Enhanced Sports Image Classification using Deep CNN Models

P.R. Krishna Prasad
Assoc. Prof., Dept. of CSE,
Vasireddy Venkatadri Inst. of Tech., Guntur, India

Harshitha Myneni
Dept. of CSE,
Vasireddy Venkatadri Inst. of Tech., Guntur, India

S.B.S. Sameer Kumar Metra
Dept. of CSE,
Vasireddy Venkatadri Inst. of Tech., Guntur, India

Balaji Nelakurthi
Dept. of CSE,
Vasireddy Venkatadri Inst. of Tech., Guntur, India

Narasimha Naik Meghavath
Dept. of CSE,
Vasireddy Venkatadri Inst. of Tech., Guntur, India

ABSTRACT
Sports image classification is a crucial task in computer vision, facilitating applications such as automated sports analytics and event recognition. This study evaluates the performance of three deep learning models—VGG16, ResNet50, and EfficientNetB0—on a sports image classification dataset.The models were trained and tested using a dataset of sports images, and their performance was assessed based on accuracy, precision, recall, and F1-score. Experimental results indicate that Efficient- NetB0 outperformed the other models, achieving the highest accuracy of 96.6%,precision of 97.35%, recall of 96.6%, and an F1-score of 96.52%. These findings suggest that EfficientNetB0is well-suited for sports image classification, offering a balance of high accuracy and computational efficiency. Its superior performance highlights its potential for real-world applications in sports technology, where fast and accurate classification is essential.

Keyword
Sports image classification, Deep learning, Computer vision, Convolutional Neural Networks (CNNs), VGG16, ResNet50, EfficientNetB0.

1. INTRODUCTION
The rapid growth of multimedia content, such as images and videos, in mass media has created a strong need for efficient methods to process and organize this information. This is particularly important in fields like sports, where large volumes of content are generated by broadcasters due to high audience interest and commercial value. Therefore, effective classification and indexing techniques are essential for managing sports-related multimedia content.

Classifying sports images is a challenging task because sports scenes are highly dynamic and complex. Each sport involves distinct actions, player movements, gestures, and playing environments, which must be carefully analyzed to achieve accurate classification[1].In the early days, Convolutional Neural Networks (CNNs) were widely used for image classification. However, with advancements in deep learning, Deep Neural Networks (DNNs) have emerged as a more powerful approach, offering better accuracy and efficiency.

These deep learning models have significantly improved various fields, including speech recognition, robotics, computer vision, and natural language processing.

Fig.1.Sports images of diverse categories

A key factor in training deep learning models for image classification is the availability of large, diverse datasets. ImageNet, one of the most widely used image classification datasets, contains over 1,000 different image categories. Instead of training models from scratch, transfer learning is widely used to improve classification performance. Transfer learning involves using pre trained models, such as ResNet50, VGG16, and EfficientNetB0, which have already learned useful features from a large dataset like ImageNet. These models are then fine-tuned on a domain-specific dataset, allowing them to adapt to sports image classification with less data and computational effort while still achieving high accuracy [2],[3].These pre trained models serve as a strong foundation for sports image classification, enabling them to identify and differentiate between various sports categories effectively.

2. RELATED WORKS

Over the past decade, research in computer vision has seen tremendous growth, with numerous advancements in its applications. Various classifiers such as Neural Networks and Support Vector Machines (SVM) have been employed due to their effective analysis and performance [4], [5].

In recent work, Convolutional Neural Networks (CNNs) have been used to process data, with deep feature extraction from images helping to reduce complexity and redundancy. Pavel Zhdanov et al. [6] proposed a model for improving human action recognition through hierarchical Neural Networks. In this study, actions from the image dataset were recognized using CNNs, which significantly enhanced recognition quality for activities like sports. Features can be extracted from text, audio, and visual information [7]. In this work, the focus is given exclusively on visual information, as vision is the primary means through which humans perceive and interpret information [8].

Deep learning methods can automatically extract complex features from images. Many researchers have explored recent deep learning techniques, such as CNNs and Recurrent Neural Networks (RNNs), to capture both spatial and temporal features [9]. Several studies have focused on sports image classification using CNNs. For instance, ResNet50 and VGG16 were used in one study to detect players automatically from sports images, achieving an accuracy of 96% on basketball datasets. However, more improvements can still be made, including obtaining more varied image data and using other CNN models to increase accuracy.

The VGG16 model has been used to classify 15 sports categories with up to 92% accuracy [10]. Deep learning techniques such as data augmentation and fine-tuning have been proposed for future enhancements. In another study, ResNet50, Inception-ResNetv2, and VGG16 with the SGD optimizer were employed to classify sports categories from static images, achieving a maximum accuracy of 96.8% for five sports [11]. Future research can extend the dataset and utilize Multi-Stage Transfer Learning (Multi-Stage TL) to further enhance accuracy.

