# DEEP LEARNING - BASED METHOD FOR RECOGNIZING GESTURES

**Mrs.T.Sarika,Assistant Professor, Department of Information Technology, CMR Engineering College, Hyderabad, Telangana,**
**E-Mail-id sarika.tuniki@cmrec.ac.in**

**S.Hanuman ,178R1A1220@cmrec.ac.in,**
**V.Sriram ,178R1A1253@cmrec.ac.in,**
**B.Bhavana,178R1A1202@cmrec.ac.in**

## ABSTRACT

Human-computer interaction is becoming increasingly prevalent around us as a result of the swift advancement of science and technology. A new branch of study called human motion analysis and recognition based on attitude sensors has significant advantages and practical improvements over motion recognition based on video. In this study, we provide a brand-new approach based on temporal gesture recognition. The characteristics of gestures are retrieved and categorised using recurrent neural networks and their variation networks by examining the kinematics of gestures. Over 98% accuracy was attained using the procedures across 16 experimenters. The outcomes demonstrate the algorithm's speedy and precise ability to recognise motions.

**Key words** : gesture recognition, recurrent neural networks, and attitude sensor

## I. INTRODUCTION

There are numerous different types of human movement expressions, but the most prevalent are gestures [1] [2] [3] [4]. Pattern recognition research is beginning to focus on gesture recognition based on attitude sensors. First, the signals produced by human motion are captured by the attitude sensor, which then sends the data to the mobile device. Preprocessing is next followed by feature extraction and selection. Based on retrieved attributes, we categorise and identify human movements. Human motion information may be analysed using two different methods: sensor-based human pattern recognition and vision-based human motion pattern recognition [5][6].

Although the first technology was developed earlier and the theory is comparatively developed, the visual based human motion pattern recognition method has the issue of being overly dependent on the external environment, necessitating the collecting of motion data adequate background lighting conditions. Given the clear advantages of attitude sensor-based motion pattern recognition over vision-based human motion pattern recognition, which are unaffected by environment or light, and the many research advances made to date, gesture recognition based on attitude sensor has captured the attention of academics [7].

These conventional classification techniques are still popular among researchers working in the field of sensor-based motion identification, and their efficiency has been demonstrated in real-world use. In order to improve sensor-based motion recognition, Daniel Olgu [8] and colleagues devised the HMM algorithm in 2006.Three accelerometers were mounted to their right hand. Letting the experimenter kneel, walk, stand, crawl, and lie down is permitted,92.13% of people are recognised. Piero Zappi [9] et al. used a Bayesian classifier in 2007 with 19 accelerometers worn on both arms to identify car repairs with an identification rate of up to 98%. In 2008, Jhun-Ying Yang [10] and coworkers picked seven test subjects to examine seven different types of everyday activities: standing, sitting, walking, jogging, vacuuming, washing clothes, and brushing. In order to identify the closest neighbour algorithm, they employed an artificial neural network classifier, which achieved recognition rates of 95.24% and 87.17%. A similar year, Zhenyu He [11] et al.67 people held a handset while 17 different motions were examined using a built-in three-axis accelerometer in a prepared handset. The support vector machine classifier was utilised to get an 87.36% detection rate. In 2010, Yu-Jin Hong [12] et al. Utilizing three triaxial accelerometers fixed to the thighs, lumbar, and forearms of 15 people, a decision tree classifier was used to identify 18 everyday activities, with a 92.58% identification rate for standing, sitting, walking, running, and waving. A neural network-based hidden Markov model identification approach is proposed by Zhu Chun and Sheng Weihua [13].

Historically, the motion was frequently determined by the physical properties of the angle and acceleration signals or by the geometric properties of the acceleration signals (period, peak, trough) [14]. The recognition result wasn't good enough. Therefore, we suggested a deep learning technique to identify the behavioural categorization. As a framework for biological neural network modelling, deep learning provides excellent feature extraction and classification capabilities and has significant research relevance. Regardless of methodology or application direction, it is crucial to recognise the human body motion using sophisticated intelligence algorithm of deep learning and the information of human body movement acquired.

