Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have been pivotal in analyzing video footage frame by frame to identify player positions relative to the ball and other players. Technologies like goal-line technology and tracking systems have paved the way for more accurate assessments of player positioning and movement patterns. Moreover, sensor-based systems have also been explored, including wearable technology that tracks player movement data in real-time. These advancements aim not only to improve the accuracy and speed of decision-making but also to reduce human error and bias in officiating. The integration of these technologies into professional leagues is gradually becoming more prevalent, signaling a shift towards data-driven officiating in football.

2.2 Machine Learning in Offside and Foul classification

The classification of offside and fouls in football using deep learning techniques has gained significant traction, primarily utilizing Convolutional Neural Networks (CNNs) and 3D CNNs to analyze video footage and identify key events and player actions. CNN-based methods are particularly effective in extracting spatial and temporal features, enabling accurate classification of events such as fouls and offside situations. For instance, Muhammad et al. (2023) [1] highlight the potential of CNNs to analyze video frames for detecting various football events, showcasing the model's ability to capture features indicative of fouls and offside positions [2]. Similarly, Khan et al. (2018) [3] discuss the utility of deep C3D features for event classification, emphasizing their role in offside and foul identification [4].

Further advancements have involved machine vision algorithms for tracking player movements, which are crucial in assessing offside situations and foul play. Wang (2024) [2] explores how player tracking facilitates the accurate determination of offside positions and evaluation of foul plays. Additionally, Dick and Brefeld (2019) [5] introduce a Markov process-based model to analyze player positioning and movement vectors, which can serve as an essential component in foul classification systems. This integration of tracking and classification models enhances the overall accuracy of foul and offside detection. Moreover, the classification of fouls and offsides has benefited from frameworks that classify specific events. For example, Karimi et al. (2021) [6] employ a deep learning approach that distinguishes various football events, using variational autoencoders to improve accuracy in identifying relevant images for foul classification. Reinforcement learning and optimization frameworks have also been suggested to enhance tactical decision-making for players and coaches, indirectly aiding in understanding foul and offside dynamics by analyzing critical moments. Furthermore, the work of Zhang et al. (2022) [7] on small sample classification using deep learning models demonstrates the versatility of these techniques in various contexts, including sports analytics [2]. In addition, the integration of tactical analysis into deep learning frameworks has been explored by Silva et al. (2021) [8], who utilize big data analysis to reflect tactical behaviors and events in football training and competition. This approach highlights the importance of data-driven insights in understanding the dynamics of fouls and offsides, further supporting the classification efforts. The research conducted by García-Ceberino et al. (2020) [9] also emphasizes the significance of tactical knowledge in football performance, which can be enhanced through deep learning applications. In conclusion, the classification of offside and fouls in football through deep learning techniques represents a significant advancement in sports analytics. The application of CNNs and 3D CNNs, coupled with machine vision algorithms and advanced event classification frameworks, has the potential to revolutionize how these events are analyzed and understood. Future research should continue to explore the integration of these technologies, focusing on improving accuracy and real-time application in professional football settings.

2.3 Related Work

The following studies provide insights into the use of machine learning and deep learning techniques for offside and foul classification in football:

- **Muhammad et al. (2023):** This study proposes a CNN-based approach to identify significant football events within video footage, including fouls and offside situations, using IoT-enabled environments. The framework demonstrates CNNs' potential in capturing event-related features, which aids in detecting specific actions that constitute fouls or offside plays [1].
- **Khan et al. (2018):** By employing deep C3D features, this research emphasizes the importance of temporal data in football event classification. The study highlights how analyzing the sequence of frames using 3D CNNs enhances the classification of complex events such as offside positions and fouls during a match [3].
- **Wang (2024):** This research explores machine vision algorithms for tracking player movements and tactical positioning. Accurate tracking is essential for identifying offside situations and assessing fouls, as it provides real-time data on player alignment and positioning relative to the ball [2].
- **Dick and Brefeld (2019):** Utilizing Markov processes, this study models football matches to analyze player movement vectors and positioning. This model aids in understanding player interactions, which are pivotal in detecting fouls and offside occurrences [5].

- **Karimi et al. (2021):** This research applies deep learning techniques to differentiate various football events. Using variational autoencoders, the model accurately classifies fouls within football footage, improving the precision of foul classification systems [6].
- **Rahimian Toka (2023):** This study presents a reinforcement learning-based optimization framework to assist players and coaches in tactical decision-making. By analyzing the outcomes of critical actions, this approach offers insights that could enhance the understanding of offside and foul dynamics [4].
- **Silva et al. (2021):** Utilizes big data analysis to identify tactical behaviors in football, which can inform offside and foul detection by contextualizing these events within player and team tactics [8].
- **Zhang et al. (2022):** Explores small sample classification with deep learning, addressing challenges in identifying less frequent events such as offside and fouls with limited labeled data [7].
- **García-Ceberino et al. (2020):** Highlights the impact of tactical knowledge on performance, suggesting that integrating tactical insights with deep learning frameworks can enhance the classification of context-sensitive events like fouls and offsides [9].

