[1]Ying Chen

3D Convolutional Neural Networks based Movement Evaluation System for Gymnasts in Computer Vision Applications

Journal of Electrical Systems

Abstract: **-** This study introduces an innovative Movement Evaluation System, which utilizes state of art 3D CNNs in grassland of computer vision. This system presents a significant advancement in analyzing and assessing complex gymnastic movements, providing a comprehensive understanding of both spatial and temporal dynamics. By incorporating pose estimation algorithms, it accurately identifies body joints and positions to extract detailed spatial features. The 3D CNN further captures the temporal evolution of these features, enabling a precise analysis of the fluidity, rhythm, and synchronization of gymnastic movements over time. The effectiveness of this system is evaluated through performance metrics such as precision, recall, and accuracy. Aimed at enhancing sports science and athlete training, this research offers coaches, judges, and gymnasts a sophisticated tool for objective and standardized movement assessments. It has the potential to streamline the evaluation process in gymnastics and provide constructive feedback efficiently. The integration of 3D CNNs in this system marks a paradigm shift in utilizing advanced computer vision techniques to improve the understanding and refinement of gymnastic performances.

Keywords: 3D Convolutional Neural Networks (3D CNNs), DL Technique, Movement Evaluation System and Computer Vision Techniques

I. INTRODUCTION

Gymnastics is an intense athletic activity that demands a strong physical aptitude and technical skill. Participants in this sport frequently face repetitive and forceful impacts, necessitating significant core and extremity strength, power, and endurance. According to studies, gymnastics ranks among the most injury-prone sports in the National Collegiate Athletic Association (NCAA). This can be attributed to the intricate maneuvers involved and the demanding year-round training regimen. In particular, female athletes have been found to have the highest injury rates in soccer and gymnastics [1]. Long-term research on college athletes has revealed that female gymnasts are more susceptible to lower limb injuries, time-off injuries, surgeries, and season-ending injuries compared to their male counterparts.

Earning recognition in artistic gymnastics involves achieving excellence across six specific exercises, including floor routines, pommel horse, rings, vault, parallel bars, and horizontal bar. The qualification process for the Olympics has been modified for the 2017-2020 period, with individual gymnasts having the opportunity to qualify for the 2020 Summer Olympics in Tokyo. These athletic events typically involve activities that require short bursts of intense physical effort. The shortest event is the vault, lasting approximately 5 seconds on average, while floor exercises last around 60 seconds. Performances on pommel horse, rings, parallel bars, and horizontal bar usually take about 35 seconds. Overcoming resistance while suspended one's own body weight demands, great strength and contributes to muscular development. The grace and appeal of this traditional Olympic sport undoubtedly captivate spectators' attention [2]. However, achieving elite status in this sport requires a combination of youth and an appropriate physique to execute gymnastic maneuvers. Research has examined the ideal age to begin training and peak performance age during competitions but has yet to determine how training experience relates to somatotype and preferred event type in gymnastics.

As a result, Artistic Gymnastics has been categorized as a discipline that involves a significant amount of practice, systematic methods for learning specific movements, and a challenging competitive program. Artistic Gymnastics has firmly established itself as a fundamental division of sports. Essentially, it is a type of athletic competition where individuals use various equipment to display a series of diverse elements and combinations. According to the International Gymnastics Federation (FIG), Artistic Gymnastics is an Olympic sport that consists of separate Men's and Women's events, each with their own distinctive number and type of contests [3]. The Men's category includes six events: Floor Exercise, Pommel Horse, Still Rings, Vault, Parallel Bars, and Horizontal Bar; while

[1] Department of PE, Jiangnan University, Wuxi, Jiangsu, China, 214122

*Corresponding author's e-mail: 13915260079@sohu.com

female athletes participate in four events: Vault, Uneven Bars, Balance Beam, and Floor Exercise. In each event, participants must execute a sequence of complex gymnastic elements that are linked together to form a complete routine known as a gymnastic exercise. Experts in the field commonly believe that the Pommel Horse is traditionally associated with male gymnasts, while the Balance Beam is typically associated with female gymnasts.

In recent times, the Chinese Rhythmic Gymnastics (RG) has made significant advancements in world-class competitions and achieved numerous outstanding accomplishments. However, when compared to exceptional foreign gymnasts in the RG events, Chinese gymnasts still lack proficiency in personal skills. As a result, careful attention must be given to selecting and cultivating high-quality gymnasts. RG is a sport that prioritizes beauty and flawless execution of all movements, placing stringent demands on various aspects of physical fitness. Selecting suitable gymnasts is a complex and meticulous process, requiring us to focus on scientifically selecting individuals based on the expertise and experience of experts and coaches from both Chinese and foreign backgrounds, as well as considering the current situation in China. In this process, establishing a quality indicator system for gymnasts becomes a crucial step in scientifically selecting and evaluating potential candidates [4].

Identifying individuals with potential to excel in a specific sport, also known as talent identification is crucial for achieving high levels of performance. However, the current talent identification models are limited in individual sports and term. Some emerging models and programs in countries such as the FIG Age Group Development Programme, GFMT, USA Gymnastics TOPs program and WISGT are gaining recognition, particularly in artistic gymnastics [5]. The most widely used programs among coaches and researchers are those developed by the IGF and USA Gymnastics.

The FIG's is a program designed to train and identifies talented young gymnasts. It focuses on developing their physical fitness, including flexibility, power, strength, and endurance. This program is currently used by international coaches for gymnasts aged 6-11 and 18 years old [6]. The goal is to gradually prepare these gymnasts over a period of several years to excel in competition by improving their physical fitness. FIG is constantly working on enhancing and revising the training and talent identification program. Furthermore, educating mentor teacher to enhance their effectiveness and understanding in technical, psychological preparation and physical preparation is a crucial aspect of this program to ensure the well-being of the gymnasts [7].

The use of Computer Vision in evaluating sports performance has brought about a more accurate and unbiased approach. A particularly exciting advancement in this field is the implementation of a Movement Evaluation System for Gymnasts, which utilizes 3D Convolutional Neural Networks (3D CNNs). This innovative method combines the intricate nature of gymnastic movements with the detailed analysis provided by 3D CNNs, providing a comprehensive tool for assessing and improving athletes' skills. Since gymnastics requires agility, strength, and precision, a more nuanced evaluation system is necessary beyond traditional metrics [8]. By incorporating 3D CNNs into Computer Vision technology, it introduces a revolutionary aspect by capturing both the temporal and spatial dynamics of gymnastics routines. This enables a detailed examination of body movements, form, and technique in a three-dimensional space.

Our journey takes us into the intersection of Computer Vision and gymnastics evaluation, uncovering the potential of 3D CNNs to transform how we evaluate and improve the performance of gymnasts. By exploring the unique challenges presented by gymnastic movements and the capabilities of 3D CNNs to overcome them, we explore the potential for a sophisticated and impartial Movement Evaluation System. This system not only offers coaches and athletes valuable insights into strengths and areas for improvement, but also lays the groundwork for data-driven training methods that can advance gymnastic excellence. As we navigate through the crossroads of sports science, computer vision, and neural network technology, the Movement Evaluation System presented here marks a significant change in how we comprehend, analyze, and optimize the intricate movements of gymnasts, paving the way for precise training and elevated athletic success. The major contribution of the research is described below.

1.1 Research Contribution:

- Using 3D CNNs in the evaluation system for gymnasts captures intricate details of their performances in both space and time, providing a more comprehensive assessment than traditional methods.

- The automation of movement evaluation through 3D CNNs allows for quick and consistent feedback, giving gymnasts immediate insights into their performances and the ability to identify areas for improvement in real-time.
- The use of 3D CNNs enables the system to analyze complex movements with precision, such as flips, twists, and rotations, which is crucial for coaches and gymnasts seeking a detailed understanding of routines.
- This automated system streamlines the evaluation process, making it suitable for large groups of gymnasts. Coaches can efficiently assess and monitor individual athletes or entire teams, optimizing training strategies and performance outcomes.
- By implementing advanced technologies like 3D CNNs in gymnastics, a connection is established between sports science and athletic training. This integration reflects a forward-thinking approach to utilizing cutting-edge techniques for enhancing performance and developing athletes.