## II. ASSISTANCE WITH GESTURE RECOGNITION

Building a strong classifier is the aim of gesture recognition. The steps involved in the identification process are represented in Fig. 1 as follows: data gathering and recording, feature extraction, training, and categorization of gestures

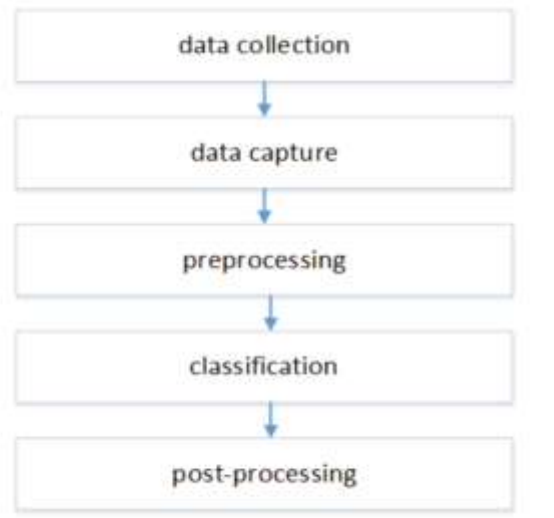


Fig. 1. System flow chart.

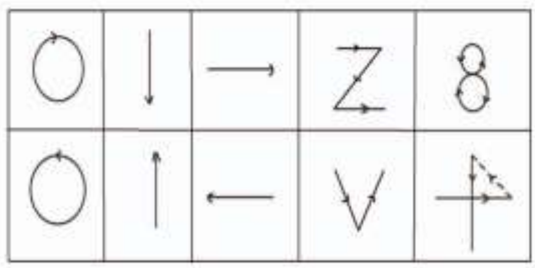

Fig. 2.Gesture List.

## III. ALGORITH FOR RECOGNITION

Since sensor data is a time series, this article chooses as its gesture recognition model the RNN, LSTM, and GRU models that excel in solving the timing problem. a short explanation of these algorithms.

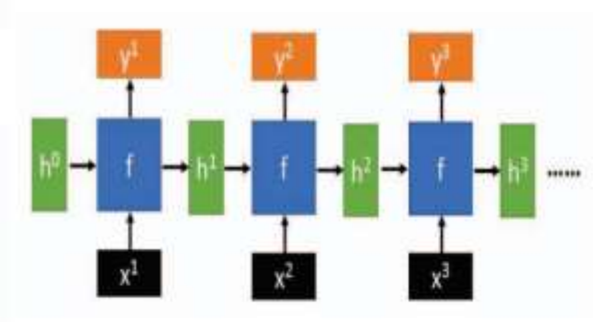


Fig. 3. RNN Network Structure

In Fig. 3, the RNN's network structure is displayed. The RNN's main function is to process sequence data. The layers in the conventional neural network model are fully linked from the input layer to the hidden layer to the output layer, while the nodes between each layer are unconnected. However, many timing issues cannot be solved by this form of neural network. Recurrent neural networks, or RNNs, are defined as systems where the current output of a series is also connected to the preceding output. The specific manifestation is that the network will memorize the previous information and apply it to the calculation of the current output. The input of the hidden layer comprises both the output of the previous hidden layer as well as the output of the input layer since the nodes connecting the hidden layers are linked. Any length of sequence data should work with RNN, theoretically. In reality, however, it is typically believed that the current state is only connected to the past few states in order to decrease complexity. Following is how the formula for this network was arrived at:

$f(w^h h^{t-1} + w^i x^t) = h^t$ (5)

$f(w^o h^t) = y^t$ (6)

The network's memory unit is represented by $x^t$ ,which stands for the input of t=1,2,3,...

$Y^t$ represents step t's result.

Where f often refers to a nonlinear activation function, such as Tanh or ReLU. A unique kind of RNN that can resolve long-term dependencies is the Long Short Term Memory Network (LSTM). An artificial neural network called Long Short-Term Memory (LSTM) is employed in deep learning and artificial intelligence. LSTM features feedback connections as opposed to typical feedforward neural networks.