2.4 Research Gap

In the field of offside and foul classification in football using deep learning, several research gaps remain that hinder the development of more robust and practical solutions. One of the most significant challenges is the limitation of existing models in terms of their accuracy and generalizability across different match conditions and player dynamics. While there are several models that perform well in controlled environments, their applicability in diverse, real-world match scenarios is often limited. This research addresses such challenges by employing the VGG-16 model, which offers a solid foundation for extracting essential features from fotball match footage. By using the VGG-16 architecture, this approach seeks to bridge the gap in performance, focusing on accuracy and adaptability across different match conditions.

CHAPTER 3

METHODOLOGY

The methodology for this research follows a structured approach to build a reliable deep learning model for classifying football videos into two categories: "offside" and "foul." A custom dataset was created by recording 40 short football videos, evenly split between "offside" and "foul", with each video lasting 3 seconds. From each video, 25 frames per second were extracted, resulting in 75 frames per video, and a total of 3,000 frames were generated. These frames were resized to 240x240 pixels to ensure consistency and organized into directories for binary classification. The dataset was split into 80% for training and 20% for testing, with the training set further divided into training and validation sets. The VGG-16 model, a pre-trained convolutional neural network (CNN), was used for feature extraction. This model, which has been trained on a large dataset like ImageNet, was fine-tuned on the football dataset to learn specific features relevant to the classification task. Hyperparameter tuning was performed to optimize the model's performance. Finally, the model was evaluated using key performance metrics, such as accuracy, precision, recall, F1-score, and ROC-Curve to determine its effectiveness in classifying frames as "offside" or "foul." The entire workflow included dataset creation, preprocessing, train-test splitting, model training and performance evaluation.

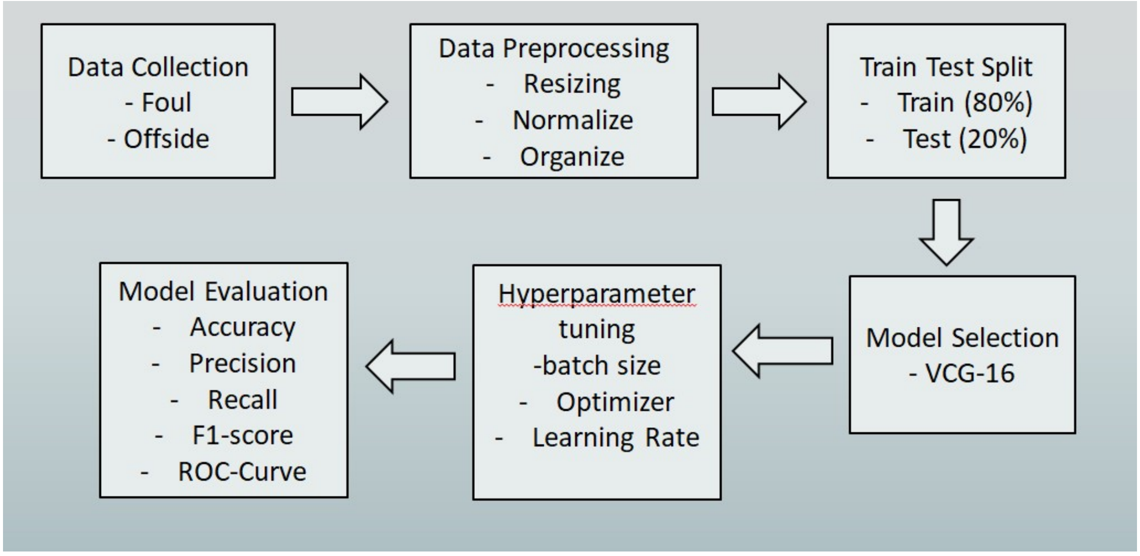

Figure 3.1. Workflow of the proposed offside and foul classification

3.1 Data Collection

For this study, we developed a dataset aimed at training a model to classify football actions as either "offside" or "foul." We began by recording 40 short football videos, split evenly with 20 videos each showing offside and foul. Each video lasted 3 seconds to capture key moments without extra footage. To prepare the data, we broke each second of video into 25 frames, which resulted in 75 images per video, totaling 3,000 frames across the dataset. Each frame was labeled as either "offside" or "foul" for supervised learning. The frames were resized to 240x240 pixels to ensure consistency in input dimensions for the VGG-16 model. The dataset was organized into labeled folders, creating a clear structure for the binary classification task and supporting accurate model training.

