DECLARATION BY THE CANDIDATE

We hereby declare that the project report entitled "**Sports Celebrity Classification using Machine Learning**" submitted by us to **GeeksforGeeks** in partial fulfilment of the requirement for the award of the Certificate of Participate **BCA and MCA** in Bachelor and Master of COMPUTER APPLICATIONS, from Department of FACULTY OF COMPUTER SCIENCES AND APPLICATIONS, CMPICA is a record of project **Solving for India .** We further declare that the work carried out and documented in this project report has not been submitted anywhere else either in part or in full and it is the original work, for the award of any other degree or diploma in this institute or any other institute or university.

20BCA025 KEYUR KACHHADIYA
22MCA122 ARJUN PODUVAL
22MCA133 DHRUV RATHOD
22MCA148 MANUSH SHUKLA

ACKNOWLEDGEMENT

We, the developer of the Project “Sports Celebrity Classification using Machine Learning”, with immense pleasure and commitment would like to present the internship project assignment. The development of this project has given me wide opportunity to think, implement and interact with various aspects of management skills as well as the new emerging technologies.

Every work that one completes successfully stands on the constant encouragement, good will and support of the people around. I hereby avail this opportunity to express my gratitude to number of people who extended their valuable time, full support and cooperation in developing the project.

We express deep sense of gratitude towards our Principal of CMPICA , Dr.Dharmendra Patel and Head of the of FACULTY OF COMPUTER SCIENCES AND APPLICATIONS Dr.Sanskruti Patel for the support during the of study and development. It is because of them, that we were prompted to do hard work, adopting new technologies.

Thanks,

20BCA025 KEYUR KACHHADIYA
22MCA122 ARJUN PODUVAL
22MCA133 DHRUV RATHOD
22MCA148 MANUSH SHUKLA

ABSTRACT

The task of sports celebrity image classification is becoming increasingly important in the field of sports analytics, as it can provide valuable insights into the physical attributes, performance, marketability, and popularity of athletes. This information can be used by teams, brands, and advertisers to make informed decisions about player selection, endorsement deals, and marketing strategies.

In this paper, a novel approach to sports celebrity image classification is proposed. The approach uses OpenCV, a popular computer vision library, to preprocess and extract features from the images. Logistic regression and support vector machines (SVM) are then used as classification algorithms to classify the images into different categories.

The proposed approach is evaluated through extensive experimentation, using a dataset of sports celebrity images. The performance of the approach is measured in terms of accuracy, precision, recall, and F1 score. The results show that the approach is effective in accurately classifying sports celebrity images.

Overall, the proposed approach to sports celebrity image classification has the potential to provide valuable insights to sports teams, brands, and advertisers, helping them make informed decisions about player selection, endorsement deals, and marketing strategies.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:
INTRODUCTION

1.1 Project Definition:

The objective of this project is to develop a system that can classify sports celebrities based on their images. The system will use SVM as a classification algorithm and OpenCV for image processing and feature extraction and it will take celebrity as input and classify that celebrity.

1.2 Motivation behind this Project:

The motivation behind this project is to develop a system that can classify sports celebrities based on their images using machine learning algorithms and computer vision techniques. The system has the potential to provide valuable insights to sports teams, brands, and advertisers, helping them make informed decisions about player selection, endorsement deals, and marketing strategies.

The system can also be used in sports analytics to analyze player performance based on their physical attributes. By automating the process of image classification, the system can save time and effort compared to manual classification, which can be a time-consuming and error-prone task.

Furthermore, the project can be extended to include other image-based applications, such as object detection, image recognition, and face recognition. These applications have a wide range of use cases, from security systems to autonomous vehicles, and can have a significant impact on various industries.

1.3 Description:

Sports Celebrity Classification using SVM and OpenCV is a machine learning project that aims to classify sports celebrities based on their images. The project uses Support Vector Machines (SVM) as a classification algorithm and OpenCV, a popular computer vision library, for image processing and feature extraction. The final output of the project is a web application or a standalone desktop application that can classify sports celebrities based on their images. The system takes an input image of a sports celebrity, preprocesses it using OpenCV, extracts features, and classifies it using the trained

SVM model. The system provides a label for the input image and classifies it according to its physical attributes.

