**HUMAN POSE DETECTION AND CLASSIFICATION USING MOVENET WITH SPATIOTEMPORAL DATA**

Abhila Anju O

Department Of Artificial Intelligence

Kongu Engineering College

Tamil Nadu, India

[abhilaanju.ai@kongu.edu](mailto:abhilaanju.ai@kongu.edu)

Saranyah V

Department Of Artificial Intelligence

Kongu Engineering College

Tamil Nadu, India

[saranyah.ai@kongu.edu](mailto:saranyah.ai@kongu.edu)

Snehavarshini S

Department Of Artificial Intelligence

Kongu Engineering College

Tamil Nadu, India

[snehavarshinis.21aid@kongu.edu](mailto:snehavarshinis.21aid@kongu.edu)

Bhuvaneshwari C

Department Of Artificial Intelligence

Kongu Engineering College

Tamil Nadu, India

[bhuvaneshwaric.21aid@kongu.edu](mailto:bhuvaneshwaric.21aid@kongu.edu)

Soundharya M

Department Of Artificial Intelligence

Kongu Engineering College

Tamil Nadu, India

[soundharyam.21aid@kongu.edu](mailto:soundharyam.21aid@kongu.edu)

***Abstract-*** The integration of Move Net, a neural network model, with spatiotemporal image data enables real-time analysis of dynamic human movements and actions. This approach extracts human pose Key points coordinates using Move Net and trains a deep learning model to classify and detect human poses and actions from image sequences. Post-processing techniques enhance model accuracy and address challenges like occlusions and lighting variations. The model's performance is evaluated using standard metrics, demonstrating its potential in applications requiring image-based real-time human pose detection and action classification. This innovative approach opens new frontiers for exploring the relationships between human actions and their underlying intentions.

By analyzing temporal patterns of human poses, we can gain insights into motivations and goals, leading to a deeper understanding of human behavior. This enhanced understanding can be used to develop more intuitive human-computer interaction systems and healthcare applications. The fusion of Move Net with spatiotemporal image data represents a significant leap forward in human pose detection and action classification, revolutionizing diverse applications, including human-computer interaction, sports analysis, and healthcare monitoring.

**Keywords -** Move Net, Spatiotemporal, Human pose detection, Action classification.

**I.INTRODUCTION**

Human pose detection and action classification using image data represent fundamental tasks within the realm of computer vision, with diverse applications ranging from gesture recognition and sports analytics to healthcare monitoring and human-computer interaction. In recent years, the evolution of deep learning models has revolutionized the field, enabling more accurate and real-time solutions. This study embarks on a journey to explore a compelling convergence: the synergy between Move Net, an advanced neural network model for real-time human pose estimation, and spatiotemporal image data. Move Net's proficiency in capturing human poses in video frames and spatiotemporal data's capacity to encompass both spatial and temporal dimensions introduce a transformative approach for understanding dynamic human behaviors. This paper delves into the methodology, encompassing the collection of image data, the preprocessing to extract pose key point coordinates, the supervised learning through annotation, and the selection and training of deep learning models. The outcome of this effort is a model primed to classify and detect human poses and actions from sequences of image frames. Post-processing techniques are also explored to refine results, especially in scenarios with occlusions and lighting variations. The research culminates in an evaluation of the model's performance using established image classification and detection metrics. By delving into the possibilities at the crossroads of Move Net and spatiotemporal image data, we seek to advance our understanding of dynamic human movements and actions, ultimately contributing to the development of highly accurate and effective human-computer interaction systems and automated monitoring solutions with far-reaching applications.

