Deep Learning-Based Sports Shot Classification System

Dr.Vinmathi M S[1,a)] Dr Sangeetha K[2,b)] Dharshika S[3,c)] Hari Priya R[4,d)]
Logeshwari S[5,e)] Dr. Kavitha Subramani[6,f)]

[1,2,3,4,5,6]*Computer Science and Engineering, Panimalar Engineering College*
Bangalore Trunk Road, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India.

[a)]*vinmathis@gmail.com*
[b)]*sangeethakalyaniraman@gmail.com*
[c)]*bsdi1246789@gmail.com*
[d)]*priyaranganathan2003@gmail.com*
[e)]*logeshwarisrinivasan5@gmail.com*
[f)]*kavitha.pec2022@gmail.com*

Abstract- This study presents an innovative deep learning architecture for sports shots classification, employing Convolutional Neural Networks (CNNs) developed using TensorFlow. Accurate shot classification is crucial for enhancing sports analytics, coaching methodologies, and broadcasting, as it yields critical insights into gameplay dynamics and athlete performance. Conventional techniques often rely on manual labeling or basic rule based system, which is insufficiently adaptable to complexities of different sports. We explore a range of CNN architectures specifically optimized for shot classification, enhancing model efficacy through strategies such as data augmentation, advanced deep learning techniques, and meticulous hyperparameter tuning. The use of TensorFlow ensures robust scalability and efficiency in both the training and deployment phases of the model. Our proposed framework shows significant promise for transforming sports analysis, broadcasting, and coaching practices, thereby enabling automated and sophisticated shot classification across diverse sporting contexts.

Keywords— *Deep learning, CNN, AlexNet, LeNet, VGG16, Sports Shot images*

I. INTRODUCTION

Advances in data collecting and analysis technology have led to a considerable evolution in sports analytics in the past few years. Accurate classification of sports shots is crucial for enhancing sports analytics, coaching methodologies, and broadcasting, as this offers vital information about gameplay dynamics and athlete performance, essential for making informed decisions. Traditional techniques for shot classification often rely on manual labeling or basic rule-based systems. While useful to some extent, these methods have significant limitations, including being labor-intensive, time-consuming, and inflexible in handling the complexities and variations inherent in different sports. Recent advancements in Convolutional Neural Networks(CNNs), a kind of deep learning, have demonstrated significant potential in tackling these issues. CNNs immediately learn and extract relevant features from data, achieving high accuracy in classification tasks and adapting to the complexities of various sports. This study presents an innovative deep learning framework for the classification of sports shots using CNNs developed with TensorFlow. Our approach includes exploring a range of optimized CNN architectures, employing data augmentation techniques, incorporating advanced deep learning methods such as dropout, batch normalization, transfer learning, and meticulously tuning hyperparameters to enhance model efficacy. By leveraging TensorFlow, we ensure robust scalability and efficiency during the model's training and deployment stages. Our proposed framework represents a noteworthy development in sports analytics, providing an opportunity to transform sports analysis, broadcasting, and coaching practices by automating and sophisticating shot classification. This, in turn, seeks to offer deeper information about athlete performance as well gameplay strategies, ultimately contributing to the evolution of sports analytics.

II. LITERATURE REVIEW

Several approacheshave been investigated in the area of sports shot recognition, indicating the development of methods in this field.The results were compiled and interpreted using Convolutional Neural Network (CNN)-based lightweight technique for re-play identification in cricket videos [1]. Linear SVM camera models with a Random Forest models were used for temporal action localisation in Cricket Stroke Extraction, another noteworthy contribution [2].However, these approaches lacked extensive consideration of diverse stroke styles, limiting their applicability across broader datasets.

In Outcome Classification in Cricket Using Deep Learning [3], a ball by ball outcome prediction method achieved a notable accuracy of 70%, although it was hindered by the absence of a standardized dataset. Such challenges have prompted researchers to explore attention mechanisms and curated datasets to overcome variations in cricketing conditions.In this field, deep learning-based methods have shown a great deal of promise. The study "Deep CNN-based Data-driven Identification of Batting Shots in Cricket" [4] utilized a combination of RNN, 3D CNN, and 2D CNN architectures to achieve high level of accuracy. However, it faced restrictions in robustness under diverse conditions and relied on small, imperfect datasets. Likewise, "CricShotClassify" [5] demonstrated a CNN-GRU hybrid architecture with a remarkable 93% accuracy rate. This method addressed misclassifications caused by extraneous frames and varying angles but could be extended further through enhanced dataset creation and advanced network designs. To show the effectiveness of hybrid deep learning models over sports action recognition, Ahmad et al. [6] suggested an optimised CNN-LSTM network to categorise five distinct cricket shots—bowled, defence, cover, pull, and reverse shots—from films. Foysal et al. [7] shown the usefulness of convolutional networks in static frame-based action identification by using a CNN-based method to categorise various cricket shots from photos. Ahmed et al. [8] employed a CNN model to classify

four distinct cricket batting strokes from videos, emphasizing the role of deep learning in automating cricket action recognition. Sen et al. [9] demonstrated the benefits of integrating convolutional and recurrent networks for sequential data processing by introducing a CNN-GRU architecture to identify batting shots from cricket films.