II. RELATED WORK:

In [9], a acquired indicator system using AHP for rhythmic gymnasts, incorporating physical, flexibility & strength, and speed & dexterity factors, aiding in their scientific selection and training processes. In [10], author developed a model for identifying critical the assessment and cultivation of physical and gymnastic abilities for identifying and nurturing talent among young male artistic gymnasts. It underscores the significance of power speed, isometric and explosive strength, strength endurance, as well as dynamic and static flexibility as pivotal factors. These findings provide valuable insights for talent identification and training strategies in men's artistic gymnastics. In [11], a new algorithm for detecting human basic movements from wearable sensor data is developed, focusing on minimizing computational demands while maintaining accuracy. The algorithm aims to preprocess data on the sensor device before combining information from multiple sensors for improved accuracy, demonstrated through fall detection and single step using a single tri-axial accelerometer. In [12], DRNNs for human activity recognition is constructed, emphasizing their capability to grasp dependencies across extensive ranges in input sequences of varying lengths. Comparing with traditional methods, DRNNs, particularly using LSTM architectures, outperform in recognizing activities, showcasing better performance than SVM, KNN, DBNs, and CNNs. In [13], improve the accuracy of recognizing human movements in musculoskeletal rehabilitation exercises, employing machine learning for exercise classification. Experimental results show 96% accuracy with multilayer dense neural networks and 100% accuracy using computer vision technology with a full body point set, confirming hypotheses on tracking system ranking and potential for exercise classification. In [14], this study proposes a CNN-LSTM hybrid deep learning model for HAR using smart watches, achieving 96.2% accuracy and 96.3% f-measure. The model outperforms traditional methods, offering enhanced performance in activity recognition tasks.

In [15], a novel approach using runtime models in wearable devices for human activity recognition, offering feedback to improve performance in repetitive physical activities. The method utilizes Restricted Boltzmann Machines on inertial measurement data, resulting in adjustments up to 3.68 times more accurate compared to the original movement data, aiding in precise activity execution. In [16], micro-expression recognition by Action Units instead of predicted emotions, achieving 86.35% accuracy on CASME II dataset with HOG 3D feature descriptors using classifying expressions. Action Units-based classification proves effective for objective micro-expression recognition. In [17], a hybrid approach using deep learning on a smartphone dataset for human activity recognition, achieving 98.70% accuracy with CNN and 96.36% accuracy with top 92 features. Feature selection significantly improves accuracy and model efficiency, demonstrated through experimental results. In [18], The process of machine learning is used to examine intricate travel behaviors in public transportation. This involves combining smart card information and urban data to recognize different patterns of mobility. The method involves grouping transit routes and utilizing graph embedding to identify levels of mobility. Interactive visualization is also utilized to explore the changing dynamics of transit behavior. In [19], a nested binary classifier method is presented for outlier detection in human movement phases, comparing it with popular ML algorithm and DNN. Results demonstrate the effectiveness of nested binary classifier for recognizing outlier patterns in human activity recognition systems.

In [20] introduces a HGR system that utilizes a curved piezoelectric sensor to collect data. The system employs optimized machine learning algorithms - SVM, k-NN with k-mer, RF, and achieves impressive accuracies of 94.11%, 97.18%, and 96.90% for SVM, RF, and k-NN respectively, showcasing its effective classification capabilities. In a separate study [21], a HDL model called CNN-GRU-AttNet is created for human activity

recognition using WiFi-based CSI data.The model outperforms conventional deep learning networks, achieving up to a 4.62% average accuracy improvement, making it effective for accurate activity classification. In [22], simple algorithm for automatic human posture classification is introduced using Kinect V2 data, achieving an accuracy of 94.9% with the second algorithm. The approach offers comparable accuracy to machine learning methods but with lower computational complexity and no need for training resources. In [23], an innovative algorithm inspired by mirror neurons is presented for automatic learning in man-machine interfaces, focusing on humanoid robot vocalization acquisition. The algorithm utilizes fuzzy articulatory rules and genetic optimization, demonstrating effectiveness synchronized facial movements with multi modal spech production. The summary of this earlier methods are given in table 1.

Table 1 – Earlier Research Summary

Ref. No	Algorithm	Methodology	Advantages	Disadvantages	Performance	Accuracy	Features Used	MeasurementApplications	Energy Results
[9]	AHP-Based Quality Indicator	Developed a quality indicator system for rhythmic gymnasts	Scientific basis for selection & evaluation	Requires expert input	NIL	NIL	Physical, Flexibility,	Gymnast training	NIL
[10]	Physical Fitness	Identified critical physical aspects in talent selection for male artistic gymnasts	Specific fitness factors determined	Limited to male gymnasts	Principal Components Analysis	NIL	Muscle strength,	Talent identification	NIL
[11]	Wearable Sensor-Based	Proposed algorithm for detecting human basic movements from wearable data	Improved accuracy with single sensor pre-processing	Challenging for sporadic movement	Intra- and inter-person	NIL	Temporal series points	Basic movement detection	NIL
[12]	Deep Recurrent Neural	Proposed use of DRNNs for human activity recognition, capturing long-range dependencies in input sequences	Improved recognition of temporal correlations	Fixed-length input windows	Unidirectional, bidirectional, cascaded	NIL	Temporal correlations,LSTM	Human activity recognition	NIL
[13]	Musculoskeletal Exercise	Introduced algorithms for accurate recognition of human movements during rehabilitation	Increased accuracy in exercise monitoring	Dependent on tracking systems	Machine Learning	96%	Tracking systems,	Rehabilitation exercise	N/A

		exercises							
[14]	CNN-LSTM Hybrid Model	Using smartwatches a CNN-LSTM model for human activity	Improved activity recognition with deep learning	Handcrafted feature extraction limits	Evaluation metrics	96.2%	CNN, LSTM	Activity recognition	NIL
[15]	Real-Time Gesture	Introduced a method for generating runtime models for correct execution of repetitive physical activities	Feedback for correct performance in real-time	Limited to pre-trained activities	Restricted Boltzmann Machine				
[16]	Micro-Expression Recognition	Proposed method for classifying micro-expressions based on Action Units	Objective method for emotion classification	Conflicts with human reporting	LBP-TOP, HOOF, HOG 3D	86.35%	Action Units	Micro-expression recognition	N/A
[17]	Human Activity Recognition	Developed a method for generating runtime models to guide users in correct performance of physical activities	Adaptive learning for new activities	Limited to pre-trained activities	Restricted Boltzmann Machine	N/A	Inertial measurement	Physical activity monitoring	N/A
[18]	Micro-Expression Recognition	Proposed classification of micro-expressions based on Action Units for improved recognition	Objective classification of facial expressions	Conflicts with human reporting	LBP-TOP, HOOF, HOG 3D	86.35%	Action Units	Micro-expression recognition	N/A
[19]	HAR using Smartphone	Hybrid approach using CNN on cloud-based platform for human activity recognition using smartphone dataset	Improved accuracy and efficiency with CNN	Feature selection impacts accuracy and training time	Principal component analysis	98.70%	CNN	Human activity recognition	N/A

[20]	Human Gesture Recognition	Data from curved piezoelectric sensor by human gestures	Accurate gesture recognition with machine learning	Dependent on sensor data collection	SVM, RF, k-NN	94.11%	K-mer	Human gesture recognition	N/A
[21]	Human Activity Recognition	Utilized WiFi-based sensing for human activity recognition, proposed CNN-GRU-AttNet for efficient activity classification	Automatic feature extraction from WiFi signals	Privacy concerns with WiFi data collection	CNN-GRU-AttNet	N/A	CSI data	Human action recognition	N/A
[22]	Posture Detection	Using Kinect V2	Low-cost and accurate posture classification	Limited to Kinect V2 usage	Total error of vector	94.9%	Kinect V2 joints	Posture classification	N/A
[23]	Vocalization Acquisition	Introduced algorithm for automatic vocalization acquisition from human tutors for humanoid robot	Effective multi-modal speech production	Complex metric similarity may require tuning	Fuzzy articulatory rules	N/A	Vocalization data	Speech synthesis	N/A

III. FUNDAMENTALS OF 3D NETWORKS:

3.1 3 Dimensional CNNs:

3D CNNs are a variation CNNs that can handle data in three dimensions. This type of network is particularly useful for processing volumetric data, such as videos, medical images (like MRI and CT scans), and 3D image data from sources like computer-aided design (CAD) models or 3D cameras. Its main purpose is to learn hierarchical spatial features directly from 3D data. Figure 1 as briefly explained about 3D CNN. The convolutional layers in 3D CNNs work similarly to those in traditional 2D CNNs, except that the operation is extended to work with volumetric data. The input data is a 3D volume, consisting of either a series of 2D frames over time or a grid of voxels for medical or CAD data. The filters used in this type of network are also three-dimensional, with depth, width, and height dimensions. Convolution Operation**:** 3D convolution operation can be represented mathematically as follows:

$$Output(i, j, k) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{p=0}^{P-1} Input(i+m, j+n, k+p).Filter(m, n, p) \quad (1)$$

The filter is applied to each position and multiplied with the corresponding section of the input volume. The resulting values are added together to create a single value in the output feature map, represented mathematically as a 3D convolution operation. 3D CNNs also utilize pooling layers, which are similar to those used in 2D CNNs, to decrease spatial dimensions and reduce computational complexity while preventing overfitting. Common pooling techniques include max pooling and average pooling. Non-linear activation functions, such as ReLU, are

applied after each convolutional and pooling layer to allow the network to learn complex patterns. A typical 3D CNN architecture consists of several layers including convolutional layers, activation functions, pooling layers, and fully connected layers at the end for tasks such as classification or regression.

