Reliability of Low Framerate in Deep Learning Based Visual Human Gait Analysis for Identification

Salisu Ibrahim Yusuf *
salisu.yusuf@nileuniversity.edu.ng

Steve Adeshina*+
steve.adeshina@nileuniversity.edu.ng

Moussa Mahamat Boukar*
musa.muhammed@nileuniversity.edu.ng

*Department of Computer Science,
+ Department of Electrical and Electronics Engineering
Nile University of Nigeria

Summary

Extracting images from video dataset determines input and computation demand in training a machine learning model, using less framerate cut down on cost, however, higher framerate leads a better video quality hence better result is expected. In this paper, we investigated the influence of framerate on accuracy in identification of human through gait using the convolutional recurrent neural network model. Primary dataset was collected from at 60fps for 26 persons each having 12 to 16 instances. Frames are extracted at varying rates 60, 30, and 15 fps. We measured and comparatively analyze the accuracy and loss in train and validation set using each instance of datasets on the same model. The outcome shows similar results for each framerate and slightly better validation accuracy was recorded for 15fps

Keywords: *computer vision, human identification, framerate, deep learning, video dataset preprocessing, CNN-LSTM, gait analysis.*

1. Introduction:

Gait is complex system that shows the behavior human and animals in motion while sustaining their center of gravity in different internal and external conditions [1]. The study of gait has gained the attention of researchers for millennia [2]; information from variations in gait due some factors is relevant to different fields such as medicine [3] [4], robotics [5] [6], psychology, security [7], sports; therapy [8], etc. are used for serval purposes. Earlier research were based on measurable attributes of the gait, with recent advancements in computer performance, adoption of machine learning on visual data opens up wider range of solutions to problems that computer visions will be applied to [9].

To effectively discuss the factors influencing decision on features to be extracted for gait analysis, it is important to discuss the walking mechanism of human.

A gait cycle comprises sequence of two steps left leg proceeding left or vice versa, in a survey by Alharthi et al [10], they categorize the gait into sequence of event based on leg's activities as shown in figure 1, in their survey gait is divided into two phases: Stance and swing phases. The latter is when the feet is in contact with ground comprising 60% of the cycle time while earlier is when the feet is off the ground.

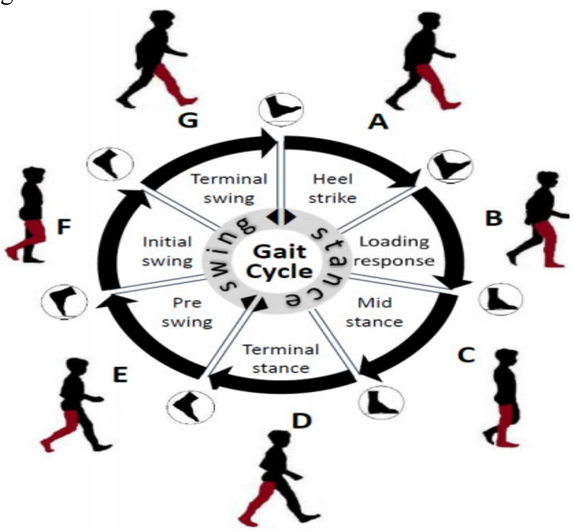

Figure 1: Gait Cycle [10]

The resulting effect of this actions can be observed at the upper part of the body. At A, the head is at its lowest, and the body weigh balanced between the two feet, at B, the head is rising and reaches its pick at C shifting he weight of the body to one leg, at D, it is at the same level as A, from D to G cycle repeat for the second step, it takes two step to make a complete gait cycle, through with the head makes two cycle, cycle may be distinct resulting from difference in condition of left and right sides of the body.

In a situation where person limps for one side, the step cycle for the affected side becomes shorter, therefore the head cycle becomes relatives shorter. At varying weight of load, the flight time on legs is reduced to maintain balance of the body.

Manuscript received January 5, 2022
Manuscript revised January 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.1.54

Advancements in sensors and actuators such as gravity sensors from mobile and wearable devices [11], floor sensors, motion detectors, visual sensors (cameras) sound and vibration sensors, accelerometer [12] etc. allows different type of data from different perspective, hence variety of data leading to different approaches to gait analysis.