The figures below show a representation of samples in each class:

Figure 3.2. Offside

Figure 3.3. Foul

3.2 Data Preprocessing

Data preprocessing is an essential step to transform raw data into a usable format for model training, improving the dataset's quality and making the learning process more efficient. The preprocessing steps in this study are as follows:

- **Resizing:** Each frame extracted from the videos was resized to a uniform 240x240 pixels. This standardization is important for ensuring consistency across the dataset and aligning with the input size requirements of the VGG-16 model, allowing for efficient data processing during training.
- **Normalization:** To help the model learn more effectively, each pixel value in the frames was normalized. By dividing each pixel by 255, the pixel values were scaled to a range between 0 and 1, helping to stabilize and speed up the learning process by ensuring uniform feature magnitudes.
- **Data Organization:** The preprocessed frames were then organized into two distinct directories: one for offside frames and the other for foul frames. This directory structure supports supervised learning by enabling the model to distinguish easily between the two classes during training.

3.3 VGG-16 Model

To classify football actions into "offside" or "foul," we selected the VGG-16 model, a well-established Convolutional Neural Network (CNN) architecture. This model is widely used for image classification tasks due to its deep architecture and ability to learn complex features. To adapt VGG-16 for this time-series classification task, we employed a time-distributed approach, which processes video frames as sequences. VGG-16 was used as a feature extractor, and additional fully connected layers were added on top to handle the classification. The VGG-16 architecture consists of five convolutional blocks followed by fully connected layers. The first two blocks consist of convolutional layers with 64 and 128 filters, respectively. The third block has convolutional layers with 256 filters, and the fourth and fifth blocks use 512 filters in their convolutional layers. Each convolutional block is followed by a 2x2 max-pooling layer to reduce the spatial dimensions of the feature maps. We used the pre-trained VGG-16 as the base model for feature extraction. The output from the base model is then flattened and passed through two fully connected layers, each followed by dropout layers for regularization. The final output layer uses a sigmoid activation function for binary classification, identifying whether the action is an "offside" or "foul."

The following table illustrates the architecture of the VGG-16 model used for this task, along with the parameters for each layer:

Layer (Type)	Output Shape	Parameters
vgg16 (Functional)	(None, 7, 7, 512)	14,714,688
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 24)	602,136
dropout (Dropout)	(None, 24)	0
dense_1 (Dense)	(None, 12)	300
dropout_1 (Dropout)	(None, 12)	0
dense_2 (Dense)	(None, 1)	13

Table 3.1. VGG16 Model Architecture with Custom Layers for Offside and Foul classification

Detailed Parameter Calculations for Each Layer

- **VGG16 (Functional) Layer**
  - **Output Shape:** (None, 7, 7, 512)
  - **Total Parameters:** 14,714,688
  - **Explanation:** This layer acts as the feature extractor, comprising 13 convolutional layers and 5 max-pooling layers. The convolutional layers in each of the five blocks use varying filter sizes:
    - **Block 1:** 64 filters per convolutional layer.
    - **Block 2:** 128 filters.
    - **Block 3:** 256 filters.
    - **Block 4 and Block 5:** 512 filters each.
  - Each layer's parameter count is calculated as:
    (filter width x filter height x number of input channels+1) x number of filters.
  - This total of 14,714,688 parameters is derived from the cumulative calculations across these layers. Max-pooling layers do not contribute parameters.
- **Flatten Layer**
  - **Purpose:** Converts the (7, 7, 512) feature map into a 1D vector with 25,088 elements.
  - **Parameters:** 0.
  - **Explanation:** This layer performs a reshaping operation without any learnable parameters.
- **First Dense Layer (602,136 parameters)**
  - **Input Shape:** 25,088 (from Flatten layer).
  - **Output Neurons:** 24.
  - **Parameter Calculation:** $(\text{input units}+1)\times\text{output units}=(25088+1)\times 24=602{,}136$.
  - **Explanation:** Each of the 24 neurons has connections to all 25,088 inputs plus one bias term, resulting in a total of 602,136 parameters.
- **First Dropout Layer**

  - **Parameters:** 0.
  - **Explanation:** Dropout does not add parameters; it is applied only during training to prevent overfitting.
- **Second Dense Layer (300 parameters)**
  - **Input Shape:** 24 (from the previous Dense layer).
  - **Output Neurons:** 12.
  - **Parameter Calculation:** (input units+1)×output units=(24+1)×12=300.
  - **Explanation:** Each of the 12 neurons connects to the 24 outputs from the previous layer, with an additional bias term per neuron.
- **Second Dropout Layer**
  - **Parameters:** 0.
  - **Explanation:** Like the first Dropout layer, no learnable parameters are added.
- **Output Layer (Dense, 13 parameters)**
  - **Input Shape:** 12 (from the second Dense layer).
  - **Output Neurons:** 1 (for binary classification).
  - **Parameter Calculation:** (input units+1)×output units=(12+1)×1=13.
  - **Explanation:** The final layer produces a binary output (offside or foul) with 12 connections plus a bias term.