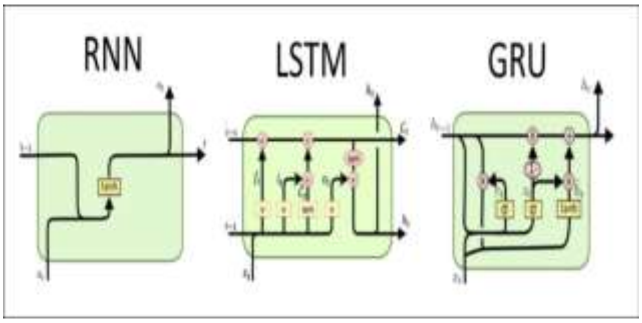


Fig. 4. Network Structures

We also examine the Gated Recurrent Unit, another RNN version, in this post (GRU). GRU keeps the impact of LSTM while simplifying the topology, resulting in Fig. 4. Another well-known network structure is GRU. Fig. 4 depicts the network structure of the LSTM.

These three algorithms are the focus of this essay, and it compares them. To train the preprocessed gesture data individually, use these three techniques. We will receive the three models' accuracy and loss.

## IV. RESULTS

10 gestures were examined in this paper. Figure 2 illustrates the definition of 10 gestures. The arrow's direction denotes the action's direction. 16 experimenters (8

males and 8 females) who met the criteria for the gesture definition were chosen, and each gesture was performed ten times in the manner and intensity of each person's habit in order to assess how well the algorithm could adjust to individual variances. A total of 1,600 sets of data samples were gathered, with 160 sets of data being received for each of our activities.400 data sets were chosen at random as test samples, while 1200 data sets were chosen as training samples. Using two layers and 128 nodes each, we employ a neural network. With this article, we have gained a fundamental understanding of how RNN, LSTM, and GRU units vary from one another. According to how both layers, LSTM and GRU, operate, LSTM is more accurate on a bigger dataset whereas GRU utilises fewer training parameters, uses less memory, and executes more quickly. In this study, three network models are used to train model parameters on training data and then test them on test data. The results of iteration include test accuracy, training loss, and test loss, which are depicted in Figs. 5, 6, and 7. The results of the RNN, LSTM, and GRU models are shown in Table I of this paper as the recognition results for each gesture movement. It can be concluded from the figures and table that the average recognition rates of RNN, LSTM, and GRU are 98%, 99.75%, and 99.75%, respectively. RNN has a lower overall recognition rate. RNN modelling converges gradually. Of these, 100% is the greatest and 95.92% is the lowest RNN recognition rate for each activity. The greatest and lowest recognition rates for each LSTM movement are 100% and 97.30%, respectively. In fact, we discover that the RNN has the quickest convergence rate and the toughest training of all the LSTMs. The following gestures are represented by the numbers in Table I: "V," "Counter clockwise rotation," "Clockwise rotation," "Right," "Up," "Left," "Z," "7," "8," "+," and "09" are all examples of rotations.

TABLE I
RNN, LSTM, GRU MODEL RESULTS

| Method | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RNN | 0.9744 | 1.0000 | 0.9796 | 0.9762 | 1.0000 | 0.9677 | 1.0000 | 0.9796 | 0.9744 | 0.9592 | 0.9800 |
| LSTM | 1.0000 | 0.9730 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9975 |
| GRU | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9756 | 0.9975 |

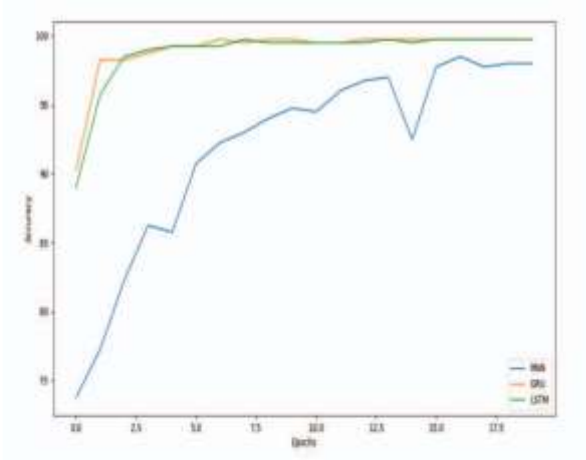

Fig. 5. Accuracy

# V. CONCLUSION

This paper builds and executes a gesture sensor-based gesture recognition method. To gather gesture signals, use the MPU6050 high-precision gyro.First, The motion signal is initially identified, gathered, and normalised. In this study, 10 movements were chosen for experimentation on gesture recognition, and 16 individual data were gathered. The construction of the RNN, LSTM, and GRU models enables gesture recognition.

Comparing the three models' experimental findings. The findings demonstrate that these techniques are efficient for real-time hand gesture recognition, particularly for complicated movements.