Gupta et al. [10] developed a method for cricket stroke classification by leveraging spatial and motion sequences, reinforcing the importance of spatiotemporal information in shot identification.In order to classify cricket strokes, Azhar et al. [11] proposed a transformer model derived from DenseNet, proving the effectiveness of transformer architectures in sports analytics. Devanandan et al. [12] utilized a Random Forest model for identifying cricket shots from images, presenting a machine learning approach as an alternative to deep learning for sports classification tasks. Thomas et al. [13] and Wang et al. [14] present comprehensive surveys on both contemporary and past research in the field. While datasets such as Sports-1M and UCF Sports [15] are among the biggest one for sports analysis, they lack sufficient data to effectively train models for a specific sport, as they do not encompass all types of events within any single sport.A motion-estimation method was presented in [16] for the classification of cricket shots. In contrast to deep learning models, this approach's accuracy was lower yet it identified eight angle-based classes to identify various shots. The off drive and hook shot had the highest reported accuracy rates, with 63.57% and 53.32%, respectively. In [17], cricket shots were categorised using wearable technology. The objective was to develop a method for quality assessment.The objective was to develop a system for evaluating quality that took into account factors like shot direction and foot position. Five layers of action classification were achieved by hierarchical representation using (k-NN) k-nearest neighbours, (SVM) support vector machines, and decision trees (DT). The best-performing classifiers' class-weighted F1 score was 88.30%. Nevertheless, this study did not evaluate actual match situations; it simply used films of training sessions.

For both left- handed and right-handed batsmen, Semwal et al. [18] created an example to identify batting strokes such the straight drive, lofted on drive, and cover drive. They achieved 83.098% accuracy for right-handed shots and 65.186% accuracy for left-handed shots using a pretrained CNN and a multiclass SVM. However, some shots, including the left cover drive, left straight drive, and right straight drive, were misclassified due to a lack of data.A method for categorising cricket shots into the run, four, six, and out occurrences—was presented in [19]. An SVM was used to classify the individual shots that were extracted from the movie.The accuracy of the suggested model was 87.8%, while it performed worse for six-hit occurrences. Using an AlexNet CNN-based technique, Rafiq et al. [20] presented based on transfer learning system for sports video summarisation that obtained 99.26% accuracy on a smaller dataset that contained only cricket sequences.

III.   METHODOLOGY

The sports shot images were thoroughly collected for this research to create an extensive dataset. A total of 4000 image data of the sport short were collected which were classified by the type of sport short. A total of 4,000 sports shot images were collected and categorized into 12 classifications based on different stages and types of shots. This classification is crucial because it provides a diverse set of deep learning model's training data, which helps the capacity of the model to recognize and categorize sports shots in images. In particular, there are 400 images for testing and each of the eight classifications-address, impact, finish, mid-downswing, mid-backswing, and mid-followthrough, toe-up, and top has 599 images for training. The dataset, which is part of the leg glance flick and sweep class, includes 600 training pictures and 299 testing images. For the drive and pull shot classes, there will be 299 testing images and 599 training images. Therefore, having such image datasets would enable the model to examine and predict the image in the future.

In the research study, several DL models were used to predict and analyze the Sport shot image data of the various sport for its classification. Figure 1 shows the various DL models used and the methodology of the research. These include the CNNs, VGG16, AlexNet and LeNet. These models have shown better results in the sports domain on the image dataset. The CNN model captured the hierarchical and spatial feature in the image as the foundational architecture. The AlexNet and LeNet model had deeper architecture to do better feature extraction and achieve better accuracy. VGG16 had a sophisticated architecture network with efficient computation.
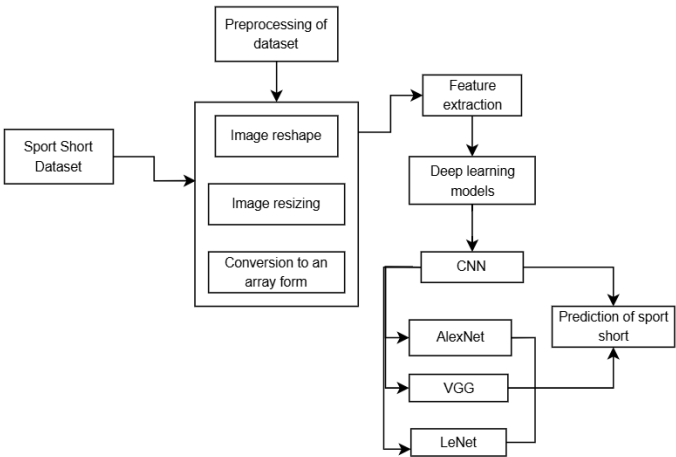