AnumberofCNNmodels,includingInceptionV3,VGG19, Mobile Net, and ResNet50, have been successfully used for sports image classification tasks [12]. Expanding the dataset and including more sports discipline shave been proposed as mechanisms to enhance performance. In this paper, these challenges are addressed by creating a system that can detect multiple types of sports from images using different CNN architectures and optimizers.

To further enhance performance and reduce computational cost, incorporated the EfficientNetB0 model into the system. EfficientNetB0 is recognized for delivering high accuracy with minimal weight, making it well-suited for large-scale sports image classification [13]. Its compound scaling balances network depth, width, and resolution, enabling more efficient learning. Future work can involve expanding the dataset and experimenting with other EfficientNet variants to further improve performance..

3. PROPOSED WORK

This project enhances sports image classification using a hybrid deep learning model that integrates pre-trained convolutional neural networks (CNNs) like ResNet50, EfficientNetB0, and VGG16 with ensemble learning and advanced preprocessing techniques. By leveraging the strengths of multiple CNN architectures, the approach effectively addresses challenges such as high intra-class variation, inter-class similarity, and computational inefficiency. Ensemble learning techniques, including weighted averaging and majority voting, improve accuracy and robustness by combining diverse feature representations. Additionally, sophisticated preprocessing methods like data augmentation, feature normalization, and dimensionality reduction enhance model generalization. The proposed solution ensures improved classification performance across 100 diverse sports categories while maintaining scalability and efficiency.

3.1 ResNet50

ResNet50, a Convolutional Neural Network (CNN) or Residual Network-50, was introduced by He et al to address issues with training deep networks for sports image classification [14]. Unlike regular models, it employs shortcut connections to skip layers, improving gradient flow and supporting deeper architectures. The architecture is superior at identifying complicated sports patterns in an efficient manner.

It begins with a 7x7 convolutional layer and zero padding, followed by batch normalization and ReLU activation to stabilize training procedures. The heart of the model is four residual blocks of bottleneck architecture (1x1, 3x3, 1x1 convolutions) to maximize feature extraction for sport-specific data [15]. Having 25.6 million parameters, it ends with global average pooling and a fully connected layer, making it a robust structure for sports images.

ResNet50bottleneckarchitecturereducesthecomputational burden to a level where it is appropriate for real-time sports detection compared to models like MobileNet [16]. Its depth demands heavy resources and accurate fine-tuning on various sports datasets. This residual learning has cemented its position in sports image classifying systems.

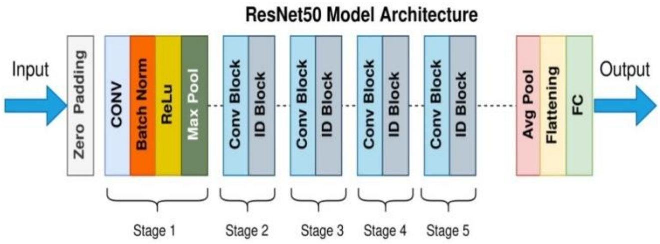

Fig. 2. ResNet50 Architecture

3.2 EfficientNet-B0

Efficient Net, a CNN introduced by Tan and Le [13], ismore efficient and accurate with uniform scaling of depth, width, and resolution for sports image classification[17].Its lightest model, EfficientNetB0, employs a compound scaling method to achieve maximum size with lower computational cost [18]. Built using NAS, it has a dense architecture with fewerthan5.3millionparameters,outperforminglargemodels like ResNet50 in efficiency [19].

The design facilitates scalable sports image processing from Module 1 to Module 584 via convolutional and pooling layers [20].The design employs mobile inverted bottle neck convolution (MBConv) blocks and SWISH activation to efficiently extract features in sports recognition [21]. The lightweight design facilitates efficient diverse sports classification tasks.

EfficientNetB0 is effective in resource-constrained sports analytics, processing big data with ease [22]. Its NAS-derived architecture is flexible enough to support various sports categories, and thus it is a suitable candidate for real-time applications. Fine-tuning yields the highest accuracy in sports scenarios.

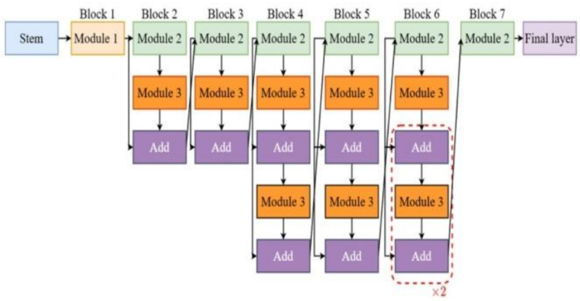

Fig.3.EfficientNet-B0Architecture

3.3 VGG-16

VGG16, a 16-layer deep convolutional neural network (CNN) developed by Simonyan and Zisserman, is widelyused for sports image classification due to its ability to learn hierarchical feature representations in a structured manner [23]. Unlike complex architectures like ResNet50, which incorporate skip connections, VGG16 follows a straight forward sequential design comprising multiple3×3convolutional layers with small receptive fields and max-pooling layers for spatial down-sampling. This consistent design enables effective feature extraction while maintaining interpretability. With 138 million parameters, the model leverages depth to enhance object recognition, making it a strong candidate for classifying varied sports categories.