Studies have shown that gait is a source of biometric data gotten from an individual from the farthest distance [13]; however, a person's gait is influenced by several factors that are grouped into external and internal factors. Internal factors deals with the person's biology like health status, age [14], emotional status: [15] among others gait [16] [17] while the external component are factors on or around the targeted persons environment from clothing, foot wear [18], sloppiness of the flow [19] [20], load etc., these variations are studied by researchers for various classification problems.

Graphical data processing has high computation demand compare to text data, researchers are proposing various way of optimizing computer resources through parallelism, algorithm optimization and preprocessing of the input data to reduce size of input. Some convert images to grey scale, resize the image or crop from the images

Video file is made of sequence of images compressed with audio. Computer vision method deal with analysis of individual image as part of the video to extract features and patterns of relationships between the identified features of interest among corresponding images or frames. Video quality is influenced by image size (images resolution) and number of frames per second (framerate). To analyze a video in computer visions, images are extracted from the video. [18]

The size of input is directly proportional to computational resources and processing time required for training a machine model, hence, minimizing the computation cost.

Most researchers on gait analysis use open dataset with are often extracted images with undefined framerate.

In rest of the paper we discussed gait analysis in section II, section III Existing work highlighting related works from literatures, the description of CRNN architecture was discussed in section IV, section V explained how the experiment was setup and dataset description, finally discussion and analysis of result in section VI, conclusion and future work was covered in section VII.

2. Gait Analysis

2.1 Methods of Gait Analysis.

Several methods were proposed for gait analysis but are generally classified into two major categories, the Model Based and Model Free Gait Analysis [10].

1) *Model-based*: this method involves the using the mathematical model to formulate an abstraction of how a person walks showing how all the parts of the person are synchronized while walking or performing an action.
2) *Model free:* however analysis is based on studying a sub structure or selected features of the gait without visualizing the whole.

Both method are widely used in several areas as mentioned earlier.

2.2 Machine Learning Vs Algorithmic Approach.

Algorithmic solutions to problem will requires a complete definition of features to be extracted how they are to be measured, stored and computed. This approach can be used on model or model free based analysis.

The machine learning approach falls under model free approach, being a black-box approach details of features selected and representation are defined by the model.

Various deep learning methods have been employed to analyze gait videos for different purposes. Computer visions applies machine learning principles on visual data (image) or video (sequence of images) in solving related problems. These processes mostly include analysis of frames of images in a video, hence, the number of images extracted is directly proportional to computational cost of training the model.

3. Existing Work

Several research in areas of applied gait analysis have been carried out to investigate the possibility of using low framerate videos dataset to achieve a goal, in clinical application of gait, optimal low framerate possible to extract features for medical diagnosis and analysis.

In a review by [22], they concluded that one of the technical source of variability in performance framerate, [23] assessed the reliability of gait assessment for teleassessment for rehabilitation at different framerates and bandwidth for elderly based on performance oriented mobility assessment gait POMA-G by individuals (raters).

In the area of identification, studies have explored the possibility of using low framerate to analyze gait sequence on algorithmic based analysis to reconstruct gait sequence or select frames from sequence that match specific phases in gait sequence in fig 1,

[24] Akae et al proposed an improved low framerate gait analysis where frames are selected from sequence to match a specific phase of gait and manifold problem is experienced where frame matches the previous frame hence appearing still.

Closest study to this was by Imoto et al [25], they analyzed their novel algorithm for gait identification based on dynamic features at different framerate, however, this research focuses on deep learning model and reliability of their analysis was not ascertained in the report.

To our knowledge, researchers have recognized framerate as an influencing factor to computational and storage demand but an investigation have not been carried out to ascertain the tradeoff is using low framerate in deep learning approach

4. Model Description: Convolutional Recurrent Neural Network

The Convolutional Neural Network CNN is a deep learning model with celebrated success in autonomous feature extraction from data like images. Features extracted from CNN is used for effective classification or prediction of problems, it comprises of two steps convolutions and pooling as shown in *figure 2.* A CNN consists of one or more convolutional layers, each layer made of multiple filters. The architecture of CNN captures different features such as edges, shapes, and texture by leveraging the 2-dimensional spatial structure of an image using these filters.