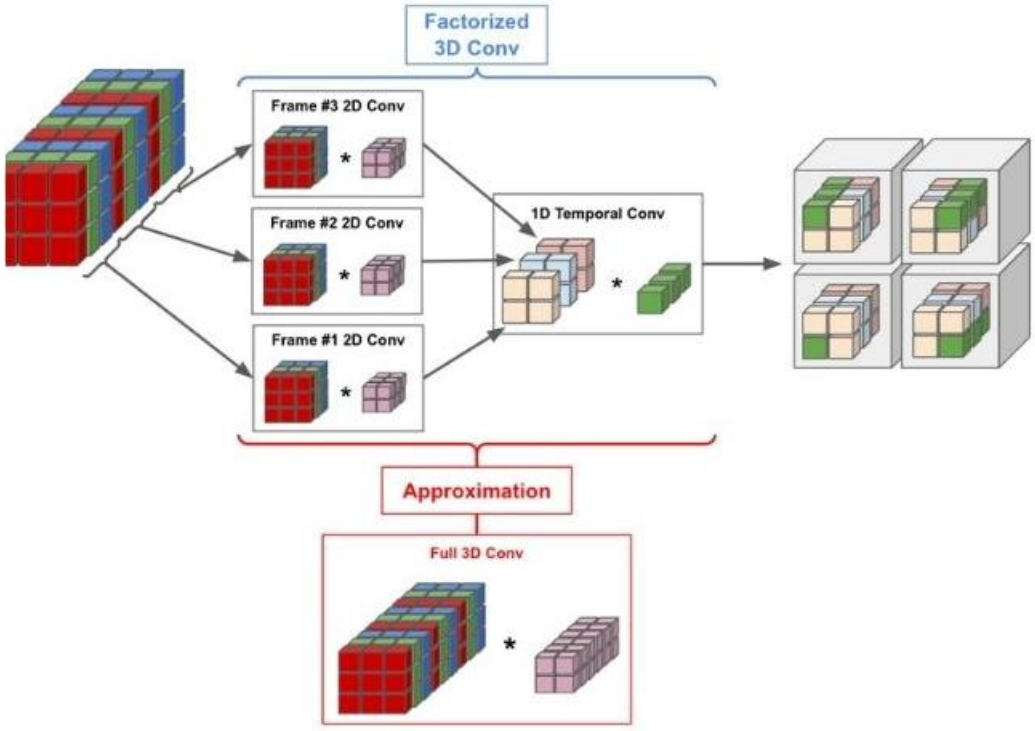

Figure 1 - 3D Convolutional Neural Networks

Training a 3D CNN involves carrying out forward propagation to calculate the output, comparing it to the actual labels using a dropping function, and using enhancement methods like SGD, RMSProp to backpropagate the error and update the weights. Some challenges and considerations when working with 3D CNNs include their high computational complexity due to larger input dimensions and more parameters, potential overfitting due to model complexity which can be mitigated with techniques like dropout, regularization, and data augmentation, and special attention needed for data preprocessing in areas such as medical imaging where steps like normalization and resizing are crucial. 3D CNN have significant capabilities in handling and deriving characteristics from volumetric data, such as 3D images and videos. By applying the concepts of 2D CNNs to the third dimension, they make it possible to create models that can comprehend patterns in both space and time within a 3D environment. These networks have a wide range of uses in fields like medical imaging, video analysis, robotics, and others. Figure 2 is architecture of 3D networks.

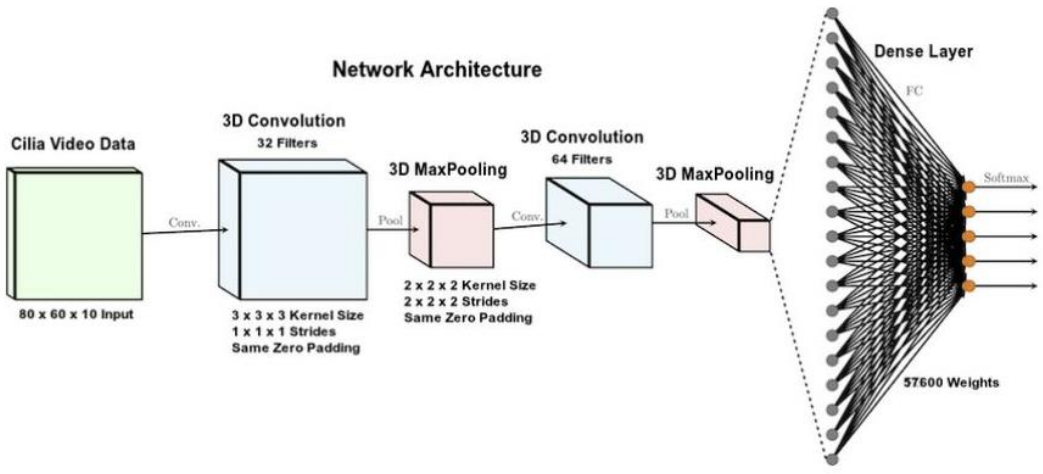

Figure 2 - 3D Network Architecture

3.2 Movement Evaluation System:

The "Movement Evaluation System" is a tool or method used to evaluate and examine human movement patterns. It is commonly utilized in fields such as sports science, physical therapy, biomechanics, and rehabilitation to comprehend, measure, and enhance the way individuals move. It typically involves collecting movement data using sensors like cameras, accelerometers, gyroscopes, or force plates and analyzing it in terms of joint angles, velocity, acceleration, and other relevant parameters. Figure 3 and Table 2 is discussed about movement analysis and Equation in Movement Analysis. The system includes components such as data acquisition through sensors, pre-processing to filter out unwanted noise from raw data, kinematics analysis focusing on joint angles and positions using equations, and dynamics analysis involving forces and torques using principles like Newton's laws of motion.

Table 2 - Equation in Movement Analysis

Equation	Description
$r(t) = (x(t), y(t), z(t))$	Position vector representing a point in 3D space over time t.
$v(t) = \frac{dv(t)}{dt} = (\frac{dv_x(t)}{dt}, \frac{dv_y(t)}{dt}, \frac{dv_z(t)}{dt})$	Velocity vector rate of change of position with respect to time.
$v(t) = \frac{dv(t)}{dt} = (\frac{dv_x(t)}{dt}, \frac{dv_y(t)}{dt}, \frac{dv_z(t)}{dt})$	Acceleration vector rate of change of velocity with respect to time.
$\theta(t) = \cos^{-1}(\frac{u_1.u_2}{\|u_1\| \|u_2\|})$	Angle between two connected body segments at time t.
$\omega(t) = \frac{d\theta(t)}{dt}$	Angular velocity rate of change of joint angle with respect to time.
$\alpha(t) = \frac{d\omega(T)}{dt}$	Angular acceleration rate of change of angular velocity with respect to time.
$F = ma$	Force vector representing the interaction causing a change in motion, where m is mass.
$\tau = rXF$	Torque vector representing the rotational force applied around an axis.