In the realm of human pose detection and action classification, spatiotemporal image data holds immense potential for capturing the nuances of dynamic human movements. Unlike traditional image data that provides a static representation of a scene, spatiotemporal data encapsulates both spatial and temporal dimensions, enabling the analysis of human actions across time. This inherent advantage aligns seamlessly with the capabilities of MoveNet, a deep learning model specifically designed for real-time human pose estimation. By leveraging the synergy between MoveNet's expertise in capturing human poses and spatiotemporal data's ability to encode motion patterns, we can unlock a deeper understanding of human behaviour, paving the way for more sophisticated and effective human-computer interaction systems and automated monitoring solutions. The convergence of MoveNet and spatiotemporal image data presents an unprecedented opportunity to bridge the gap between static pose estimation and dynamic action recognition. By leveraging MoveNet's ability to accurately capture human poses in individual frames and spatiotemporal data's ability to capture motion patterns across time, we can construct a comprehensive representation of human movements. This enhanced representation enables the development of models that can not only identify individual poses but also recognize sequences of poses that form distinct actions. This capability holds immense potential for applications that require real-time understanding of human movements, such as gesture-based interaction systems, sports analysis, and motion-controlled gaming.

**II.LITERATURE SURVEY**

Advances in human posture estimation in recent times have been facilitated by deep learning, namely with the use of recurrent models like PoseNet and Convolutional Neural Networks (CNNs). The precision of a project based on deep learning for human posture estimation Different learning methods may be used based on the quality of the dataset, training parameters and the design of the model. Modern models frequently attain accuracy rates of more than 76%, with certain methods claiming even more accuracy [1]. A new approach that makes use of instance-aware attention addresses the identification and grouping of keypoints for multi-person pose estimation. This method efficiently detects and organises important points.by adding systems for paying attention that take specific cases into account inside a setting. The algorithm improves accuracy by emphasising relevant keypoints, enabling reliable posture estimation for several individuals. State-of-the-art methods in multi-person pose estimation often achieve accuracy rates exceeding 90% [2]. A variety of methods are included in the Computer Vision and Image Understanding algorithm for deciphering visual data. Its versatility and ability to glean valuable information from images propels computer vision forward, and its high degree of accuracy in automating visual perception tasks makes it indispensable across a range of domains eyesight. Modern algorithms often attain high accuracy rates, frequently exceeding 81%, for tasks like object identification, picture segmentation, and scene interpretation [3].

By adding diversity to the training dataset, image augmentation enhances the model's capacity to generalise across various postures. Optimising hyperparameters helps to maximise performance by fine-tuning model parameters. By utilising transfer learning, the algorithm may leverage the information obtained from models that have already been trained. This method produces reliable and precise human posture detection, which has uses in ergonomics, sports analysis, and healthcare. Similar approaches used in state-of-the-art procedures usually result in excellent accuracy rates, often exceeding 80% [4]. Robust pose estimation is ensured by utilising MediaPipe Pose's pre-trained capabilities. The optimisation technique improves accuracy and refines outcomes using a humanoid model suitability for stances akin to human. This technique works well for a variety of uses such as motion analysis, virtual reality, and animation. Its strength is achieved by combining a potent pre-trained model with a customised optimisation strategy, producing accurate and pertinent information for the context estimates of human stance cutting edge position estimate algorithms, when properly set up and refined, frequently very excellent accuracy rates—often more than 80% [5]. The system effectively recognises different sitting positions using an unsupervised data-driven framework, which eliminates the need for labelled training data. That strategy makes use of data-driven methods to adjust and effectively identify the postures of wheelchair users setting. Its power comes from its capacity to learn and adjust to a variety of sitting postures without requiring a lot of manually marked up. Optimisation and fine-tuning are necessary phases to Verify if the algorithm's accuracy complies with the practical applications for assistive technology and healthcare [6].

It reliably records human postures using methods like bone tracking and keypoints identification. The programme then uses this posture information to identify and categorise different activities. Applications include anything from gaming to surveillance and video analysis. Biomechanics in sports. It works by creating a bridge between the gap between temporal dynamics and spatial pose information, allowing thorough examination and study of human activity. Modern methods that combine pose estimation with Action recognition has a high accuracy rate—it frequently surpasses 90% [7]. It methodically investigates and assesses several methods for estimating human posture, including but not restricted to traditional computer vision strategies, hybrid models, and deep learning architectures. That survey is an important tool for practitioners and scholars, directing the choice of suitable methods depending on particular prerequisites. Its strength is in providing a thorough overview of human position estimate techniques, promoting a better comprehension of the developments in the field. Metrics for accuracy are usually connected to particular posture estimation methods or applications covered in the survey [8].