Figure. 1. Methodology of the research

Finally, AlexNet prevents the vanishing gradient exception in bigger neural networks so that we can go deeper without the loss of their performance. The implementation of the model involved various stages such as data preprocessing to make data more normalized and augmentation to maximize the quality and variance in the image dataset. These images were then trained and validated, where the high classes were learned manually by the model. Validation involved examining the model to ensure generalization of the data unseen during training. Finally, after evaluating the model's performance based on F1-Score, accuracy, and recall. Predicting the sport shot classes and improving player performance are the goals of this study. Since players' performance raises their scores, it is necessary to improve results.

IV.   DEEP LEARNING MODELS

Sports shot categorization is a difficult task due to the dynamic movements, varying angles, and environmental conditions that affect shot recognition. Models for deep learning, particularly about Convolutional Neural Networks (CNNs) have proved notable improvements in recognizing different shot types across sports. This study employs AlexNet, VGG16, and LeNet to classify sports shots, comparing their accuracy to decide which model is most suited for this task.

In sports shot classification, AlexNet excels at recognizing body posture, ball trajectory, and shot execution. AlexNet is a CNN model designed for high-accuracy image classification. Each of its five convolutional layers is succeeded by max pooling and ReLU activation layers, which enable the extraction of spatial and motion features.

The Visual Geometry Group at the University of Oxford developed the sophisticated CNN architecture known as VGG16. It accepts 224x224-pixel images as input, maintaining consistency for the ImageNet competition. The model uses 3x3 convolutional filters to capture spatial movements and ReLU activation to speed up training. Three fully connected layers come after the VGG16 convolutional layers, with 4096 channels in the first two and the third layer having 1000 channels for classification. Local response normalization is typically not used in VGG16 to avoid increased memory usage and training time.

LeNet is one of the simplest convolutional neural network introduced by Yann LeCun in 1989, commonly referred as LeNet-5. LeNet, originally designed for digit recognition, serves as a lightweight baseline model for sports shot classification. Its two layers of convolution are then subsampling (pooling) layers enabling efficient extraction of features, making it useful when dataset size is limited or computational resources are constrained.

Table 1 represents in-depth analysis of the three model architectures-AlexNet, VGG16, and LeNet which were employed in this study. In order to prevent overfitting, AlexNet uses eight layers with 11x11, 5x5, and 3x3 filters. Max-pooling is used after the first two convolutional layers, dropout is used in fully connected layers and ReLU activation is used in all levels except output. With 16 layers and uniform 3x3 filters, VGG16 uses fully connected layers for classification, max-pooling after convolutional layers, and ReLU activation in hidden layers. LeNet, which has five layers, uses 5x5 and 3x3 filters. Tanh activation was used at first, but ReLU was later used for superior performance. Subsampling is used to minimize spatial dimensions and effectively extract features.

Table 1. Comparison of model architecture.

Model	Layers	Filter Size	KeyFeatures
AlexNet	8	11×11, 5×5, 3×3	Max-pooling, Dropout, ReLU
VGG16	16	3×3	ReLU, Max-pooling, Fully Connected
LeNet	5	5×5, 3×3	Tanh, Subsampling

By leveraging VGG16 domain-specific adaptability, AlexNet's deep feature extraction, and LeNet's efficiency, this study provides a comprehensive approach to sports shot classification. Figure 2 represents that the highest-performing model (AlexNet) is selected for final implementation, ensuring accurate recognition of diverse shot types, aiding in performance analysis, coaching, and automated sports analytics.

