OPEN ACCESS 118

ISSN 2356-5462
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 11, No. 2, Dec 2025. pp. 118-135
doi: 10.21108/ijoict.v11i2.9622

Pose-Based Action Recognition in Tennis Using MediaPipe and LSTM

Walid Hanif Ataullah[1]*, Isa Mulia Insan[2], Sheina Fathur Rahman[3]

[1,2,3]*School of Computing, Telkom University*
Bandung, Indonesia

* walidhanifataullah@student.telkomuniversity.ac.id

Abstract

Pose recognition in tennis is an essential aspect for analysing playing techniques and evaluating athlete performance. This study develops a tennis pose recognition system that integrates MediaPipe for pose feature extraction with Long Short-Term Memory (LSTM) networks for movement classification. The research dataset consists of 2,010 images of tennis movements across four categories: backhand, forehand, ready position, and serve, annotated in COCO format. MediaPipe successfully extracted pose landmarks from 1,782 images (88.7%), generating 33 pose landmarks flattened into a 99-dimensional feature vector. The LSTM model is designed with a 3-layer LSTM architecture and 2 dense layers, trained using a stratified train-test split with an 80:20 ratio. Model evaluation uses various metrics including accuracy, precision, recall, and F1-score. The results show that the system achieves 90.20% accuracy, with the best performance in the ready position category (F1-score: 91.28%) and the lowest in the forehand category (F1-score: 88.89%). The model demonstrates good computational efficiency with a memory footprint of 714.39 KB, enabling deployment on mobile devices. This study contributes to the development of automated sports analysis systems and demonstrates the feasibility of integrating MediaPipe-LSTM for real-time tennis pose recognition applications.

Keywords: MediaPipe, LSTM, Deep learning, Computer vision, Pose Classification

I. INTRODUCTION

The recent proliferation of computer vision and machine learning technologies has catalysed a paradigm shift in sports analytics, particularly within precision-oriented disciplines such as tennis, golf, and gymnastics. These advanced computational tools provide a framework for objective, data-driven performance evaluation, fundamentally altering traditional coaching methodologies, injury prevention strategies, and tactical planning [1], [2]. The integration of these technologies empowers coaches and athletes with deep, real-time insights, fostering an environment of continuous improvement and optimized performance. Key benefits include enhanced performance analysis through detailed biomechanical breakdowns, modernization of training programs with personalized feedback, and proactive injury prevention by monitoring athlete fatigue and movement patterns [3], [4], [5]. This technological evolution has established a new standard in sports science, where decisions are increasingly informed by robust empirical data rather than subjective observation alone [6].

A cornerstone of this transformation is human pose estimation (HPE), a critical subfield of computer vision focused on identifying and tracking anatomical keypoints to analyse movement. Frameworks such as MediaPipe, OpenPose, and MoveNet have become instrumental in this domain, each offering a unique balance of accuracy, processing speed, and implementation simplicity. MediaPipe, developed by Google, has distinguished itself through its high efficiency and accessibility, demonstrating robust performance on low-

Received on 18 Jul 2025. Revised on 28 Aug 2025. Accepted and Published on 1 Dec 2025.

power devices and making it ideal for real-time applications [7]. Studies have reported its high accuracy, achieving 92-96% in tracking weightlifting exercises and maintaining reliability even in challenging conditions [8], [9]. While OpenPose excels in multi-person pose estimation, crucial for team sports analytics, it often entails greater computational overhead and implementation complexity [10], [11]. Conversely, MoveNet offers the fastest processing speeds, making it a prime choice for applications demanding minimal latency [12]. For this research, MediaPipe was selected for its optimal blend of high accuracy, real-time capability, and user-friendly implementation, which aligns with the goal of developing an accessible and effective sports analysis tool.

Following feature extraction via pose estimation, the classification of complex, sequential movements necessitate a sophisticated temporal modelling approach. Long Short-Term Memory (LSTM) networks, a specialized type of recurrent neural network (RNN), have proven exceptionally effective for this task due to their inherent ability to capture long-range dependencies in sequential data. The architecture has been successfully implemented across numerous sports for action classification, consistently delivering high performance. For instance, in Taekwondo, an LSTM-based model achieved a remarkable 99.10% accuracy in real-time technique assessment [13]. Similarly, in tennis, a Bidirectional LSTM (Bi-LSTM) model combined with transfer learning reached 96.72% accuracy, outperforming previous methods [14]. Furthermore, advanced variants like Attention-based Bi-LSTM have demonstrated state-of-the-art results, achieving 99.87% accuracy on benchmark sports datasets [15]. The consistent success of LSTM and its variants in accurately classifying nuanced athletic movements based on pose data underscores their suitability for developing a high-fidelity action recognition system [16], [17].

Despite these technological advancements, the persistence of specific research gaps highlights unresolved challenges in the practical application of these systems. These unresolved challenges culminate in a significant problem: the absence of an accessible, accurate, and real-time system for tennis motion analysis tailored to the needs of specific communities, such as in Indonesia.

The core research problem stems from three interconnected issues. First, there is a distinct lack of implemented pose recognition systems within the Indonesian sports landscape, creating a barrier to adopting modern, data-driven coaching techniques for local athlete development. Second, existing models often struggle with the technical challenge of accurately differentiating between visually similar tennis strokes (e.g., certain phases of forehand and backhand), a crucial requirement for providing meaningful and granular feedback to players. Finally, this problem is exacerbated by a reliance on limited or non-localized datasets, which can negatively impact a model's performance and its ability to generalize to athletes and conditions within a specific regional context.

In response to this multifaceted problem, this study sets forth a clear objective to develop and evaluate a targeted solution. The primary aim of this research is to design, implement, and validate an integrated tennis action recognition system by synergizing the MediaPipe framework for real-time pose feature extraction with a deep Long Short-Term Memory (LSTM) network for robust temporal classification. This study will focus on accurately classifying four fundamental tennis movements: backhand, forehand, ready position, and serve. The evaluation will prioritize not only high classification accuracy but also computational efficiency to ensure the model's viability for practical, real-world deployment.

The novelty of this research lies in three key areas. First, it proposes a specific methodological integration of MediaPipe's lightweight architecture with a tailored LSTM network, aiming to strike an optimal balance between high-fidelity motion classification and computational efficiency for a resource-constrained environment. Second, this study is among the first to develop and validate such a system specifically within the Indonesian context, directly addressing the local need for accessible and advanced sports technology. Finally, its distinct emphasis on creating a model with a small memory footprint provides a practical contribution by bridging the gap between theoretical academic research and tangible coaching tools, making the system potentially deployable on widely available mobile devices.

II. LITERATURE REVIEW

A. Computer Vision

Computer Vision is an interdisciplinary field within artificial intelligence dedicated to enabling machines to interpret and understand the visual world in a manner akin to human perception [18], [19]. The field has undergone a remarkable evolution, transitioning from traditional, manually crafted algorithms for image processing to modern deep learning-based approaches [20], [21]. Early methods were heavily reliant on handcrafted features, which, while effective for well-defined problems, lacked the flexibility to generalize across varied and complex visual scenarios [22], [23]. The integration of machine learning marked a significant shift towards data-driven models, yet it was the advent of deep learning, particularly Convolutional Neural Networks (CNNs), that truly revolutionized the domain [24]. Inspired by the human visual cortex, deep learning models can automatically learn hierarchical features directly from raw pixel data, dramatically enhancing performance in tasks such as image classification, object detection, and semantic segmentation [20], [25], [26].