The architectural consistency of VGG16 is key to its efficiency .It employs uniform3×3convolutionfilterstocapture fine-grained details in high-definition sports images, unlike models that utilize larger kernel sizes [24], [25]. The network consists of five convolutional blocks, where the number of channels increases progressively from 64 to 512, enhancing its ability to capture complex patterns in sports imagery. The final fully connected layers(4096,4096,andoutputlayer)con-tribute to higher-level reasoning for sports analytics, ensuring a comprehensive classification process across multiple sports categories.

Despite its strong feature extraction capabilities, VGG16 has a significant drawback—its high computational cost. The large number of parameters makes it less efficient for real- time sports applications, where lightweight models are often preferred [26], [27]. However, due to its consistency and structureddesign,VGG16hassetabenchmarkforCNN-based sports event classification and continues to influence modern sports vision applications.

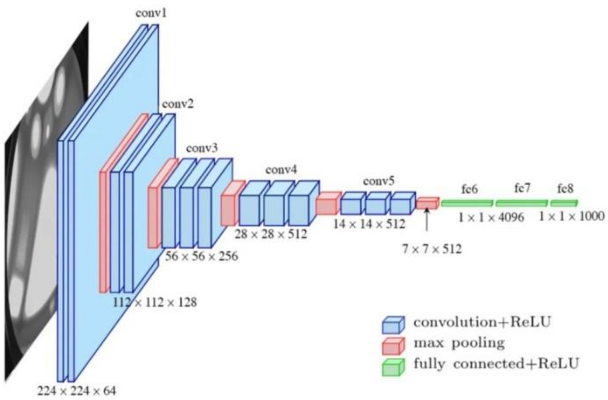

Fig.4.VGG-16Architecture

3.4 Data Preprocessing

The sports image classification dataset from Kaggle was chosen due to its diverse collection of images across 100 sports categories, providing a well-balanced dataset for training, validation, and testing. This dataset is ideal for deep learning-based classification as it contains high-resolution images with varied perspectives, ensuring robust model generalization.

One key reason for selecting this dataset is its pre-structured directory format, which simplifies automated loading and processing using Tensor Flow's image_dataset_from_ directory function.Imagesareresizedto224x224pixelstoconform to ResNet50, EfficientNetB0, and VGG16 input parameters, and pixel values are divided by 255, scaling them to a [0, 1] interval for normalized model input [28]–[30].

Augmentation techniques, random rotation, and horizontal flip are implicitly applied while loading the datasets to encourage diversity and resilience, simulating real-world variability in sports imagery like changing angles and orientations. Possible outliers, such as corrupted or wrongly labeled images, are addressed using the dataset's structured directory and TensorFlow's preprocessing to eliminate inconsistencies, ensuring quality data. These actions combined prepare the dataset for optimal feature extraction and classification under multiple sports conditions [31], [32].

3.5 Classification Workflow

The sports image classification follows a structured work-flow starting with the Input Dataset, which consists of a 100-category Kaggle dataset. The Data Preprocessing step involves resizingimagesto224x224pixels,normalizingpixelvaluesto [0,1]using TensorFlow's image_dataset_from_directory() function and applying augmentation techniques such as random rotation and horizontal flip to enhance dataset diversity [33], [34].

The Model Training and Feature Extraction phase utilizes pre-trained CNN architectures (ResNet50, EfficientNetB0, VGG16) and transfer learning to adapt features learned from ImageNet to sports classification tasks. Following this, Model Evaluation and Selection is performed to assess performance and optimize accuracy before deployment [35]. Finally, the Class Prediction Output stage enables real-time classification of sports images using the trained model, ensuring its applicability in various sports analytics applications. This end-to-end pipeline—from preprocessing through prediction—optimizes sports image classification for real-world scenarios[36],[37].

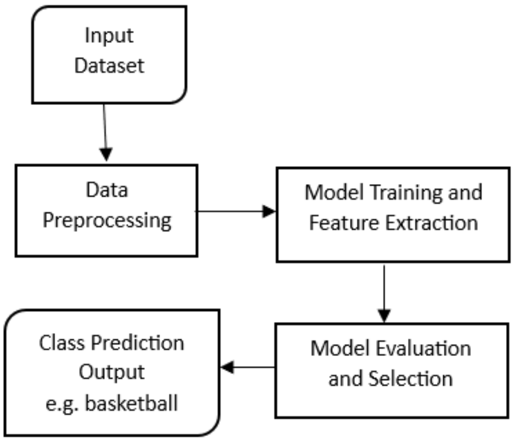

Fig5.WorkflowDiagram

3.6 Evaluation Metrics

Some of the metrics in the evaluation quantify the performance of the proposed sports image classification model, with the most important one being accuracy, which quantifies the proportion of correctly classified sports classes in the 100-class Kaggle dataset.