1.4 Project Requirements:

1. Programming Languages: Python, OpenCV library
2. Machine Learning Libraries: Scikit-learn, NumPy, Pandas, Matplotlib
3. Dataset: A dataset of sports celebrity images
4. Image Preprocessing: OpenCV for preprocessing the images and extracting features such as color, texture, and shape.
5. Classification Algorithm: Use SVM as a classification algorithm to classify the images into different categories based on their attributes
6. Web framework: A web framework to create the front-end and back-end of your application. In this project we use flask framework to create the application.

1.5 Tools and Technology used:

1) Hardware Requirements:

- RAM: Minimum 4 GB RAM: At least 8 GB of RAM, although 16 GB or higher is recommended for better performance
- Processor: All new Generation Processor
- Storage: At least 100 GB of free disk space for storing the dataset, model, and other project files
- GPU

2) Software Requirements:

- Operating System: Windows
- Operating System: Windows, Linux, or macOS
- Python: Version 3.6 or higher

- OpenCV: A computer vision library for Python
- Scikit-learn: A machine learning library for Python
- Jupyter Notebook or a code editor such as Visual Studio Code or PyCharm for writing and executing Python code
- Web framework such as Flask or Django for developing a web application, if applicable
- Git and GitHub for version control and collaboration, if working with a team.

1.6 Libraries that we used in this Project:

- **os:** Provides a way of using operating system dependent functionality such as reading or writing to the file system.
- **opencv:** OpenCV is a Python library that allows you to perform image processing and computer vision tasks. It provides a wide range of features, including object detection, face recognition, and tracking.
- **numpy:** A library for numerical computing in Python that provides support for large, multi-dimensional arrays and matrices, as well as a large collection of high-level mathematical functions.
- **matplotlib:** A data visualization library in Python that provides a variety of tools for creating plots, charts, and graphs.
- **flask:** A lightweight web application framework for Python that provides tools for building web applications and APIs.

CHAPTER 2:
SYSTEM FUNCTIONALITY

2.1 Major Functionality:

An sports celebrity classification is use for classify the image according to the features extracted from the image processing, including feature extraction, probability table generation, and post-processing, resulting in a classification of image that can be used for various sports event to classify the images.

