ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20240506001521
Original scientific paper

Innovative Edge Computing for Real-Time Video Surveillance and Taekwondo Training Enhancement

Nithya S.*, Samaya Pillai Iyengar, Poobalan A., Parameswari A.

Abstract: The constantly growing volume of data created globally makes it impossible for the centralised cloud computing method to provide low-latency, high-efficiency surveillance camera services. In order to alleviate transmission pressure, the load on the main cloud server, and the end to end latency of the video surveillance system, a distributed computing architecture is developed that immediately analyzes peripheral video data. By lowering the probability of tracker drift or malfunction in the videos, the suggested Enhanced Multiple Instance Learning with Whale Optimization Technique (EMIL-WOM) enables the classifier to extract the features with lower computing costs and shorter computation times. For various scenarios, the optimised neural network generates computation models, which are then logically placed in edge devices. The level of Taekwondo is chosen to address the uneven teaching quality for the goal of real-time analysis as society develops. To solve the teaching challenges in the Taekwondo learning process and improve the calibre of Taekwondo, the researcher conducted particular study in relation to the tactile learning theory. This research uses scientific and technological resources as a guide to assess technical actions and strategies and apply them to specific educational experiments for testing. This work analyses and recommends a way for making innovative services based on the edge computing paradigm. This experimental technique eliminates several interoperability and service scalability issues with conventional design. The suggested EMIL-WOM achieves 96.5% accuracy, 56.1% computational complexity, 32.4% *RMSE*, 24.1% *RAE*, 30% *MAE*, and 45.3 seconds of response time when compared to existing approaches.

Keywords: edge computing; enhanced multiple instance learning; feature extraction; video surveillance; whale optimization method

1 INTRODUCTION

Taekwondo is an ancient martial art that originates in Korea and is organized and scientific. It is currently a popular international sport [1]. Due to its highly dynamic motions and emphasis on physical training, taekwondo differs from other martial arts in terms of their physical requirements. It is also a martial art that places a strong emphasis on learning how to use hands, fists, elbows, knees, feet, and any other part of the body to execute techniques. In a conference room or other learning environment, pupils get Taekwondo instruction from their teacher. The instructor demonstrates and explains the movements before the students follow pattern [2]. In order to enhance and obtain excellent performance, the trainer is crucial to Taekwondo training because it is his or her job to lead and instruct the students on how to carry out each technique correctly. Training would be challenging when there are huge groups involved, according to [3]. All students may not be able to visibly follow the trainer's performance because they are positioned too distant or they can be obstructed through other students because Taekwondo training often comprises a big group of learners [4]. In addition, the students struggle with a shortage of time to attend training courses, and instructors are occasionally unavailable [5]. As a result, the students engaged in self-directed training anytime they desired to enhance their Taekwondo skills. Since the trainer is not present during this type of training, the learners can do it whenever and wherever they like. Self-directed learning is commonly carried out with the use of already available supplemental learning materials, such DVD/CD (video), YouTube (online video sites), websites, and books. The currently available supplemental materials, however, have a number of drawbacks [6]. Video is rarely used as an instructional medium [7]. The loss of a third dimension and feedback is a major issue with the use of videos. Although the movies can be fast-forwarded, paused, or rewound, they cannot capture all of the movements of the body. However, using books alone is insufficient because they lack interactive features and contain static material that is difficult to follow when moving. These videos may be kept in the cloud, often known as cloud computing. It combines the capabilities of distributed computing, parallel computing, and grid computing to provide users with scalable computational resources and storage regions. It can achieve this by utilising a storage cluster made up of several low-cost devices and a computational cluster made up of numerous ordinary servers. Cloud technology is gradually becoming a network application trend by providing a new software mode that allows computational power to flow throughout the network and provides computational resources services that are comparable to those offered by supercomputers through combination. Supercomputer programmes have a very high barrier to entry since they need a very specialised and costly investment. Cloud computing, on the other hand, is a low-cost substitute for high performance computing services as it links shared PCs and servers to create computer clusters over the Internet [9]. The distributed storage of a huge amount of training data and the simultaneous participation of several cloud nodes in neural network computation are both capabilities of the cloud computing environment. The neural network dependent on a cloud computing environment significantly reduces the preparation time similar to the neural network operating systems on a single computer and has the capability to retain and analyze enormous amounts of training data [10]. In order to increase the efficiency of network information storage while addressing the enormous volume of data stored in the network, this research represents a platform for cloud computing using a neural network. Depending on this, the following are the contributions of this paper.

The whale optimization technique works in conjunction with Enhanced Multiple Instance Learning (EMIL-WOM), which makes it easier to choose characteristics and categorise data while spending minimal time and money on computing. It also lessens the probability that the tracker will drift or fail. A weighted average for data association and state updating approach. Greater real-time processing speed is taken into consideration with enhanced precision and accuracy. Including a comprehensive benchmark against several leading techniques in edge computing and video analysis, with detailed performance metrics, would provide a clearer context for the strengths and weaknesses of EMIL-WOM and help validate its superiority in practical applications.

The EMIL-WOM (Enhanced Multiple Instance Learning with Whale Optimization Method) system shows promising results in specific scenarios by effectively addressing video-based human motion recognition challenges. It reduces system latency and enhances user privacy by processing data at the edge. However, to broaden its applicability, a detailed analysis of its scalability across diverse applications and environments is necessary. Integrating the EMIL-WOM system into existing setups can be facilitated through a modular design, ensuring compatibility with various hardware and software platforms, and adopting standard communication protocols. Comprehensive training and support for system administrators and users will further ensure smooth implementation and operation. By addressing these areas, the EMIL-WOM system can be refined to perform reliably across a broader range of applications and environments, enhancing its effectiveness in human motion recognition tasks while maintaining user privacy and system efficiency.