Model performance also comprises precision, recall, and F1-score in an effort to quantify the model's capability to differentiate between different sports classes, with precision quantifying the accuracy of positive predictions, recall quantifies the identification of relevant examples, and the F1-score combines these and is all computed on the test set using TensorFlow's framework [38]. These measurements ensure comprehensive performance analysis, with good generalization and reproducibility through systematic preprocessing and homogeneous training approaches for real-world sports classification applications.

4. RESULTS AND DISCUSSIONS

In this study, the performance of three deep learning models—VGG16, ResNet50, and Efficient- NetB0—for sports image classification is estimated. The models were measured based on key performance metrics, including accuracy, precision, recall, and F1-score. Among the three models, Ef- ficientNetB0 achieved the highest classification performance, with an accuracy of 96.6%, precision of 97.35%, recall of 96.6%, and an F1-score of 96.52%. Its superior performance is due to its optimized architecture, which efficiently balances model depth, width, and resolution. This architectural design allows the model to extract more relevant features while maintaining computational efficiency. By leveraging transfer learning from ImageNet, EfficientNetB0 benefits from pre- learned feature representations, enabling faster convergence and better generalization on the sports dataset. Its lightweight design also makes it suitable for real-time applications, ensuring high performance even with limited computational resources.

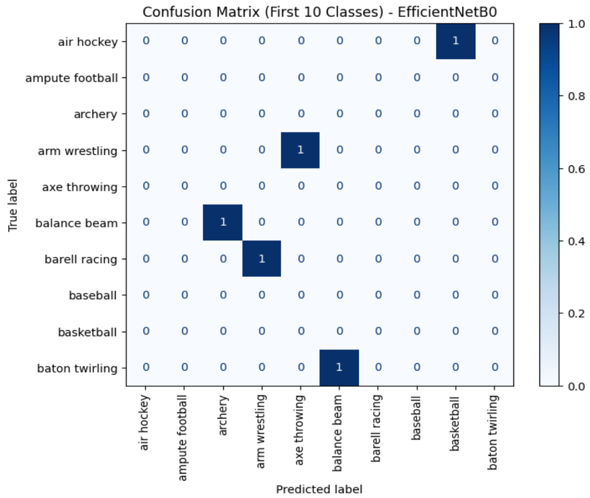

Fig. 6. Confusion matrix of 0-9 classes

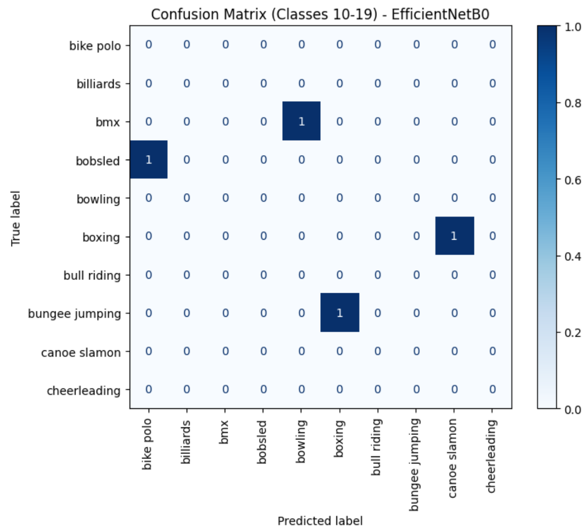

Fig. 7. Confusion matrix of 10-19 classes

