Gesture Recognition for Enhancing Human Computer Interaction

Sangapu Sreenivasa Chakravarthi[1]*, B Narendra Kumar Rao[2], Nagendra Panini Challa[3], R Ranjana[4] & Ankush Rai[5]

[1]Department of CSE, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai 601 103, India

[2]AIML, School of Computing, Mohan Babu University, Tirupati 517 102, India

[3]School of Computer Science and Engineering, VIT-AP University, Amaravati 522 237, India

[4]Department of IT, Sri Sairam Engineering College, Chennai– 600 044, India

[5]AI Division, Faculty of Electronics and Information Technology, Warsaw University of Technology, 00-661 Warsaw, Poland

Gesture recognition is critical in human-computer communication. As observed, a plethora of current technological developments are in the works, including biometric authentication, which we see all the time in our smartphones. Hand gesture focus, a frequent human-computer interface in which we manage our devices by presenting our hands in front of a webcam, can benefit people of different backgrounds. Some of the efforts in human-computer interface include voice assistance and virtual mouse implementation with voice commands, fingertip recognition and hand motion tracking based on an image in a live video. Human Computer Interaction (HCI), particularly vision-based gesture and object recognition, is becoming increasingly important. Hence, we focused to design and develop a system for monitoring fingers using extreme learning-based hand gesture recognition techniques. Extreme learning helps in quickly interpreting the hand gestures with improved accuracy which would be a highly useful in the domains like healthcare, financial transactions and global business

Introduction

The goal of gesture recognition in computer programming and language technology is to define, describe and understand human gestures using computational models. Gesture detection and tracking is a method of image processing application as, it stands as a computer vision standard. Several motion detection algorithms have been studied and presented subsequently, where it is evidently found that hand tracking has a variety of uses, such as motion graphics and capturing technology, human-computer interface, and social and behavioural analysis.[1] Between a device and a person, gestures can perhaps be leveraged as a form of communication.[2–5] It is significantly distinctive from the conventional equipment approaches and enables gesture detection for human-computer interaction. Through the recognition of a gesture or movement of the body or limbs, gesture recognition can reveal the user's stated purpose.

For hand motion detection and tracking, a variety of sensors and detecting gloves are employed. Rather than employing more expensive sensors, basic digital webcams can understand, analyse and track gestures. The major goal of our work is to create a method for hand movement tracking in the real environment with which the cursor actions could be controlled without involving the mouse directly. This is a wonderful use now days because of the pandemic people fear the virus that can transmit by touching the physical objects so to have a solution for such problem we can implement this in public places and can reduce the spread of the bacteria.

Gesture and voice help Controlled Virtual Mouse uses Hand Gestures and Voice Commands to make human-computer interaction simple.[6] Almost no direct contact is required with the computer. All the I/O operations are controlled via static and dynamic hand gestures, as well as voice assistance. The suggested gesture recognition method can be utilized in developing virtual mouse which in turn solve issues such as avoiding human touch on devices. This is extremely helpful during the pandemic situations, where we can reduce the physical devices in public places, which results in less chances of spreading the viruses and the interaction or operations are still being carried out either utilizing a built-in camera or a webcam.

―――――――
*Author for Correspondence
E-mail: ss_chakravarthi@ch.amrita.edu

Bots usually respond to questions using computational linguistics (Natural Language Processing) and systems, voice recognition techniques hold conversations audibly. Voice assistants perform this audibly, but text-based interfaces require robots to process text, evaluate it, and map out a response. Simply to put, by clicking on call-to-action buttons or typing out your query, you should talk to voice assistants aloud. The proposed technology could be developed to handle both keyboard and mouse capabilities, which is a future goal of HCI.

Literature Survey

Urban intelligence is a novel concept that is leading a plethora of infrastructure changes in IOT connected cities.[7] Human Computer Interaction (HCI) being the interface connecting individuals and IoT is critical. It specifically play important role to create the link between digital technologies and their use in smart metropolises. The use of surface electromyogram (sEMG) to identify human hand motions is a current interest of the researchers in sEMG research because, they are found to be a useful effective HCI tool.

A novel technique of gesture recognition was developed using a skeletonization approach and a CNN with the ASL database.[8] Vision-based gesture recognition systems are being extensively investigated in the context of human interaction. However, the impact of recognition is heavily reliant on the recognition algorithm's performance. In tricky situations, the skeletonization strategy and the CNN in the recognition algorithm decrease the influence of recording angle and surroundings on recognition and increase the accuracy of gesture detection. Virtual Mouse using Hand Gestures for Gesture-controlled laptops and PCs have recently attracted a lot of interest.[9] Leap motion is the term for this technique. By waving our hand in front of our computer or laptop, we may manage a variety of features. Digital presentations would be enhanced on using interactive, audio, and video input methods, which promotes the human computer interaction to next level.

Medical pictures of the liver and chest X-rays of several human organs are segmented using the fuzzy based technique.[10] The enhanced approach employs a threshold segmentation algorithm to aid in the automated selection of seed points and improve region growth rules, followed by morphological post-processing to enhance the segmentation result. Gesture tracking and recognition is an image processing. Needless to say, a variety of gesture recognition systems have been presented in recent years. Also, it demonstrated in several other works that the ensemble of extreme learning is found effective in gesture recognition systems and applications.[11] Hand tracking can also be used for motion capture, human-computer interaction, human behaviour analysis, reality education applications and smart health care systems are a few applications just to name.[12,13] A variety of sensors and detecting gloves are used to detect and track hand movements. Rather than employing more expensive sensors, basic web cameras can recognize and track gestures. The major goal is to create a method to track the fingers movement and also to control the cursor actions without physically touching the mouse. However, the disadvantages of the existing system are:

- The word co-occurrence matrix is used to train the detecting glove model, which requires considerable memory to store.
- Reconstruct the co-occurrence matrix, possibly depends on the number of modifications in the hyper-parameters being done in the co-occurrence matrix. However, it is time consuming.
- The sensors for virtual mouse are very expensive.