2 RELATED WORKS

A focus of research in computer vision is visual behavior detection, which has applications in driverless vehicles, human-computer interaction, and surveillance cameras. In order to fully utilise the spatial characteristics in the video, it is important to learn how to extract useful features. Different classes can be made based on various feature extraction techniques: behaviour recognition methods based on deep neural networks [13] and behaviour recognition methods based on traditional classifiers [11].

The precision of a categorization procedure developed by Xie et al. [14] relying on Levenberg-Marquardt backpropagation (LMBP) algorithm oriented neural network strategy was assessed using a wide range of permutations of target movement, camera view, subframe location, and picture modality. This methodology employs one-dimensional spatial information obtained from consecutive frames as input. Taekwondo trajectory analysis *d*. Due to the fact that Multitarget Detection is crucial for robots, Zhang et al. created the multiobjective PSO method. It assists with fuzzy PID controller's membership function and fuzzy rules optimization. The optimization outcomes are compared and examined after obtaining the vector sets for the different optimization targets. Computer-aided instructional methods are introducing changes to contemporary higher education at an unpredictably rapid rate in college and university classrooms. Long Short Term Memory (LSTM), which was recommended by Nurahmadan et al. [16], was used to help this teaching approach. Multilayer Perceptron Approach minimizes the computing time required by LSTM, which is a problem when attempting to predict Taekwondo motions. The DAS (Dynamic Able Success) club provided the core information for this study, which included two kicks and two blocks. Hailong et al. [17]'s objective is to improve the effectiveness of Taekwondo instruction delivered online by implementing artificial intelligence algorithms. The model helps perform the motion guidance and exercises using the simulation approach and can improve the students' movement by identifying the Taekwondo movement features of the students. Data loss is a significant problem when teaching online, but Zhong et al. [18]'s novel graph convolution model addresses it by automatically analysing the technical actions of karate athletes, tracing their trajectory, and calculating action frequency statistics. The shortcomings of conventional tactical intelligence analysis techniques are well compensated for by the technologies.

The convolutional neural network is a multilayer network, just like the conventional BP neural network. The convolutional neural network differs because each layer's neurons are located in two-dimensional space and are typically targeted at picture signals. A new encoder-decoder-reconstructor structure was developed by Ji et al. [19] for video labeling that collected the data in both videos and captions. This CNN must be integrated with Various Instance Learning. On the basis of the encoder-decoder-reconstructor framework, a unique multi-instance multi-label dual learning approach (MIMLDL) was created for video captions. The restriction of choosing a fixed k value is eliminated by the Distributionally Robust Optimization (DRO) constraint introduced by Sapkota et al. in [20]. In order to accurately obtain the covariance framework of high-dimensional data, we improve the GP kernel to corrected basis functions by utilising a deep neural network to find adaptive basis functions. This ensures the predictive power over high-dimensional data, which is typical in MIL and includes videos and images.

Despite significant advancements in human behavior identification and a wealth of scientific research findings, the following possible issues still require immediate attention: The behavior judgement difficulty of Long-term video still faces numerous challenges because the duration of videos that are typically shown is significantly longer. All of the data sets utilised for recognising human behaviour are dependent on the video level. It cannot properly detect the existence of continuous video picture frames if it solely relies on classic CNN. Because recognising human behaviour is a perpetual process and some action films contain video segments that are readily misunderstood, it is important to take the video frame series' constraints before and after. For certain movement patterns, there is greater interference, which significantly lowers judgmental accuracy.

3 PROPOSED METHOD

This section presents a deep learning-based video control scheme by combining the cloud computing and collaborative learning models. This concept is made up of four components, as shown in Fig. 1 a cloud server, fog nodes, an enterprise server, and a camera. On the cloud server, open data sets are computed, open neural network models are created, and network models are distributed to business servers. Only after building a neural network model, does the cloud server get the network weight from each corporate server, analyses it collectively, and modifies the results under the part on neural networks, which also covers accuracy evaluation, action detection from frames, and data collection.

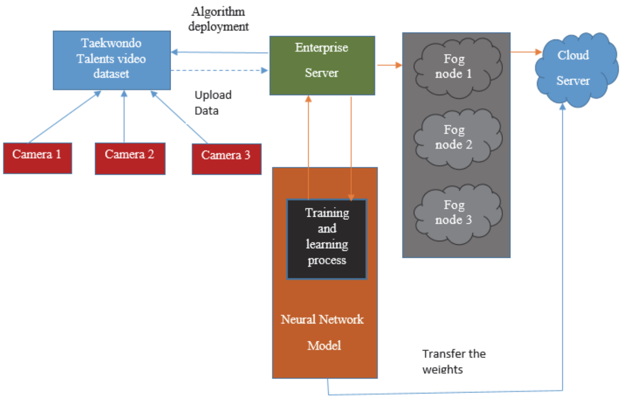

Figure 1 System architecture of storage of video and its processing steps via neural network

Taekwondo unit method: picture frames were collected for data collecting in order to create TUHAD. Taekwondo professionals from TUHAD completed each unit technique. The multidimensional image sequences were collected using a mono RGB-D camera. The gathered frames can be preprocessed using the Whale Optimization-based feature extraction and Advanced Stochastic Gradient Descent (A-SGD) approach. TUHAD was used to educate the key frame-based Enhanced Multiple Instance Learning activity recognition framework. The most effective input configuration for taekwondo activity recognition was then determined by testing the identification model's accuracy with different input configurations.