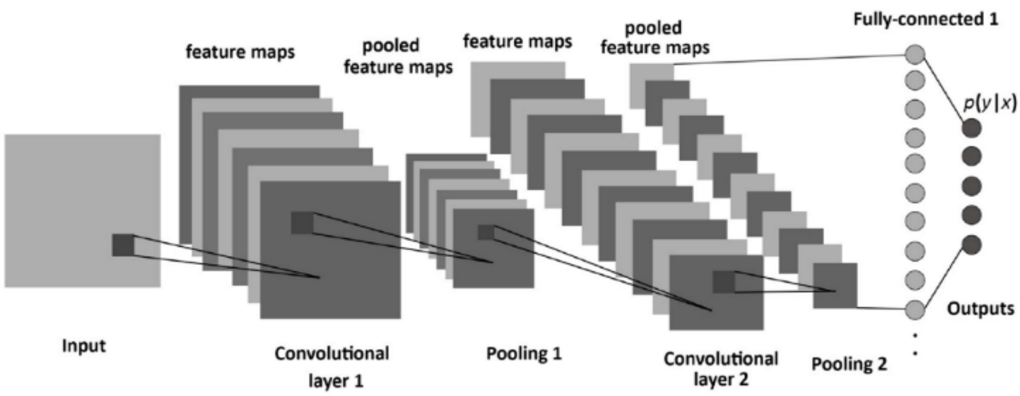

Figure 2: CONVOLUTION RECURRENT NUERAL NETWORK

Recurrent Neural Network RNN has a competitive victory analyzing sequential datasets such as words in sentences, sequence of images from video. Natural Language Processing NLP problems like translation, DNA sequence analysis [26]and video analysis problems like action recognition.
Traditional recurrent neural networks (RNNs, Figure 3.4) model temporal dynamics by mapping input sequences to hidden states, and hidden states to outputs via the following recurrence equations 1 and 2 where g is an element-wise non-linearity, such as a sigmoid or hyperbolic tangent, x_t is the input, $h_t \in \mathbb{R}^N$ is the hidden state with N hidden units, and y_t is the output at time t. For a length T input sequence $<x_1,x_2,...,x_T>$, the updates above are computed sequentially as h_1 (letting $h_0 = 0$), $y_1, h_2, y_2, ..., h_T, y_T$.

$$h_t = g(W_{xh}x_t + W_{hh}h_{t-1} + b_h \quad (1)$$

$$y_t = g(W_{hz}h_t + b_{z)} \quad (2)$$

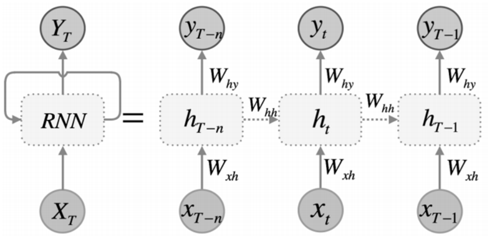

Figure 3 Traditional Recurrent Neural Network RNN

Though RNNs have proven successful on tasks such as speech recognition and text generation, it can be difficult to train them to learn long-term dynamics, likely due in part to the vanishing and exploding gradients problem that can result from propagating the gradients down through the many layers of the recurrent network, each corresponding to a particular time step. LSTMs provide a solution by incorporating memory units that explicitly allow the network to learn when to "forget" previous hidden states and when to update hidden states given new information. As research on LSTMs has progressed, hidden units with varying connections within the memory unit have been proposed. We use the LSTM unit as described in (Figure 3 – 5),

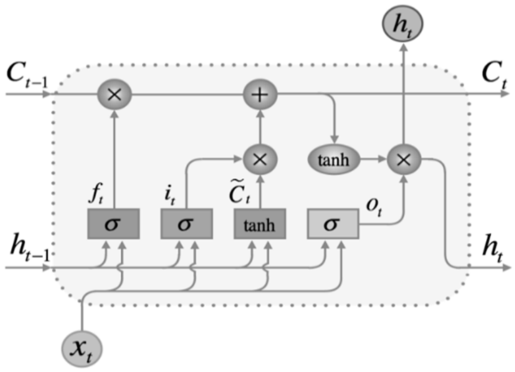

Figure 4: LSTM

A slight simplification of the one described. Letting $\sigma(x) = (1 + e^{-x})^{-1}$ be the sigmoid non-linearity which squashes real-valued inputs to a [0,1] range, and letting $\tanh(x) = \frac{e^x - e^{-e}}{e^x + e^{-x}} = 2\sigma(2x) - 1$ be the hyperbolic tangent nonlinearity, similarly squashing its inputs to a $[-1,1]$ range, the LSTM updates for time step t given inputs x_t, h_{t-1}, and c_{t-1} are:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f \quad (4)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o \quad (5)$$