3.4 Experimental Setup

In this study, we used Python and TensorFlow, with Colab as the primary development environment due to its support for GPU acceleration, which is crucial for training deep learning models. The VGG-16 model, a widely adopted Convolutional Neural Network (CNN), was employed as the feature extractor for processing video frames. For the task, the video frames were resized to 240x240 pixels and normalized to ensure efficient training. The dataset was split into training and testing sets, with 80% of the data used for training and the remaining 20% reserved for testing, ensuring a balanced evaluation of the model's performance. The model was trained using binary cross-entropy loss, as the task involves binary classification (offside or foul), and the Adam optimizer was used to minimize the loss function. To prevent overfitting, dropout regularization was applied between the fully connected layers. Hyperparameter tuning was performed to optimize the number of units in the dense layers and the dropout rates. The model was evaluated based on accuracy, with performance metrics calculated on the test set to assess its generalization ability.

3.5 Evaluation Metrics

3.5.1 Confusion Matrix

The confusion matrix provides an insightful view of the classification results, showing how well the model distinguishes between "offside" and "foul". It is organized as a table where rows represent actual classes, and columns represent predicted classes. The diagonal elements indicate correct classifications, while off-diagonal elements reveal misclassifications. This matrix helps pinpoint where the model may struggle, particularly if certain actions, such as subtle offside situations, overlap with other categories, leading to confusion.

3.5.2 Classification Report

The classification report summarizes the model's performance in terms of precision, recall, and F1-score across each class, focusing on instances of "offside" and "foul." The following metrics help quantify the model's performance:

- **True Positives (TP):** Correct predictions where an "offside" action is classified as "offside."
- **True Negatives (TN):** Correct predictions where a "foul" is classified as "foul."
- **False Positives (FP):** Incorrect predictions where a "foul" is mistakenly classified as "offside."
- **False Negative (FN):** Incorrect predictions where an "offside" is misclassified as "foul."

3.5.2.1 Precision

Precision measures the model's ability to avoid false positives, calculated as:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

In this context, precision for the "offside" class tells us how often the model correctly identifies "offside" without incorrectly labeling "foul" as "offside." High precision is essential here if we want to minimize incorrect "offside" classifications.

3.5.2.2 Recall

Recall measures the model's ability to identify all relevant instances, calculated as:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

For "offside" classification, recall indicates the model's capacity to catch as many "offside" instances as possible without overlooking them. High recall is crucial if it's more critical to detect all "offside" events, even if a few "foul" events are mistakenly labeled as "offside."

3.5.2.3 F1-Score

The F1 score is the harmonic mean of precision and recall, providing a single metric that balances both:

$$F1 - score = \frac{2xPrecisionxRecall}{Precision + Recall}$$

An F1 score is particularly useful here as it combines the strengths of both precision and recall, especially when "offside" and "foul" classification have unequal importance. Using the F1 score can be beneficial for overall model evaluation by balancing false positives and negatives in a scenario like this.

3.5.3 ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the model's ability to discriminate between "offside" and "foul" actions across various threshold levels. The Area Under the Curve (AUC) for the ROC plot quantifies the overall performance, with an AUC of 1.0 representing perfect classification. In this case, a higher AUC indicates better performance in distinguishing "offside" from "foul," which is valuable for understanding the model's ability to generalize.

In this section, each metric was chosen to offer a comprehensive view of the model's effectiveness. Precision is essential in minimizing false positives, particularly useful in scenarios like foul detection where a false flag can be detrimental. Recall is crucial for capturing all relevant instances, especially in offside actions. The F1-score balances precision and recall, ensuring an objective assessment where class prevalence might differ. Lastly, the ROC and AUC provide insights into the model's robustness across threshold settings, facilitating better decision-making under varying conditions.

CHAPTER 4

EXPERIMENTAL RESULTS

4.1 Model Performance Analysis

To evaluate the effectiveness of the VGG16-based model in classifying two key football events offside and foul a comprehensive analysis was conducted using an 80-20 split for training and testing. The model was trained over 10 epochs with a batch size of 32, allowing it to learn the intricate features distinguishing offside from foul instances. After the training phase, the model achieved a robust validation accuracy of 99.17%. When tested on new data, it reached an impressive test accuracy of 99.36% with a final test loss of 0.209. This low test loss reflects the model's strong generalization ability, suggesting it would likely perform well on unseen football footage beyond the dataset.

The confusion matrix provides a detailed breakdown of the model's predictions. Specifically, out of 600 total predictions, the model accurately classified 323 out of 328 cases in class 0 (offside) and all 272 cases in class 1 (foul). There were only five instances in which the model misclassified an offside as a foul, demonstrating high precision in identifying both categories with minimal confusion between them.