3.1 Preprocessing of Video Frames

Here, two optimization aspects of the Advanced Stochastic Gradient Descent (A-SGD) technique are the Implicit User Queries Structure and the Explicit Online Cloud or Explicit Picture in Online/Offline Storage Device. The following approach significantly facilitates the frames by utilizing these two processing characteristics. We concentrate on instances where the cost function $f(x)$ has an L - Lipschitz gradient $\left\|\nabla f(x)-\nabla f(y)\right\|<L\left\|x-y\right\|$ and it is λ - strongly convex i.e., for all $x, y \in R^d$ [17] as given in Eq. (1).

$$f(x) \geq f(y)+\left(f(y), x-y\right)+\frac{\lambda}{2}\left\|x-y\right\| \tag{1}$$

We represent the collection ${}_{i=1}^{n}\{f_i\}$ of functions that make up f by F. For our method, we maintain an additional parameter $\alpha_t^i \in R^d$ for each $f_i \in F$. We use A^t to denote ${}_{i=1}^{n}\{\alpha_i\}$. Because the algorithm does not describe the subroutine, it is important to note that this is still abstract. This function is the essential $\{\alpha_i\}$ updating mechanism. Schedule_update is triggered every m iterations in the case of A-SGD. Thus, A^t remains constant for m iterations, and at the mth iteration, all α_i are updated to the present iteration. For any t in the range $[T]$, at changes at thec t-th iteration in A-SGD. This change is limited to a single A^t element and is determined by the function it's index (the one selected at iteration t). The A-SGD update in so far as each iteration only modifies one of the α_i. But rather than i^t, the update for A^{t+1} is based on i^{t+1}. A pair of sequence $\{\varphi_K(w)\} n_{k=0}^{\infty}$ and $\{\alpha_K\} n_{k=0}^{\infty} \cdot \alpha_K \geq 0$ denotes an estimate sequence of function $f(w)$ if $\alpha_K \to 0$, $w \in R^d$ and all $k \geq 0$ given in Eq. (2) and Eq. (3) as:

$$\varphi_K \leq \left(\alpha_{K+1}\right)=\left(1-\varphi_K\right) f(w) \tag{2}$$

$$\begin{aligned} &\alpha_{K+1}(w)=\left(1-\varphi_K\right)\varphi_K(w)+ \\ &+\alpha_K\left(f\left(v_K\right)+\left(\nabla f\left(v_K\right)\right)\right), w-v_K+\frac{\mu}{2} \end{aligned} \tag{3}$$

where (i) f is distinctly convex; (ii) v_K is a Random function on R^d; (iii) $\{v_K\} x_{k=0}^{\infty}$ is a Random sequence on R^d; (iv) $\{x_{k=0}^{\infty}$: $\alpha_K \in (0, 1)$, $\propto k=0$ $\alpha k = \propto$; $(v) \lambda_0 = 1$. Note that $\nabla^2 \varphi_0(w)=\gamma I$ where I is persona matrix. Hence, prove that $\nabla^2 \varphi_0(w)=\gamma I$ for all $k \geq 0$. In fact, if it is accurate for some k, then, according to the iteration technique of $\{\varphi_k\}$, as given in Eq. (4)

$$\Delta^2 \varphi_{K+1}(w)=\left(1-\alpha_K\right)\Delta^2 \varphi_K(w)+\mu\alpha_K I=\mu I \tag{4}$$

The conversion of the simple scene format into the split tokenized scene format depends on certain user-requested process features. After that, two separate procedures named Frame Match (TM) and Visual Effects are applied to the input video file in tokenized scene format. We use a method by choosing a thresholding technique on that parameter in ability to detect potential shortcuts:

$$T_p = \mu \cdot \sigma \tag{5}$$

where σ indicates the standard deviation and μ is the mean value of the APE of the size of a series of frames, p. It takes values in the range of three to five, depending on our traditional analysis. Because they are GLCM-based, the Thibaluat characteristics offer a new spatial-temporal connection that can be used to characterise the region of interest (ROI) in a sequence of frames. This allows for the enhanced identification of shot boundaries by capturing minute textures and structures in frames. The frame patterns gradually alter in the gradual transitions in accordance with changes in the values of the Zijdenbos similarity index and the relationships between the frames. The angular metric is another name for this angle, which is a member of the inner product group. The formula for the Zijdenbos similarity coefficient, $zij(sim)$, is as follows.

$$zij(sim)[a,b] = \frac{2(a \cdot b)}{|a| + |b|} = \\ = \frac{x(a)x(b) + y(a)y(b) + z(a)z(b)}{\sqrt{x(a)^2 + y(a)^2 + z(a)^2} \cdot \sqrt{x(b)^2 + y(b)^2 + z(b)^2}} \tag{6}$$

where, B stands for the claimed vector obtained in the most recent pixel, and AIS is the estimated vector based on the most recent observations. Assume that $a = (xa, ya, va)$ and $b = (xb, yb, vb)$. The three components of each of these vectors are x, y, and v, where x and y represent the xy coordinates and v is the velocity. In Eq. (1), the pixel's coordinates are xb, yb; its velocity is vbis; and its estimated locations are xa, ya, depending on the pixel that was reached earlier than the most recent one.