$$g_t = tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c \quad (6)$$

$$c_t = f_t . c_{t-1} + i_t . g_t \quad (7)$$

$$h_t = o_t . \tanh(c_t) \quad (8)$$

In addition to a hidden unit $h_t \in \mathbb{R}^N$, the LSTM includes an input gate $i_t \in \mathbb{R}^N$, forget gate $f_t \in \mathbb{R}^N$, output gate $o_t \in \mathbb{R}^N$, input modulation gate $g_t \in \mathbb{R}^N$, and memory cell $c_t \in \mathbb{R}^N$. The memory cell unit c_t is a sum of two terms: the previous memory cell unit c_{t-1} which is modulated by f_t, and g_t, a function of the current input and previous hidden state, modulated by the input gate i_t. Because it and f_t are sigmoidal, their values lie within the range [0,1], and i i_t and f_t can be thought of as knobs that the LSTM learns to selectively forget its previous memory or consider its current input. Likewise, the output gate o_t learns how much of the memory cell to transfer to the hidden state. These additional cells seem to enable the LSTM to learn complex and long-term temporal dynamics for a wide variety of sequence learning and prediction tasks. Additional depth can be added to LSTMs by stacking them on top of each other, using the hidden state $h(-1)_t$ of the LSTM in layer `−1 as the input to the LSTM in layer.

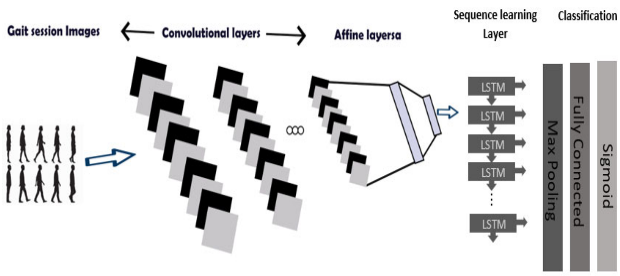

Figure 5: C-RNN (LSTM)

Loading images on the RNN will be very expensive as storage demanding even with LSTM, however, a concatenation of the two architecture to form the Convolutional Recurrent Neural Network (CRNN) will be more efficient, with the convolution as encoders loading extracted feature onto the RNN as decoders to learn patterns in in the sequence of extracted features.

5. Experiments

5.1 Dataset

To investigate the influence of framerate on reliability of accuracy result a high resolution data was captured at 60fps at distance of 5 meters from 26 subject between the ages of 45 and 20 with majority in their early to mid-20s. 14 of the subject were male and 12 females. Each subject was has 12 to 16 instance carrying any load of choice walking in different directions totaling 315 clips at varying lengths.

Subjects were briefed on the research and willingly volunteered with all ethics observed.

3 sets of dataset were prepared with varying framerate, 60fps, 30fps and 15fps, implies reducing the size of input by 50% sequentially, hence reducing the computational and storage demand accordingly.

5.2 Evaluation

Cross validation in deep learning models gives to true behavior of the model, for classification problem such as the one discussed in this paper, 70% of the data is used for training and 30% for validation, the model was not exposed to the data, it is to use knowledge learned from the training to classify to test the true accuracy of the model. The greater the difference the less reliable, also it shows and control overfitting.

The three instances of dataset prepared for this study were used to identical models on identical computer with similar hyper parameter, the accuracy and loss for train and validation where measured compared.

6. Results and Discussion

The result of the experiment is found on figures 8 – 13, with 8 and 9 from 15 fps, 30 fps in figures 10 and 11 , figures 12 and 13 for 60 frames showing Accuracy as loss. Trainnig time was ralative to the ratio of input sizes as observed but not timed.

All three experiment show similar accuracy after 100 epochs with 15fps doing slighly better at 82.08% with 69.77 validation at 98th epoch, at the 100th epoch all three samples had 81.1321% test accuracy with validation of 72.093, 70.3704 and 60.4938, for 15fps, 30fps and 60fps, we compute the absolute differnce between the traain accuracy and validations accuracy as shown in figure 12, showing a very identical pattern in the learning with 15fps having more lower differnce as the minimum points.