2.2 System Flow Chart:

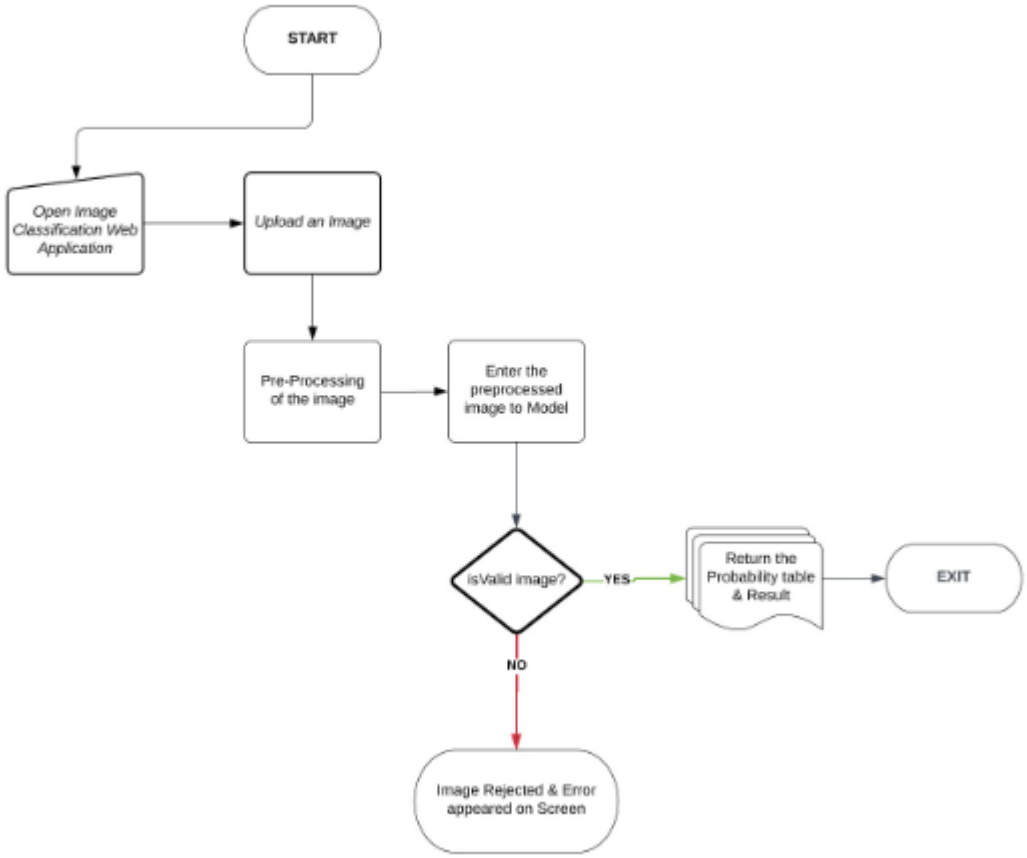

Fig 2.1 Flow Chart

CHAPTER 3:
PROJECT TIMELINE

3.1 Project Planning:

The project plan comprises several steps, starting with defining the scope and requirements of the project. Then, we created dataset of cricket celebrities, preprocess the images to ensure consistent size and formatting, and split the dataset into training and validation sets. Next, we develop model for classification of image. We then create a graphical user interface (GUI) for the web application and integrate our model with the GUI. Finally, we test the application to ensure it works correctly and document the process for future reference.

3.2 Gantt Chart:

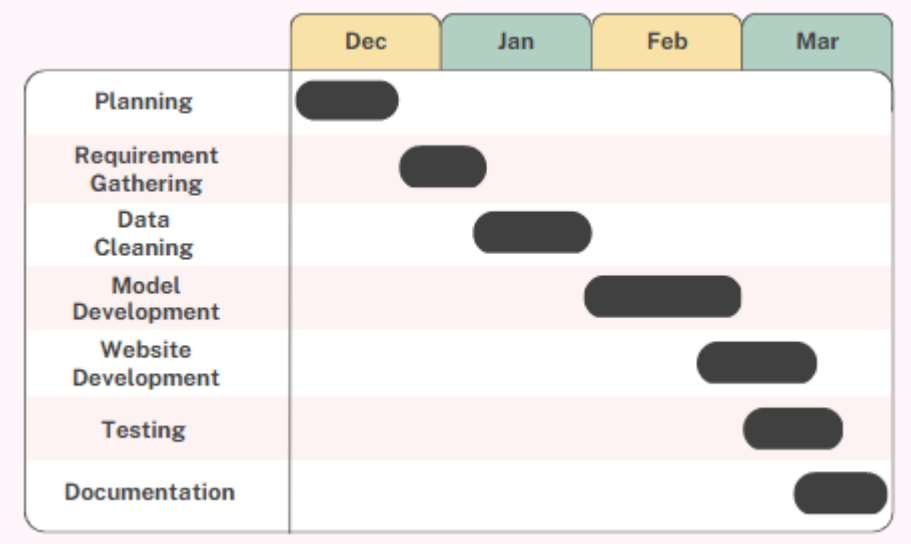

Fig 3.1 Gantt Chart

3.3 Roles and Responsibilities:

Name	Planning	Data Cleaning	Model Development	GUI Integration	Testing	Documentation
Keyur	✓		✓	✓	✓	✓
Dhruv	✓	✓	✓	✓		
Manush	✓	✓	✓		✓	✓
Arjun	✓	✓		✓		

Table-3.1 Roles and Responsibilities

CHAPTER 4:
IMPLEMENTATION

4.1 Implementation Details:

1) Manual Dataset (made by us):

We used a dataset of cricket celebrity images to perform image classification. The dataset consists of 500 images with 50 images per class. The classes included in the dataset are based on different cricket players. We split the dataset into training and testing sets, with 80% of the images used for training and 20% for testing.