The following succinct description of the shot boundary recognition system:

Step 1. Initially, if the scaled frame difference $d_{\log}(i)$ is greater than a max-threshold thmax then the present frame is chosen to user shot frame $d_{\log}(i) \geq th_{\max}$.

Step 2. Then we estimate the unique definition difference value $bd_{\log}(i)$, $fd_{\log}(i)$ as: $bd_{\log}(i) = d_{\log}(i) - d_{\log}(i-1)$.

$$fd_{\log}(i) = d_{\log}(i+1) - d_{\log}(i) \tag{7}$$

The estimated difference value $bd_{\log}(i)$, $fd_{\log}(i)$ must be greater than a k - threshold k_{global}, $bd_{\log}(i) \geq k_{\text{global}} \,\&\&\, fd_{\log}(i) \geq k_{\text{global}}$.

Setp 3. Finally, in the Zijdenbos similarity index the disparity between every estimated frame's distance $bfd_{\log}(i)$ is defined as:

$$bfd_{\log}(i) = \sqrt{bd_{\log}(i) + fd_{\log}(i)} \tag{8}$$

It also needs to be higher than a global threshold th_{global}.

3.2 Frame extraction by Whale Optimization Method

Key frame extraction uses elements such colours (especially the colour histogram), edges, forms, and optical flow. Here, weused the whale optimization technique, which consists of three steps: hunting for prey, encircling prey, and spiraling prey. The application's exploration phase is one of them, and it involves looking for prey. The algorithm's exploitation stage also involves prey that is encircled and prey that spirals in a logarithmic pattern.

Exploitation Phase: When humpback whales hunt, they first locate their target and circle them. The following equations illustrate the mathematical formula of diminishing encirclement:

$$D = \left| C \cdot X^*_{(t)} - X_{(t)} \right| X_{(t+1)} = X^*_{(t)} - A \cdot D \tag{9}$$

where t indicates the current number of iterations, X is the position vector, X^* displays the current location of the best successful solution, | | stands for the true operation, and indicates an element-by-element multiplication. Two parameters, A and C, are assessed in the manner described below:

$$A = 2 \cdot a \cdot r - a \text{ and } C = 2 \cdot r \tag{10}$$

The value of a is evaluated by $a = 2 - t\left(\frac{2}{MaxIter}\right)$ and $MaxIter$ is the large number of iterations. Spiral updating positions, which work in tandem with the preceding diminishing encircling to form the humpback whales' bubble-net assault approach, are another technique utilised in the exploitation phase. The expressions are:

$$\boldsymbol{D'} = \left| X^*_{(t)} - X_{(t)} \right| X_{(t+1)} = D' \cdot e^{b1} \cdot \cos(2\pi l) + X^*_{(t)} \tag{11}$$

where b is a constant that controls the logarithmic spiral's shape, and l is a random variable in the range [1, 1]. During the exploitation phase, diminishing encircling and spiral upgrading positioning are applied together. Here is the mathematical model:

$$X_{(t+1)} = \begin{cases} X^*_{(t)} - A \cdot Dp > 0.5 \\ D' \cdot e^{b1} \cdot \cos(2\pi l) + X^*_{(t)} p \geq 0.5 \end{cases} \tag{12}$$

where p is a value at random inside [–1, 1]. It shows a 50% probability of choosing the decreasing encircling strategy or the spiral-shaped mechanism to change the whales' location during an optimisation stage. The optimal solution at the time is X^*, which is determined by evaluating each member of the randomly initialized population against the fitness function. After that, until the stop criteria is fulfilled, the algorithm is continued. When $|A| < 1$ or when $|A| \geq 1$, search agents adjust their place at each iteration in accordance with either a randomly chosen person or the best result so far. Based on p, the WOA algorithm decides whether to use spiral or circular movement to extract features.

3.3 Enhanced Multiple Instance Learning

Let $X \in x$ be the instance-level random variable for the input and $Y_X \in y_x$ is the instance output, where the space of X is xCR^d, and d is the dimension of the feature vector, also, $Y_X = \{0, 1\}$ Consider $B \in b$ is the random variable input at the frames level and $Y_B \in y_b$ is the frames-level output.

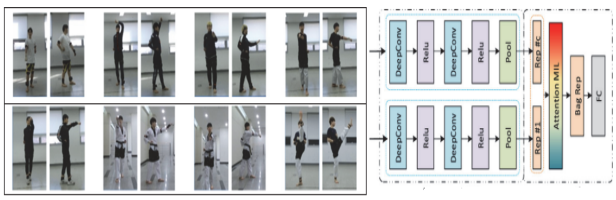

Figure 2 Architecture of EMIL with input as eight taekwondo actions in TUHAD

The binary classifier can be specifically defined for instance-based binary classification problems when the training samples are i.i.d.:

$$f(x) = \operatorname{sign}(g(x)) \in \{+1, \ -1\} \tag{13}$$

where $g(\cdot)$ is a mapping function $g{:}R^d \to R$.