Convolutional neural networks (CNNs) or more sophisticated models like PoseNet, OpenPose, or Hourglass networks are frequently used in this method because they can capture complicated spatial connections and connections in human postures. Its ability to use deep learning to automatically deduce specific and human poses that are appropriate for the setting using visual data, providing a broad variety of real-world uses. cutting-edge deep learning Pose estimation methods are capable of great accuracy rates, frequently above 95%, particularly in the case of extensive and varied training datasets [9]. The initial purpose of ViTs was picture categorization; they have now been modified to capture spatial interdependence in human postures. Its potency is seen in the simple incorporation of Vision Transformers, demonstrating their effectiveness in tasks other than image categorization, especially when it comes to estimating human stance. Vision Transformers have shown to be quite effective in numerous tasks using computer vision, however the precision for this particular project should be ascertained by a thorough analysis of pertinent datasets, taking into account the subtleties of estimating human stance [10].

**III.PROPOSED METHODOLOGY**

**A. DATASET**

In this project, the utilization of the MoveNet architecture is explored to enhance yoga pose detection through deep learning. MoveNet, a convolutional neural network known for its efficiency in extracting features from images, is deemed suitable for the task of accurately detecting keypoints and poses in yoga images. Its architecture, characterized by lightweight connections, facilitates the learning of intricate features, making it an effective tool for yoga pose detection. To address challenges related to keypoints and pose detection, a novel adaptive approach is incorporated into the methodology. This approach dynamically adjusts the number of training epochs and batch sizes based on the impact of each batch on the model's ability to accurately detect keypoints and poses. By setting a threshold based on validation accuracy, the training process terminates when the model's performance in detecting yoga poses begins to deteriorate. This adaptive strategy effectively mitigates issues related to keypoints and pose detection, enhancing the model's precision in identifying yoga poses. Furthermore, a gradual validation accuracy estimation technique is introduced to improve the accuracy of the model in keypoints and pose detection. This method evaluates the model's performance more accurately by quantifying the difference between the actual keypoints and pose labels and the predicted keypoints and pose probabilities. The strategy for improving MoveNet-based yoga pose detection relies on the incorporation of the validation accuracy estimate in tandem with the adaptive keypoints and pose detection mechanism.

The dataset includes 5 picture categorization categories. Chair, cobra, dog, tree, and warrior, where each position specifies the pose of each yoga practise.

FIG 1

FIG 2

FIG 3



FIG 4

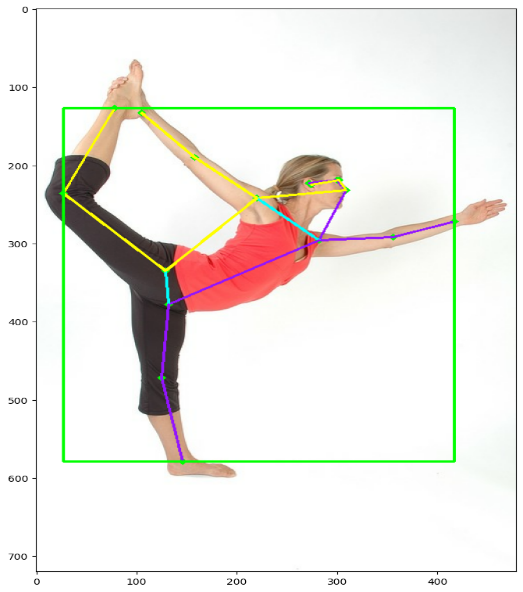
FIG 5

**B. TRAIN-TEST SPLIT**

In our study, we employed a train-test split methodology for dataset partitioning. The training dataset consisted of 200 images per class, and we worked with a total of 5 classes. This resulted in a comprehensive training set with a collective size of [200 images/class \* 5 classes = 1000 images]. For the evaluation of model performance, we reserved a separate test dataset, containing approximately 100 images per class, ensuring a robust assessment across all categories. This meticulous train-test split allowed us to train our models effectively on diverse data while rigorously evaluating their generalization capabilities on previously unseen samples.