A fundamental subdomain within computer vision, particularly relevant to human movement analysis, is pose estimation. This technology aims to detect and localize the key joints of the human body from images or videos to reconstruct a skeletal model that represents a person's posture [27], [28], [29]. Pose estimation techniques are broadly categorized into marker-based and markerless approaches. While traditional marker-based systems offer high accuracy, they are intrusive and can impede natural movement, making them impractical for real-world sports analysis. In contrast, markerless techniques leverage computer vision algorithms to analyze visual features directly, allowing for an unobtrusive and naturalistic assessment of movement [30]. The advancement of deep learning has enabled markerless pose estimation to be performed in real-time, which is crucial for applications requiring immediate feedback, such as athletic coaching and interactive fitness systems [27], [31].

The application of markerless pose estimation has become a valuable tool in sports performance analytics, offering deep insights into athletic technique and strategy. It is used to dissect the shooting posture of basketball players [32], evaluate the swing quality of baseball batters [33], and analyse the complex stroke dynamics of tennis players [3]. Despite its transformative potential, the technology faces significant technical challenges in the dynamic context of sports. High-speed movements common in athletic activities often result in motion blur, which complicates accurate tracking [34]. Furthermore, object occlusion, where athletes are partially or fully obscured by others, remains a complex problem, particularly in team sports [35]. Finally, variable and uncontrolled lighting conditions in sports venues introduce additional complexity, affecting the consistency and reliability of visual data analysis [36]. Addressing these challenges is crucial for advancing the practical application of computer vision in sports.

B. Deep Learning and Machine Learning

Machine Learning (ML), a fundamental subfield of artificial intelligence, equips systems with the ability to autonomously learn from data and improve their performance over time without being explicitly programmed [37], [38]. It operates on three main paradigms: supervised, unsupervised, and reinforcement learning. Supervised learning, in particular, is highly relevant for classification and prediction tasks, as it trains models on labelled datasets to map input data to a correct output [39], [40]. The effectiveness of this paradigm is well-documented, with algorithms like Support Vector Machines and Random Forests consistently achieving high accuracy in domains ranging from medical diagnostics to image recognition [41], [42].

Deep Learning (DL) represents an advanced and powerful evolution of machine learning, distinguished by its use of multi-layered neural network architectures inspired by the human brain's structure [43], [44]. The core strength of deep learning lies in its capacity for hierarchical feature learning, where the model automatically and progressively extracts increasingly complex representations from raw data. Lower layers might identify simple features like edges and colours, while deeper layers learn to recognize intricate patterns such as textures, shapes, and eventually entire objects [45], [46]. This automated feature engineering process is a significant advantage over traditional ML methods, which often require extensive and time-consuming manual feature extraction [47]. Consequently, deep learning models exhibit superior performance in complex pattern recognition tasks involving high-dimensional data, such as image and sequence analysis [48].

The efficacy of deep learning techniques, especially Convolutional Neural Networks (CNNs), in multi-class visual recognition is strongly supported by empirical evidence. Modern architectures consistently demonstrate state-of-the-art performance across diverse and challenging datasets. For instance, a ResNet model has achieved an accuracy of 99% in classifying grape varieties, while an Adapted Deep Convolutional Neural Network (ADCNN) reached 99.73% accuracy in hand gesture recognition [49]. Further studies using networks like Inception-v1 on medical imaging datasets have reported high performance not only in accuracy but also across comprehensive metrics including AUC, precision, and recall, validating their classification robustness [50]. This consistent delivery of high-accuracy results, often enhanced by techniques like data augmentation, solidifies the role of deep learning as the foundational technology for solving complex real-world visual classification problems [51], [52].

C. Long Short-Term Memory

Recurrent Neural Networks (RNNs) are designed to process sequential data, yet standard architectures are often hindered by the vanishing gradient problem, which severely limits their ability to capture and learn long-term dependencies. To address this limitation, Long Short-Term Memory (LSTM) networks were introduced as an advanced solution, offering a unique design that excels at modeling temporal sequences [53], [54]. The central element of the LSTM is its memory cell, which is specifically designed to preserve information over long horizons. This memory cell is controlled by three distinct gating mechanisms: the input gate, the forget gate, and the output gate. Together, these gates regulate the flow of information by determining what to store, what to discard, and what to output, thereby enabling the network to selectively retain relevant signals while filtering out noise. This architecture ensures stable gradient flow during training, which is essential for learning long-range dependencies that traditional RNNs typically fail to capture [55], [56]. Consequently, LSTMs are inherently suitable for tasks where temporal order plays a critical role, such as sequential prediction and event classification [57], [58]. In order to illustrate this mechanism more clearly, the general structure of the LSTM is presented in Figure 1, which depicts the internal dynamics of the gating system and its interaction with the memory cell.

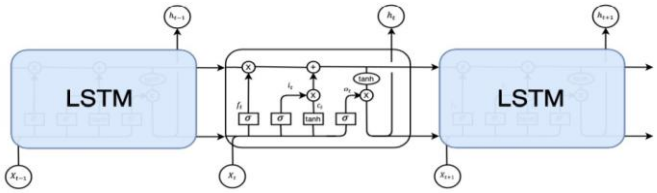

Fig. 1. Long Short-Term Memory Architecture

The figure demonstrates how information flows through the LSTM unit: the input gate controls which new information enters the cell state, the forget gate decides which information should be erased from memory, and the output gate regulates which part of the information is propagated to the next layer. This design allows the LSTM to effectively mitigate the vanishing gradient problem by maintaining and updating information across many time steps, thereby providing the network with the ability to capture both short-term and long-term dependencies within sequential data.

The practical effectiveness of this architecture has been validated across a wide range of applications involving both prediction and classification. In the field of Natural Language Processing (NLP), LSTMs have achieved remarkable success, particularly in sentiment analysis tasks using social media data. Their ability to preserve contextual information across sequences enables a nuanced understanding of textual narratives, which is crucial for accurately capturing sentiment in dynamic online conversations [59], [60]. Empirical studies report that LSTM-based models outperform traditional classifiers, with accuracies reaching up to 84% on Twitter datasets [61], [62]).

Beyond NLP, LSTMs have been widely applied to time-series forecasting, especially in domains characterized by volatility and non-linear dependencies such as financial markets. Their capability to learn complex temporal patterns in historical data allows them to surpass conventional statistical models like ARIMA

in predicting stock prices and indices [63], [64]. Several studies highlight that LSTM models can achieve high prediction accuracy, with reported results of up to 91.97% on stock market test datasets (Agarwal et al., 2024). Moreover, they exhibit notable robustness during periods of high volatility, a condition under which traditional models typically fail, thereby showcasing their strength in capturing intricate, non-linear temporal dynamics.

The adaptability of LSTM models extends further into other forecasting domains where the understanding of temporal dependencies is essential. In the energy sector, LSTM models have been successfully implemented to predict electricity load demand by learning complex temporal cycles, including daily, weekly, and seasonal patterns—an ability crucial for efficient power grid management and resource distribution (Kouziokas, 2019). Similarly, in the telecommunications industry, LSTMs have been applied for revenue forecasting, which is highly challenging due to fluctuating market conditions and customer behaviour. By analysing large volumes of historical data, LSTMs are able to uncover long-term trends and subtle behavioural dynamics, producing significantly more accurate earnings projections than traditional methods [65]. These diverse applications highlight the versatility of LSTM networks and confirm their position as one of the most effective deep learning architectures for sequential data analysis across different domains between 2020 and 2025.