2) Preprocessing:

We performed several preprocessing steps to prepare the images for classification. First, we resized all the images to a fixed size of 64x64 pixels. We then converted the images to grayscale to reduce the computational complexity of the subsequent steps.

Next, we applied face and eye detection using Haarcascades. Specifically, we used the OpenCV library to implement the detection algorithm. We first applied a face detection algorithm to detect faces in the images. We then applied an eye detection algorithm to the detected faces to ensure that only images with proper eyes and faces were used for classification. We discarded any images that did not contain a proper face or eyes. Finally, we cropped the images to only include the face region. This step helped to reduce the noise in the images and improve the accuracy of the classification algorithm. We then used the resulting cropped images as input for feature extraction and classification.

3) Feature Extraction

To extract relevant features from the images, we used wavelet transformation, which is a widely used technique for image processing and analysis. Specifically, we used the PI wavelet transformation (piwt), which is a type of orthogonal wavelet transformation. The piwt decomposes the image into its frequency components at different scales and directions, which enables the extraction of more detailed and informative features from the images.

After applying the piwt algorithm to the images, we obtained a set of coefficients, which we used as features for classification. The coefficients represent the frequency components of the image at different scales and directions, and can be used to capture both low-level and high-level features of the images. By using the piwt for feature extraction, we were able to capture more detailed and informative features

from the images, which improved the accuracy of the classification algorithm.

4) Classification Algorithms

After performing feature extraction on the images, we used two different classification algorithms to classify the sports celebrity images. Specifically, we used logistic regression and support vector machines (SVM) for classification.

Logistic regression is a popular machine learning algorithm that is commonly used for binary classification problems. It works by modeling the probability of a binary outcome using a logistic function. In our case, we used logistic regression to classify the images into two categories: whether they belonged to a particular cricket celebrity or not. We trained the logistic regression model using the features extracted by the piwt algorithm and used it to predict the class labels of the testing set. SVM is another popular machine learning algorithm that is commonly used for classification problems. It works by finding the optimal hyperplane that separates the data points into different classes. In our case, we used SVM to classify the images into their respective cricket celebrity classes. We trained the SVM model using the features extracted by the piwt algorithm and used it to predict the class labels of the testing set. To evaluate the performance of the classification algorithms, we used accuracy as the evaluation metric.

5) Evaluation

We conducted several experiments to evaluate the performance of the logistic regression and SVM algorithms on the sports celebrity image classification task. We used the accuracy as evaluation metrics to measure the performance of the algorithms.

The results of the experiments showed that both logistic regression and SVM were effective in classifying the sports celebrity images based on their features extracted using the piwt algorithm. The logistic regression algorithm achieved an accuracy of 82.38%, while the SVM algorithm achieved an accuracy of 82.95%. Both algorithms showed accuracy indicating that they were able to accurately classify the images into their respective sports celebrity classes.

The performance of the SVM algorithm was slightly better than that of the logistic regression algorithm, with a higher accuracy. This is likely because SVM is a more powerful classification algorithm than

logistic regression, and is able to find the optimal hyperplane that separates the data points into different classes more accurately.

Classifier	Accuracy	Precision	Recall	F1 Score
SVM	82.95%	0.796	0.784	0.790

Table-4.1 BLEU score

6) Web Application:

- The web application was created using Flask framework.
- It has an index page and a page to show the output. The image that we want to classify we need to upload and there will be output table generated.