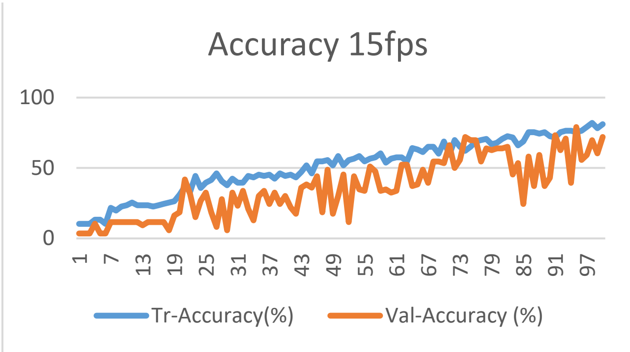

Figure 6: Accuracy 15fps

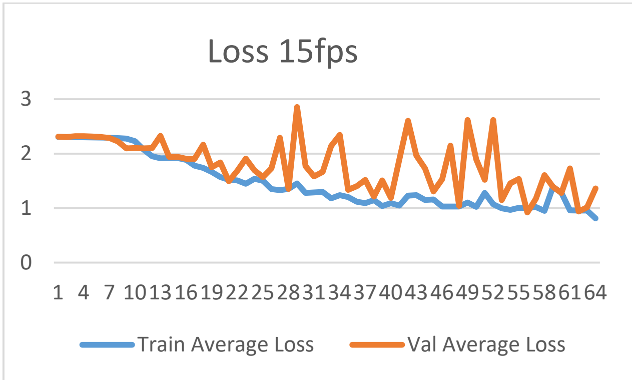

Figure 7: Loss 15fps

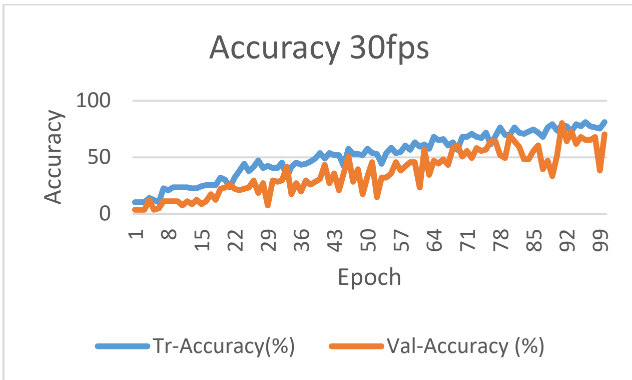

Figure 8: Accuracy 30fps

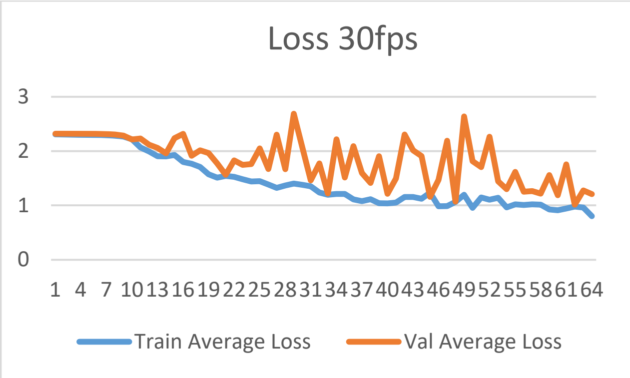

Figure 9: Loss 30fps

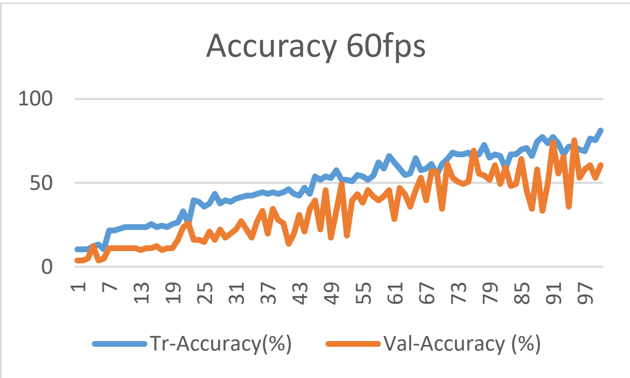

Figure 10: Accuracy 60fps

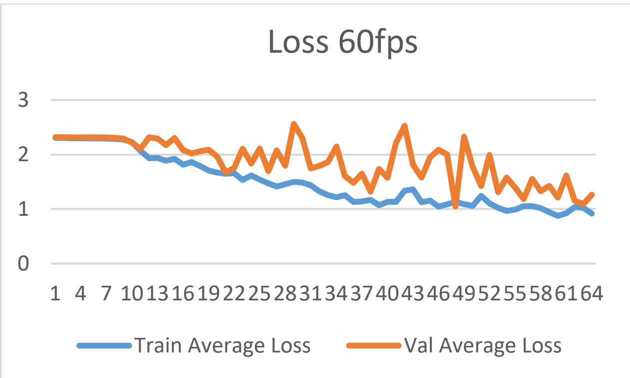

Figure 11: Loss 60fps

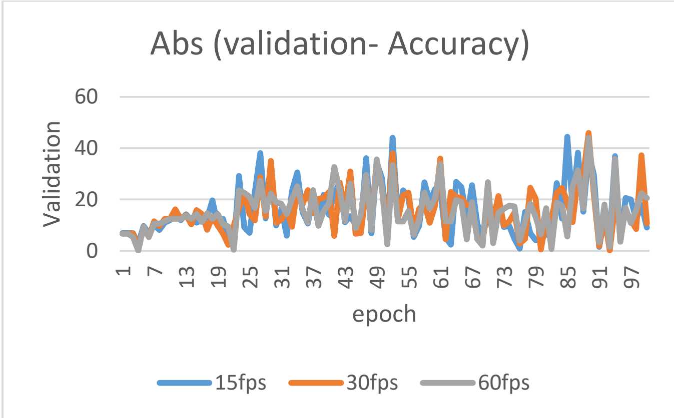

Figure 12: Absolute difference of test Accuracy and Val Accuracy

7. Conclusion

In conclusion, in this research we set out to investigate the tradeoff in terms of reliability of result in reducing framerates from 60fps to 30fps then 15fps in human gait identification, we collected primary dataset of 315 clips from 26 subjects at 60fps, images sequences extracted at 60pfs, 30fps and 15fps. We implemented a convolutional recurrent neural network deep learning model, 3 prepared dataset were used the train model with 30% of dataset used for valition. The outcome shows there is no tradeoff with respect to accuracy, however improvement in the reliability of result was observed to be increasingly better from 60fps to 30fps and 15fps.

8. Future Work

In future work, we will consider investigation optimal framerate in gait classification problems like of action recognition, clinical diagonosis among and more to generalize optimal framerate for all gait problems, lower framerates than 15 will also be analysed. Evaluation metrics like f1 score, precision and recall will be used to further measure effect of framerates on gait analysis.

Acknowledgement