D. MediaPipe Integration with Deep Learning

Google's MediaPipe framework has emerged as a versatile and powerful solution for real-time, on-device pose estimation, distinguished by its high efficiency, accuracy, and model flexibility [66], [67]. It is optimized for high-speed processing, achieving impressive frame rates even on resource-constrained hardware, which makes it an ideal foundation for interactive applications [68]. The framework's models have consistently demonstrated high accuracy across various domains, such as fitness tracking, where it has reached 92-96% accuracy in classifying exercises, and in human action recognition, where it serves as a robust feature extractor [8].

The integration of a specialized feature extraction tool like MediaPipe with deep learning models for classification, such as Long Short-Term Memory (LSTM) networks or Convolutional Neural Networks (CNNs), creates a powerful hybrid approach. This synergy leverages the strengths of each component: MediaPipe excels at providing precise and detailed spatial landmark data, which is then fed into a deep learning model to interpret the temporal context or classify the pose. This combination significantly enhances system accuracy and robustness. For instance, a hybrid system integrating MediaPipe with a Bidirectional LSTM (BiLSTM) for human activity recognition achieved an impressive accuracy of 97.05% on the KTH dataset, showcasing the effectiveness of this layered approach [69].

Recent studies provide extensive evidence of this successful integration across various applications. In the field of Human Action Recognition (HAR), the combination of MediaPipe for keypoint extraction and an LSTM network for temporal modelling has yielded state-of-the-art results, achieving 92.40% accuracy on the UCF101 dataset and 86.8% on Kinetics 400 [70]. This methodology has also been successfully applied to more specific tasks, such as real-time Indian Sign Language recognition, where MediaPipe Holistic extracts comprehensive hand and body landmarks that are subsequently processed by an LSTM network (Shinde et al., 2024). Similarly, in sports and fitness analytics, this hybrid model is used for yoga pose analysis and gym exercise classification, where MediaPipe's pose data is classified by models like CNN-GRU or EfficientNetV2 to provide real-time feedback [71]. More advanced integrations even represent human poses as graphs, which are then analysed by Graph Neural Networks (GNNs) for dynamic sports action recognition. In this approach, key points serve as nodes and skeletal connections as edges, allowing GNNs to effectively model the spatial relationships between body parts. This method has shown exceptional performance in recognizing complex activities within challenging datasets, such as those from Olympic sports [72], as it provides a richer, more structured representation of human posture compared to raw coordinate sequences. In summary, the integration of MediaPipe's high-fidelity feature extraction with the sophisticated classification capabilities of deep learning networks has proven to be a highly effective and versatile strategy. This approach creates a robust pipeline capable of transforming raw visual data into meaningful, actionable insights, applicable across a wide spectrum of domains from elite sports analytics to clinical rehabilitation and interactive human-computer interfaces. The synergy between precise, efficient landmark detection and powerful temporal or spatial classification models

paves the way for the continued development of advanced, accessible, and impactful applications for human movement analysis.

E. LSTM in Temporal Prediction and Classification

Long Short-Term Memory (LSTM) networks have demonstrated profound effectiveness in both prediction and classification tasks involving sequential and time-series data. Their architectural design is specifically engineered to overcome the limitations of traditional Recurrent Neural Networks (RNNs), primarily by addressing the vanishing gradient problem and adeptly managing long-term dependencies [53], [54], [55]. The core of this capability lies in LSTM's unique gating mechanisms—the input, forget, and output gates—which intelligently regulate the flow of information through the network's memory cells. This allows the model to selectively retain relevant historical data over extended sequences while discarding irrelevant information, a crucial function for accurately identifying complex temporal patterns [58].

The practical application of this architecture is extensive and highly successful. In the domain of temporal classification, LSTMs have become a cornerstone for sentiment analysis of social media data. By processing text sequentially, LSTMs can capture the evolving context and nuanced dependencies between words, which is essential for interpreting informal language, slang, and sarcasm. Empirical studies consistently report superior performance, with LSTM-based models achieving high accuracy, precision, and recall metrics that significantly outperform conventional machine learning algorithms on challenging datasets like Twitter [60], [61], [62]. Performance is often further enhanced by advanced variants like Bidirectional LSTMs (Bi-LSTMs) and the integration of attention mechanisms, which allow the model to focus on the most influential parts of the text [73], [74].

Similarly, in the realm of time-series forecasting, LSTMs have proven to be an indispensable tool, particularly for modelling complex, non-linear, and volatile data such as stock market prices. Unlike traditional linear models like ARIMA which struggle with such dynamics, LSTMs can effectively learn intricate, long-term patterns from historical data [63], [64]. Numerous comparative studies have validated their superiority in financial forecasting, showing consistently lower error rates and higher predictive accuracy, even during periods of high market fluctuation [75]. The robustness of LSTMs is further amplified in hybrid models, which combine them with other statistical or deep learning techniques to achieve even greater precision [76]. In summary, the proven ability of LSTMs to manage long-range dependencies and model complex temporal dynamics makes them a versatile and powerful tool for both classification and prediction tasks across diverse fields.

III. RESEARCH METHOD

This research uses quantitative experimental design with a system development approach. The research was conducted in several systematic stages including dataset collection, data preprocessing, feature extraction using MediaPipe, LSTM model design and training, and system performance evaluation. This approach was chosen to ensure the validity and reliability of the research results and allow replication by other researchers.

The independent variable in this study is the pose feature extracted using MediaPipe from the tennis motion images, while the dependent variable is the tennis motion classification consisting of four categories: backhand, forehand, ready position, and serve. Control variables include lighting conditions, image resolution, and subject distance from the camera which are standardized to ensure data consistency.

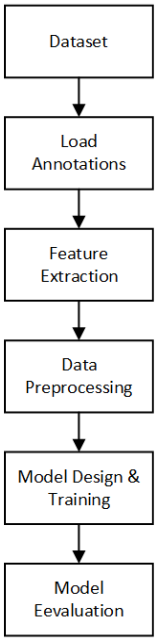

Fig. 2. Research Flow

A. Dataset and Data Collection

The research dataset consists of 2,010 tennis motion images divided into four motion categories: backhand (500 images), forehand (500 images), ready position (510 images), and serve (500 images). The data was collected in COCO (Common Objects in Context) format with key point annotations that have been labeled using JSON format. Each image has a resolution of 1280x720 pixels and is saved in JPEG format with high quality to ensure feature extraction accuracy.

The dataset structure is organized in separate directories for each motion category, with corresponding JSON annotation files. Each annotation contains pose key point coordinate information in [x, y, visibility] format for 18 pre-annotated key points of the human body. The data is stored in Google Drive with an organized folder structure.

B. Pose Detection and Feature Extraction

MediaPipe was configured with the following parameters for the extraction of pose landmarks:

- static_image_mode = True for static image processing
- model_complexity = 1 for balance between accuracy and speed
- enable_segmentation = False to focus on pose detection
- min_detection_confidence = 0.5 as minimum detection threshold

The pose detection process is performed using MediaPipe Pose which is able to detect 33 pose landmarks from each image. Each image was converted from BGR to RGB format using OpenCV before being processed by MediaPipe. The detection results are normalized (x, y, z) coordinates in the range 0-1, where z represents the depth relative to the hip plane.

The feature extraction process is done in several systematic stages, Image Loading and Preprocessing: Each image was converted from BGR to RGB format using OpenCV. Pose Detection: MediaPipe processes the image to extract 33 pose landmarks. Coordinate Normalization: Landmarks are normalized relative to the image dimensions. Feature Vector Creation: Landmarks are flattened into a 99-dimensional vector (33 landmarks × 3 coordinates). Statistical Normalization: Application of z-score normalization to standardize the data distribution. Each successfully extracted feature is then normalized using z-score normalization to standardize the data distribution and improve model training performance.