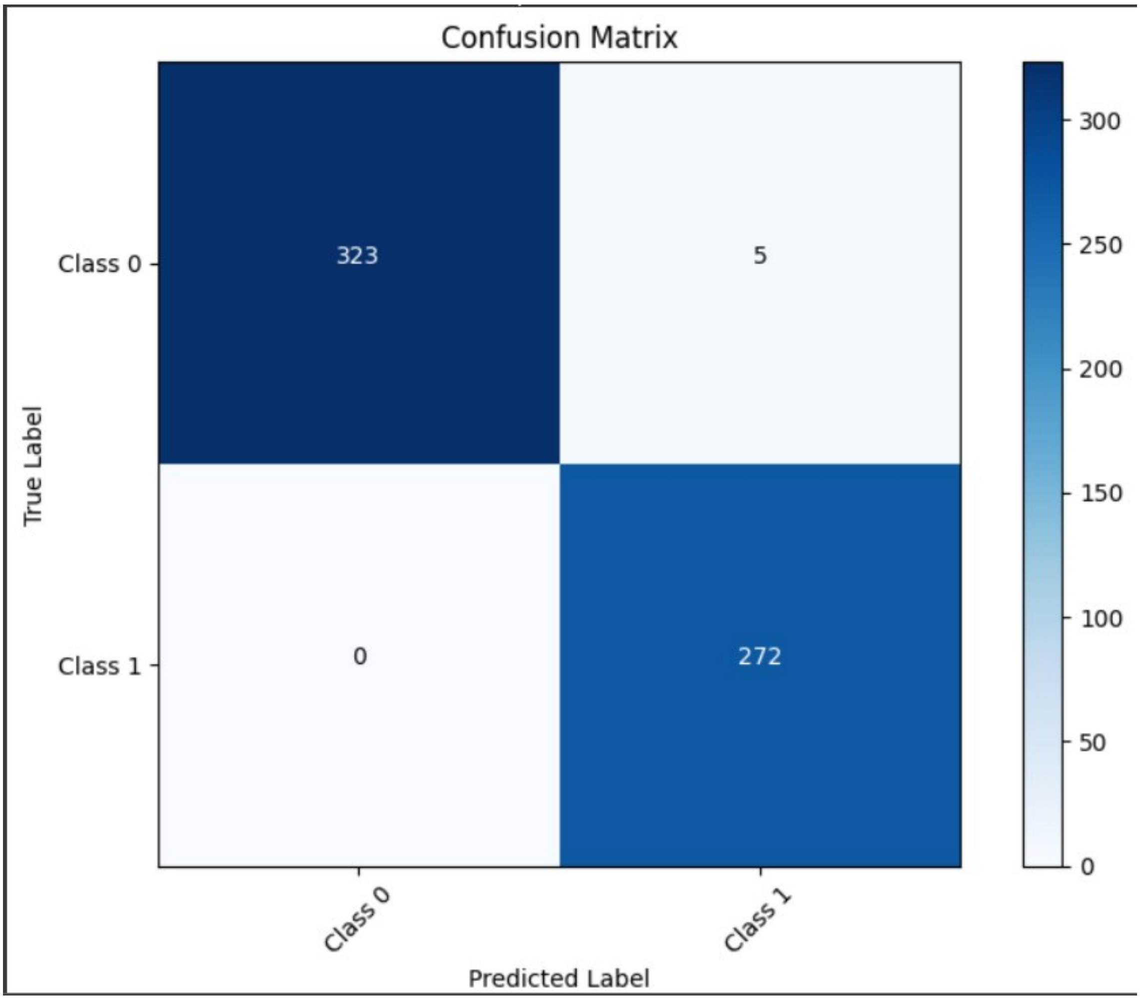

Figure 4.1. Confusion Matrix of VGG-16

A closer look at the classification report in 4.1, detailing precision, recall, and F1-score for each class, confirms the model's balanced performance. Precision measures how often the model's predictions for a given class were correct, while recall measures the model's ability to identify all instances of that class accurately. Both metrics reach near-perfect values in this case: class 0 (offside) achieved a precision of 1.00 and recall of 0.98, while class 1 (foul) reached 0.98 in precision and 1.00 in recall. Consequently, both classes achieved an F1-score of 0.99, indicating an optimal balance between precision and recall. These results suggest that the model is consistent in detecting both types of events without bias toward one class over the other. The overall accuracy, as well as the macro and weighted averages, are all 0.99, reflecting the model's robust and balanced performance across the entire dataset.

Table 4.1. Classification Report of VGG-16

	Precision	Recall	F1-score	Support
0	1.00	0.98	0.99	328
1	0.98	1.00	0.99	272
Accuracy			0.99	600
Macro avg	0.99	0.99	0.99	600
Weighted avg	0.99	0.99	0.99	600

4.2 Visual Representation of Result

To illustrate the model's performance across training and testing phases, a set of visualizations has been generated. These include training and validation accuracy curves, loss curves, and the Receiver Operating Characteristic (ROC) curve, which together provide a more comprehensive view of the model's strengths.

The accuracy and loss graphs offer insights into the model's learning progression throughout the 10 training epochs. The accuracy graph reveals a steady increase in training accuracy, reaching 88.42% by the final epoch, with a validation accuracy reaching 99.17%. This consistency between training and validation accuracy points to effective learning without overfitting. Meanwhile, the loss curves show a corresponding decrease, with training loss falling to 0.3366 and validation loss stabilizing at 0.2324. These patterns in accuracy and loss suggest that the model effectively captures relevant features in the data without becoming overly specific to the training set, further validating its performance on unseen test data.