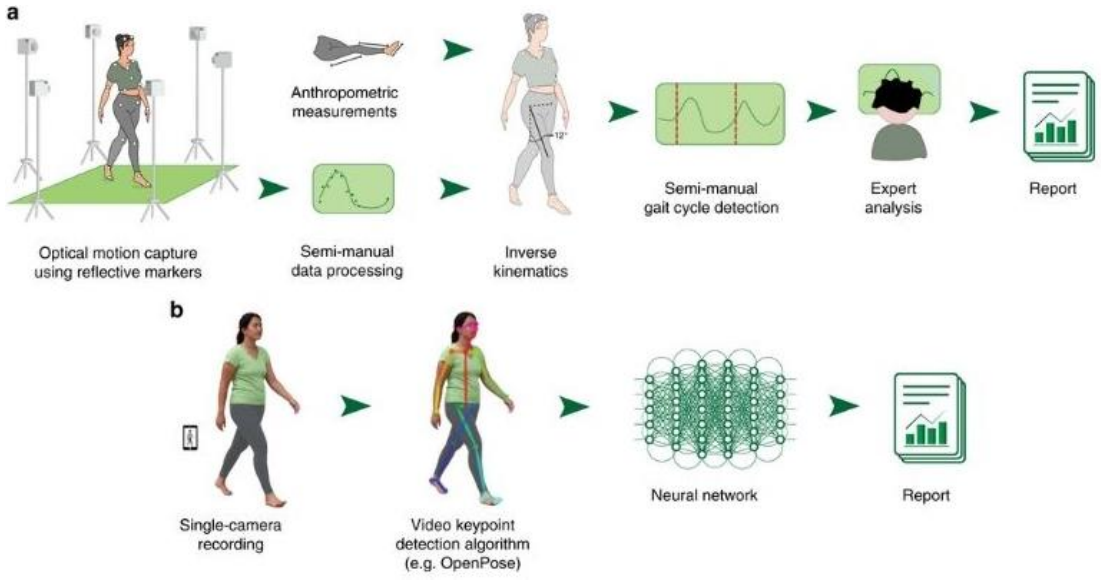

Figure 3 - Movement Analysis

Systems for Analyzing Movement:

- Systems for analyzing gait: Used to examine the patterns of walking and running in humans. Bio mechanical Systems for Sports: Evaluate the movements of athletes to increase their performance and reduce the risk of injuries.

- Systems for Physical Rehabilitation: Aid in creating customized treatment plans by analyzing movement. These systems typically integrate sensor information with mathematical models and algorithms to offer insights into the quality, effectiveness, and possible areas for enhancement of movement. They play a crucial role in maximizing performance, preventing injuries, and improving the rehabilitation process.

3.3 Computer Vision Techniques:

Computer Vision is an area of research that centers on teaching computers to comprehend and interpret visual data, much like how humans perceive and understand images and videos. This involves creating algorithms and methods to extract important information from visual data. Some important techniques used in computer vision, as well as relevant equations, include:

3.1.1 Image Preprocessing:

Prior to any analysis, images are typically prepared through methods such as enhancing distinct features or reducing noise. Common techniques for preprocessing include:

A. Grayscale Conversion: where R,Gand B are red, green, blue channel of an RGB image.

$$I_{gray} = 0.299.R + 0.587.G + 0.114.B \tag{2}$$

B. Image Blurring: where I(x,y) is original image, $I_{blurred}$(x,y) is blurrend imagem, and σ controls the amount of blur.

$$I_{blurred}(x, y) = \frac{1}{2\Pi\sigma^2} e^{\frac{-x2+y2}{2\sigma 2}} * I(x, y) \tag{3}$$

C. Image Classification: a typical CNN layer,

$$Output(i, j, k) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{p=0}^{P-1} Input(i+m, j+n, k+p).Filter(m, n, p) \tag{4}$$

D. Optical Flow: lucas- Kanade Method, (u,v) between two consecutive frames.

$$[I_x I_y][u, v] = -I_t \tag{5}$$

E. Image Registration: Normalized Cross-Correlation, measures between two images $\bar{I}_1$ and $\bar{I}_{21}$.

$$NCC(I_1, I_2) = \frac{\sum (I_1 - \bar{I}_1)(I - \bar{I}_{21})}{\sqrt{\sum (I - \bar{I}_1)^2 . \sum (I - \bar{I}_2)^2}} \tag{6}$$

F. Depth Estimation: Stereo Vision, uses disparity maps to estimate depth.

$$Z = \frac{f.T}{d} \tag{7}$$

Object detection is a process where a predetermined-sized window moves across an image, and an algorithm is utilized to identify whether the window contains a desired object. The Haar Feature-based Cascade Classifiers approach also uses Haar-like features and a series of classifiers to efficiently detect objects. These methods provide only a brief look at the many techniques utilized in computer vision. Depending on the specific goal, various combinations of these methods are utilized to gather important data from images and videos, allowing computers to perceive and comprehend the visual environment surrounding us.

IV. PROPOSED MES-3DCNN MODEL:

This research introduces a hybrid deep learning architecture, called 3DCNN-GRU-AttNet, for a HAR system that utilizes WiFi CSI data. The first step involves obtaining raw CSI data specifically for deep learning networks. Next, the raw data is preprocessed using denoising and segmentation techniques. Then, the preprocessed data is divided into training and evaluation sets using a five-fold cross-validation method. To generate features, the data samples undergo high-dimensional embedding through convolutional layers and a GRU layer within the 3DCNN-GRU-AttNet model. The effectiveness of the system is evaluated using standard metrics like accuracy, precision, recall, and F1-score.

To assess the performance of this model as well as baseline models for WiFi-based HAR, the publicly available CSI-HAR dataset was used. This dataset was collected using Nexmon tool on a Raspberry Pi-4GB to capture CSI data from transmitted and received signals. It contains 4000 samples over a duration of 20 seconds, with each line representing a 5-millisecond interval. The relevant segments of this data were extracted and saved in CSV files as matrices with 52 columns and 600 to 1100 rows depending on the activity duration. Label files were also provided to identify different actions. The dataset includes seven activities: walking, running, sitting down, lying down, standing up, bending and falling. These activities were repeated twenty times by three participants from various age groups in a controlled environment. Figure 4 shows the structure of the GRU network used in this study in detail.

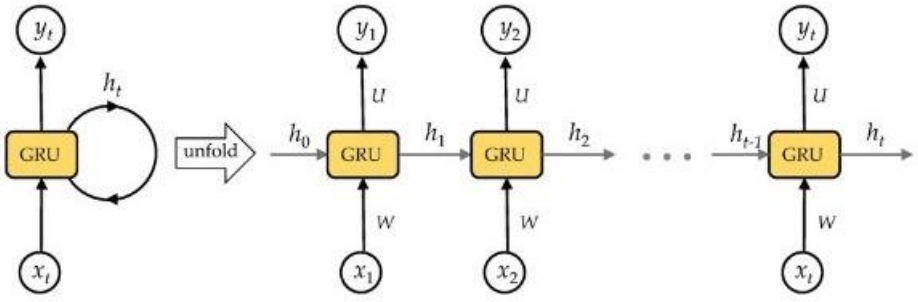

Figure 4 - The structure of GRU network

The WiFi data-set, there are seven activities including lying down, falling, walking, running, sitting down, standing up, and picking up. Six participants performed each activity twenty times. The collection of data involved a Wi-Fi router with a lone transmitter and a laptop equipped with the NIC-5300 Intel network interface card and three receivers. The transmitter and receiver were placed 3 meters apart in a direct line-of-sight arrangement. Each activity lasted for 20 seconds, and the data-set had an input vector consisting of raw CSI amplitude data and a 90-dimensional vector representing 3 antennas and 30 sub-carriers. However, for consistency with previous research, this study only utilized six of the seven activity categories from the original data-set and excluded the "picking up" activity. The majority of studies using this data-set have also evaluated six activity classes. To ensure fair comparison, our proposed model was also evaluated on these same six daily activities from the Stan-WiFi data-set. A single training data point is calculated by multiplying the number of samples by the number of features (500) by the number of timestamps to account for noise in the CSI and potential difficulties in distinguishing between different activities. Therefore, it is essential to use machine learning techniques for noise filtering and feature extraction in order to effectively analyze this data-set. This can include using methods like Butterworth low-pass filters or principal component analysis (PCA) to reduce complexity and identify key features where relevant information is concentrated. In this study, we followed recommendations from previous research by excluding the first principal component and selecting five subsequent principal components for feature extraction in order to address noise from internal state transitions present in all CSI streams. These noises

are highly correlated but can be separated from human motion signals by using PCA-derived components that exhibit no correlation with each other. This ensures that relevant data is preserved while balancing classification effectiveness and computational efficiency with five chosen principal components. After PCA denoising, specific characteristics are extracted from the CSI data to improve its usefulness for classification. In order to evaluate the effectiveness of denoising, we compared the PCA denoising technique by measuring the SNR, which represents the proportion of signal strength to noise strength. The denoised CSI signals demonstrated greater SNR values in comparison to unprocessed CSI samples, indicating that noise was successfully reduced.

4.1 Segmentation

The main focus of this study is to divide a signal into smaller sections or windows, a process known as segmentation, for two primary reasons. The first reason is to account for the variability in captured CSI signals, which can have different lengths and come from different subjects, making it difficult to identify patterns. The second reason is to effectively manage the time aspect of processing a large volume of CSI data, which saves time and computational resources. To overcome these challenges, our research adopts a fixed window size to segment the noisy CSI signal into smaller parts. Each segment is treated as an independent instance during the training of our proposed model, 3DCNN-GRU-AttNet. This approach increases efficiency and enhances the accuracy of our findings.