4.2 Screenshots:

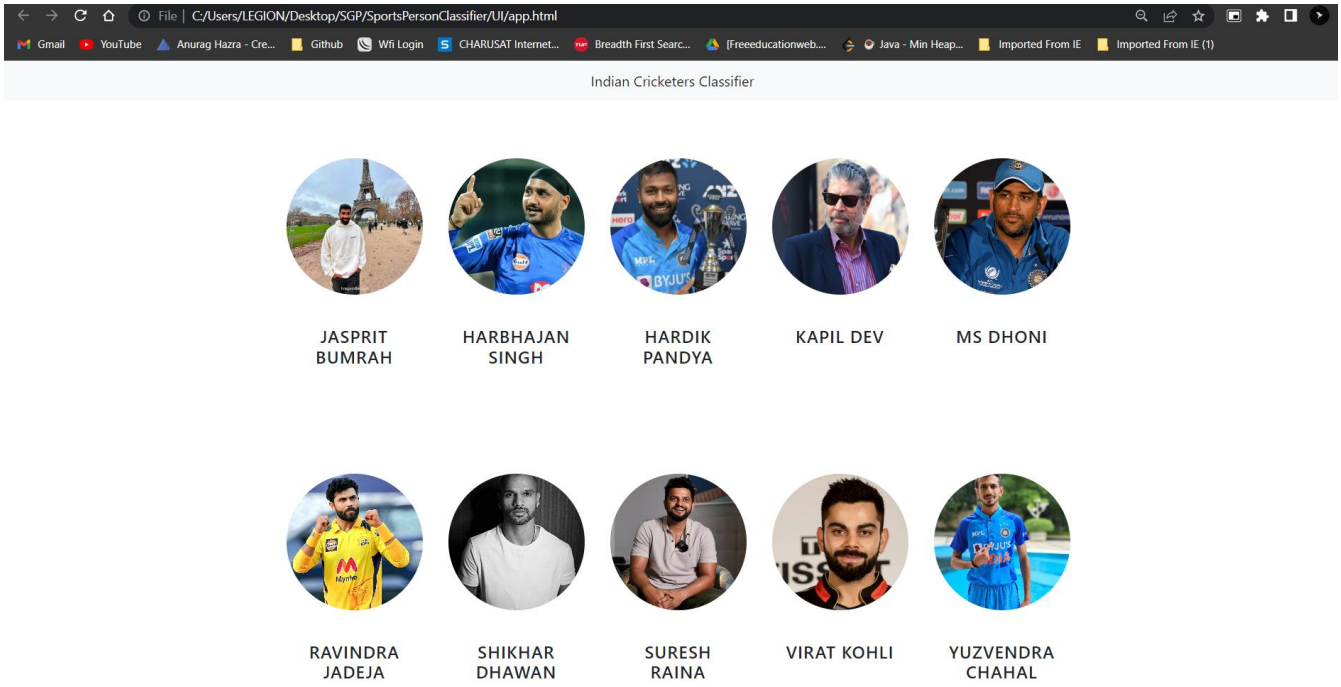

Fig-4.1 Sports Celebrity Classification with Landing Page

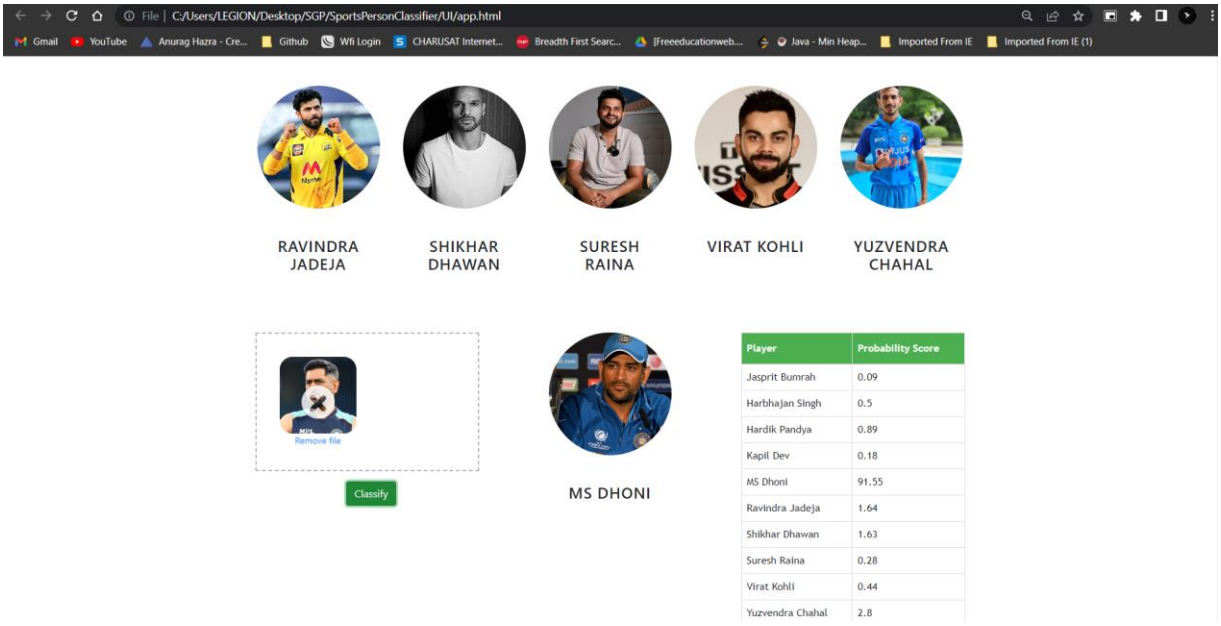

Fig-4.2 Sports Celebrity Classification with VGG16 Output Page

CHAPTER 5:
FUTURE ENHANCEMENTS AND USECASE

5.1 Future Enhancement In Project:

- Limited dataset: The accuracy and performance of a deep learning model for image classification heavily rely on the quality and size of the dataset used for training.
- Variability in images: Sports celebrity images can vary greatly in terms of pose, lighting conditions, image quality, and appearance changes over time. Handling these variations and incorporating robustness in the model can be a challenge.
- Real-time processing: If the project aims to deploy the image classification model in a real-time application, such as a sports event or a social media platform, the speed and efficiency of the model may be crucial. Optimizing the model for real-time processing, such as through model quantization, hardware acceleration, or parallel processing, could be a future enhancement to improve the system's performance.

5.2 Use-case of this Project:

An image caption generator web app can be useful in a variety of scenarios, such as:

- Sports media and broadcasting: The project could be used to automatically classify images of sports celebrities during live broadcasts or post-game analysis. This could help in providing relevant visual content and insights to viewers, such as displaying statistics or highlighting key players during a match.