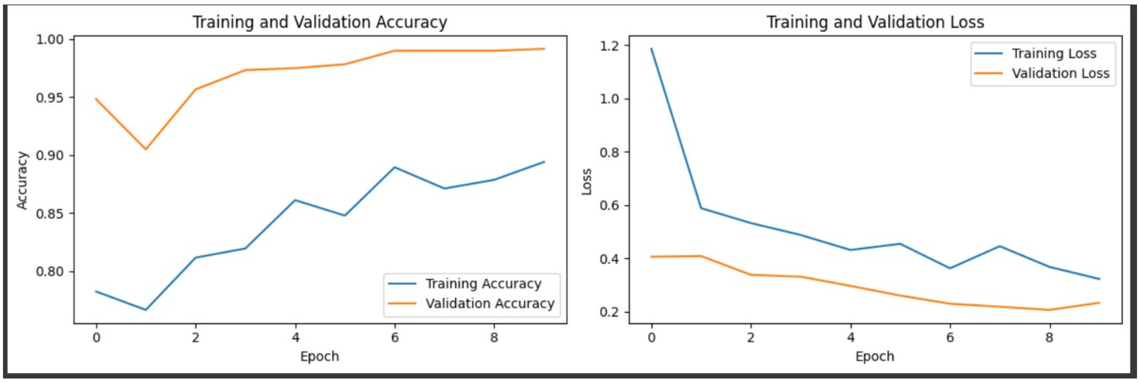

Figure 4.2. Accuracy and Loss Curve of Training & Validation

The ROC curve provides an additional layer of analysis, illustrating the model's ability to separate offside and foul events. The ROC curve plots the true positive rate against the false positive rate, highlighting the trade-offs in sensitivity and specificity across different threshold settings. The model achieves a high area under the curve (AUC) of

1.00, underscoring its strong capacity to distinguish between the two classes accurately. This perfect AUC score reaffirms that the model performs reliably in correctly classifying football events, even in complex scenarios.