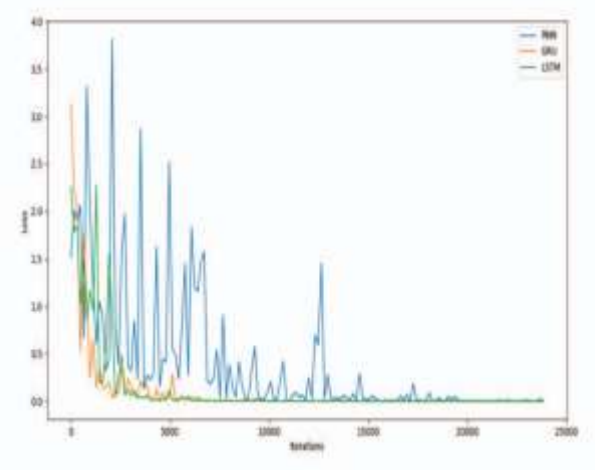

Fig. 6. Training Loss

The technique integrates the gyroscope's output data, does away with device posture restrictions, achieves gesture detection regardless of device posture, and has a high degree of recognition accuracy. End-to-end gesture recognition is still challenging today, though.

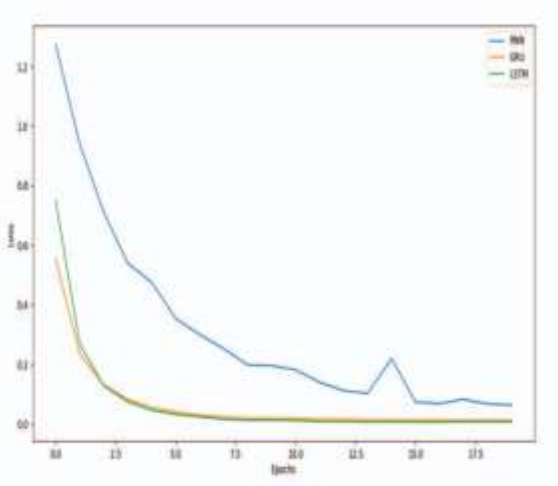

Fig. 7. Test Loss

## VI. REFERENCES

[1] Choi Jiho,Song Kyohyun,Lee Seongil. Enabling a Gesture based Numeric Input on Mobile Phones [C]. IEEE International Conference on Consumer Electronics, Xi-an, China, 2011,pp.151-152.
[2] Zhenyu He, Lianwen Jin, Lixin Zhen, Jiancheng Huang. Gesture recognition based on 3D accelerometer for cell phones interaction. IEEE Asia Pacific Conference, 2008,pp.217-220.
[3] Jhun-Ying Yang, Jeen-Shing Wang, Yen-Ping Chen. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classiers[C].Pattern Recognition Letters, 2008,pp.2213-2220.
[4] Piero Zappi, Thomas Stiefmeier. Activity recognition from on-body sensors by classier fusion: sensor scalability and robustness[J]. IEEE Intelligent Sensors, Sensor Networks and Information,Dec,2007,pp.281-286.
[5] Daniel Olgu, Alex Pentland. Human Activity Recognition: Accuracy across Common Locations for Wearable Sensors[J].IEEE 10th International Symposium on Wearable Computers, Montreaux,Switzerland, October, 2006.
[6] Malik S,Ranjan A,Balakrishnan R. Interacting with Large Displays from a Distance with Vision-Tracked Multi-Finger Gestural Input [C]. Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology Seattle, 2005,pp.43-52.
[7] Cao X, Balakrishnan R. VisionWand: Interaction Techniques for Large Displays Using a Passive Wand Tracked in 3D [C]. Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, Vancouver, 2003,pp.173-182.
[8] Bouten C V, Koek K T M, Verduin M. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity[J].IEEE

Tramaction 011 Biomedical Eng, 1997,pp.136-147.

[9] K. Vaananen and K. Boehm, Gesture Driven Interaction as a Human Factor in Virtual Environments,An Approach with Neural Networks, Virtual Reality Systems[M]. R. Earnshaw, M. Gigante, H. Jones. chapter 7, pp. 93-106. Academic Press,1993.

[10] C. Maggioni. A Novel Gestural Input Device for Virtual Reality. IEEE Virtual Reality Ann[J]. Seattle, Wash, 1993,pp.118-124.