Figure. 2. Accuracy of each model
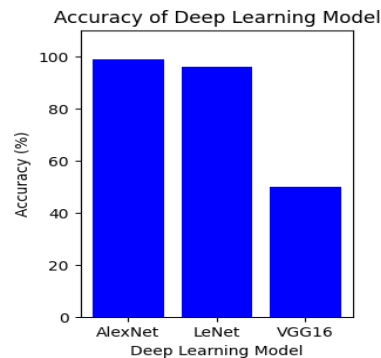
V. PREPROCESSING OF DATASET

Data preprocessing is essential to ensuring the success of deep learning models for sports shot classification. Since raw images often contain variations in quality, resolution, and lighting conditions, careful preprocessing strengthens the model's capacity to extract meaningful features, improving classification accuracy and overall performance.

The first step involves collecting and cleaning the datasets. Sports action images are gathered from multiple sources, including publicly available datasets and manually labeled images. These images capture various shot types performed by athletes in different sports. During the cleaning process, low-quality, blurry, or corrupted images are removed to prevent inaccuracies in feature extraction. Ensuring high-quality data is crucial, as noisy or distorted images can negatively impact model training and lead to misclassifications. Once the dataset is cleaned, normalization is applied to bring all images to a standard intensity range. Pixel values are scaled between 0 and 1 to maintain consistency and improve training efficiency.

Without normalization, variations in pixel intensities may cause instability during training, leading to poor generalization on unseen data. This step also helps the deep learning model process images more effectively by ensuring that features extracted from different images remain comparable. To maintain uniformity in input dimensions, all images are resized to a fixed resolution. Since deep learning models operate with predefined input sizes, resizing ensures compatibility without distorting image features. Maintaining aspect ratios where possible prevents loss of important structural details in the images. By standardizing input dimensions, the model can efficiently learn distinguishing characteristics across different shot types. Given the limited availability of sports action datasets, data augmentation is applied to artificially expand the training set. Transformations such as rotation, flipping, scaling, and brightness adjustments introduce variability while preserving essential features. Augmentation reduces overfitting, allowing the model to generalize better to new images. Since athletes perform similar shots under different lighting conditions, angles, and orientations, augmentation ensures that the model learns from diverse representations rather than memorizing specific examples. To further refine image quality, noise reduction techniques are employed. Sports images, especially those extracted from live-action videos, often contain motion blur and background clutter. Applying Gaussian blurring and median filtering helps remove unwanted artifacts while retaining essential action details. Enhancing contrast ensures that the model focuses on the player's movement and shot mechanics rather than distractions in the background. This preprocessing step improves the clarity of extracted features, leading to more precise classification results.

After preprocessing, a stratified sample technique is used to divide the dataset into training, validation, and test sets. This ensures that each shot type is proportionally represented across all sets, preventing bias toward majority classes. Images are utilized to identify patterns in the training set, and the validation set aids hyperparameter tuning. In contrast, the test set evaluates how well the finished model performs on unknown data. Proper dataset splitting ensures that the trained model remains robust and performs well across different scenarios. Following preprocessing, multiple deep learning models are trained and evaluated on the processed dataset. The best model for classifying sports shots is identified using performance criteria such as accuracy, precision, recall, and F1 score. The model achieving the highest accuracy on the test set is selected for final deployment. By ensuring a structured preprocessing pipeline, the classification system becomes more reliable, improving its ability to recognize and differentiate between various sports actions.

VI. FEATURE EXTRACTION

This is a crucial step in cricket shot classification, as it enables the deep learning model to identify and differentiate between different shot types based on player movements, bat trajectory, and ball position. This process involves applying convolutional neural networks (CNNs) such as AlexNet, VGG16, and LeNet to extract meaningful patterns from the video frames. These architectures improve the ability of the model to depict both low-level and high-level spatial features, improving classification accuracy.

AlexNet, a convolutional neural network, is widely used for image classification tasks. Five convolutional layers and max-pooling layers make up this system, which gradually reduces spatial dimensions while retaining essential features. The first convolutional layer utilizes large 11×11 filters to capture broad spatial patterns, while deeper layers use smaller 3×3 filters to refine the details. Activation functions such as ReLU speed up training by introducing non-linearity, helping the network to learn complex patterns in cricket shots. Dropout layers randomly deactivate neurons during training in order to minimize overfitting, making sure the model performs adequately in generalizing to unknown video frames. Furthermore, the retrieved characteristics are processed for categorization by fully linked layers after the network, which finally gives the cricket shot a label.