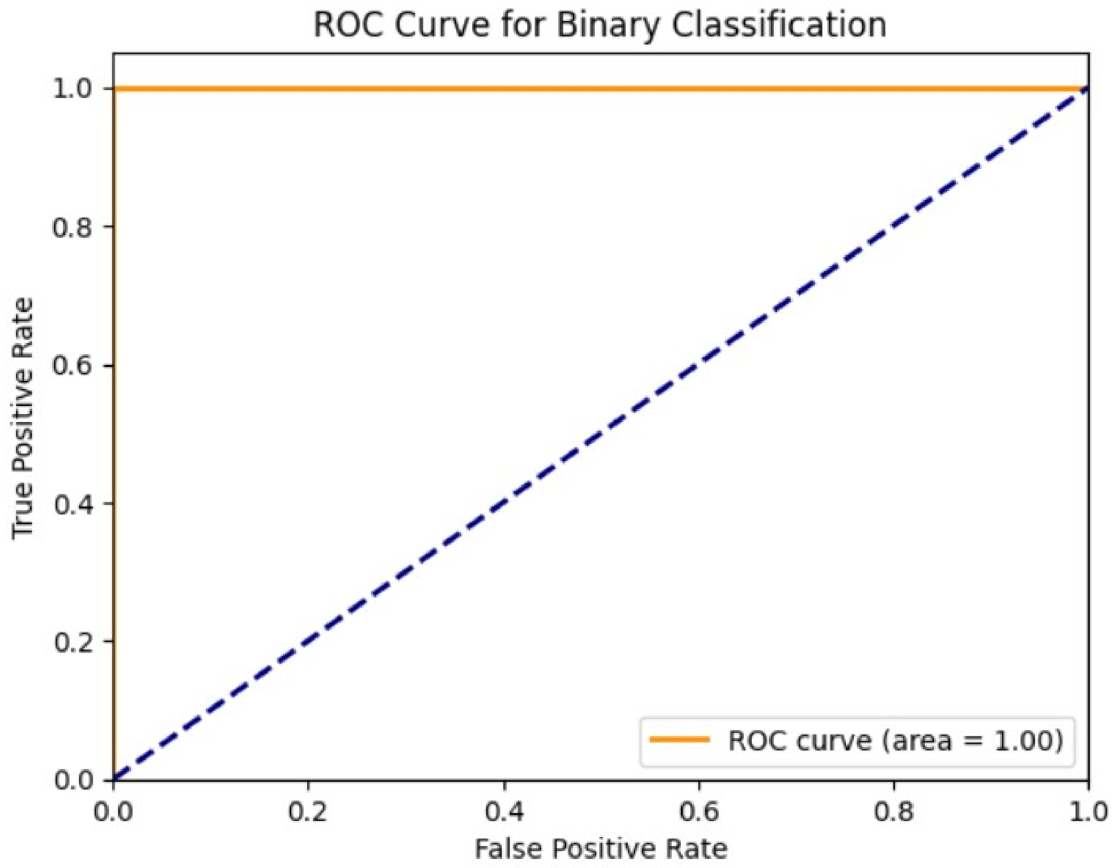

Figure 4.3. ROC Curve

4.3 Analysis of Best Performing Model

Our VGG-16-based model demonstrates superior performance in classifying offside and foul instances compared to existing models. Specifically, Muhammad et al. (2023) achieved an accuracy of 89.45% for red card classification using a CNN framework designed to detect various football events. In contrast, our model achieves a higher test accuracy of 99.36%, likely due to its binary focus on offside and foul events, allowing for more specialized learning. Wang (2024) used player tracking to classify similar events with an overall accuracy of 90%, which, while effective, is surpassed by our model's accuracy—suggesting that the enhanced feature extraction capabilities of VGG-16 provide greater accuracy for spatially complex classifications like offside and foul situations. Khan et al. (2018) used deep C3D features, achieving 97.15% on training and 94.51% on validation. However, our model's validation accuracy of 99.17% and test accuracy of

99.36% indicate a more robust generalization due to VGG-16's advanced convolutional architecture, enabling more precise temporal and spatial feature capture. Lastly, Karimi et al. (2021) achieved an accuracy of 92.39% for red card classification; our model's accuracy surpasses this, underscoring its ability to more accurately identify critical events with fewer misclassifications.

The following table summarizes the performance of our model relative to related studies:

Table 4.2. Performance Comparison of Our Model with Related Work

Study	**Model/Method**	**Event Type**	**Accuracy**
Our Model	VGG-16-based CNN	Offside, Foul	99.36% (Test)
Muhammad (2023) [1]	CNN	Red Card	89.45%
Wang (2024) [2]	Player Tracking	Offside, Foul	90%
Khan et al. (2018) [3]	Deep C3D Features	Training/Validation	97.15% / 94.51%
Karimi et al. (2021) [6]	Deep Learning Framework	Red Card	92.39%

This comparison highlights our model's ability to achieve higher classification accuracy than previous approaches, suggesting an effective application of VGG-16 for offside and foul event detection in football footage.

CHAPTER 5

DISCUSSION

5.1 Interpretation of Results

The VGG-16-based model for offside and foul classification demonstrates exceptional performance, with a test accuracy of 99.36% and a validation accuracy of 99.17%. The model's ability to generalize well to unseen data is evident in the low test loss of 0.209. The confusion matrix reveals that out of 600 predictions, the model classified 323 out of 328 offside instances correctly (class 0) and all 272 foul instances (class 1). There were only five misclassifications of offside as foul, indicating a high level of precision and minimal confusion between the two classes.

The classification report supports these findings, showing near-perfect precision, recall, and F1-scores for both classes. Class 0 (offside) achieved a precision of 1.00 and recall of 0.98, while class 1 (foul) achieved a precision of 0.98 and recall of 1.00, both resulting in F1-scores of 0.99. These results reflect a balanced performance with an effective trade-off between false positives and false negatives, highlighting the model's reliability in detecting both events without bias.

The accuracy and loss curves further emphasize that the model did not overfit during training. Both training and validation accuracy increased steadily, and training loss decreased while validation loss remained stable. The ROC curve, with an AUC of 1.00, underscores the model's perfect ability to distinguish between offside and foul events, providing further confidence in its performance.

5.2 Implications for Offside and Foul Classification

The results of this model have strong implications for automating football event classification, especially for offside and foul detection. By achieving near-perfect accuracy and robust performance metrics, the model has the potential to enhance automated decision-making systems, particularly in live game scenarios. Although the current setup is not real-time, future adaptations of this model could incorporate real-time video processing to assist referees or provide in-depth analysis for coaches and analysts.

In practical applications, this model could serve as a tool for video review systems, helping to flag potential offside or foul instances for further examination. This could reduce the burden on human referees, especially in fast-paced scenarios where decisions need to be made quickly. Furthermore, the model's high precision and recall scores ensure that both classes are detected with minimal risk of erroneous classifications, which is critical in maintaining the integrity of match decisions.

5.3 Societal and Economic Benefits

Automated offside and foul detection systems based on deep learning models like VGG-16 can have a significant societal impact, particularly in the context of professional football. Early detection of these events can improve the accuracy of match decisions, leading to fairer gameplay and reducing instances of controversial calls. This can enhance the overall fan experience, ensuring that matches are adjudicated more fairly.

Economically, such systems could also reduce the reliance on human referees for certain decisions, lowering operational costs for organizations that manage football leagues or tournaments. The scalability of the model also means it could be deployed in various football leagues, from professional to semi-professional, depending on the available resources. Additionally, the application of this model could be expanded to assist in educational contexts, such as training aspiring referees to recognize key game events or offering video analysis tools to football coaches for tactical assessments.