**C. PROPOSED METHODOLOGY**

The methodology for yoga poses classification using MoveNet encompasses several key steps. Initially, a diverse dataset is collected, comprising images of individuals in various yoga poses, with a focus on ensuring representation across different poses for robust model training. Subsequently, the dataset undergoes preprocessing, involving resizing images to a consistent resolution, normalizing pixel values, and data augmentation to enhance model generalization. MoveNet Pose Detection is applied to identify human poses in the images, extracting keypoints that represent different body parts and capture their spatial relationships. (FIG 6)



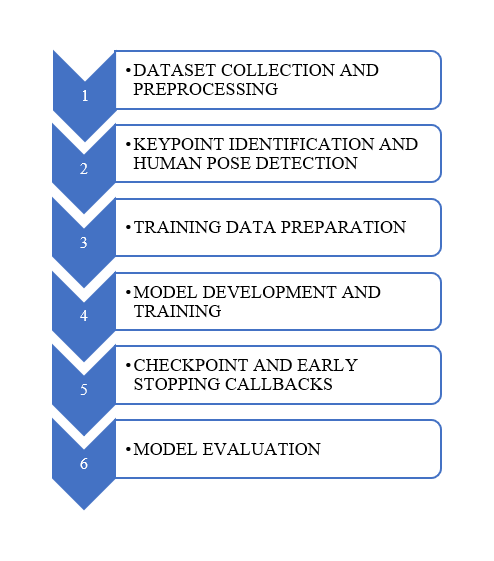
KEYPOINTS DETECTION USING

MOVENET ALGORITHM

The detected keypoints are then translated into a suitable format for machine learning, potentially involving vectorization or the use of a pose descriptor in the Pose Representation stage. Following this, a labeled dataset is created, associating pose representations with corresponding yoga pose labels, and the dataset is split into training and validation sets for model training and evaluation.

The subsequent Model Training phase involves training the neural network using the labeled training dataset, employing an appropriate loss function, optimization algorithm, and potentially introducing techniques like dropout or batch normalization to prevent overfitting. An Adaptive Mechanism is implemented, featuring an adaptive early stopping mechanism that dynamically adjusts training parameters based on validation accuracy, mitigating overfitting during model training. Following model training, Checkpoint and Early Stopping Callbacks are introduced, optimizing the training process.

The model is then evaluated on the validation set, allowing for fine-tuning as needed, and its performance is assessed on a separate test set to gauge generalization to new, unseen data. Additionally, a Post-Training Analysis and Optimization step involves refining the model's performance based on the evaluation results. The methodology concludes with the introduction of a Gradient Estimation Technique, such as gradient-based methods, to enhance the accuracy assessment during training. Overall, this comprehensive methodology ensures a systematic approach to yoga pose classification, leveraging MoveNet and incorporating adaptive strategies for effective model training and evaluation.



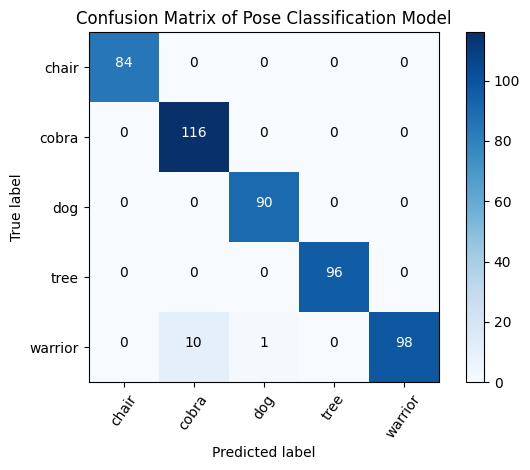
**IV. RESULT ANALYSIS**

The various performance metric such as precision, recall and F1 – score value is obtained for the prediction. The Performance table given below depicts the performance of the developed model. (TABLE 1)