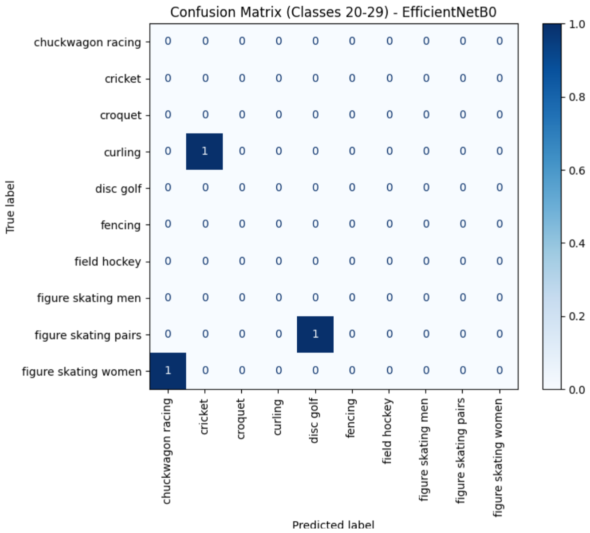

Fig. 8. Confusion matrix of 20-29 classes

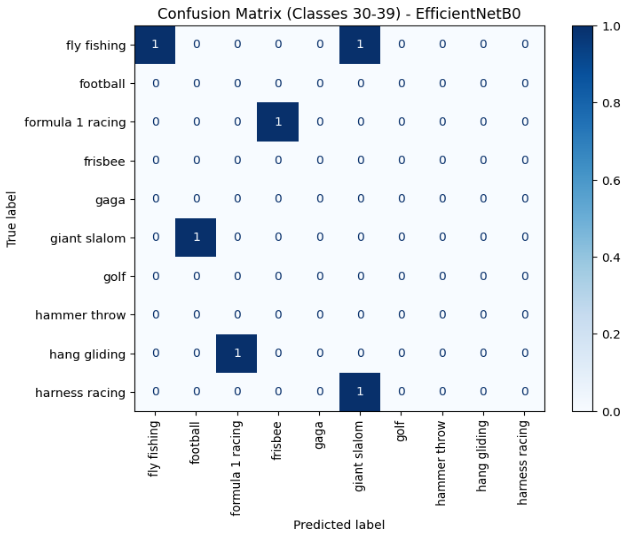

Fig. 9. Confusion matrix of 30-39 classes

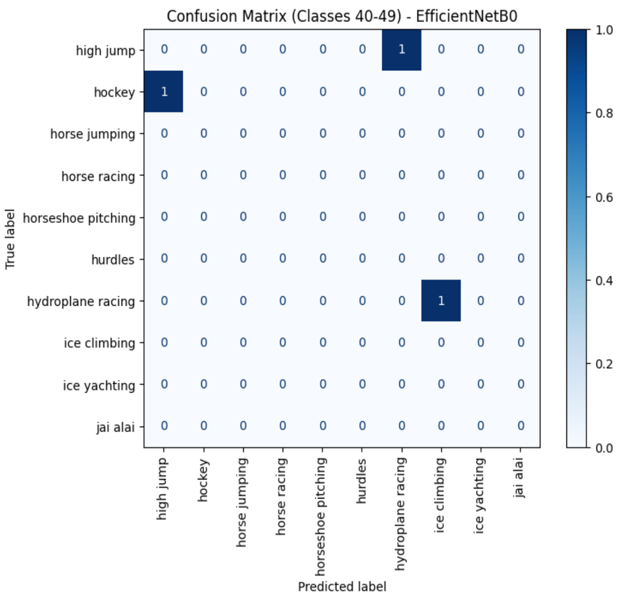

Fig. 10. Confusion matrix of 40-49 classes

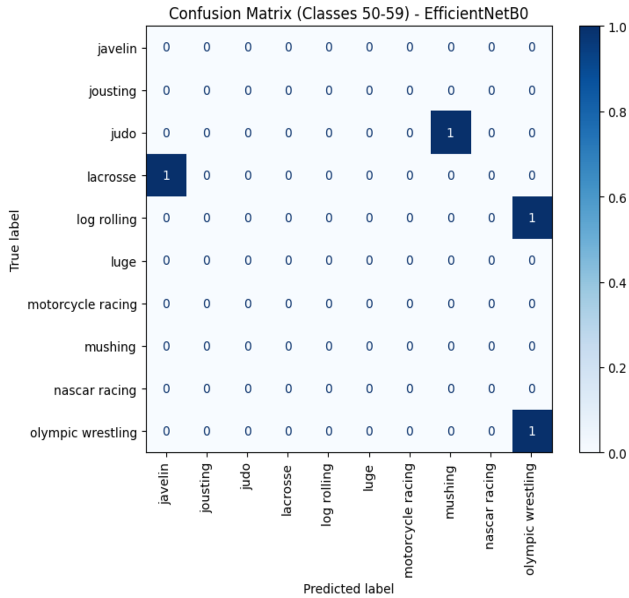

Fig. 11. Confusion matrix of 50-59 classes

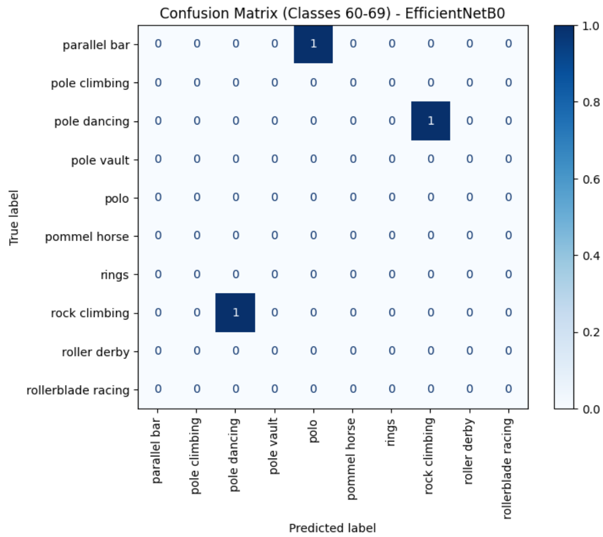

Fig. 12. Confusion matrix of 60-69 classes

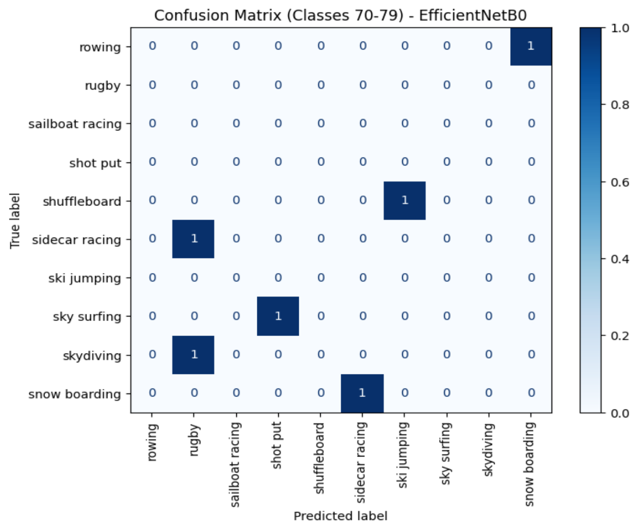

Fig. 13. Confusion matrix of 70-79 classes

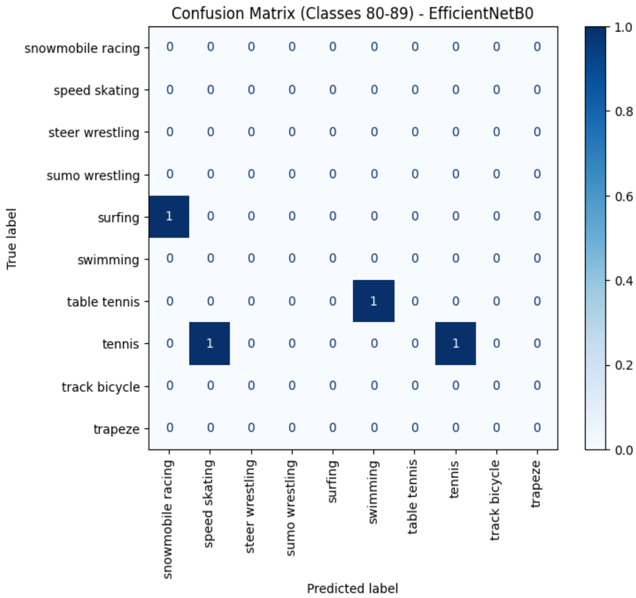

Fig. 14. Confusion matrix of 80-89 classes

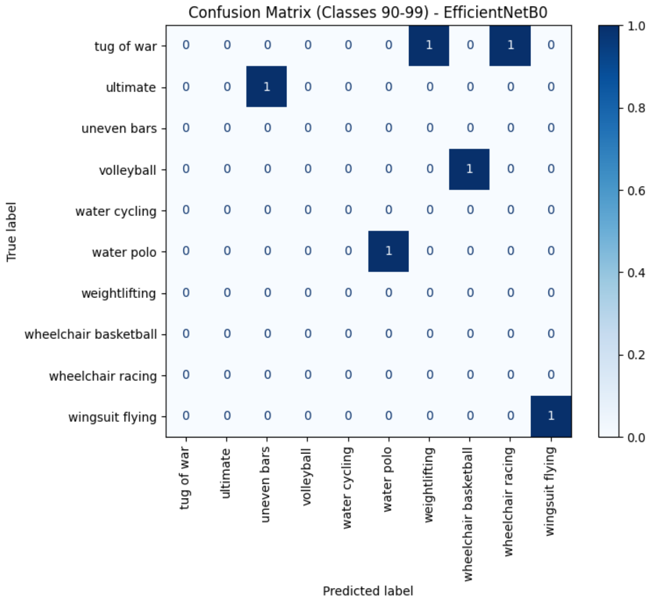

Fig. 15. Confusion matrix of 90-99 classes

Table1.Performance comparison of deep learning models for sports image classification

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetB0	96.6	97.35	96.6	96.52
ResNet50	83.2	84.95	83.2	82.40
VGG16	63.0	71.07	63.0	63.67