C. Data Preprocessing and Feature Extraction

Data preprocessing includes several important steps to prepare the training data. Label encoding was performed using LabelEncoder from scikit-learn to convert motion categories into numerical format, followed by one-hot encoding using to_categorical for compatibility with softmax output. The data was divided using stratified sampling to maintain the proportion of each class with a division of 80% for the training set (1,425 samples) and 20% for the testing set (357 samples) with random state 42 for reproducibility.

Data reshaping was done to adjust the LSTM input to the format (samples, timesteps, features) where timesteps=1 since static images were used. This resulted in input dimensions (batch_size, 1, 99) that fit the designed LSTM architecture.

D. LSTM Architecture Design

The LSTM model is designed with the following architecture:

- Input Layer: Receives pose features with dimensions (batch_size, timesteps=1, features=99)
- LSTM Layer 1: 128 units with return_sequences=True and dropout 0.3
- LSTM Layer 2: 64 units with return_sequences=True and dropout 0.3
- LSTM Layer 3: 32 units with a dropout of 0.3
- Dense Layer 1: 64 units with ReLU activation and dropout 0.4
- Dense Layer 2: 32 units with ReLU activation and dropout 0.3
- Output Layer: 4 units with softmax activation for multi-class classification

The model was compiled with Adam optimizer (learning rate 0.001), loss function categorical crossentropy, and accuracy metrics. Training is done with batch size 32, maximum 50 epochs, equipped with EarlyStopping (patience=10) and ReduceLROnPlateau (factor=0.5) callbacks for training process optimization.

E. Model Evaluation

Model evaluation uses various metrics for comprehensive analysis of system performance including accuracy as the proportion of correct predictions out of total predictions, precision as the proportion of true positives out of total positive predictions per class, recall as the proportion of true positives out of total actual positives per class, F1-Score as the harmonic mean of precision and recall, and confusion matrix for detailed analysis of classification per class to identify pattern misclassification. Evaluation was conducted on the separated test set with performance analysis per class to identify the strengths and weaknesses of the model in each tennis movement category, thus providing an in-depth understanding of the model's ability to distinguish between backhand, forehand, ready position, and serve movements.

IV. RESULTS AND DISCUSSION

A. Feature Extraction Results with MediaPipe

The pose feature extraction process using MediaPipe was successfully performed on a dataset consisting of 2,010 tennis motion images. Of the total processed images, MediaPipe successfully extracted pose landmarks in 1,782 images (88.7%), while 228 images (11.3%) failed the extraction process due to various factors such as low image quality, incomplete poses, or suboptimal lighting conditions. Table 1 shows the distribution of feature extraction success per motion category, where ready position has the highest success rate (92.7%) and serve has the lowest success rate (79.6%).

TABLE I
MEDIAPIPE FEATURE EXTRACTION SUCCESS DISTRIBUTION

Movement Category	Total Images	Successful Extraction	Extraction Failure	Success Rate
Backhand	500	458	42	91.6%
Forehand	500	453	47	90.6%
Ready Position	510	473	37	92.7%
Serve	500	398	102	79.6%
Total	**2,010**	**1,782**	**228**	**88.7%**

Analysis of the feature extraction quality shows that MediaPipe generates 33 pose landmarks with normalized (x, y, z) coordinates. The results of extracting pose landmarks for each motion category show that MediaPipe is able to detect pose key points with good accuracy under optimal image conditions. The extracted pose features are then flattened into 99-dimensional vectors ready to be used as input for the LSTM model.

B. Data Preprocessing and Distribution Analysis

The data preprocessing stage produces a final dataset of 1,782 samples distributed in four classes with a relatively balanced proportion.

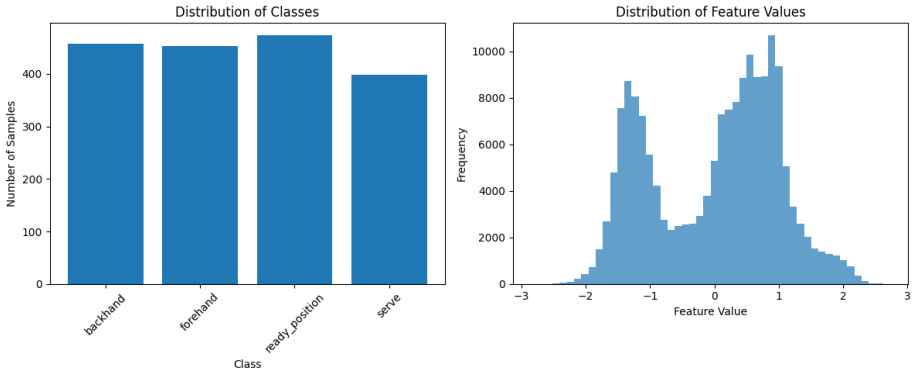

Fig. 3. Dataset Class Distribution After Preprocessing

Figure 3 displays the class distribution after preprocessing, where ready position has the highest number of samples (473 samples) and serve has the least number of samples (398 samples). The normalization process using z-score successfully standardizes the feature distribution, which is shown in the second Figure with a histogram of the feature value distribution before and after normalization.

Data division using stratified sampling resulted in a training set of 1,425 samples (80%) and a testing set of 357 samples (20%). The distribution of samples per class in the training set and testing set, ensured a balanced proportion for each movement category. Stratified sampling successfully maintained a proportional representation of each class in both data subsets.

C. LSTM Model Training Results

Training the LSTM model was performed for a maximum of 50 epochs using EarlyStopping and ReduceLROnPlateau callbacks to optimize the learning process.

The model reaches convergence at the 28th epoch with EarlyStopping activated as there is no increase in validation loss for 10 consecutive epochs. Figure 4 shows the learning curve in the form of training loss and validation loss graphs during the training process, while the second Figure displays the training accuracy and validation accuracy curves.

The learning curve analysis shows that the model learns well without experiencing significant overfitting. Training loss decreased consistently from 1.386 in the first epoch to 0.306 in the last epoch, while validation loss decreased from 1.298 to 0.306. Training accuracy increased from 25.2% in the first epoch to 90.2% in the last epoch, with validation accuracy reaching 90.2%. The model showed good stability with minimal gap between training and validation metrics.

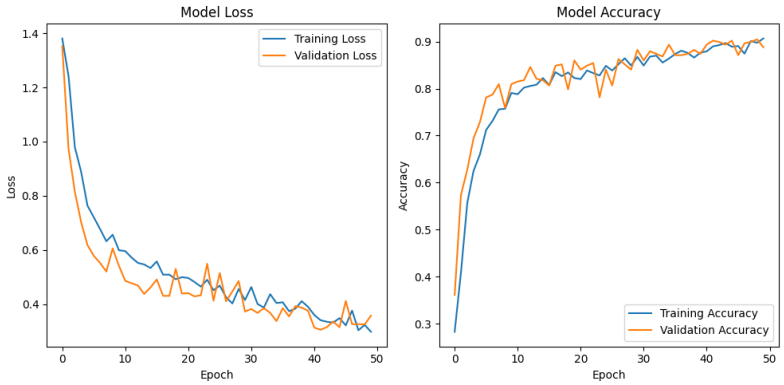

Fig. 4. Kurva Learning - Training Loss dan Validation Loss

D. Visual Analysis of Model Testing Results

To test the model's performance visually and practically, testing was conducted using random image samples from each tennis movement category. Testing is done by taking 3 random images from each category (backhand, forehand, ready position, and serve) to see the model's ability to make predictions on data that has never been seen before.