We will like to acknowledge the efforts of our esteem colleague Dr. Muhammad Suleiman, Mr. Yusuf Abass Aleshinloye, and all the staff of computer science department, Nile University of Nigeria for their contributions through relentless criticism and encouragement to complete this research.

References

[1] S. W. and G.-V. D., "Walking dynamics from mechanism models to parameter optimization,," in *2011 Symposium on human body dynamics*, 2011.

[2] C. Patrick and R. Arun, "Biometric recognition by gait: A survey of modalities and features," *Computer vision and Image Understanding,* vol. 167, pp. 1-27, 2018.

[3] M.-d.-l.-H. A, G. Z. B. and M. Z. A., "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications,," *Sensors, Basel, Switzerland,* vol. 14, no. 2, pp. 3362-3394, 2014.

[4] E. Abdulhay and K. N. a. E. V. N. Arunkumar, "“Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease," *Future Gener. Comput. Syst,* vol. 83, pp. 366-373, 2018.

[5] V. K. Geetanjali and H. P. Varsha, "A Study of Vision based Human Motion Recognition and Analysis," *International Journal of Ambient Computing and Intelligence,* vol. VII, no. 2, pp. 75-92, December 2016.

[6] H. G.C. and P. A.S., "Human Motion Recognition in Real Time surveillance System: A review," *Journal of Applied Science,* pp. 2793-2798, 2010.

[7] M. M. Ali and S. C. Muayad, "Human Gait identification System," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* vol. 6, no. 10, pp. 1574-1594, 2017.

[8] J. P. Feng, J. R. B. Wick, E. P. D. P. Bompiani and M. M. Aiona, "Applications of gait analysis in pediatric orthopaedics," *Current Orthopaedic Practice,* vol. 27, no. 4, pp. 455-464, 2016.

[9] A. Pannu, "Artificial Intelligence and its Application in Different Areas," *International Journal of Engineering and Innovative Technology (IJEIT),* vol. 4, no. 10, pp. 79-84, 2015.