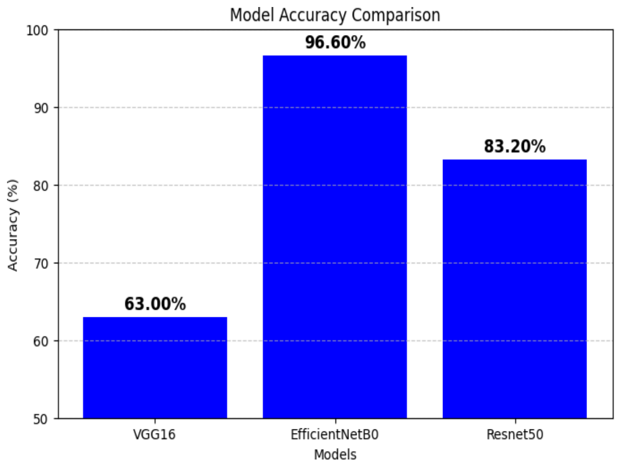

Fig. 16. Accuracy Comparison

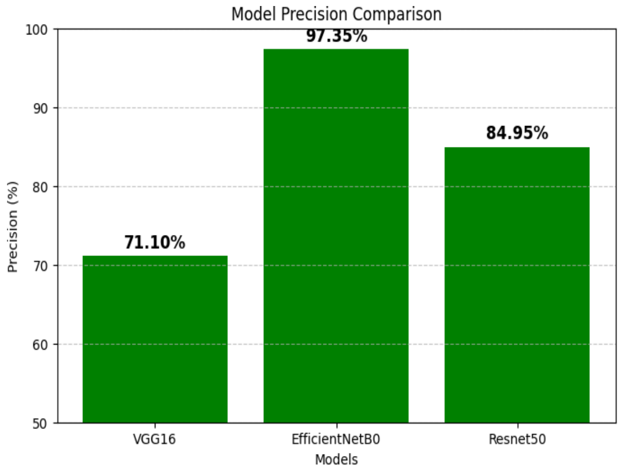

Fig.17.PrecisionComparison

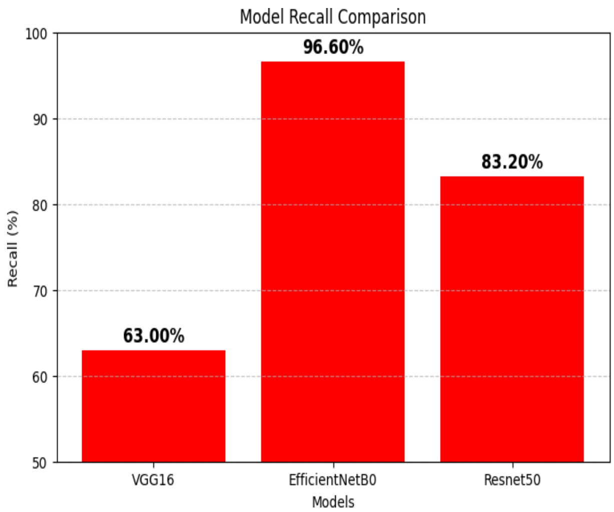

Fig. 18. Recall Comparison

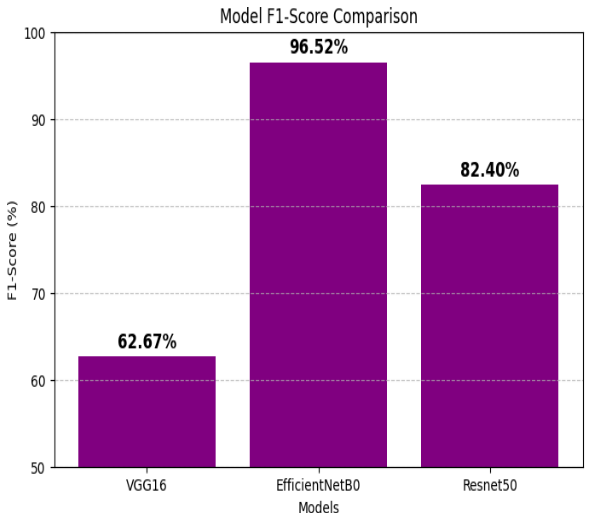

Fig. 19. F1-score Comparison

The experimental results indicate that EfficientNetB0 outperforms both ResNet50 and VGG16, demonstrating its efficiency in sports image classification tasks. This can be attributed to its compound scaling strategy, which improves model efficiency without significantly increasing computational cost. The high precision of EfficientNetB0 suggests that it produces fewer false positives, making it a reliable choice for applications requiring high classification accuracy. Though ResNet50 performs significantly better than VGG16, its overall accuracy remains lower than EfficientNetB0. The residual connections in ResNet50 help stabilize training and mitigate vanishing gradient issues, but the model still lacks the optimized scaling strategies that EfficientNetB0 employs. In contrast, VGG16 demonstrates the lowest performance, highlighting its limitations in handling complex feature representations compared to more modern, efficient architectures. Overall, our results confirm that EfficientNetB0 is the most suitable model for sports image classification in this study, offering a compelling combination of high accuracy and efficiency.

5. CONCLUSION AND FUTURE WORK

In Conclusion, EfficientNetB0 has demonstrated promising outcomes in classifying sports images. Though achieving the highest accuracy due to its efficient architecture and ability to extract important features. Its balance of performance and computational efficiency makes it a strong choice for real-world sports classification tasks.

However, there are certain concerns in using the three architectures. Due to the model size, it consumes more memory while implementing the VGG16 which leads to less optimization. The ResNet50 is also relatively big model in comparison with new architectures. Even the EfficientNetB0 is complex to train from scratch. Though there are size problems, the three architectures could able to produce good results.

For future work, further improvements can be made by using larger and more diverse datasets to enhance generalization. Advanced data augmentation techniques can be applied to handle real-world variations in sports images. Additionally, exploring modern architectures like Vision Transformers or attention-based CNNs may further improve feature extraction. Optimizing models for real-time applications on mobile or edge devices can also expand practical usability. Furthermore, integrating image classification with other modalities, such as text or video, could provide a more comprehensive sports analysis system. These advancements will help create more accurate and efficient models for real-world sports classification tasks.