Fig. 5. Tennis Pose Backhand

The Backhand category performed very consistently with 100% accuracy and an average confidence score of 0.9950, indicating that the model has very high confidence in identifying backhand movements. Figure 5 displays the prediction results for 3 backhand samples that were correctly classified, with confidence scores ranging from 0.9936 to 0.9970.

Fig. 6. Tennis Pose Forehand

The Forehand category achieved 100% accuracy but with a wider variety of confidence scores (average 0.8794), where one sample showed a relatively low confidence score (0.6498) compared to the other two

samples that achieved confidence above 0.98. Figure 6 shows an example of a forehand prediction with confidence variations that reflect the complexity of this motion detection.

Fig. 7. Tennis Pose Ready Postion

The Ready Position category also achieved 100% accuracy with an average confidence score of 0.8718. Figure 7 shows the prediction results which indicate that even though one sample has a relatively low confidence score (0.6577), the model still manages to classify correctly as it still has the highest probability compared to other categories.

Fig. 8. Tennis Pose Serve

The Serve category faced the biggest challenge with only 1 out of 3 images successfully processed, resulting in a testing accuracy of 33.3%. Two images failed the pose detection stage by MediaPipe, demonstrating the complexity of serve poses which often involve extreme movements or poses that are difficult to detect. Figure 8 shows an example of a successful serve prediction with a high confidence score (0.9661).

The testing results show a pattern consistent with the model evaluation on the test set, where the backhand category has the highest and most consistent confidence score, while the serve category has difficulty in the preprocessing stage (pose detection). The probability distribution for each prediction shows that the model generally assigns a very low probability to the wrong category, indicating a clear decision boundary between categories. These visual tests validate the robustness of the model under real-world conditions and provide insight into the characteristics of each movement category from the perspective of the developed system.

E. Model Performance Evaluation

Evaluation of the model on the testing set resulted in satisfactory performance with a test accuracy of 90.20%. Table 2 presents the results of the comprehensive evaluation of the model using various metrics, including precision, recall, and F1-score for each movement category. The model showed the best performance in the serve category with an F1-score of 89.82% and the lowest performance in the forehand category with an F1-score of 88.89%.

TABLE 2
HASIL EVALUASI PERFORMA MODEL PER KATEGORI

Movement Category	Precision	Recall	F1-Score	Support
Backhand	92.13%	89.13%	90.61%	92
Forehand	95.00%	83.52%	88.89%	91
Ready Position	89.00%	93.68%	91.28%	95
Serve	85.23%	94.94%	89.82%	79
Accuracy			90.20%	357
Macro Avg	90.34%	90.32%	90.15%	357
Weighted Avg	90.59%	90.20%	90.20%	357

The model performed very well in classifying the ready position category with an accuracy of 93.68%, indicating that ready poses have distinctive characteristics and are easily distinguished from other movements. The serve category also shows a high recall (94.94%) despite having a relatively lower precision (85.23%), indicating that the model tends to over-predict the serve category but manages to detect almost all actual serve samples.

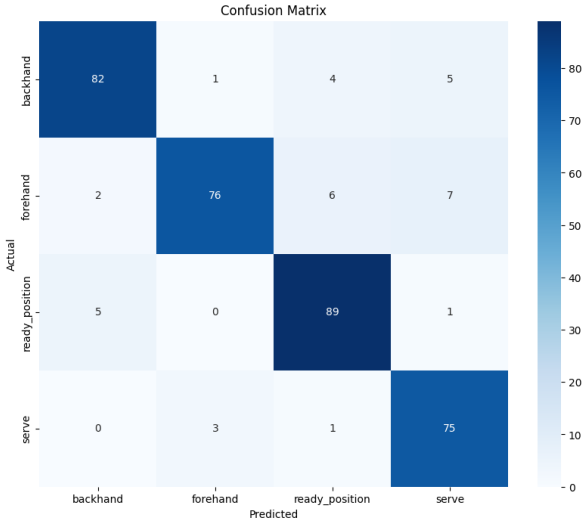

Fig. 9. Confusion Matrix of the Model Classification Results

Figure 9 displays the confusion matrix, which provides deep insight into the model's classification pattern for each movement category. The confusion matrix analysis shows that the model has good discriminative ability in distinguishing the four categories of tennis moves. The highest misclassification rate occurs between the forehand and backhand categories (8 cases), which is understandable since both movements have relatively similar pose characteristics in some movement phases.

F. Performance Analysis per Movement Category

The backhand category showed a balanced performance with a precision of 92.13% and a recall of 89.13%. The model successfully classified 82 out of 92 backhand samples correctly, with 10 misclassified samples. Error analysis showed that misclassification occurred mainly with the forehand (5 cases) and ready position (3 cases) categories. This suggests that some backhand poses have similarities with transition poses or forehand preparation poses.

The forehand category has the highest precision (95.00%) but the lowest recall (83.52%), indicating that the model is very selective in predicting the forehand but tends to miss some of the actual forehand samples. Out of 91 forehand samples, the model successfully classified 76 samples correctly. Pattern misclassification

showed that 15 forehand samples were misclassified, with 8 samples predicted as backhand and 7 samples predicted as other categories.

Ready position performed best overall with the highest recall (93.68%) and second highest F1-score. The model correctly identified 89 out of 95 ready position samples. The high performance in this category indicates that ready poses have distinctive and consistent feature characteristics, making them easily distinguishable from other dynamic movements such as forehand, backhand, or serve.

The serve category showed the highest recall (94.94%) but the lowest precision (85.23%), indicating that the model is very sensitive in detecting serve movements but also tends to be false positive for this category. The model correctly identified 75 out of 79 serve samples, with only 4 missed samples. However, the model also predicted 13 non-serve samples as serves, which contributed to the relatively lower precision.

G. Limitations and Future Work

This study has several limitations that warrant consideration. The dataset employed was relatively small by deep learning standards, which may restrict the generalizability of the findings across broader and more diverse contexts. In addition, the reliance on static images with a single timestep limited the ability of the Long Short-Term Memory (LSTM) network to capture temporal dependencies, thereby reducing its core advantage in sequence modeling. Failure cases, particularly in the classification of the 'serve' motion, were acknowledged but not examined in depth; potential factors such as occlusion, rapid limb movements, or unfavorable camera perspectives may have contributed to these errors. Furthermore, the evaluation methodology relied solely on a single train-test split, without the inclusion of cross-validation or statistical significance testing, which limits the robustness of the reported results.

Future work is expected to address these limitations to enhance the robustness and applicability of the proposed system. Expanding the dataset and transitioning from static images to video sequences is anticipated to fully exploit LSTM's temporal modeling strengths, enabling a more comprehensive analysis of tennis movements. A systematic investigation of misclassification cases, especially in complex 'serve' motions, is also expected to provide deeper insights into the causes of errors and improve the reliability of pose extraction. Finally, subsequent evaluations are expected to incorporate more rigorous validation methods, such as k-fold cross-validation and statistical significance testing, thereby ensuring the stability and generalizability of the model's performance.

V. CONCLUSION

This research successfully developed a tennis pose recognition system that integrates MediaPipe for pose feature extraction with Long Short-Term Memory (LSTM) networks for tennis motion classification. The system is able to classify four categories of tennis moves (backhand, forehand, ready position, and serve) with 90.20% accuracy on a dataset of 1,782 images. MediaPipe showed effectiveness with an 88.7% extraction success rate, resulting in 33 pose landmarks that were flattened into a 99-dimensional feature vector. The LSTM model with 3 layer LSTM architecture and 2 dense layers achieved balanced performance with the highest F1 score on ready position (91.28%) and the lowest on forehand (88.89%).