$$\min_{\alpha\in R^d, \beta\in R} \frac{1}{2}\|\alpha\|^2 + C\sum\nolimits_{i=1}^{N} \max \boxed{f(\ )} \left\{0, 1 - y_i\left(\alpha^T \varnothing(X_i) + \beta\right)\right\} \tag{14}$$

where $C > 0$ is a penalty parameter, $\beta \in R$ is a bias parameter, $\alpha \in R^m$ is an m - dimensional classifier weight to be learned, $x_i \in R^d$ is the d - dimensional instance feature vector, $\varphi : R^d \to R^m$ and y_i is the instance label. The initial term of Eq. (14) is the L_2 regularization and the next term is the hinge loss term, defined as $g(x) = \max(0, 1-x)$ The hinge loss function is not distinguishable, which should be noted. Employing the instances' highest possible score to deduce the frames' label from their labels

$$\min_{\alpha\in R^d, \beta\in R} \frac{1}{2}\|\alpha\|^2 + C\sum\nolimits_{j=1}^{N_B} \max_{x\in B_j}\left\{0, 1 - y_{B_j}\left(\max\left(\alpha^T \varnothing(X_i) + \beta\right)\right)\right\} \tag{15}$$

where N_B is the total number of the frames. The primary presumption is that abnormal frames always have a greater anomaly score than typical frames:

$$\max_{X\in B_a} f(x) > \max_{X\in B_n} f(x), \text{ for any } B_a \in B_n \ \& \text{ any } B_n \in b_n \tag{16}$$

where B_a and B_n are the given ensembles of failure and normal video frames, accordingly, and $f(\cdot)$ is the evaluated anomaly score for an instance in a frames. The first term of Eq. (16) denotes the instance (fragment) that, in a set of anomalous frames (clip), has the greatest arbitrary number and is most likely to be an aberration. The video fragment with the greatest abnormality score in a given normal video is represented by the second term of Eq. (16), which is most likely a normal occurrence. It comprises the aberrant video segments' temporal smoothness and the sparsity term.

$$\begin{aligned} &L(B_a, B_n) = \\ &= \max\left(0{,}1 - \max\left(0{,}1 - \max_{X\in B_a} f(x) + \max_{X\in B_n} f(x) + 1\right) + \right. \\ &+ \pi_1 \sum\nolimits_{x\in B_a} \cup f(x)^2 + \pi_2 \sum\nolimits_{x\in B_n} \cup f(x)^2 + h \end{aligned} \tag{17}$$

where π_1 and π_2 are hyper-parameters that control the amount of trade-off, and the $\Delta f(\cdot) : R^d \to [-1, 1]$ is the discrete gradient function, defined as:

$$\Delta f(x_i) \ = f(x_i) - f(x_i + 1) \tag{18}$$

Consider $B_a = \{x_1, ..., x_n\}$ where n is the number of occurrences in a video frame; in our suggested solution, which was put through its paces in the experiments, $n = 32$. The L_1 regularization sparse, the final factor in Eq. (18), highlights the fact that anomalies in aberrant footage only last for a shorter amount of time. It is clear that the positive frames are penalized with scoring low in the loss function's initial term. The formula, which is mostly made up of four sections, illustrates how the loss function of EMIL is enhanced. Loss center, loss of bounding box size Loss scale, loss of assurance Loss conf, and loss in classification Loss class.

3.4 Action Recognition

When evaluating the human body and developing a human body movement model, it is important to appropriately consider the variability in the human body as well as the randomness of the movement process. All joints are combined and integrated, and each group of three joints has a shared portion when the Part Dividing layer receives the athlete's bone characteristics. The result of combining all the parts is a channel matrix. Graph convolution is then used to retrieve the three-channel spatiotemporal properties from such channel matrices. In order to compile the Taekwondo competitors' actions, we stacked seven input units.

Table1 Taekwondo action evaluation criteria

Score	Requirement
1	Attack with your hands to the upper or centre portion; attack with your feet to the same region; and play perfect defence.
2	Attack the top area with your feet, leap into the air, and use your hands to strike it while keeping one foot above the ground. Then, jump towards the middle section and attack with your feet.
3	Leap up, turn around, and hit the elevated part; leap up 180 degrees or higher; jump up and strike the middle portion; or jump up and do a greater than 180-degree rotation.
4	Jump up 180 degrees and then instantly turn around to land on the top area; land on the center section by turning around 360 ° or more.
5	Jump up to 360 degrees, pivot, and strike to the top part.

Each input unit consists of a ReLU activation layer, a BatchNorm layer, and a partial perceptual layer. Each unit's convolutional layer is completed with a step size of 1 and in accordance with specifications. Taekwondo uses an accurate and powerful blow to the effective scoring region while using the permitted methods in order to determine the winner of the match. Headbutts are valued at three points, spin kicks and back kicks two points each, other methods one point, and the referee's reading receives no extra points. The areas of the face that can be targeted, the belly, and both ribs all contribute to a method's maximum three-point score. The knockdown would be counted if the authorized approach is utilized to strike a non-valid scoring area that is shielded by safety equipment. Tab. 1 lists the exact grading guidelines.

4 PERFORMANCE ANALYSIS

The effectiveness of the Enhanced Multiple Instance Learning with Whale Optimization Method (EMIL-WOM) is evaluated using criteria like accuracy, precision, recall,

$F1$-score, and AUC-score. Three cutting-edge techniques, including Levenberg-Marquardt backpropagation (LMBP) [14], Long Short Term Memory (LSTM) [16], and Multi-label Dual Learning Approach (MIMLDL) [19], are evaluated with some of these parameters.

Dataset description-TUHAD: Taekwondo Technique Unit the Human Action Dataset includes almost 100,000 image sequences of 10 professionals using the eight fundamental taekwondo methods. Every movement was recorded on film from both the front and the side. Konkuk University's Institutional Review Board approved the study procedure (7001355-202004-HR-375). The experiment to collect action data for TUHAD included 10 professional taekwondo poomsae demonstrators (or equivalents) of various heights and physical characteristics. The subjects were not subjected to any clothing restrictions.