VGG (Visual Geometry Group) networks, specifically VGG16 and VGG19, are designed to improve feature extraction by using very small convolutional filters of size 3×3. Unlike AlexNet, which starts with larger filter sizes, VGG relies on a uniform structure with multiple stacked convolutional layers to extract intricate spatial features. The deeper architecture allows the network to capture finer motion details in cricket shots, making it particularly effective for distinguishing visually similar shots like cover drives and square cuts. By maintaining the same filter size throughout the network and increasing the depth, VGG16 enhances the hierarchical learning of features. Max-pooling layers gradually lower spatial dimensions while maintaining significant shot-related properties, and all hidden layers employ the ReLU activation function. At the end, the fully connected layers divide the retrieved representations into several shot types.

LeNet is one of the earliest convolutional neural network, provides a structured yet simpler approach to feature extraction. For classification, the model comprises three completely linked layers, two average-pooling layers, and convolutional layers. LeNet efficiently captures essential patterns with fewer parameters, making it computationally lightweight compared to AlexNet and VGG16. The use of tanh activation in earlier versions was later replaced with ReLU for improved training efficiency. Despite being a smaller network, LeNet remains useful for recognizing fundamental shot patterns, making it a strong baseline model for comparison in cricket shot classification.

Each of these CNN architectures contributes uniquely to feature extraction in cricket shot classification. AlexNet excels in learning both broad and detailed features, making it suitable for complex datasets with diverse shot variations. VGG16's deep and structured architecture refines spatial feature extraction, making it ideal for recognizing intricate shot mechanics. LeNet, though simpler, offers an efficient and lightweight alternative for applications requiring lower computational resources.

VII. RESULT AND DISCUSSION

The different deep learning models' performance is evaluated using the Sport shot Dataset, which consisted of 4000 Sports shots sourced from [source, e.g., Kaggle's Sport shot dataset, 2024]. This dataset included images from twelve distinct classifications and it is divided into four different classifications based on the shots. To make training and assessing models easier, the dataset was separated into training (60%), validation (20%), and testing (20%) sets. Preprocessing included resizing the images to 224x224 pixels and normalizing them to [mean and standard deviation]. In order to reduce overfitting to enhance the generalization of the model, data augmentation methods include flipping, rotation, and random cropping were applied during training. The models were subsequently evaluated on the test set, with performance metrics summarized to demonstrate the comparative efficacy of the architectures.

After training the models, they were thoroughly evaluated using a test dataset to determine their ability to classify sports shots accurately as shown in figure 2. The AlexNet model demonstrated the highest accuracy, achieving 99.70% for training and 96.15% for validation, making it the most effective model for sports shot classification. Its deep architecture, use of ReLU activation, and dropout regularization contribute to its high performance by efficiently extracting features while preventing overfitting. The LeNet model, known for its simplicity and computational efficiency, achieved 96.47% accuracy. LeNet-5, originally introduced by YannLeCun in 1989, comprises two subsampling layers, two fully linked layers, and three convolutional layers. Despite its architecture's relative narrowness compared to AlexNet, it effectively captures spatial features using convolution and tanh activation while maintaining a lower computational cost. The VGG16 model, although deeper than LeNet, yielded an accuracy of 50%, indicating that its increased depth might not be optimal for sports shot classification in this dataset. While VGG16 is known for using small 3×3 convolutional filters to capture fine-grained spatial details, its complexity might require more fine-tuning or a larger dataset for improved performance. These results highlight the effectiveness of AlexNet in sports shot classification due to its superior feature extraction capabilities and training stability. The structured evaluation of these CNN architectures ensures that the model chosen for deployment can accurately classify various sports shots, enhancing automated analysis in sports analytics.

Table. 2. Metrics scores of each model

Model	Precision	Recall	F1 Score
AlexNet	96 %	95 %	95.5 %
LeNet	90 %	89 %	89.5 %
VGG16	52 %	51 %	51.5 %

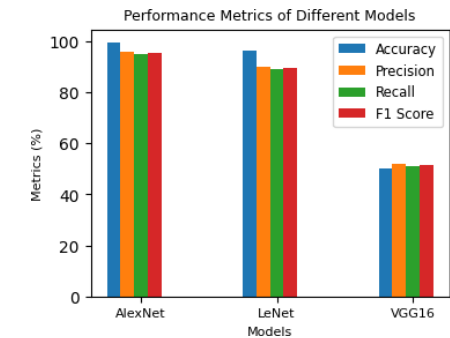**Figure. 3.** Performance Metrics of Different Models

Figure 3 represents the value of the performance scores, such as precision, recall, and F1 score for DL models in the study. Precision is defined as the proportion of genuine positive predictions among all of the model's positive predictions. Sensitivity is another name for recall, which is the percentage of true positive models among all positive events in the dataset. However, the F1 score measures the harmonic mean between precision and recall. As such, F1 score is a balanced measure of the models accuracy since it considers both false positives and false negatives. The AlexNet model in Convolutional Neural Network exhibited high performance in all the models,as shown in table 2. Thus, the model has a high degree of capturing and classifying accurate sport shot since the measures for precision and recall were close. The Lenet model is slightly less compare to AlexNet, and one of the last performing models was the VGG16 model, which had slightly low values as compared to the Alexnet, and Lenet. Therefore, the best model was Alexnet with high values across the metric.