The main contribution of this research is the demonstration of successful integration between pose estimation and deep learning for tennis motion recognition applications in the Indonesian context. Results show the feasibility of implementation for real-world applications such as coaching assistance and performance analysis, with computational efficiency that allows deployment on mobile devices (memory footprint 714.39 KB). This research provides a foundation for the development of automated sports analysis systems and opens up opportunities for more advanced application development with dataset expansion, data augmentation, and video data integration to optimally utilize temporal aspects.

ACKNOWLEDGMENT

We thank Orvile for uploading the "Tennis Player Actions Dataset" so we can use it for this research.

REFERENCES

[1] T. Sampaio, J. P. Oliveira, D. A. Marinho, H. P. Neiva, and J. E. Morais, "Applications of Machine Learning to Optimize Tennis Performance: A Systematic Review," Appl. Sci., vol. 14, no. 13, 2024, doi: 10.3390/app14135517.

[2] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions," Appl. Sci., vol. 12, no. 9, 2022, doi: 10.3390/app12094429.

[3] F. Boscolo, F. Lamberti, and L. Morra, "Analyzing the Performance of Deep Learning-based Techniques for Human Pose Estimation," in 2024 IEEE International Workshop on Sport Technology and Research, STAR 2024 - Proceedings, 2024, pp. 193 – 198. doi: 10.1109/STAR62027.2024.10635956.

[4] Y. Li and J. Hou, "Real time monitoring of motion posture using a motion accelerometer based on sensors and cellular thermodynamic analysis," Therm. Sci. Eng. Prog., vol. 57, 2025, doi: 10.1016/j.tsep.2024.103136.

[5] Z. He et al., "Real-Time Accurate Determination of Table Tennis Ball and Evaluation of Player Stroke Effectiveness with Computer Vision-Based Deep Learning," Appl. Sci., vol. 15, no. 10, 2025, doi: 10.3390/app15105370.

[6] K. S. Gill, V. Anand, S. Malhotra, and S. Devliyal, "Sports Game Classification and Detection Using ResNet50 Model Through Machine Learning Techniques Using Artificial Intelligence," in 2024 3rd International Conference for Innovation in Technology, INOCON 2024, 2024. doi: 10.1109/INOCON60754.2024.10511858.

[7] I. Jayaweerage et al., "Motion Capturing in cricket with bare minimum hardware and optimised software: A comparison of MediaPipe and OpenPose," in 2024 1st International Conference on Software, Systems and Information Technology, SSITCON 2024, 2024. doi: 10.1109/SSITCON62437.2024.10796169.

[8] A. P. Jyothi, A. Anurag, G. Gagan, S. Uday, and J. Pragathi, "Pose Fit: ML Powered Fitness Application," in IEEE International Conference on Signal Processing and Advance Research in Computing, SPARC 2024, 2024. doi: 10.1109/SPARC61891.2024.10828671.

[9] M. Rivera, D. Huamanchahua, and C. Flores, "Comparative Analysis of state-of-art pre-trained Human Pose Estimation models in underwater condition," in 2024 IEEE Colombian Conference on Communications and Computing, COLCOM 2024 - Proceedings, 2024. doi: 10.1109/COLCOM62950.2024.10720259.

[10] L. Liu, Y. Dai, and Z. Liu, "Real-time pose estimation and motion tracking for motion performance using deep learning models," J. Intell. Syst., vol. 33, no. 1, 2024, doi: 10.1515/jisys-2023-0288.

[11] B. Jo and S. Kim, "Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices," Trait. du Signal, vol. 39, no. 1, pp. 119 – 124, 2022, doi: 10.18280/ts.390111.

[12] J.-L. Chung, L.-Y. Ong, and M.-C. Leow, "Comparative Analysis of Skeleton-Based Human Pose Estimation," Futur. Internet, vol. 14, no. 12, 2022, doi: 10.3390/fi14120380.

[13] P. Barbosa, P. Cunha, V. Carvalho, and F. Soares, "Deep Learning in Taekwondo Techniques Recognition System: A Preliminary Approach," Lect. Notes Mech. Eng., pp. 280 – 291, 2022, doi: 10.1007/978-3-031-09385-2_25.

[14] Z. Chen, Q. Xie, and W. Jiang, "Hybrid Deep Learning Models for Tennis Action Recognition: Enhancing Professional Training Through CNN-BiLSTM Integration," Concurr. Comput. Pract. Exp., vol. 37, no. 6–8, 2025, doi: 10.1002/cpe.70029.

[15] X. Zhang, "Computer Vision Technology in Sports Training Using Attention-based Bidirectional Long Short- Term Memory," in 2nd IEEE International Conference on Data Science and Information System, ICDSIS 2024, 2024. doi: 10.1109/ICDSIS61070.2024.10594412.

[16] P. Enosh, J. L. Alekhya, P. Sai Nithin, and Y. Devika, "Human Action Recognition Using Pose-Guided Graph Convolutional Networks and Long Short-Term Memory," in 2024 International Conference on Electrical, Electronics and Computing Technologies, ICEECT 2024, 2024. doi: 10.1109/ICEECT61758.2024.10739243.

[17] B. Erfianto, B. Purnama, and I. R. Wirawan, "Time Series Classification of Badminton Pose Using Long Short-Term Memory with Landmark Tracking," J. Electron. Electromed. Eng. Med. Informatics, vol. 7, no. 1, pp. 27 – 37, 2025, doi: 10.35882/jeeemi.v7i1.488.

[18] S. H. Ali and H. Aygun, "Air-Drawing," in 3rd International Informatics and Software Engineering Conference, IISEC 2022, 2022. doi: 10.1109/IISEC56263.2022.9998215.

[19] N. Chaithra, J. Jha, A. Sayal, V. Gupta, and A. Gupta, "A Paradigm Shift towards Computer Vision," in Proceedings - IEEE International Conference on Device Intelligence, Computing and Communication Technologies, DICCT 2023, 2023, pp. 54 – 58. doi: 10.1109/DICCT56244.2023.10110300.

[20] P. Agrawal, R. Bose, G. K. Gupta, G. Kaur, S. Paliwal, and A. Raut, "Advancements in Computer Vision: A Comprehensive Review," in 2024 International Conference on Innovations and Challenges in Emerging Technologies, ICICET 2024, 2024. doi: 10.1109/ICICET59348.2024.10616321.

[21] N. O'Mahony et al., "Deep Learning vs. Traditional Computer Vision," Adv. Intell. Syst. Comput., vol. 943, pp. 128 – 144, 2020, doi: 10.1007/978-3-030-17795-9_10.

[22] M. Laad, R. Maurya, and N. Saiyed, "Unveiling the Vision: A Comprehensive Review of Computer Vision in AI and ML," in 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, ADICS 2024, 2024. doi: 10.1109/ADICS58448.2024.10533631.

[23] Y. Mao and R. Mao, "Research on Some Key Technologies of Deep Learning in the Field of Computer Vision," in 2024 3rd International Conference for Innovation in Technology, INOCON 2024, 2024. doi: 10.1109/INOCON60754.2024.10511442.

[24] D. Zhu, L. Song, F. Yuan, and Q. Yang, "Research Status of Damage Identification Algorithm Based on Deep Learning," E3s Web Conf., 2021, doi: 10.1051/e3sconf/202123304039.