Table 2 Taekwondo event data detection and evaluation

Taekwondo action	Usage frequency	Score frequency	Total score	Action recognition Accuracy / %
Front kick	1358	25	57	89
Cross kick	1457	57	56	85
Push kick	345	54	87	86
Downward kick	568	67	43	84
Single leg kick	786	13	68	82
Hook kick	458	18	98	81
Back kick	645	43	64	84
Backspin sick	32	47	23	83

Table 3 Comparison of accuracy and computation complexity

Parameter = accuracy					Parameter = computation complexity				
frames	LMBP	LSTM	MIMLDL	EMIL-WOM	frames	LMBP	LSTM	MIMLDL	EMIL-WOM
10	75	85	75	92	10	81	88	77	91
20	74	84	74	91	20	79	77	88	89
30	75	85	75	90	30	89	87	83	89
40	76	82	76	91	40	82	82	85	92
50	75	83	75	90	50	71	83	84	81

Incorporating a section that focuses on the design and usability of the edge devices, along with feedback from actual users (such as Taekwondo instructors and students), would offer insights into the system's effectiveness and areas for improvement in real-world settings. Conducting user experience tests and gathering qualitative data on the interface could inform further refinements.

Tab. 2 shows the Taekwondo event data detection and evaluation and Tab. 3 shows the comparison of accuracy and computational complexity for proposed EMIL-WOM method with existing methods of LMBP, LSTM, and MIMLDL.

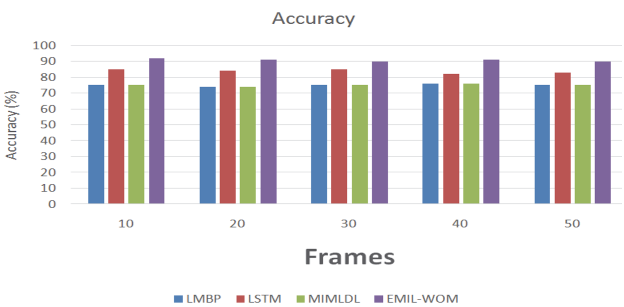

Figure 3 Comparison of accuracy

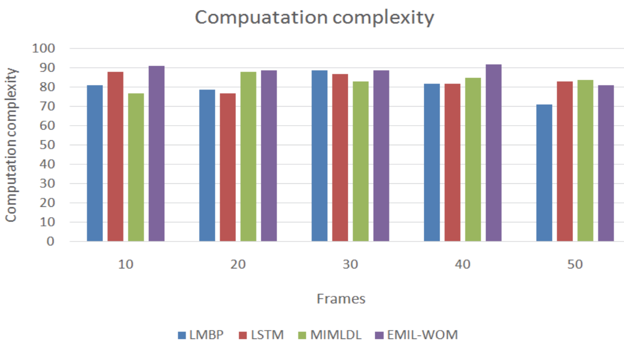

Figure 4 Comparison of computation complexity

Fig. 5 depicts the performance of $RMSE$ between the developed EMIL-WOM and the conventional LMBP, LSTM, and MIMLDL. The X and Y axes, accordingly, display the number of frames and percentage values. The proposed EMIL-WOM methodology, in contrary, obtains 32.4% of $RMSE$, which is 24%, 12.6%, and 24.9% less than the LMBP, LSTM, and MIMLDL techniques.

Fig. 6 displays the RAE comparison between the proposed EMIL-WOM and the conventional LMBP, LSTM, and MIMLDL. The X and Y axes, respectively, display the number of frames and percentage values. The proposed EMIL-WOM methodology, conversely, only obtains 24.1% of RAE, which is 23.3%, 23.9%, and 25% less than the LMBP, LSTM, and MIMLDL methods.

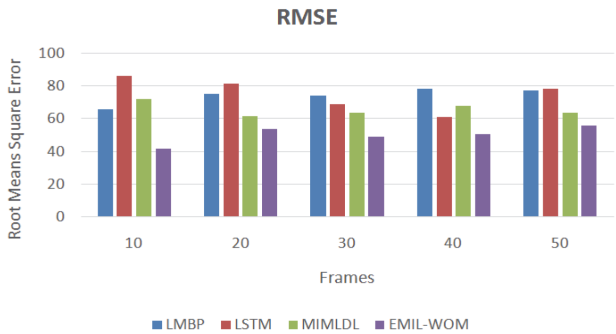

Figure 5 Comparison of *RMSE*

Fig. 5 depicts the $RMSE$ comparison between the suggested EMIL-WOM and the conventional LMBP, LSTM, and MIMLDL. The X and Y axes, accordingly, display the number of frames and percentage values. The proposed EMIL-WOM methodology, in contrary, obtains 32.4% of $RMSE$, which is 24%, 12.6%, and 24.9% less than the LMBP, LSTM, and MIMLDL techniques.

