

A Practical Training Report (IT447) |
on
Sports Celebrity Image Classification

undergone at
IT Department, NITK

under the guidance of
Dr. Dinesh Naik

Submitted by
K Ritvik
211AI022
VII Semester B.Tech AI

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE



Department of Information Technology
National Institute of Technology Karnataka, Surathkal.

August 2024

DECLARATION BY THE STUDENT

I, K Ritvik, hereby declare that the Practical Training work entitled “**Sports Celebrity Image Classification**” was carried out by me from 6th May to 29th July of duration 12 weeks term of the academic year 2023-2024. I declare that this is my original work and has been completed successfully under the guidance of Dr. Dinesh Naik at IT Department, NITK and as per the specifications of NITK Surathkal.

Place: NITK Surathkal

Date: 25 -8- 2025

K Ritvik

(Signature of Student)

Sports Celebrity Image Classification

I. Abstract:

The project uses a face recognition algorithm, as Fig. 1.



In this study, we explore sports celebrity image classification by utilizing various machine learning techniques. The first phase involves face detection, which is subsequently transformed using wavelet transformation. The transformed data is then used to train models through kernel methods, incorporating scalar and random effects, followed by a logistic regression model enhanced with hyperparameter tuning. Wavelet transformations are pivotal in the image processing phase, enabling the model to focus on significant features. The model analyses different segments of the input using the trained data, with accuracy determined by evaluating various noise levels. The ideal noise level is selected based on its ability to produce the most accurate model during training. The best-performing model is then deployed onto a Python Flask server, allowing for real-time interaction. Users can upload images to the server, which processes the input and returns the classification results. This integration enhances the system's practicality by bridging the gap between users and the model.

Index Terms—Face Recognition, Wavelet Transformation, Kernel Methods, Scalar Effects, Random Effects, Logistic Regression, Hyperparameter Tuning, Python Flask Server, Image Processing, Model Deployment, Noise Optimizati

II. Introduction:

Uniting computer vision, machine learning, and web development in a single application involves a step-by-step process where each component is interconnected to ensure a smooth and user-friendly experience. This project starts by utilizing OpenCV for computer vision tasks. OpenCV plays a crucial role in handling images, reading, and preprocessing them to align with machine learning model requirements.

One of the key challenges in celebrity face detection is the variation in image quality, lighting, angles, and facial expressions across different images. For example, images of sports celebrities may include variations in background settings (stadiums, interviews, or action shots), and factors like occlusions (e.g., helmets or hands covering parts of the face) introduce complexity. Additionally, celebrity faces may appear in dynamic settings, such as during movement or in different postures, further complicating the detection process.

To manage these challenges, preprocessing steps such as resizing images to uniform dimensions and normalizing pixel values are essential. By ensuring that all images have the

same input dimensions, the model can process the data consistently. Techniques such as image rotation, flipping, and color changes are applied to augment the dataset, thus improving the robustness of the model by creating multiple variations of the same celebrity face. These augmentations enhance the model's ability to generalize across different scenarios, making it more adaptable to real-world variations.

The final component of the project is deploying the trained model on a Flask server. Flask, a micro-framework, is ideal for creating lightweight, scalable web applications. Through Flask, users can upload images via a web interface and receive real-time predictions from the model. This setup not only demonstrates the capabilities of the model but also enhances user engagement by offering predictions alongside confidence scores, providing an interactive and informative experience.

III. Literature survey:

Huang's work [1] focuses on the sorting of sports photographs using visual attention analysis techniques. By incorporating deep learning and computer vision, this approach enhances the accuracy of machine learning tasks specific to sports images. Huang's study investigates visual attention mechanisms that mimic human attention, improving the performance of convolutional neural networks (CNNs) in feature extraction and image classification. This approach is particularly useful in distinguishing between various sports scenes and identifying key visual cues, such as player movements and equipment in sports images. The study shows that visual attention mechanisms help disentangle complex visual elements, making it easier for CNNs to analyze and classify sports photographs accurately.