[25] D. Ma, "Recent advances in deep learning based computer vision," in Proceedings - 2022 International Conference on Computers, Information Processing and Advanced Education, CIPAE 2022, 2022, pp. 174 – 179. doi: 10.1109/CIPAE55637.2022.00044.

[26] S. Sumit, S. Bisht, S. Joshi, and U. Rana, "Comprehensive Review of R-CNN and Its Variant Architectures," Int Res J Adv Engg Hub, 2024, doi: 10.47392/irjaeh.2024.0134.

[27] R. N. Anand and S. Palaniswamy, "Multi Person Pose Estimation and 3D Pose Detection Animation," in 2023 3rd International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2023, 2023. doi: 10.1109/SMARTGENCON60755.2023.10442905.

[28] L. Chen, T. Liu, Z. Gong, and D. Wang, "Movement Function Assessment Based on Human Pose Estimation from Multi-View," Comput. Syst. Sci. Eng., vol. 48, no. 2, pp. 321 – 339, 2024, doi: 10.32604/csse.2023.037865.

[29] Meharaj-Ul-Mahmmud, M. A. Ahmed, S. M. Alam, O. T. Imam, A. W. Reza, and M. S. Arefin, "Human Posture Estimation: In Aspect of the Agriculture Industry," Lect. Notes Networks Syst., vol. 514 LNNS, pp. 479 – 490, 2022, doi: 10.1007/978-3-031-12413-6_38.

[30] O.-H. Kwon, J. Tanke, and J. Gall, "Recursive Bayesian Filtering for Multiple Human Pose Tracking from Multiple Cameras," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12623 LNCS, pp. 438 – 453, 2021, doi: 10.1007/978-3-030-69532-3_27.

[31] A. Amini, H. Farazi, and S. Behnke, "Real-Time Pose Estimation from Images for Multiple Humanoid Robots," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 13132 LNAI, pp. 91 – 102, 2022, doi: 10.1007/978-3-030-98682-7_8.

[32] J.-C. Hsu and M.-H. Su, "Application of Skeleton Image Detection in Basketball Free Throw Posture Research," in Proceedings - 2024 16th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2024, 2024, pp. 324 – 328. doi: 10.1109/IIAI-AAI63651.2024.00067.

[33] Y.-C. Li, C.-T. Chang, C.-C. Cheng, and Y.-L. Huang, "Baseball Swing Pose Estimation Using OpenPose," in 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence, RAAI 2021, 2021, pp. 6 – 9. doi: 10.1109/RAAI52226.2021.9507807.

[34] F. Zheng, D. Z. Al-Hamid, P. H. J. Chong, C. Yang, and X. J. Li, "A Review of Computer Vision Technology for Football Videos," Inf., vol. 16, no. 5, 2025, doi: 10.3390/info16050355.

[35] P. J. Devi, A. Sumani, C. M. Chandra, B. B. Thrisha, and A. S. Vamshi, "Enhancing Athletic Performance:2D Human Pose Estimation Using Deep Neural Networks for Movement Analysis," in Proceedings of International Conference on Circuit Power and Computing Technologies, ICCPCT 2024, 2024, pp. 1115 – 1123. doi: 10.1109/ICCPCT61902.2024.10672943.

[36] T. Luczak et al., "A survey of technical challenges in computer vision via machine and deep learning for human pose estimation," in IISE Annual Conference and Expo 2022, 2022. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85137173921&partnerID=40&md5=d0e596e60e7c8afd27cc340675e506f6

[37] D. R. Gopagoni, P. V Lakshmi, and A. Chaudhary, "Evaluating Machine Learning Algorithms for Marketing Data Analysis: Predicting Grocery Store Sales," Lect. Notes Networks Syst., vol. 134, pp. 155 – 163, 2021, doi: 10.1007/978-981-15-5397-4_17.

[38] V. Gupta, V. K. Mishra, P. Singhal, and A. Kumar, "An Overview of Supervised Machine Learning Algorithm," in Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022, 2022, pp. 87 – 92. doi: 10.1109/SMART55829.2022.10047618.

[39] R. Sawhney, V. Tiwari, D. Kirti, and V. R. Vadi, Disease detection system: Supervised learning to detect diseases. 2024. doi: 10.4018/979-8-3693-6577-9.ch003.

[40] R. K. Stephen and T. Archana, Optimization, machine learning, and fuzzy logic: Machine learning fundamentals and introduction to machine learning. 2025. doi: 10.4018/979-8-3693-7352-1.ch001.

[41] S. A. Alomari et al., Supervised Learning: Teaching Machines with Labeled Data. 2025. doi: 10.1201/9781003516385-4.

[42] S. Sharma and H. K. Soni, "Discernment of potential buyers based on purchasing behaviour via machine learning techniques," in Proceedings of 2020 IEEE International Conference on Advances and Developments in Electrical and Electronics Engineering, ICADEE 2020, 2020. doi: 10.1109/ICADEE51157.2020.9368935.

[43] Q. Yi, S. Ling, G. Chen, and L. Liu, "Research on Computer Vision Technology Based on BP-LSTM Hybrid Network," Appl. Math. Nonlinear Sci., 2023, doi: 10.2478/amns.2021.2.00270.

[44] F. Zulkernine, M. Gasmallah, H. Isah, J. Lam, S. Mahfuz, and S. Khan, "A hands-on tutorial on deep learning for object and pattern recognition," in CASCON 2019 Proceedings - Conference of the Centre for Advanced Studies on Collaborative Research - Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, 2020, pp. 386 – 387. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087416922&partnerID=40&md5=47341fc06b6a9a94ca761a69054e719a

[45] P. M. Cheng et al., "Deep learning: An update for radiologists," Radiographics, vol. 41, no. 5, pp. 1427 – 1445, 2021, doi: 10.1148/rg.2021200210.

[46] M. Sahu and R. Dash, "A survey on deep learning: Convolution neural network (cnn)," Smart Innov. Syst. Technol., vol. 153, pp. 317 – 325, 2021, doi: 10.1007/978-981-15-6202-0_32.

[47] R. Indrakumari, T. Poongodi, and K. Singh, "Introduction to Deep Learning," EAI/Springer Innov. Commun. Comput., pp. 1 – 22, 2021, doi: 10.1007/978-3-030-66519-7_1.

[48] M. Kavitha and K. Akila, "Amplifying document categorization with advanced features and deep learning," Multimed. Tools Appl., vol. 83, no. 26, pp. 68087 – 68105, 2024, doi: 10.1007/s11042-024-18483-7.

[49] H. Cecotti, A. Rivera, M. Farhadloo, and M. A. Pedroza, "Grape detection with convolutional neural networks," Expert Syst. Appl., vol. 159, 2020, doi: 10.1016/j.eswa.2020.113588.

[50] F. Parvin and M. Al Mehedi Hasan, "A Comparative Study of Different Types of Convolutional Neural Networks for Breast Cancer Histopathological Image Classification," in 2020 IEEE Region 10 Symposium, TENSYMP 2020, 2020, pp. 945 – 948. doi: 10.1109/TENSYMP50017.2020.9230787.

[51] M. Agarwal, K. S. Gill, M. Kumar, R. Rawat, and K. R. Chythanya, "Image Classification of Nespresso Capsules by the Use of Convolutional Neural Networks Through Deep Learning," in 2024 2nd International Conference on Computer, Communication and Control, IC4 2024, 2024. doi: 10.1109/IC457434.2024.10486655.