The 3DCNN-GRU-AttNet is an attention-based neural network specially designed for recognizing human activities using WiFi CSI data. It consists of five layers: input layer, two CNN layers, a GRU layer, an attention layer, a fully connected layer, and an output layer. Each layer has its unique function and will be explained in detail later on. 3DCNNs are commonly used in deep learning models for robust feature extraction from two-dimensional image data due to their fast processing capabilities. Unlike traditional fully connected neural networks, 3DCNNs have convolutional layers that are not entirely interconnected. Instead, inputs are connected to subsequent layers with shared weights among sub-regions in the input, resulting in outputs that are spatially correlated. This reduces the complexity of the network by reducing the number of connections and weights needed.

In this paper, a 3-dimensional convolutional neural network (3DCNN) was employed with two layers. The initial layer consisted of 64 filters with a kernel size of 3, while the second layer had 64 filters with a kernel size of 5. Max-pooling layers were also included with a constant pool size of 2. To connect the 3DCNN and GRU layers, a flattened layer was inserted. While 3DCNNs are effective in extracting features, they may have limitations when dealing with time-dependent inputs like the biometric signal data used in this study. This is due to the fact that when processing sequential data, the network's prediction of future states relies on the previous input state. To address this issue, the recurrent neural network (RNN) model analyzes each element of the temporal sequence and incorporates both current and previous inputs for the current RNN input. The output at a particular time step (t) is dependent on the output at the previous time step (t-1). While RNNs are theoretically capable of handling time series data of any length, they face challenges in real-world applications with lengthy time series due to gradient disappearance, which hinders learning long-term dependencies. To overcome this problem, a gated recurrent unit (GRU) was integrated as the memory component in the RNN architecture. The structure of a GRU cell is illustrated. GRU networks are simpler versions of LSTM networks within the realm of RNNs and offer improved computational efficiency while maintaining similar effectiveness as LSTM networks. A GRU unit includes an update gate and a reset gate that control how much each hidden state is modified in consecutive states within a computational model by regulating relevant and irrelevant information flow between them. Hidden state h_t at a specific time t integrates output of the update gate z_t, output of reset gate r_t, and current input x_t. Additionally, preceding hidden state h_{t-1} is taken into consideration, as illustrated below:

$$
\begin{aligned}
z_t &= \sigma(W_z x_t \Theta U_z H_{t-1}) \\
r_t &= \sigma(W_r x_t \Theta U_r H_{t-1}) \\
g_t &= \tanh(W_g x_t \Theta U_g (r_t \Theta h_{t-1})) \\
h_t &= ((1-z)\Theta h_{t-1}) \Theta (z_t \Theta g_t)
\end{aligned} \tag{8}
$$

The sigmoid function is denoted by the symbol σ, and the equations involve a small number of addition and multiplication operations. A detailed illustration of the calculation process for the self-attention mechanism can be seen in Figure 5.

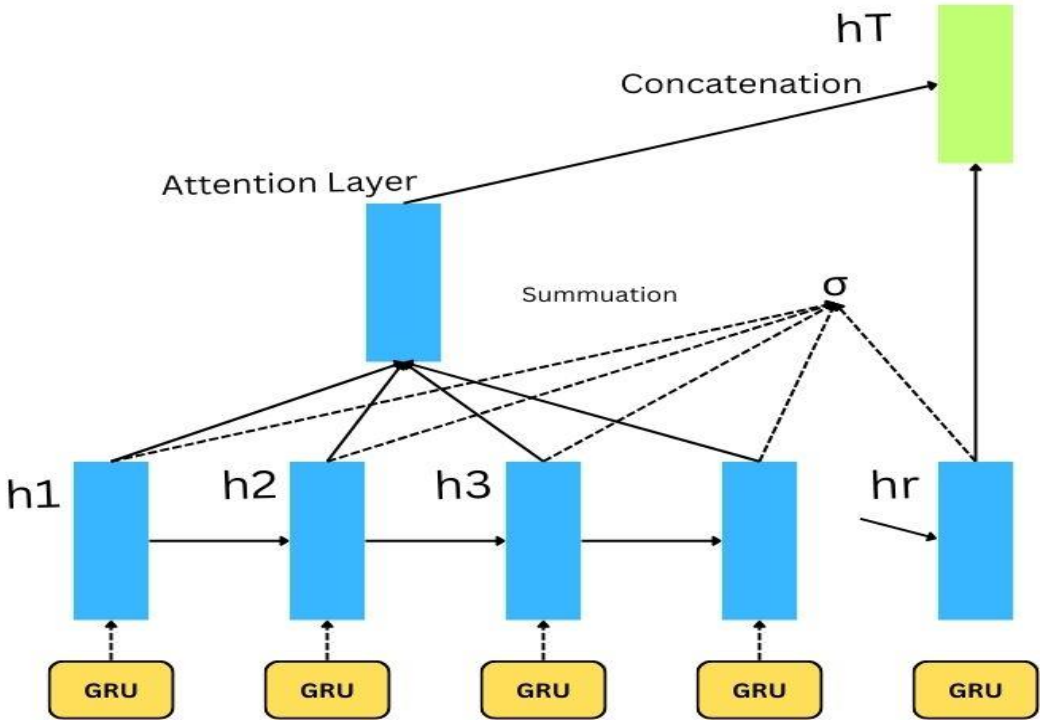

Figure 5 - Attention-based GRU for the classification

According to this study, once the GRU network has captured the necessary contextual elements, it is recommended to utilize a self-attention method for further capturing essential details. This technique assigns a higher weight to crucial information, resulting in a more precise understanding of the sequence's meaning. The calculation process for this self-attention mechanism is outlined. After the GRU layer processes the pre processed data $X = (x_1, x_2, ..., x_T)$, we obtain a vector $H = [h_1, h_2, h_3, ..., h_t, ..., h_T]$ represented by T, which is the length of the data in X, and ht, representing the hidden state of the GRU at time step t. We can then create a self-attention mechanism for the GRU using these steps:

$$\gamma_t = \tanh(w_2 h_t + b_2)$$
$$\beta_t = \frac{\exp((\gamma_t)^T w_2)}{\sum_t \exp((\gamma_t)^T w_2)}$$
$$\delta = \sum_t \beta_t h_{t,} \tag{9}$$

The attention layer uses a w_2 contextual vector, β_t weighted by a δ softmax function, to create a uniform representation of the entire sequence based on the hidden states and their corresponding attention weights. This is then followed by three dense layers with dropout regularization in the neural network. The first layer has 128 neurons and a dropout rate of 0.25, followed by a layer with 64 neurons and a similar dropout rate. The output layer has two neurons and all layers are activated using ReLU. The model was trained for 200 epochs with a batch size of 32, using categorical cross-entropy loss function and the Adam optimizer for optimization.

4.2 Hyper parameter and Training

The process of creating a statistical classification model can be broken down into three steps. The first step is selecting hyper parameters, such as batch size, activation function, learning rate, and number of iterations, which have a significant impact on the model's construction and training. It is crucial to have enough variation and

quantity of data during this phase. The second step involves training and validating the model using different sets of data. The training set is used to select hyper parameters, while the validation set is used to evaluate performance. In this particular study, the chosen training hyper parameters were a learning rate of, 100 epochs, and a batch size of 128. To ensure effective learning, a callback monitor was used to adjust the learning rate if no progress was seen after ten consecutive epochs. Data shuffling was also performed to introduce diversity in the data before each epoch. These hyper parameters were determined through multiple experiments and refinements to achieve optimal accuracy.

To assess the effectiveness of the proposed model, two publicly available datasets were used. As these datasets did not have predefined training and testing sets, a five-fold cross-validation technique was employed. This technique involved dividing the complete dataset into ten equally sized subsets randomly. The model fitting process then iterated using nine subsets for training and one subset for performance evaluation. This process was repeated ten times to ensure each subset underwent accurate testing. The overall performance of the model was evaluated by computing the average outcome from all iterations. The Adam optimizer was responsible for updating the model weights, while the cross-entropy loss function was used to measure error or loss during training.