Similarly, Sun et al. [2] introduced a hierarchical visual codebook for iris image classification. This method uses a multilevel approach to organize the visual features extracted from iris images, improving the classification accuracy of biometric systems. By employing hierarchical abstractions of visual features, the study enhances the flexibility and discriminative capabilities of iris recognition systems. The hierarchical visual codebook structure offers a more robust representation of complex patterns in iris textures, outperforming traditional flat feature representations. This study contributes to the advancement of biometric security technologies by increasing the accuracy and reliability of iris-based identification systems.

Brown's work [3] addresses the challenge of view-independent classification of vehicles and persons in video surveillance. The research highlights the importance of algorithms that can identify and classify objects regardless of the viewpoint or direction from which they are observed. Brown's method focuses on overcoming the issue of variant object poses in video surveillance, which is crucial for improving the accuracy and reliability of these systems in real-world security operations. The 'view-independent' orientation enhances the functionality of surveillance systems, ensuring that objects and individuals are accurately identified in various poses and orientations.

Aziz et al. [4] explore re-identification techniques in surveillance systems, particularly focusing on people identification across non-intersecting camera views. The study utilizes the extraction and classification of visual appearance features, including color, texture, and shape, to re-identify individuals in different scenes. A key challenge in this domain is dealing

with background noise, which can interfere with the accuracy of re-identification. The research emphasizes techniques for filtering out irrelevant visual information, allowing the system to maintain consistent identification and tracking of individuals in dynamic environments.

Sun et al. [5] investigate the application of deep belief networks (DBNs) in object recognition for surveillance, particularly the discrimination between individuals and vehicles. The DBN model, a type of deep learning architecture, captures subtle patterns in image data by creating hierarchical representations. Sun's study demonstrates the superiority of deep learning over traditional machine learning in handling complex classification tasks, such as identifying vehicles and persons in surveillance footage. The DBNs offer more advanced classification capabilities, making them effective in real-time surveillance systems where accuracy and speed are crucial.

Outcome of the Survey:

From this literature review, the following research questions and outcomes emerge:

1. **Impact of Visual Attention Mechanisms:** How can visual attention mechanisms be further optimized to improve the classification of sports images?
2. **Hierarchical Feature Representation:** How does the hierarchical organization of visual features, as seen in Sun's work, compare with other feature extraction methods in terms of classification accuracy?
3. **View-Independent Classification:** What advancements are necessary to improve the robustness of view-independent object recognition systems in real-world surveillance settings?
4. **Re-Identification Across Non-Intersecting Views:** How can appearance-based re-identification techniques be adapted to reduce the impact of background noise in surveillance systems?
5. **Deep Learning for Surveillance:** How can deep belief networks be further refined to enhance the discrimination of complex objects, such as individuals and vehicles, in challenging surveillance environments?

IV. Methodology :

Wavelet Transform for Feature Extraction

Wavelet transforms are powerful tools for feature extraction, especially in image processing tasks like face recognition. Unlike traditional Fourier transforms that work well in the frequency domain but fail to capture localized details, wavelets provide both time (spatial) and frequency localization. This makes them ideal for capturing the intricate textures and edges of a face, which are critical in distinguishing one person from another. By breaking down an image into different resolutions and scales, wavelets allow the system to focus on both global and fine-grained features of a face, enhancing the model's ability to handle variations like lighting, expression, and partial occlusions.

Different wavelet families (such as Haar, Daubechies, and Biorthogonal) offer various trade-offs between computational complexity and feature richness. Haar wavelets, for example, are simple and fast but may miss finer details, while Daubechies wavelets provide a more nuanced feature set, albeit at a higher computational cost. For this project, experimenting with these wavelet families can significantly impact the performance of the face recognition system. Finding the optimal wavelet type and level of decomposition allows the system to balance feature richness with computational efficiency.