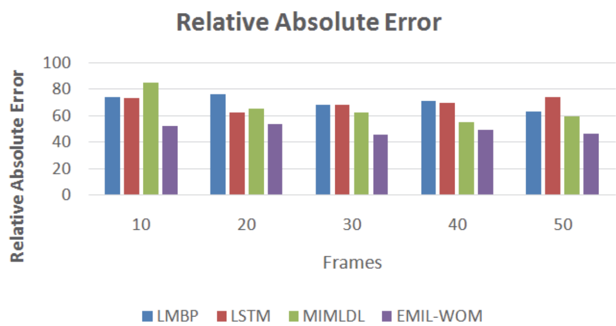

Figure 6 Comparison of *RAE*

Fig. 6 displays the RAE comparison between the proposed EMIL-WOM and the conventional LMBP, LSTM, and MIMLDL. The *X* and *Y* axes, respectively, display the number of frames and percentage values. The proposed EMIL-WOM methodology, conversely, only obtains 24.1% of *RAE*, which is 23.3%, 23.9%, and 25% less than the LMBP, LSTM, and MIMLDL methods.

Tab. 4 shows the comparison of *MAE* and Response time (sec) for proposed EMIL-WOM method with existing methods of LMBP, LSTM, and MIMLDL.

Table 4 Comparison of *MAE* and Response time

Parameter = Mean Absolute Error (*MAE*)					Parameter = Response time / sec				
frames	LMBP	LSTM	MIMLDL	EMIL-WOM	frames	LMBP	LSTM	MIMLDL	EMIL-WOM
10	74	84	54	55	10	64	76	80	54
20	73	76	61	42	20	65	87	79	45
30	79	73	79	52	30	73	75	80	42
40	75	87	76	44	40	62	79	76	49
50	79	89	75	57	50	81	78	79	44

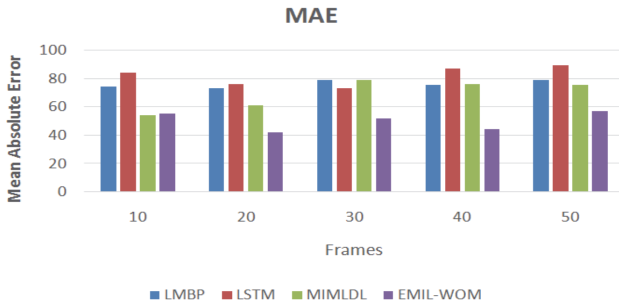

Figure 7 Computation of *MAE*

Fig. 7 displays the performance of *MAE* between the proposed EMIL-WOM and the conventional LMBP, LSTM, and MIMLDL. The *X* and *Y* axes, accordingly, display the number of frames and percentage values. The proposed EMIL-WOM methodology, in contrary, only obtains 30% of *MAE*, which is 31.7%, 35%, and 37% less than the conventional LMBP, LSTM, and MIMLDL techniques.

Fig. 8 compares the response times of the suggested EMIL-WOM with those of the conventional LMBP, LSTM, and MIMLDL. The *X* and *Y* axes, accordingly, display the number of frames and percentage values. The proposed EMIL-WOM methodology, in comparison, provides a response time of 45.3 seconds, which is 30 seconds, 32 seconds, and 35 seconds faster than the LMBP, LSTM, and MIMLDL methods.

Tab. 5 shows the overall comparative analysis of suggested method EMIL-WOM with existing methods of LMBP, LSTM, and MIMLDL.

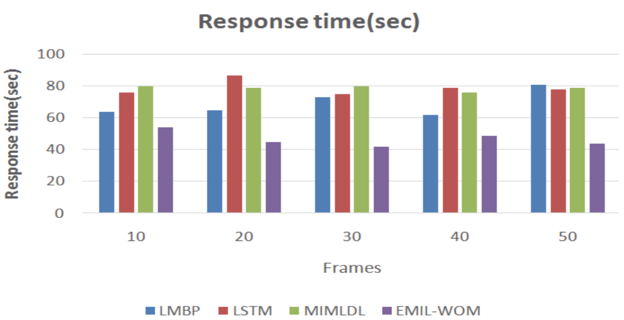

Figure 8 Comparison of response time

Table 5 Overall comparative analysis

Parameters	LMBP	LSTM	MIMLDL	EMIL-WOM
Accuracy (%	78.5	74.5	85	96.5
Computation complexity / %	87.4	89	83	56.1
RMSE / %	56.4	45	57.3	32.4
RAE / %	47.4	48	49.1	24.1
MAE / %	61.7	65	67	30
Response time / sec	75.6	78	73.3	45.3

5 CONCLUSION

In the field of human motion recognition, extracting features from videos is a crucial component, unlike recognition approaches. The non-rigid structure of the human body and the presence of obstacles during motion make video-based human movement capture particularly difficult. Conventional distributed cloud computing video surveillance technology has high system latency and transmits a large amount of network data, which could compromise consumer privacy. To address these issues, this paper suggests using an Enhanced Multiple Instance Learning with Whale Optimization Method (EMIL-WOM) technique to distribute video analysis on an edge development board. This model is designed to gather information from the video surveillance system at the computing edge. By developing the neural network model on the enterprise's servers, the need for data transmission is eliminated, thus protecting user privacy and significantly reducing the strain on cloud computing and storage servers. Future studies will focus on network security methods for effective data transmission. Future studies should focus on adaptability to different conditions (e.g., lighting, weather), robustness in crowded areas, and efficiency in larger networks. Enhancements in integration with existing surveillance infrastructure are also crucial. Additionally, exploring advanced network security methods is essential to protect data during transmission and storage.