[10] A. Alharthi, S. YUNAS and K. Ozanyan, "'Deep Learning for Monitoring of Human Gait: A ReviewResearch output: Contribution to journa," *IEEE Sensors Journal,* vol. 19, no. 21, pp. 9575-9291, 2019.

[11] M. T. Hoang, Q. V. Vo, D. T. Nguyen and D. C. , "Gait Identification Using Accelerometer on Mobile Phone," in *2012 International Conference on Control, Automation and Information Sciences (ICCAIS)IEEE*, IEEE, 2012.

[12] O. D. Mohammad, "Accelerometer-Based Gait Analysis, A survey," in *NISK-2010 conference.*, 2010.

[13] S. o. s. q. a. g. recognition, "Zongyi Liu;Malave, L; Sarkar, S.," pp. 704-711, 2004.

[14] J. Woo, S. Ho, J. Lau, S. Chan and Y. Yuen, "Age-Associated Gait Changes in the Elderly: Pathological or Physiological?," *Neuro epidemiology,* vol. 14, pp. 65-71, 1995.

[15] T. B. a. D. B. I. Birch, "The identification of emotions from gait," *Sci. & Jus,* vol. 56, no. 5, pp. 351-356, 2016.

[16] I. Wakako, I. Tome, T. Tadao, S. Ikuya, B. M. Katarzyna, T. Kawaguchi, Y. Tabara, T. Nakayama, M. Fumihiko and I. Noriaki, "Are there different factors affecting walking speed and gait cycle variability between men and women in community-dwelling older adults?," *Aging Clinical and Experimental Research,* vol. 29, no. 2, pp. 215-221, 2017.

[17] B. D. De and M. Iosa, "The Walking Brain: Factors Influencing Human Gait," *EC PSYCHOLOGY AND PSYCHIATRY,* vol. 7, no. 12, pp. 960-963, 2018.

[18] A. H. Md, M. Yasushi, W. Junqiu and Y. Yasushi, "Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control," *Pattern Recognition,* vol. 43, no. 6, pp. 2281-2291, 2009.

[19] C. Wen-Ruey, C. Chien-Chi, F. Mary and M. Simon, "Gait adaptation on surfaces with different degrees of

slipperiness," *Applied Ergonomics,* vol. 59, no. a, pp. 333-341, 2017.

[20] S. Jeffrey and A. Panagiotis, "On the effect of walking surface stiffness on inter-limb coordination in human walking: toward bilaterally informed robotic gait rehabilitation," *Journal of NeuroEngineering and Rehabilitaition,* vol. 13, no. 32, 2016.

[21] A. Jerry, S. Chen, W. Aosen, L. Jeanne, L. Zhinan and X. Wenyao, "A Pervasive and Sensor-Free Deep Learning System for Parkinsonian Gait Analysis," in *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* , Nevada, 2018.

[22] M. Alexandria, E. Arezoo and A. Jan, "Two-dimensional video gait analysis: A systematic review of reliability, validity, and best practice considerations," *INTERNATIONAL SOCIETY FOR PROSTHETICS AND ORTHOTICS,* pp. 1-18, 2020.

[23] V. Kavita, A. Kristopher, L. Lawrence R., C. Kevin, C. K. Gerald and H. Helen, "Teleassessment of Gait and Gait Aids: Validity and Interrater Reliability," *American Physical Therapy Association 2020,* vol. 100, no. 4, pp. 708-717, 2020.

[24] A. Naoki, M. Yasushi and Y. Yasushi, "Gait recognition using periodic temporal super resolution for low frame-rate videos 2011," in *International Joint Conference on Biometrics (IJCB)*, 2011.

[25] I. Daisuke, K. Kenji, T. Ken'ichi, K. Kenyo and M. Hirabayashi, "A Novel Human Identification Method by Gait using Dynamics of Feature Points and Local Shape Features," in *7th European Workshop on Visual Information Processing (EUVIP)* , Tampere, 2018.

[26] Y. A. Abass and S. A. Adeshina, "Deep Learning Methodologies for Genomic Data Prediction," *Jourbal of Artifitial Intelligence for Medical Sciences,* 2021.

[27] D. Fani, G. Yao and Y. Guang-Zhong, "From Emotions to Mood Disorders: A Survey on Gait Analysis Methodology," *IEEE Journal of Biomedical and Health Informatics,* vol. 23, no. 6, pp. 2302-2316, 2019.