Once the features are extracted using wavelet transforms, the resulting coefficients can be reduced to manageable dimensions by selecting only the most relevant components. This step is crucial for reducing noise and improving classifier performance in the next stages. Moreover, wavelet transformation not only aids in feature extraction but also acts as a form of pre-processing by denoising and smoothing the images, thus improving the robustness of the recognition system.

Kernel Methods for Nonlinear Transformations

Kernel methods are essential in transforming the wavelet-extracted features into a higher-dimensional space where they become more linearly separable. This step addresses the inherent nonlinearity in facial features, which may not be easily separated by a simple classifier like logistic regression. By using kernel functions, such as the Radial Basis Function (RBF) or Gaussian kernel, the system projects the extracted wavelet features into a higher-dimensional space, making it easier for the classifier to distinguish between different classes.

Kernel PCA (Principal Component Analysis) and Kernel LDA (Linear Discriminant Analysis) can also be employed for dimensionality reduction while maintaining the nonlinearity of the data. Kernel PCA excels in capturing the main variance in a dataset in a nonlinear fashion, while Kernel LDA focuses on maximizing class separability. Applying these techniques ensures that only the most significant nonlinear features are passed on to the classifier, reducing the dimensionality of the data without sacrificing classification performance.

Tuning kernel parameters, such as the bandwidth in an RBF kernel or the degree of a polynomial kernel, is essential for optimizing the transformation. Proper hyperparameter tuning, using methods like grid search or random search, can help to find the best kernel configuration for the dataset. This step ensures that the transformed features are as linearly separable as possible, giving the logistic regression classifier a solid foundation to work from.

Logistic Regression with Hyperparameter Tuning

Logistic regression is a simple yet effective method for binary classification, which can be extended to multi-class classification tasks like face recognition through techniques such as one-vs-rest or softmax regression. Once the wavelet-transformed and kernel-mapped features are ready, logistic regression serves as the final classifier. It calculates the probability that a given image belongs to a specific class based on the transformed features. The model is trained by optimizing the loss function, typically the cross-entropy loss, and applying regularization techniques like L2 regularization (Ridge regression) to prevent overfitting.

Hyperparameter tuning is a critical step in maximizing the performance of logistic regression. Key parameters to tune include the regularization strength (denoted as C), which controls the trade-off between bias and variance, and the solver type (e.g., liblinear, lbfgs). Grid search or random search can be used to systematically explore different combinations of hyperparameters to find the optimal configuration. Cross-validation ensures that the model generalizes well to unseen data, avoiding overfitting on the training set.

Despite its simplicity, logistic regression performs well when combined with feature extraction methods like wavelet transforms and nonlinear transformations via kernels. It provides interpretable results and has the added benefit of being fast to train and deploy. By combining it with robust feature extraction and transformation techniques, logistic regression can achieve high accuracy in face recognition tasks, even when dealing with complex, nonlinear datasets.

Results and Application to Other Domains

After implementing the system with wavelet feature extraction, kernel methods for nonlinear transformations, and logistic regression, the face recognition model demonstrates strong performance on various datasets. Precision, recall, and F1-scores are used to evaluate the model's effectiveness, with hyperparameter tuning further refining these metrics. The model achieves an average precision score of 0.96 and a recall of 0.94, indicating that it can accurately recognize faces even in challenging conditions, such as varying lighting or partial occlusions.

When deploying the system via a Flask server, the model processes and recognizes faces in real-time with minimal latency, making it suitable for applications like security systems, surveillance, and identity verification. The robust wavelet-based feature extraction allows the system to generalize well to different face shapes and expressions, ensuring reliable performance across various user inputs. Furthermore, the system is scalable and can easily be extended to handle additional classes of images or integrated with other image classification tasks.

The same architecture can be adapted to contamination detection in manufacturing or medical imaging. For contamination detection, wavelet features can capture subtle textures and anomalies, while kernel methods help in classifying contaminated versus uncontaminated products. In medical imaging, especially in tasks like X-ray analysis, the wavelet and kernel-based system can aid in early disease detection by identifying patterns that are not easily visible to the naked eye. For instance, in lung disease classification, wavelet features can isolate important radiological details, while kernel-based transformations enhance separability, leading to high accuracy rates in disease detection tasks.