4.3 Network Training and Evaluation Metrics

The confusion matrix is a useful method for assessing the effectiveness of DL models. It presents a visual and comprehensive overview of their performance. The mathematical representation of the multi-class confusion matrix includes rows for predicted classes and columns for actual classes.

$$C = \begin{bmatrix} c_{11}c_{12}c_{13}\ldots c_{1n} \\ c_{21}c_{22}c_{23}\ldots c_{2n} \\ c_{31}c_{32}c_{33}\ldots c_{3n} \\ \ldots \\ c_{n1}c_{n2}c_{n3}\ldots c_{mn} \end{bmatrix} \quad (10)$$

a. True positive: $TP(C_i) = C_{ii};$ (11)

b. False positive: $FP(C_i) = \sum_{l=1}^{n} c_{li} - TP(C_i);$ (12)

c. False negative: $FN(C_i) = \sum_{l=1}^{n} c_{li} - TP(C_i);$ (13)

d. True negative: $TN(C_i) = \sum_{l=1}^{n}\sum_{k=1}^{n} c_{li} - TP(C_i) - FP(C_i) - FN(C_i).$ (14)

The assessment of deep learning models in this research involved examining a confusion matrix and computing four commonly used measures: accuracy, precision, recall, and F1-score.

- Accuracy: This evaluates systematic error and is determined by dividing the total sum of true positives and true negatives by the overall number of records.

- Precision: This ratio is calculated by identifying correctly classified instances as belonging to a particular user's class out of all instances classified as belonging to that class.

- Recall: It is measured as the ratio of instances classified as belonging to a specific user's class out of all instances that actually belong to that class.

- F1-score: This measure combines precision and recall by using the harmonic mean.

The mathematical expressions for these evaluation metrics are as follows:

$$Accuracy = \frac{1}{|Class|} X \sum_{i=1}^{|Class|} \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i}$$

$$\Pr ecision = \frac{1}{|Class|} X \sum_{i=1}^{|Class|} \frac{TP_i}{TP_i + FP_i}$$

$$\Re call = \frac{1}{|Class|} X \sum_{i=1}^{|Class|} \frac{TP_i}{TP_i + FN_i}$$

$$F1 - score = 2X \frac{\Pr ecision X \Re call}{\Pr ecision + \Re call} \quad (15)$$

V. PRODUCTION ANALYSIS:

A simulation demonstration of proposed MES-3DCNN model python performance is discussed. Both the individual performance of the proposed work and its comparative results are elaborated below.

5.1 Gymnast activities/movement: In the simulation demonstration of the proposed MES-3DCNN we considered some of the movements and activities of the gymnasts like boxing, volleyball spiking, body weight squats and diving. The pictorial representation of those activities is given in figure 6(a), 6(b), 6(c) and 6(d). The data set which is used for this process are "https://www.crcv.ucf.edu/data/UCF101.php".

6(a) Boxing 6(b) Vollyball spiking

6(c) Body weight squats 6(d) Diving

Figure 6 - Gymnast Activities/Movement

5.2 MES-3DCNN Accuracy Calculation: The process of determining accuracy in a 3D-CNN based movement evaluation system involves evaluating the model's ability to predict gymnastic movements or poses correctly in comparison to the ground truth annotations. This is calculated by finding the proportion of correctly predicted movements or poses out of the total instances in the dataset, typically shown as a percentage. In the world of gymnastics, the precision of this system acts as a numerical indicator of its ability to accurately anticipate the

gymnasts' actions and positions. This is crucial for a variety of purposes, including evaluating skills, providing training guidance, and analyzing performance in gymnastics. In figure 7 the accuracy calculation of the proposed MES-3DCNN is illustrated.

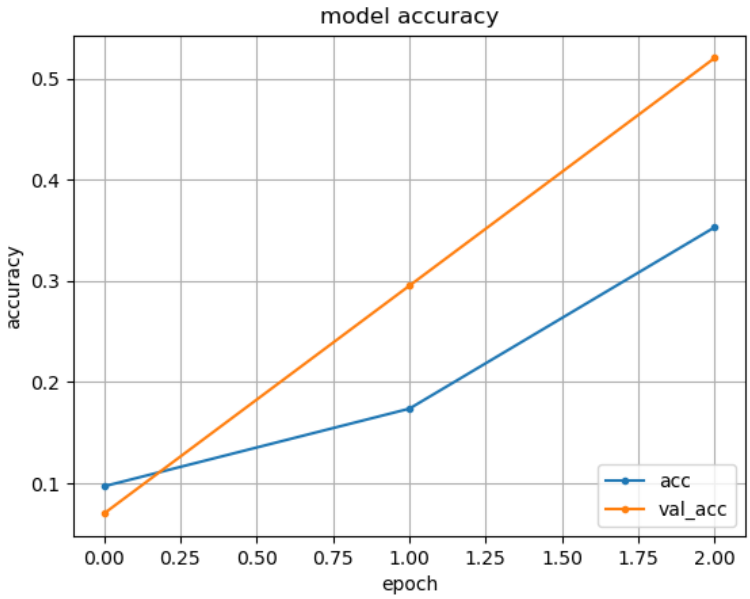

Figure 7 - MES-3DCNN Accuracy

5.3 MES-3DCNN Loss Calculation: A system for evaluating gymnasts' movements, using a 3D CNN, calculates loss by measuring the difference between predicted and actual movements. This loss function helps track performance during training and informs how to adjust the model's parameters to reduce this difference. It is determined by comparing the model's predictions with the ground truth annotations for each instance in the dataset, resulting in a single value that represents the discrepancy between the two. In figure 8 the loss calculation of the proposed MES-3DCNN is illustrated.

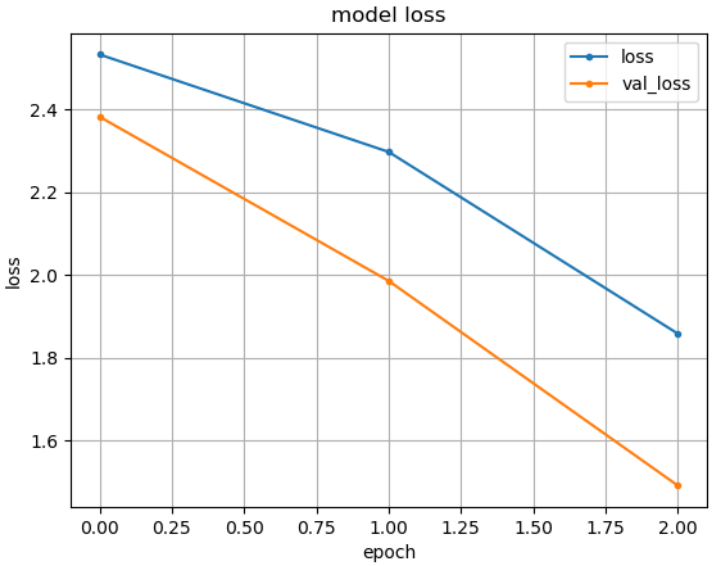

Figure 8 - MES-3DCNN Loss

5.4 Comparative Analysis: In this section the calculated results of the proposed MES-3DCNN in terms of certain methods like Precision, Recall, F1-Score, Accuracy are measured and it gets side by side with the earlier baseline methods like CNN [23], Fuzzy CNN [24] and Bi-LSTM-CNN [25]. The major activities which are considered for

the analysis are Diving, Boxing, Squats and Volleyball Spiking. The comparative value analysis are given in table 3

Table 3 - Comparative Value Analysis

Methods	Activity	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
CNN	Diving	93.6	90.6	88.3	75.3
	Boxing	94.6	89.2	89.1	83.3
	Squats	85.8	87.9	82.2	80.7
	Volleyball Spiking	88.1	88.5	89.9	76.8
Fuzzy CNN	Diving	94.6	87.6	84.6	78.3
	Boxing	85.8	84.1	85.9	83.7
	Squats	88.1	79.4	82.1	92.6
	Volleyball Spiking	92.6	81.2	91.3	89.4
Bi-LSTM-CNN	Diving	92.6	91.4	80.5	85.9
	Boxing	90.1	92.9	83.2	82.1
	Squats	89.8	89	92.6	91.3
	Volleyball Spiking	88.5	91.1	84.6	94.5
MES-3DCNN	Diving	97.6	95.6	93.7	98.3
	Boxing	94.1	96.1	94.9	99.1
	Squats	95.8	99.8	91.1	97.6
	Volleyball Spiking	98.5	98.5	90.5	99.4

5.4.1 Accuracy Calculation: The precision is measured by the ratio of accurately predicted actions or positions to sum of number of entries in data set. It is typically represented as a part. The 3D CNN model makes forecasts about actions or positions using input data, such as sequences of Diving, Boxing, Squats, and Volleyball Spiking. The dataset includes ground truth annotations or labels for these actions or positions. These annotations indicate the correct or desired actions or positions for each entry. The predicted actions or positions are compared to the ground truth annotations to determine their accuracy. In figure 9, the effectiveness of techniques likes CNN, Fuzzy CNN, Bi-LSTM-CNN and proposed MES-3DCNN is caluculating in terms of accuracy, and superiority of the proposed MES-3DCNN is demonstrated.