6 REFERENCES

[1] Fong, S. S., Ng, S. S., & Chung, L. M. (2013). Health through martial arts training: Physical fitness and reaction time in adolescent Taekwondo practitioners. *Health, 5*(6A3), 1-6. https://doi.org/10.4236/health.2013.56A3001

[2] Chan, J. C., Leung, H., Tang, J., K., & Komura,T. (2011). A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies, 4*(2), 187-195. https://doi.org/10.1109/TLT.2010.27

[3] Witte, K., Emmermacher, P., Bandow, N., & Masik, S. (2012). Usage of virtual reality technology to study reactions in karate-kumite. *International Journal of Sports Science and Engineering, 6*, 17-24.

[4] Gorini, A. & Riva, G. (2014). Virtual reality in anxiety disorders: the past and the future. *Expert Review of Neurotherapeutics, 8*(2), 215-33. https://doi.org/10.1586/14737175.8.2.215

[5] Tao, G., Archambault, P. S., & Levin, M. (2013). Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis.*Virtual Rehabilitation (ICVR), 2013 International Conference.* https://doi.org/10.1109/ICVR.2013.6662084

[6] Shikalepo, E. E. & Hautemo, A. M. (2021). Open Education and Self-directed Learning in Adult, Professional and Vocational Education in Africa. *Radical Solutions for*

Education in Africa, Springer, Singapore, 105-128. https://doi.org/10.1007/978-981-16-4099-5_6
[7] Bayram, L. (2012). Use of online video cases in teacher training. *Procedia-Social and Behavioral Sciences, 47*, 1007-1011. https://doi.org/10.1016/j.sbspro.2012.06.770
[8] Lin, T. C., Jun-Zhe, Y., & Yu, C. S. (2019). Development of a transmission network fault location platform based on cloud computing and synchrophasors. *IEEE Transactions on Power Delivery, 14*(29), 1-6. https://doi.org/10.1109/TPWRD.2019.2929290
[9] Abdulhameed, A. (2018). Evaluating distributed IoT databases for edge/cloud platforms using the analytic hierarchy process. *Journal of Parallel and Distributed Computing, 124*, 41-46. https://doi.org/10.1016/j.jpdc.2018.10.008
[10] Pooja, D. & Arun, S. (2020). A new software development paradigm for intelligent information systems. *International Journal of Intelligent Information and Database Systems, 13*(2-4), 356-375. https://doi.org/10.1504/IJIIDS.2020.10031608
[11] Mabrouk, A. B. & Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications, 91*, 480-491. https://doi.org/10.1016/j.eswa.2017.09.029
[12] He, J., Zhang, C., He, X., & Dong, R. (2020). Visual recognition of traffic police gestures with convolutional pose machine and handcrafted features. *Neurocomputing, 390*, 248-259. https://doi.org/10.1016/j.neucom.2019.07.103
[13] Rezaei, F. & Yazdi, M. (2021). Real-time crowd behavior recognition in surveillance videos based on deep learning methods. *Journal of Real-Time Image Processing, 18*(5), 1669-1679. https://doi.org/10.1007/s11554-021-01116-9
[14] Xie, X. & Tan, F. (2022). Realization of Intelligent Scoring System of Taekwondo Protective Gear under the Application of Neural Network BP Model. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2022/5902983
[15] Zhang, L. & Li, M. (2022). Taekwondo Trajectory Tracking Based on Multitarget Detection Algorithm. *Mobile Information Systems*. https://doi.org/10.1155/2023/9872423
[16] Nurahmadan, I. F. & Pradnyana, I. W. W. (2021). Utilization of Pose Estimation and Multilayer Perceptron Methods in the Development of Taekwondo Martial Arts Independent Learning. *2021 International Conference on Informatics, Multimedia, Cyber and Information System,* 267-272. https://doi.org/10.1109/ICIMCIS53775.2021.9699268
[17] Hailong, L. (2021). Role of artificial intelligence algorithm for taekwondo teaching effect evaluation model. *Journal of Intelligent & Fuzzy Systems, 40*(2), 3239-3250. https://doi.org/10.3233/JIFS-189364
[18] Zhong, J. & Xu, J. (2022). Video Tactical Intelligence Analysis Method of Karate Competition Based on Convolutional Neural Network. *Discrete Dynamics in Nature and Society*. https://doi.org/10.1155/2022/6204173
[19] Ji, W. & Wang, R. (2021). A multi-instance multi-label dual learning approach for video captioning. *ACM Transactions on Multimidia Computing Communications and Applications, 17*(2s), 1-18. https://doi.org/10.1145/3446792
[20] Sapkota, H., Ying, Y., Chen, F., & Yu, Q. (2021). Distributionally robust optimization for deep kernel multiple instance learning. *International Conference on Artificial Intelligence and Statistics*, 2188-2196.
[21] Lee, J. & Jung, H. (2020). Tuhad: Taekwondo unit technique human action dataset with key frame-based cnn action recognition. *Sensors, 20*(17), 4871. https://doi.org/10.3390/s20174871

Contact information:

Nithya S., Assistant Professor
(Corresponding author)
Department of Information Technology,
PSNA College of Engineering and Technology (Autonomous),
Dindigul-624622
E-mail: snithya@psnacet.edu.in

Samaya PILLAI IYENGAR, Assistant Professor
Symbiosis Institute of Digital and Telecom Management,
Symbiosis International (Deemed University),
Pune, Maharashtra, India
E-mail: samaya.pillai@sidtm.edu.in

Poobalan A., Assistant Professor
Department of Computer Science and Engineering,
University College of Engineering,
Dindigul, India
E-mail: apoobal@gmail.com

Parameswari A., Assistant Professor
Department of ECE,
Adithya Institute of Technology,
Coimbatore
E-mail: parameswari_a@adithyatech.com