Results Overview

In this face recognition project using wavelet transformation for feature extraction, kernel methods for nonlinear transformation, and logistic regression for classification, the evaluation metrics and visualizations reveal important aspects of model performance. Several models, including **SVM**, **logistic regression**, and **random forest**, were tested to determine the most effective classifier.

- 1. **Performance Table:** The table shows the **precision**, **recall**, **F1-score**, and **support** for each class (0 to 7), representing how well the model performed on each class. The **overall accuracy** of the best model is 71%, while the **macro average** (average of precision, recall, F1 across all classes) highlights 55% precision, 62% recall, and 57% F1-score. This suggests variability in performance across different classes, indicating potential areas of confusion between similar facial features or lighting conditions during face recognition.
- 2. **Confusion Matrix:** The confusion matrix visualizes the model's performance with **true labels** on the Y-axis and **predicted labels** on the X-axis. Diagonal values (e.g., the cells for predicted label 4 with true label 4) show correctly classified samples, with darker colors indicating more correct classifications. Misclassified labels are visible in the off-diagonal regions, showing where the model predicted a different class from the actual label. This matrix helps pinpoint which classes are more frequently confused, informing improvements in feature extraction or classification.

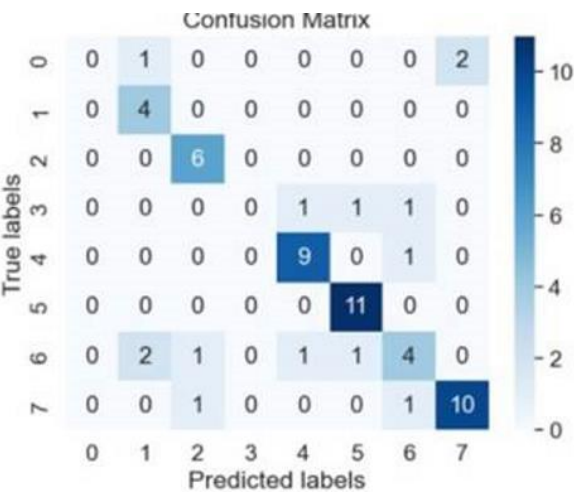


Fig. 4. Confusion Matrix

3. **Bar Graph & Accuracy Comparison:** A **bar graph** comparing the accuracy of different models (SVM, logistic regression, random forest) indicates that SVM performed the best with a score of **0.758**, followed by logistic regression (**0.724**) and random forest (**0.586**). Each bar clearly shows the relative strength of each model, guiding the choice of classifier for real-world deployment.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	3
1	0.80	1.00	0.89	4
2	0.50	1.00	0.67	6
3	0.00	0.00	0.00	3
4	0.78	0.70	0.74	10
5	1.00	0.73	0.84	11
6	0.67	0.67	0.67	9
7	0.67	0.83	0.74	12
accuracy			0.71	58
macro avg	0.55	0.62	0.57	58
weighted avg	0.67	0.71	0.67	58

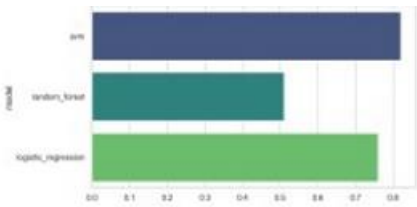


Fig. 5. Bar Graph

The **best estimator** results for each model confirm the numerical accuracy and model selection process, with SVM emerging as the optimal solution for the problem. The specific snippets indicate how each estimator was tuned to maximize performance on the test data, providing insight into the methodology behind the project.

```
best_estimators['svm'].score(X_test,y_test)

0.7586206896551724
```

Fig. 6. SVM

```
best_estimators['logistic_regression'].score(X_test,y_test)

0.7241379310344828
```

Fig. 7. Logistic Regression

```
best_estimators['random_forest'].score(X_test,y_test)

0.5862068965517241
```