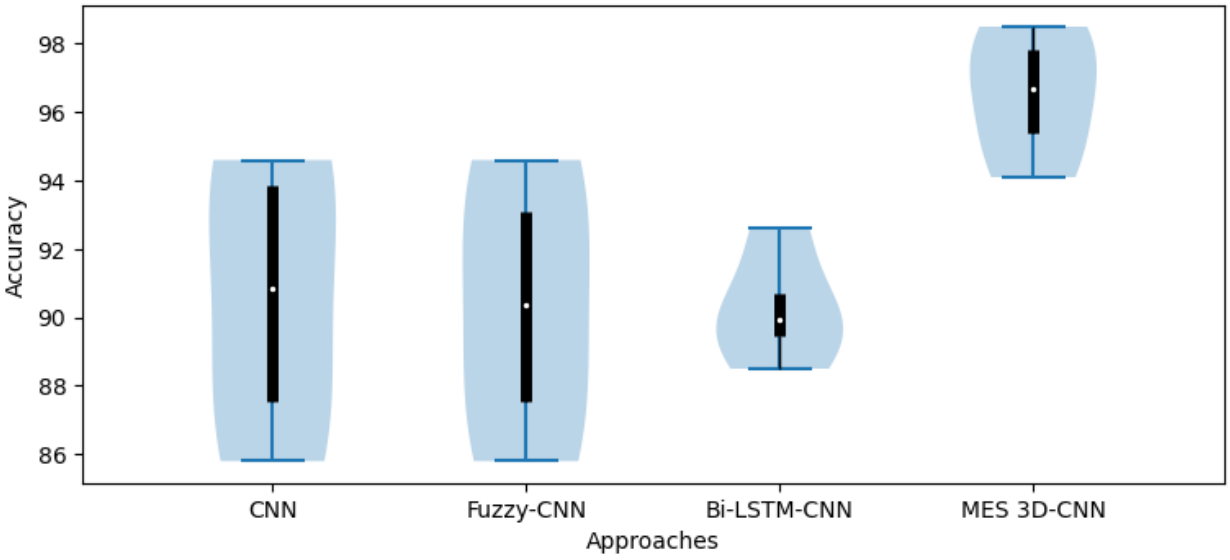

Figure 9 - Accuracy Calculation

5.4.2 Precision Calculation: The precision calculation is a crucial aspect of a 3D CNN movement evaluation system used for gymnasts, as it determines the model's ability to accurately predict gymnastic movements and poses. This calculation indicates the percentage of accurately identified positive instances (true positives) among all instances that the model classified as positive (true positives + false positives). It reflects the level of confidence in the model's classification of specific gymnastic movements, such as Diving, Boxing, Squats, and Volleyball Spiking. A higher precision indicates fewer false positive predictions, which is desirable for accurately capturing identified movements or poses. In figure 10, the performance of precision is calculated with respect to the presented methods like CNN, Fuzzy CNN, Bi-LSTM-CNN and proposed MES-3DCNN and as it proves that the proposed MES-3DCNN performed when compared with others.

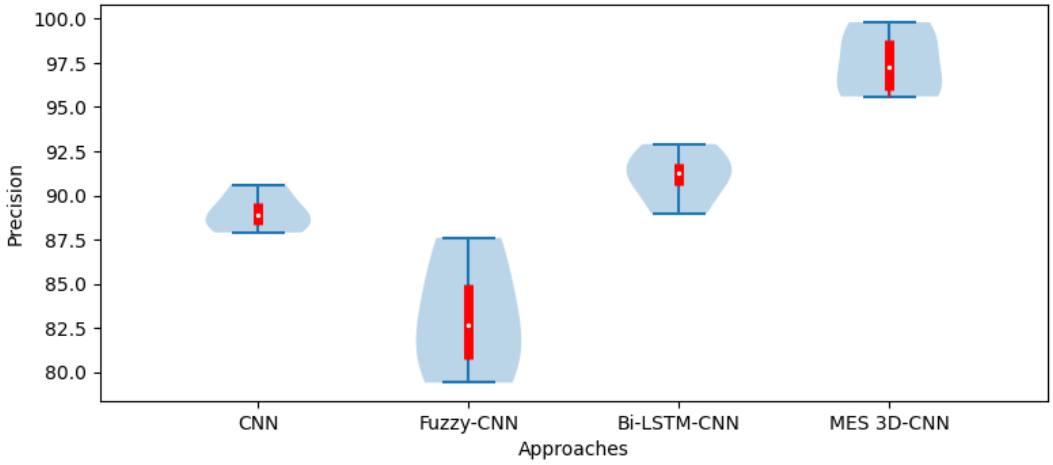

Figure 10 - Precision Calculation

5.4.3 Recall Calculation: In a system that evaluates movement using a 3D-CNN, calculating recall is crucial for determining the model's ability to accurately detect all instances of a particular gymnastic movement in a dataset that focuses on Diving, Boxing, Squats, and Volleyball Spiking. In figure 11, the performance of recall is calculated with respect to the methods like CNN, Fuzzy CNN, Bi-LSTM-CNN and proposed MES-3DCNN and as it proves that the proposed MES-3DCNN performed when compared with others.

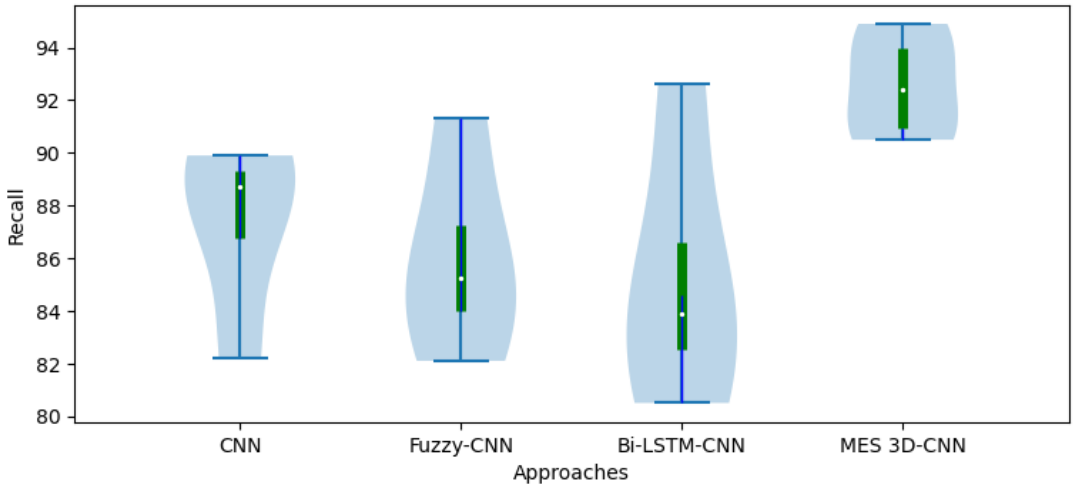

Figure 11 – Recall Calculation

Recall calculate section of correctly identified positive cases out of all positive cases, indicating how well the model captures instances of the gymnastic movement in gymnastic performances. For gymnasts being evaluated by this system, recall helps gauge the model's effectiveness in identifying and capturing all occurrences of specific movements within the dataset. A higher recall suggests that the model successfully captures a greater portion of actual instances of the gymnastic movement, which is important for ensuring that no relevant occurrences are overlooked.

5.4.4 F1-Score Calculation: The F1-score, utilized in a 3D CNN-based system for evaluating movements, is a comprehensive measure that takes into account both precision and recall. It serves as a single metric to assess the effectiveness of the model. This score strikes a balance between precision and recall by considering false positives and false negatives. It is particularly valuable when classes are not evenly distributed or when the consequences of incorrect classifications vary. In the context of assessing gymnastic movements, the F1-score plays a crucial role in determining the overall accuracy of the 3D CNN model in identifying specific movements from a dataset that includes Diving, Boxing, Squats, and Volleyball Spiking. A higher F1-score indicates a better balance between precision and recall, indicating a stronger and more reliable model for recognizing movements. The results from Figure 12 illustrate that the MES-3DCNN model outperformed other methods such as CNN, Fuzzy CNN, and Bi-LSTM-CNN in terms of F1-score evaluation.