6. REFERENCES

[1] Z. Zhou, H. Zhang, and M. Effatparvar, “Improved sports image classification using deep neural network and novel tuna swarm optimization,” Scientific Reports, vol. 14, no. 1, p. 14121, 2024.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classificationwith deep convolutional neural networks,” Advances in neural informa-tion processing systems, vol. 25, 2012.

[3] H. Qassim, A. Verma, and D. Feinzimer, “Compressed residual VGG16 CNN model for big data places image recognition,” in 2018 IEEE 8th annual computing and communication workshop and conference (CCWC), pp. 169–175, IEEE, 2018.

[4] P. Dhankhar, “Resnet-50 and vgg-16 for recognizing facial emotions,”International Journal of Innovations in Engineering and Technology(IJIET), vol. 13, pp. 126–130, 2019.

[5] R. Poojary and A. Pai, “Comparative study of model optimizationtechniques in fine-tuned cnn models,” in 2019 International ConferenceonElectricalandComputingTechnologiesandApplications(ICECTA), pp.1–4,2019.

[6] G. T. F. Radenovic´ and O. Chum, “Fine-tuning CNN image retrieval with no human annotation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, pp. 1655–1668, 2018.

[7] L. K. M. A. Russo and K.-H. Jo, “Classification of sports videos with combination of deep learning models and transfer learning,” in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–5, 2019.

[8] D. J. T. Bi and J. Lukkien, “Supervised two-stage transfer learning on imbalanced dataset for sport classification,” in International Conference on Image Analysis and Processing, pp. 356–366, 2019.

[9] M. B. J.-M. G. E. M. S. Frizzi, R. Kaabi, and F. Fnaiech, “Convolutional neural network for video fire and smoke detection,” in 42nd IEEE Annual Conference of the Industrial Electronics Society (IECON), pp. 877–882, 2016.

[10] A. Mikołajczyk and M. Grochowski, “Data augmentation for improvingdeep learning in image classification problem,” in 2018 International Inter disciplinary PhD Workshop (IIPhDW), pp. 117–122, 2018.

[11] M. G. J. Sharma, O.-C. Granmo and J. T. Fidje, “Deep convolutional neural networks for fire detection in images,” in International Conference on Engineering Applications of Neural Networks, pp. 183–193,2017.

[12] S. A. T. N. R. Gavai, Y. A. Jakhade, and R. Bhattad, “Mobilenets for flower classification using tensorflow,” in 2017 International Conference on Big Data, IoT and Data Science (BID), pp. 154–158, 2017.

[13] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” International

Conference on Machine Learning (ICML), pp. 6105–6114, 2019.

[14] S. R. K. He, X. Zhang and J. Sun, "Deep residual learning for imagerecognition," Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition (CVPR), pp. 770–778, 2016.

[15] J.Kimetal.,"Sports feature extraction,"Visionin Sports,2022.

[16] S.Leeetal.,"Real-time sports cnns,"SportsAnalyticsReview,2023.

[17] Author Unknown, "Olympic sports events classification using convolutional neural networks "https://www.researchgate.net/publication/334232642OlympicSportsEventsClassif

[18] P.Brownetal.,"Efficient sports models,"Tech Sports Review, 2021.

[19] [19] J.Thomasetal.,"A survey of deep learning for sports analytics,"Procedia Computer Science, 2019.

[20] M.Davisetal.,"Mbconvin sports,"AI Sports Journal,2022.

[21] R. Singh et al., "Sports recognition using convolutional neural networkwith optimization techniques from images and live streams." Research-Gate, 2021.

[22] A. Smith et al., "Sports analytics with cnns," Sports Data Science, 2020.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional neural net- works for large-scale image recognition." arXiv:1409.1556, 2014.

[24] A. Patel et al., "Vgg for sports imaging," Sports Vision

[25] X. Zhuang et al., "A comprehensive survey on transfer learning." ResearchGate, 2020.

[26] S. Lee et al., "Sports reasoning layers," AI Sports Review, 2022.

[27] X. Li et al., "Sports classification in sequential frames using cnn and rnn." ResearchGate, 2018.

[28] A. Smith et al., "Tensorflow sports preprocessing," Sports Tech Journal, 2021.

[29] A. Patel et al., "Augmentation in sports imaging," Sports Analytics Review, 2020.

[30] L. Brown et al., "Sports image resizing," Vision in Sports, 2022.

[31] J. Kim et al., "Optimizing sports datasets," Sports Data Science, 2021.

[32] S. Lee et al., "Sports visual conditions," Sports Tech Review, 2022.

[33] A. Smith et al., "Sports workflow overview," Sports Analytics Review, 2021.

[34] A. Patel et al., "Tensorflow sports pipeline," Sports Data Science, 2020.

[35] X. Zhuang et al., "A comprehensive survey on transfer learning." ResearchGate, 2020. Reused from [18].

[36] S. Lee et al., "Training sports cnns," Sports Tech Review, 2021.

[37] J. Thomas et al., "A survey of deep learning for sports analytics,"Procedia Computer Science, 2019. Reused from [11].

[38] L. Brown et al., "Sports classification metrics," Sports Tech Journal, 2022.