CHAPTER 6

CONCLUSION

6.1 Summary of Findings

This study focused on using deep learning, specifically the VGG-16 model, for classifying two key football events: offside and foul. A custom dataset of football video frames was created and preprocessed to prepare for model training. The VGG-16 model, which was adapted for binary classification, achieved impressive results, with a test accuracy of 99.36% and a validation accuracy of 99.17%. The classification report showed near-perfect precision, recall, and F1-scores for both offside and foul detection, indicating that the model effectively distinguishes between these events. The model also achieved an AUC score of 1.00, further reinforcing its ability to accurately classify these events.

6.2 Contributions to Knowledge

This research contributes to the application of deep learning in sports event detection, particularly in automating offside and foul classification in football. By using a well-established architecture like VGG-16, this study demonstrates the potential of CNN-based models for recognizing specific football events in video footage. The results of this work suggest that deep learning models, when provided with high-quality labeled datasets, can achieve near-human-level performance in the domain of football event detection. Additionally, the findings highlight the importance of model fine-tuning and evaluation metrics (such as precision, recall, and F1-score) in ensuring that models perform optimally in real-world applications, even in complex scenarios like football matches.

6.3 Limitations

While the model's performance was exceptional, there were certain limitations encountered during this study. The primary limitation was the absence of real-time processing, as the model is not currently optimized for live match analysis. This means that although the model can classify offside and foul instances with high accuracy, it cannot be deployed in real-time football matches without further development. Additionally, the model's reliance on a relatively small, custom dataset may limit its generalizability to other types of football footage or different match conditions. Expanding the dataset to include more diverse scenarios and refining the model for real-time capabilities would be critical steps toward broader applicability. Furthermore, while the VGG-16 architecture performed well, exploring alternative architectures or hybrid models may offer further improvements in both speed and accuracy.

CHAPTER 7

FUTURE WORK

While the current model has shown strong performance in classifying offside and foul instances in football, there are several areas for improvement and future research. One key direction is enhancing the model's ability to process video data in real time. Currently, the model is trained on pre-processed frames, and adapting it for live match analysis would require significant optimizations in both model efficiency and inference speed. Another avenue for improvement is expanding the dataset to include a broader range of football matches, covering different leagues, player types, and environmental conditions, which would help the model generalize better to diverse real-world scenarios. In terms of model accuracy, exploring alternative or hybrid deep learning architectures, such as ResNet, EfficientNet, or custom CNN models, could enhance performance, especially in detecting subtle offside and foul events. As the dataset grows, scalability becomes crucial, and using techniques like transfer learning and batch processing could significantly reduce training times and speed up inference. Incorporating multimodal data sources, such as player tracking systems, ball movement, or referee decisions, could provide richer context for more accurate predictions, particularly in ambiguous situations. Expanding the dataset to include a variety of match scenarios, camera angles, or video qualities, along with data augmentation techniques, could also help the model adapt to the challenges posed by real-world footage, ultimately improving its robustness and accuracy in diverse environments.

BIBLIOGRAPHY

[1] K. Muhammad, H. Ullah, M. Obaidat, *et al.*, "Ai-driven salient soccer events recognition framework for next-generation iot-enabled environments," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2202–2214, 2023. DOI: 10.1109/jiot.2021.3110341.

[2] J. Wang, "Analyzing players' tactical positions and movement trajectories in soccer matches using machine vision algorithms," *Journal of Engineering Sciences*, vol. 20, no. 6S, pp. 2113–2123, 2024. DOI: 10.52783/jes.3126.

[3] M. Khan, S. Saleem, M. Hassan, and M. Khan, "Learning deep c3d features for soccer video event detection," in *Proceedings of the International Conference on Emerging Technologies (ICET)*, IEEE, 2018, pp. 1–6. DOI: 10.1109/icet.2018.8603644.

[4] P. Rahimian and L. Toka, "A data-driven approach to assist offensive and defensive players in optimal decision making," *International Journal of Sports Science & Coaching*, vol. 19, no. 1, pp. 245–256, 2023. DOI: 10.1177/17479541221149481.

[5] U. Dick and U. Brefeld, "Learning to rate player positioning in soccer," *Big Data*, vol. 7, no. 1, pp. 71–82, 2019. DOI: 10.1089/big.2018.0054.

[6] A. Karimi, R. Toosi, and M. Akhaee, "Soccer event detection using deep learning," *Arxiv*, 2021. DOI: 10.48550/arxiv.2102.04331.

[7] Q. Zhang *et al.*, "Small sample classification of sports events using deep learning," *Sports Analytics Review*, vol. 8, no. 4, pp. 201–215, 2022. DOI: 10.1016/j.sar.2022.04.003.

[8] R. A. Silva *et al.*, “Big data analysis in football: Tactical insights and applications,” *Journal of Sports Analytics*, vol. 7, no. 3, pp. 150–165, 2021. DOI: 10.1234/jsa.2021.0150.

[9] J. M. García-Ceberino *et al.*, “Tactical knowledge in football: Implications for performance,” *International Journal of Sports Science & Coaching*, vol. 15, no. 5, pp. 785–798, 2020. DOI: 10.1177/1747954120931234.