[52] A. R. Bushara and R. S. V. Kumar, "Deep Learning-based Lung Cancer Classification of CT Images using Augmented Convolutional Neural Networks," Electron. Lett. Comput. Vis. Image Anal., vol. 21, no. 1, pp. 130 – 142, 2022, doi: 10.5565/REV/ELCVIA.1490.

[53] J. Sen and S. Mehtab, Long-and-Short-Term Memory (LSTM) NetworksArchitectures and Applications in Stock Price Prediction. 2022. doi: 10.1002/9781119813439.ch8.
[54] H. Agarwal, G. Mahajan, A. Shrotriya, and D. Shekhawat, "Predictive Data Analysis: Leveraging RNN and LSTM Techniques for Time Series Dataset," in Procedia Computer Science, 2024, pp. 979 – 989. doi: 10.1016/j.procs.2024.04.093.
[55] M. M. Rahman, I. Jahan, S. M. Nizamuddin Shuvo, and M. Rabbul Hossain Taj, "Parallel Hybrid LSTM with Longitudinal Memory: To Handle Longer Sequential Data," in 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025, 2025. doi: 10.1109/ECCE64574.2025.11013934.
[56] A. Ardakani, Z. Ji, and W. J. Gross, "Learning to Skip Ineffectual Recurrent Computations in LSTMs," in Proceedings of the 2019 Design, Automation and Test in Europe Conference and Exhibition, DATE 2019, 2019, pp. 1427 – 1432. doi: 10.23919/DATE.2019.8714765.
[57] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," Sci. Rep., vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-86432-7.
[58] S. Taheri, B. Talebjedi, and T. Laukkanen, "Electricity demand time series forecasting based on empirical mode decomposition and long short-term memory," Energy Eng. J. Assoc. Energy Eng., vol. 118, no. 6, pp. 1577 – 1594, 2021, doi: 10.32604/EE.2021.017795.
[59] B. Prabha, S. Maheshwari, and P. Durgadevi, "Sentiment Analysis using Long Short-Term Memory," in 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023, 2023. doi: 10.1109/ICCCNT56998.2023.10306792.
[60] R. Ashtagi, R. V. Bidwe, A. Fukate, O. Kulkarni, P. Jadhav, and S. Patil, "Sentiment Analysis on YouTube Comments using Long Short-Term Memory (LSTM) Networks," in 6th International Conference on Mobile Computing and Sustainable Informatics, ICMCSI 2025 - Proceedings, 2025, pp. 785 – 788. doi: 10.1109/ICMCSI64620.2025.10883113.
[61] F. Khanum, P. S. Lakshmi, and K. Harsha Vardhan Reddy, "Sentiment Analysis Using Natural Language Processing, Machine Learning and Deep Learning," in 5th International Conference on Circuits, Control, Communication and Computing, I4C 2024, 2024, pp. 113 – 118. doi: 10.1109/I4C62240.2024.10748425.
[62] B. G. Premasudha and V. Patil, "Enhanced Sentiment Analysis of Airline Twitter Review Using Hybrid Machine Learning and Deep Learning Models," in 2024 1st International Conference on Innovations in Communications, Electrical and Computer Engineering, ICICEC 2024, 2024. doi: 10.1109/ICICEC62498.2024.10808987.
[63] Z. Jianxun, H. Shaojie, and L. Lixu, "The Application of EMD-ARIMA-LSTM Neural Network in Stocks," in Proceedings - 2023 Asia Conference on Advanced Robotics, Automation, and Control Engineering, ARACE 2023, 2023, pp. 154 – 161. doi: 10.1109/ARACE60380.2023.00031.
[64] K. Kumari, S. Das, A. Bhowmick, A. Saha, A. B. Puja, and K. Karmakar, "Stock Price Prediction Using Machine Learning," in International Conference on Big Data Analytics in Bioinformatics, DABCon 2024, 2024. doi: 10.1109/DABCon63472.2024.10919355.
[65] P. A. C. Bautista, C. P. O. Chan Shio, and P. A. R. Abu, "Telecommunications Product Revenue Time-Series Forecasting Using Target Variable Preprocessing Methods," in ACMLC 2024 - 2024 6th Asia Conference on Machine Learning and Computing, 2025, pp. 53 – 61. doi: 10.1145/3690771.3690784.
[66] A. Hayat, C. H. Li, N. Prakoso, R. Zheng, A. Wyawahare, and J. Wu, "Gesture and Body Position Control for Lightweight Drones Using Remote Machine Learning Framework," Lect. Notes Electr. Eng., vol. 1295 LNEE, pp. 17 – 42, 2025, doi: 10.1007/978-981-97-9112-5_2.
[67] Y. N. Lavanya, N. N. Rajalakshmi, K. Sumanth, S. Gowrishankar, and K. P. S. Asha Rani, "A Novel Approach for Developing Inclusive Real-Time Yoga Pose Detection for Health and Wellness Using Raspberry pi," in 7th IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2023 - Proceedings, 2023. doi: 10.1109/CSITSS60515.2023.10334109.
[68] Z. Lai and Q. Zhao, "VRA-L: A real-time system for generating expressive avatar animations and object detection based on RGB cameras and monocular videos," in ACM International Conference Proceeding Series, 2024, pp. 292 – 298. doi: 10.1145/3700906.3700954.
[69] M. F. Asghar, M. H. Ali, and J. Waleed, "Towards Improve Human Activity Recognition Using Mediapipe," in AIP Conference Proceedings, 2025. doi: 10.1063/5.0257434.

[70] U. Shandilya, V. Sharma, and D. Mishra, "Human Action Recognition Using Mediapipe Holistic Keypoints: A Deep Learning Approach," Commun. Comput. Inf. Sci., vol. 2336 CCIS, pp. 58 – 69, 2025, doi: 10.1007/978-3-031-83796-8_5.

[71] H. C. Saini, "iSmartYog: A Real Time Yoga Pose Recognition and Correction Feedback Model Using Deep Learning for Smart Healthcare," in International Conference on Smart Systems for Applications in Electrical Sciences, ICSSES 2023, 2023. doi: 10.1109/ICSSES58299.2023.10201061.

[72] N. Jlidi, O. Jemai, and T. Bouchrika, "Human Pose Estimation for Action Recognition in Sports Video Using GNN," Lect. Notes Networks Syst., vol. 1224 LNNS, pp. 105 – 114, 2025, doi: 10.1007/978-3-031-78925-0_11.

[73] M. Ati, M. U. G. Khan, and I. Kiran, "Social Media Trends Analysis using the Bi-LSTM with Multi-Head Attention," in 2022 International Conference on Electrical and Computing Technologies and Applications, ICECTA 2022, 2022, pp. 295 – 299. doi: 10.1109/ICECTA57148.2022.9990328.

[74] H. S. S. Al-Deen, Z. Zeng, R. Al-Sabri, and A. Hekmat, "An improved model for analyzing textual sentiment based on a deep neural network using multi-head attention mechanism," Appl. Syst. Innov., vol. 4, no. 4, 2021, doi: 10.3390/asi4040085.

[75] D. E. Herwindiati, J. Hendryli, and N. H. Sarmin, "A Comparison of Two Deep Learning Models on The Stock Exchange Predictions," Int. J. Adv. Soft Comput. its Appl., vol. 15, no. 2, pp. 225 – 234, 2023, doi: 10.15849/IJASCA.230720.15.

[76] S. Kulshreshtha and A. Vijayalakshmi, "An ARIMA-LSTM hybrid model for stock market prediction using live data," J. Eng. Sci. Technol. Rev., vol. 13, no. 4, pp. 117 – 123, 2020, doi: 10.25103/jestr.134.11.