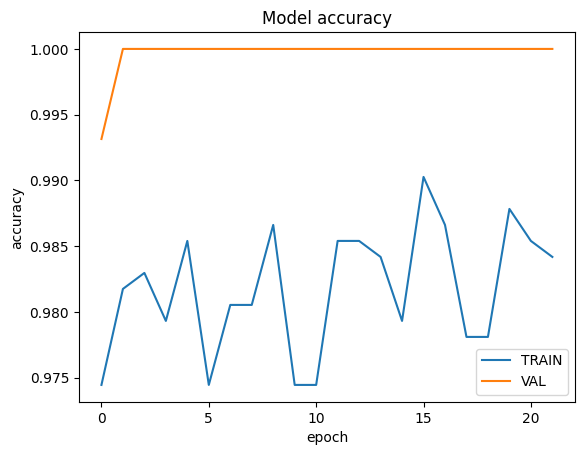
| **Type** | **Precision** | **Recall** | **F1-Score** | **Support** |
| --- | --- | --- | --- | --- |
| Chair | 1.00 | 1.00 | 1.00 | 84 |
| Cobra | 0.92 | 1.00 | 0.96 | 116 |
| Dog | 0.99 | 1.00 | 0.99 | 90 |
| Tree | 1.00 | 1.00 | 1.00 | 96 |
| Warrior | 1.00 | 0.90 | 0.95 | 109 |
| Accuracy | - | - | 0.98 | 495 |
| Macro Avg | 0.98 | 0.98 | 0.98 | 495 |
| Weighted Avg | 0.98 | 0.98 | 0.98 | 495 |

TABLE 1 (PERFORMANCE EVALUATION)

From the confusion matrix, it is identified that the model performed well on the 3 classes chair, cobra, warrior.



The comparison of model’s performance on validation data, train data during Each epoch. The validation data undergoes the accurate predicting in the model.



**V. CONCLUSION**

The objective of the proposed model is to enhance the Deep learning models to classify the images accurately in Spatiotemporal data. Employing large data sets, tuning parameters, dense layers, and dropout layers, the model performed precisely. By looking through the different categories of images, we increased the performance of classification and pose detection. This project will overcome all the difficulties in pose detection and classification.

**VI. REFERENCES**

[1]. “Researches Advanced in Human Pose Estimation Based on Deep Learning” 2022

DOI:10.1088/1742-6596/2404/1/012057

[2].” Pose-Oriented Transformer with Uncertainty-Guided Refinement for 2D-to-3D Human Pose Estimation” Feb 2023

DOI: https://doi.org/10.48550/arXiv.2302.07408

[3].” Detecting and grouping keypoints for multi-person pose estimation using instance-aware attention.” April 2023

DOI: <https://doi.org/10.1016/j.patcog.2022.109232>

[4]. “Human Posture Detection Using Image Augmentation and Hyperparameter-Optimized Transfer Learning Algorithms” 2022

DOI: <https://doi.org/10.3390/app121910156>

[5]. “Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model” 2023

DOI: <https://doi.org/10.3390/app13042700>

[6]. “Efficient Sitting Posture Recognition for Wheelchair Users: An Unsupervised Data-Driven Framework” 2023

DOI: 10.1109/MIM.2023.10146559

[7]. “Human Pose Estimation and Its Application to Action Recognition.” 2021

DOI: https://doi.org/10.1016/j.jvcir.2021.103055

[8]. “A comprehensive survey on human pose estimation approaches” 2023

DOI: DOI: 10.1109/ACCESS.2021.3139877

[9]. “Human Pose Estimation via Deep Neural Networks” 2014

DOI: 10.1109/TPAMI.2014.2313848

[10]. “ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation” 2022

DOI: https://doi.org/10.48550/arXiv.2204.12484