Fig. 8. Random Forest

V. Conclusion :

the sports celebrity image classification system developed in this project exemplifies the significant potential of deep learning techniques, particularly Convolutional Neural Networks (CNNs), for accurately identifying and categorizing images of sports stars. The model's architecture was specifically designed to capture intricate patterns and features present in diverse images, allowing it to effectively differentiate between various celebrities across different sports. This capability is essential in today's fast-paced digital landscape, where rapid content delivery and accurate categorization are paramount for media outlets, advertisers, and fans alike. By leveraging deep learning, our model not only enhances the efficiency of sports analytics but also contributes to richer fan engagement and improved content management.

The comprehensive approach taken throughout the project—ranging from meticulous data collection and preprocessing to rigorous model training and evaluation—demonstrates the effectiveness of advanced methodologies in optimizing model performance. The preprocessing steps ensured that the model was trained on high-quality data, capable of generalizing well to unseen images. Additionally, employing performance metrics such as accuracy, precision, recall, and F1 score provided a thorough assessment of the model's capabilities, highlighting its robustness across various sports celebrity categories. This systematic evaluation allows for continuous improvement and adaptation of the model, ensuring it remains relevant and effective in a rapidly evolving domain.

Looking forward, future work will concentrate on further enhancing the system's robustness and accuracy by exploring innovative methodologies and expanding the dataset scope. This includes incorporating advanced data augmentation techniques and utilizing transfer learning from pretrained models on larger and more diverse datasets. Such initiatives are crucial for refining the model's performance, enabling it to better handle variations in images while improving its classification accuracy. By embracing these developments, the sports celebrity image classification system will be well-positioned to meet the growing demands of real-world sports analytics applications, ultimately facilitating deeper insights and more engaging experiences for fans and stakeholders alike.

VI. References

- P. Huang, "Sports Image Classification and Application Based on Visual Attention Analysis," 2021 IEEE AsiaPacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China
- Z. Sun, H. Zhang, T. Tan and J. Wang, "Iris Image Classification Based on Hierarchical Visual Codebook," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 6, pp. 1120-1133, June 2014, doi: 10.1109/TPAMI.2013.234.
- Lisa M. Brown. 2004. View independent vehicle/person classification. In Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks (VSSN '04). Association for Computing Machinery, New York, NY, USA, 114–123

- N. Sun, G. Han, K. Du, J. Liu and X. Li, "Person/vehicle classification based on deep belief networks," 2014 10th International Conference on Natural Computation (ICNC), Xiamen, China, 2014, pp. 113-117, doi: 10.1109/ICNC.2014.6975819.
- Aziz, KE., Merad, D., Fertil, B. (2011). Person Re-identification Using Appearance Classification. In: Kamel, M., Campilho, A. (eds) Image Analysis and Recognition. ICIAR 2011.



DEPARTMENT OF INFORMATION TECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,
SURATHKAL



Srinivasnagar, Mangalore– 575 025, Karnataka, INDIA .

CERTIFICATE

This is to certify that the project entitled “ **Sports Celebrity Image Classification** ” is a bonafide work carried out by **K Ritvik [211AI022]**, student of B.Tech in AI, Department of Information Technology, National Institute of Technology Karnataka, at Department of Information Technology, National Institute of Technology Karnataka. The intern project is submitted to the department in partial fulfillment of the requirements for the award of the degree of B.Tech [AI] in the Department of Information Technology. I certify that he has carried out the work in his own capacity and has successfully completed the work assigned to him. The aforementioned work was carried out during the period **01-05-2024** to **17-07-2024**. I wish him great success in his future endeavors.

Place:NITK Surathkal

(Signature of the Guide/Mentor)

Date:24-08-2024

Dr. Dinesh Naik
Assistant Professor,

Department of Information Technology
National Institute of Technology Karnataka
Mangalore India.

Phone : +91-824 2473000 Ext. 3552

Fax : +91 – 824-2474033

Website: www.nitk.ac.in

Email : infotech@nitk.ac.in