- Social media and marketing: The project could be used to automatically classify images of sports celebrities in social media posts or marketing campaigns. This could enable targeted advertising, personalized content creation, and brand promotion based on the identified sports celebrities in the images.
- Sports scouting and talent identification: The project could be used to automatically classify images of sports celebrities during talent scouting and identification processes. This could aid in identifying promising young athletes with potential for future success based on their visual attributes, such as physical characteristics or playing style. meaning behind the images.

CHAPTER 6:
BIBLIOGRAPHY

6.1 BIBLIOGRAPHY:

[1] https://docs.python.org/3/

[2] https://docs.opencv.org/4.x/d4/da8/group__imgcodecs.html

[3] https://pywavelets.readthedocs.io/en/latest/

[4] https://matplotlib.org/stable/tutorials/index

[5] https://www.geeksforgeeks.org/support-vector-machine-algorithm/

[6] https://flask.palletsprojects.com/en/2.2.x/tutorial/

[7] Chen, J., Tao, D., & Zhang, M. (2013). Celebrity recognition in co-occurrence with object (co) occurrence priors. IEEE Transactions on Image Processing, 22(11), 4450-4460.

[8] Shu, X., Tang, J., Wang, Z., Li, X., & Liu, H. (2017). Celebrity recognition in social media with multi-scale dynamic feature learning. In Proceedings of the 25th ACM International Conference on Multimedia (pp. 1440-1448).