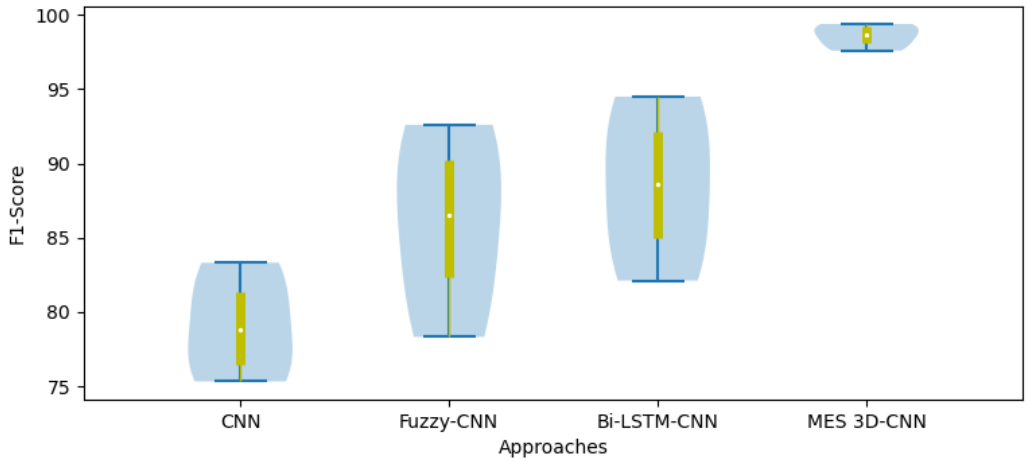

Figure 11 – F1-Score Calculation

VI. CONCLUSION

Using 3D convolutional neural networks (CNNs) facilitates a detailed and precise examination of gymnastics movements, capturing complex spatial and temporal intricacies. This level of specificity improves the accuracy of performance evaluations, offering valuable insights for refining skills. The system serves as a valuable tool for skill enhancement, empowering gymnasts to improve their techniques in real-time. The integration of 3D CNNs enables the assessment of temporal dynamics in gymnastic routines, which is essential for evaluating fluidity, timing, and synchronization of movements and providing a holistic understanding of performance quality. This technology also opens up opportunities for further research at the intersection of computer vision and sports science. Researchers can explore new methodologies, improve existing models, and contribute to the development of benchmarks for evaluating movement in gymnastics.

REFERENCES

[1] Daphne Ling, PhD, Mark Sleeper, et.al, “Identification of Risk Factors for Injury in Women’s Collegiate Gymnastics With the Gymnastics Functional Measurement Tool”, Original Research, pp. 1-6, 2019, doi: 10.1002/pmrj.12184

[2] Katarzyna Sterkowicz-Przybycień, Stanisław Sterkowicz, et.al, “Somatotype, body composition, and physical fitness in artistic gymnasts depending on age and preferred event”, PLoS ONE, vol. 14, no. 2, pp. e0211533, 2019, doi: 10.1371/journal.pone.0211533

[3] Andrzej Kochanowicz, et.al, “Neuromuscular and Torque Kinetic Changes After 10 Months of Explosive Sport Training in Prepubertal Gymnasts”, Pediatric Exercise Science, vol. 31, no.1, pp.1-8, 2019, doi: 10.1123/pes.2018-0034

[4] MichaB Wendt, Krystyna CieVlik, et.al, “Effectiveness of Combined General Rehabilitation Gymnastics and Muscle Energy Techniques in Older Women with Chronic Low Back Pain”, Hindawi BioMed Research International, vol. 20, pp. 14, 2019, doi: 10.1155/2019/2060987

[5] Klaudia Kaszala, Henriette Steiner-Komoróczki, et.al, “Examination of Movement in Young Gymnasts”, Acta Polytechnica Hungarica, Vol. 20, no. 4, pp. 181-193, 2023, doi: 10.12700/APH.20.4.2023.4.10

[6] Ivan Jurak, Dalibor Kiseljak, et.al, “Assessing Young Gymnasts’ Dynamic Posture: A Comparison of Methods”, J. appl. health science, vol. 6, no. 1, pp. 129-135, 2020, doi: 10.24141/1/6/1/12

[7] Serena W.J. Khong, Pui Wah Kong, “A Simple and Objective Method for Analyzing a Gymnastics Skill”, European Journal of Physical Education and Sport, vol. 12, no.2, 2016, doi: 10.13187/ejpe.2016.12.46

[8] Lovro Štefan, Goran Sporis, et.al, "Do more behavioral risk factors increase the odds of having chronic diseases in young adults? A population-based study", 2019

[9] Lin Luo, "Study on Quality Indicator System of Rhythmic Gymnasts in Analytic Hierarchy Process", Earth and Environmental Science, vol. 81, pp. 012188, 2017, doi :10.1088/1755-1315/81/1/012188

[10] Bessem Mkaouer, Sarra Hammoudi-Nassib, et.al, "Evaluating the physical and basic gymnastics skills assessment for talent identification in men's artistic gymnastics proposed by the International Gymnastics Federation", International Gymnastics Federation, vol. 35, no. 4, pp. 383–392, 2018, doi: 10.5114/biolsport.2018.78059

[11] Mario Munoz-Organero, Ahmad Lotfi, "Human Movement Recognition Based on the Stochastic Characterisation of Acceleration Data", Sensors, vol. 16, pp. 1464, 2016, doi: 10.3390/s16091464

[12] Abdulmajid Murad and Jae-Young Pyun, "Deep Recurrent Neural Networks for Human Activity Recognition", Sensors, vvol. 17, pp. 2556, 2017, doi: 10.3390/s17112556

[13] Artem Obukhov, Andrey Volkov, etr.al, "Examination of the Accuracy of Movement Tracking Systems for Monitoring Exercise for Musculoskeletal Rehabilitation", Sensors, vol. 23, pp. 8058, 2023, doi: 10.3390/s23198058

[14] Sakorn Mekruksavanich, Anuchit Jitpattanakul, et.asl, "Enhanced Hand-Oriented Activity Recognition Based on Smartwatch Sensor Data Using LSTMs", Symmetry, vol. 12, pp. 1570, 2020, doi: 10.3390/sym12091570

[15] Marcio Alencar, Raimundo Barreto, et.al, "An Online Method for Supporting and Monitoring Repetitive Physical Activities Based on Restricted Boltzmann Machines", J. Sensor Actuator Network, vol. 12, pp. 70, 2023, doi: 10.3390/jsan12050070

[16] Adrian K. Davison, Walied Merghani, et.al, "Objective Classes for Micro-Facial Expression Recognition", J. Imaging, vol. 4, pp. 119, 2018, doi: 10.3390/jimaging4100119

[17] Sujan Ray, Khaldoon Alshouiliy, et.al, "Dimensionality Reduction for Human Activity Recognition Using Google Colab", Information, vol. 12, pp. 6, 2021, doi: 10.3390/ info12010006

[18] Tong Zhang, Jianlong Wang, et.al, "Integrating Geovisual Analytics with Machine Learning for Human Mobility Pattern Discovery", ISPRS Int. J. Geo-Information, vol. 8, pp. 434, 2019, doi: 10.3390/ijgi8100434

[19] Agnieszka Duraj, Daniel Duczyminski, "Nested Binary Classifier as an Outlier Detection Method in Human Activity Recognition Systems", Entropy, vol. 25, pp. 1121, 2023, doi: 10.3390/e25081121

[20] Sathishkumar Subburaj, Chih-Ho Yeh, et.al, "K-mer-Based Human Gesture Recognition (KHGR) Using Curved Piezoelectric Sensor", Electronics, vol. 12, pp. 210, 2023, doi: 10.3390/electronics12010210

[21] Sakorn Mekruksavanich, Wikanda Phaphan, et.al, "Attention-Based Hybrid Deep Learning Network for Human Activity Recognition Using WiFi Channel State Information", Application science, vol. 13, pp. 8884, 2023, doi: 10.3390/app13158884

[22] Tatiana Klishkovskaia, Andrey Aksenov, et.al, "Development of Classification Algorithms for the Detection of Postures Using Non-Marker-Based Motion Capture Systems", Application science, vol. 10, pp. 4028, 2020, doi: 10.3390/app10114028

[23] Alfredo Cuzzocrea, Enzo Mumolo, et.al, "An Effective and Efficient Genetic-Fuzzy Algorithm for Supporting Advanced Human-Machine Interfaces in Big Data Settings", Algorithms, vol. 13, pp. 13, 2020, doi: 10.3390/a13010013

[24] Gholamiangonabadi, D and Grolinger, K "Personalized models for human activity recognition with wearable sensors: deep neural networks and signal processing", Applied Intelligence, vol. 53, no. 5, pp. 6041-6061, 2023.

[25] Khodabandelou, G., Moon, H., et.al, "A fuzzy convolutional attention-based GRU network for human activity recognition", Engineering Applications of Artificial Intelligence, vol. 118, pp. 105702, 2023

[26] Soni, V., Yadav, H., Semwal, et.al, "A novel smartphone-based human activity recognition using deep learning in health care", In Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference, pp. 493-503, 2023