Figure. 4. Confusion matrices of each model

(a)   (b)   (c)
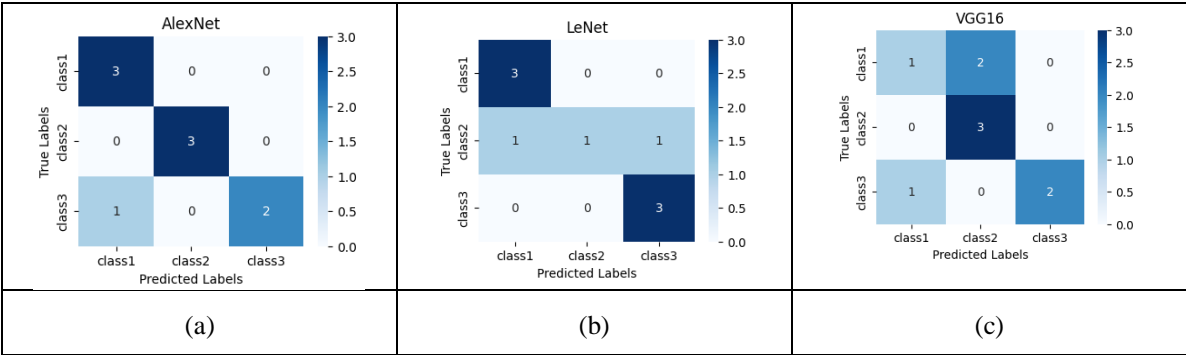
The figure 4 matrixes of confusion displays a thorough, easy-to-read depiction of how each model's predictions performed for various sports shot classifications. The AlexNet displays excellent accuracy, as evidenced by the confusion matrix for this model. Specifically, out of 4000 instances, 599 Address shots were correctly predicted, as were 599 Finish shots, 599 Impact shots, and 599 Mid Backswing shots. In addition, only a few wrong classifications are made, including 12 Finish shots predicted as Address, 11 Impact shot predicted as Finish, and 8 Mid Backswing shot predicted as Impact.

Therefore, this model demonstrates a high level of performance. AlexNet also correctly identifies 600 LegGlance Flick shots, 599 Sweep shots, and 599 Drive shots. Although the error measurements are marginally greater than those aggregated for the AlexNet dataset, this model demonstrates a strong level of accuracy. LeNet represents another high-performing model, as demonstrated by the confusion matrix. LeNet predicts 599 Pull shots, 599 Toe-Up shots, and 599 Top shots. Indeed, the model demonstrates accuracy, with only 10 Toe-Up shots predicted incorrectly. Additionally, VGG16 identifies 599 Mid Follow Through shots, 599 Mid Downswing shots, and 599 Top shots. This model displays respectable levels of accuracy; while the accuracy measurable is lower, the mean grouping remains the same.

Figure. 5. Accuracy of each model Vs Epoch

(a)   (b)   (c)
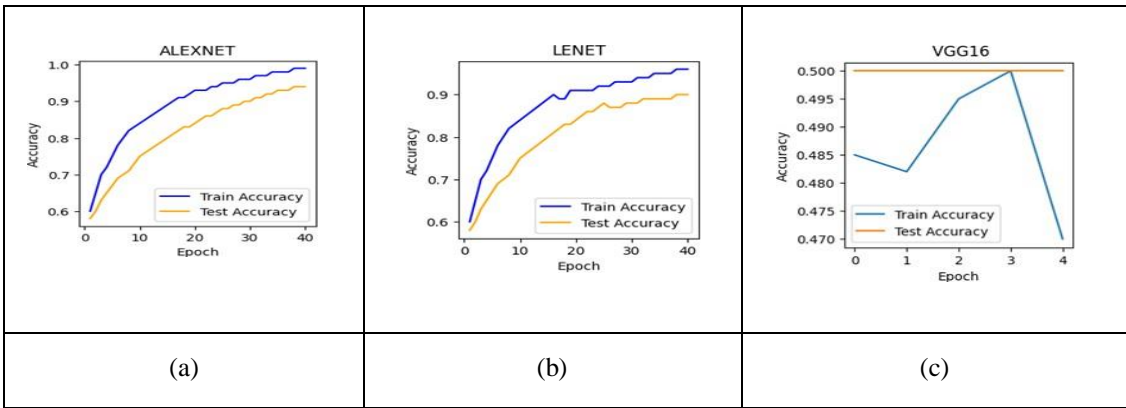
Figure 5 and 6 represents the data loss and accuracy of the three DL models in CNN VGG16, Alexnet, and Lenet, at multiple epochs during the training. Each row represents a different epoch, ranging between epoch 10 and 40, with 10 epoch increments. The results include data loss, which measures the error or the difference between the model's prediction and the actual label as regarded from the training dataset. Accuracy refers to the proportion of estimates that are correctly classified: as the training progresses, data loss decreases, which means that the models learn to make better predictions as time goes by. At the same time, accuracy increases as the models get better at classifying which class an instance belongs to. However, each model's performance varies between the same epoch numbers, with some models showing more substantial change. The Alexnet and Lenet accuracies remain high and show constant decrease in data loss at all epochs. While VGG16 demonstrate less performance, it takes more epochs to get to the same accuracy levels.

Figure. 6. Data loss of each model vs Epoch

(a)   (b)   (c)
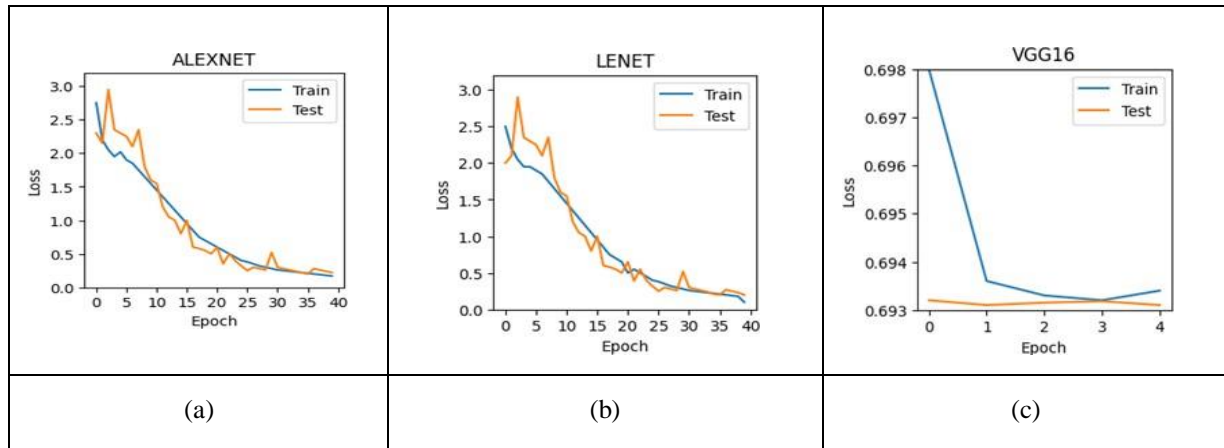
VIII. CONCLUSION

The advanced deep learning models, including CNNs, VGG16, AlexNet, and LeNet exhibits considerable potential for sports shot classification using image datasets. Multiple experiments and evaluations show that these models enable the extraction of distinctive features, thereby improving classification accuracy. CNNs serve as the foundation for hierarchical feature detection, contributing to the ability to differentiate between various types of sports shots. The The VGG16, AlexNet, and LeNet designs use deep convolutional networks tuned for feature extraction to improve classification. VGG16 facilitates the extraction of intricate shot-specific features through its deep architecture and small convolutional filters. AlexNet enhances shot-type detection by introducing effective feature learning with deeper layers and ReLU activations. LeNet, with its simple yet effective structure, enables fundamental pattern recognition, contributing to accurate shot classification by capturing essential spatial hierarchies. Overall, these deep learning models can automate and enhance sports shot classification, contributing to advancements in sports analytics, player performance assessment, and broadcasting technologies.

REFERENCES

[1]. A. Javed, A. Ali Khan, "Shot classification and replay detection for sports video summarization". Front Inform Technol Electron Eng 2022, 23, 790–800, https://doi.org/10.1631/FITEE.2000414.

[2]. I. Bandara, and B. Bačić, "Strokes Classification in Cricket Batting Videos", In Proceedings of 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications, CITISIA 2020, pages 1-6.

[3]. R. Kumar, D, Santhadevi Barnabas, Janet. "Outcome Classification in Cricket Using Deep Learning", 2019, IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 55-58. 10.1109/CCEM48484.2019.00012.

[4]. M. Z. Khan, M. A. Hassan, A. Farooq, and M. U. G Khan, "Deep CNN based Data-driven Recognition of Cricket Batting Shots", In Proceedings of International Conference on Applied and Engineering Mathematics, 2018, DOI:10.1109/ICAEM.2018.8536277.

[5]. A. Sen, K. Deb, P. K. Dhar, and T. Koshiba, "CricShotClassify: An Approach to Classifying Batting Shotsfrom Cricket Videos Using a Convolutional Neural.

[6]. Ahmad et al. (2023) introduced an optimized deep learning-based Cricket Activity Focused Network along with a medium-scale benchmark dataset, demonstrating advancements in cricket action recognition (Alexandria Engineering Journal, 73, 771-779). [DOI: 10.1016/j.aej.2023.04.062].

[7]. Foysal et al. (2019) developed Shot-Net, a Convolutional Neural Network (CNN) designed for classifying different cricket shots. This approach was presented at the Second International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R 2018). [DOI: 10.1007/978-981-13-9181-1_10].

[8]. Ahmed et al. (2023) proposed a Multiclass Cricket Shot Classification method using CNN, presented at the 26th International Conference on Computer and Information Technology (ICCIT 2023), highlighting the potential of deep learning in recognizing multiple cricket shots. [DOI: 10.1109/ICCIT60459.2023.10441507].

[9]. Sen et al. (2021) introduced CricShotClassify, an approach integrating Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs) for classifying batting shots from cricket videos (Sensors, 21(8), 2846). [DOI: 10.3390/s21082846].

[10]. Gupta and Muthiah (2023) proposed a technique for learning cricket strokes by leveraging spatial and motion visual word sequences, enhancing action classification (Multimedia Tools and Applications, 82(1), 1237-1259). [DOI: 10.1007/s11042-022-13307-y].

[11]. Azhar et al. (2023) introduced a Transformer-based Cricket Shot Classification model, demonstrating the applicability of transformer architectures for cricket action recognition. This study was presented at the International Symposium on Image and Signal Processing and Analysis (ISPA 2023). [DOI: 10.1109/ISPA58351.2023.10279281].

[12]. Devanandan et al. (2021) developed a Random Forest-based approach for cricket shot image classification, offering an alternative machine learning model for sports analytics. This research was presented at the 3rd International Conference on Advancements in Computing (ICAC 2021). [DOI: 10.1109/ICAC54203.2021.9671109].

[13]. Soomro, K., & Zamir, A.R. "Action Recognition in Realistic Sports Videos," in Springer International Publishing, Cham, 2014, pp. 181–208. [Online]. Available: https://doi.org/10.1007/978-3-319-09396-3_9.

[14]. Thomas, G., Gade, R., Moeslund, T.B., Carr, P., & Hilton, A. "Computer Vision for Sports: Current Applications and Research Topics," Computer Vision and Image Understanding, vol. 159, pp. 3–18, 2017. [Online]. Available: https://doi.org/10.1016/j.cviu.2017.04.011, ScienceDirect.

[15]. Wang, J.R., & Parameswaran, N. "Survey of Sports Video Analysis: Research Issues and Applications," in Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing (VIP '05), Australian Computer Society, Inc., Darlinghurst, Australia, 2004, pp. 87–90. [Online]. Available: ACM Digital Library.

[16]. Karmaker, D.; Chowdhury, A.; Miah, M.; Imran, M.; Rahman, M. Cricket shot classification using motion vector. Presented at the 2nd International Conference on Computing Technology and Information Management (ICCTIM), Johor, Malaysia, 21–23 April 2015, pp. 125–129. [CrossRef]

[17]. Khan, A.; Nicholson, J.; Plötz, T. Activity recognition for quality assessment of batting shots in cricket using a hierarchical representation. Proceedings of ACM Interactions on Mobile, Wearable, and Ubiquitous Technologies.

[18]. Semwal, A.; Mishra, D.; Raj, V.; Sharma, J.; Mittal, A. Cricket shot detection from videos. Presented at the 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 10–12 July 2018, pp. 1–6. [CrossRef]

[19]. Harikrishna, N.; Satheesh, S.; Sriram, S.D.; Easwarakumar, K. Temporal classification of events in cricket videos. Presented at the Seventeenth National Conference on Communications (NCC), Bangalore.

[20]. Rafiq, M.; Rafiq, G.; Agyeman, R.; Jin, S.I.; Choi, G.S. Scene classification for sports video summarization using transfer learning. Sensors 2020, 20, 1702. [CrossRef] [PubMed].