A brief study on the gesture recognition methods and challenges associated with them[14-24] is listed in Table 1. These methods/proposals have marked the significant milestone in the research of gestures recognition.

Proposed Methodology

For designing and developing such a useful application, the Extreme Learning Machine (ELM) technique is being identified and proposed. The ELM is suggested as a rapid convergence technique dependent on a single hidden layer feed forward neural network (SLFN).[25] It could unilaterally select the hidden layer elements and the features of a SLFN as shown in the Fig. 1.

Only the network's output weights must be modified during the training process using the systematised least squares procedure.[26] This results in accomplishing strong network generalisation performance while learning at an exceptionally fast rate. ELM's expression cost function is usually defined as in Eq. (1).

$$y_j = \sum_{i=1}^{n} \left(\sum_{l=1}^{k} \beta g(d_l * x_i + b_i) \right) \qquad \dots (1)$$

where, n turns to be the number of input layers, *l* turns to be the weight vector linking the hidden and

Table 1 — Key gesture recognition methods and their challenges[14–24]

| Methods	Challenges
Image processing techniques, ANN & CNN	3D gesture disparities were measured in videos.
Microsoft Kinect V2 sensor-based gesture recognition	Restricted to Microsoft Kinect v2 sensor's & its capabilities only.
Gesture recognition is based on four pressure sensors enacted with a wearable device and a computational framework.	Hardware design of the system is complex and requires simplification to achieve higher sensitivity
YOLO: You Look Only Once and other neural networks.	The primary challenge is that many approaches and concepts from multiple viewpoints are applied to get hand movements.
Point LSTM is used while maintaining spatial structures.	Dot clouds provide a precise description of the object surfaces' distances and core structural parameters.
Hand area segmentation using segmented finger image normalization and CNN classification finger identification.	For the identification of gesture patterns, a significant amount of data was required.
Discussed contemporary deep learning techniques for detecting movements and gestures in image sequences.	Described and analysed the existing works only; but focused on temporal components of the data for using gesture recognition.
Recurrent 3D convolutional neural network and spatio-temporal models.	Egocentric vision gesture detection happens to be the challenge as the gadget wearer's spontaneous head movements create considerable camera motion.
Long-term RCNN is used to categorize video sequences of hand motions.	Computing complexity of the system is high with low accuracy.
Radio frequencies were used to detect gestures, using short-range radar detection at a high frequency (60 GHz).	The design is more challenging in terms of input recognition algorithms due to the signal's unique attribute - detect movements at a discrete level.
Spatio-temporal properties-based learning using a 3D convolutional neural network (CNN).	Application specific challenge: Recognizing surgical gestures automatically is the challenging step to gain the surgical expertise.
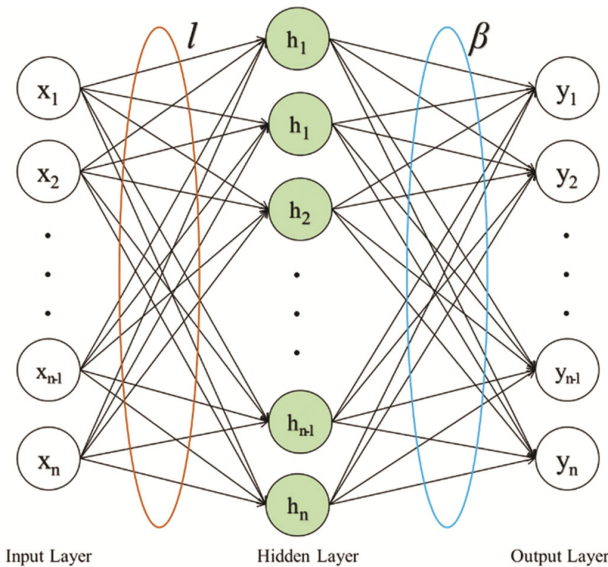
Fig. 1 — Extreme learning layers

output layers, *g(x)* happens to be the activation function, β shall be the weight vector linking the input and hidden layers, x_i turns to be the input data, b_i shall be the bias, and y_j becomes the output data of the SLFN. The below algorithm explains the ELM procedure in three steps:[25]

Step. 1. *To begin with, count the neurons in the hidden layer, initialise the weight vector l_i between the input and hidden layer, and bias b_i unilaterally;*

Step. 2. *Next, compute the hidden layer output matrix H using the identified activation function*

Step. 3. *Finally, use the theory $\beta = H+Y$ to determine the output weight vector.*

Similar to that of how voice assistance set, microphone parameters convert audio to string execute commands (input: String) such as static controls (hello, what is your name, time, search, and location) and Dynamic controls (launch gesture recognition, stop gesture recognition, copy), the proposed work identifies the dynamic controls.[27] In the driver code, lock the main thread until chatbot has taken input from GUI else take input from voice for processing the voice data and finally handle the system exit in main loop. Then, create gesture encoding, extra mapping variables and multi-handedness labels and convert media-pipe landmarks to recognizable Gestures. Further, create a function to find gesture encoding using current *finger_state*. The finger is open if *finger_state*: 1, else 0. Aiming for accuracy, the fluctuations in the finger movement are to be handled cautiously as there would be noise during the input. Now, the predefined commands are executed according to the detected gestures. Use OpenCV to capture videos and locate the hand to get cursor position.[28] Measures should be taken to ensure that the cursor is stabilized by dampening. Such final position for 5 frames is hold to determine the change in status. The flowchart of the gesture recognition process can be seen in Fig. 2.
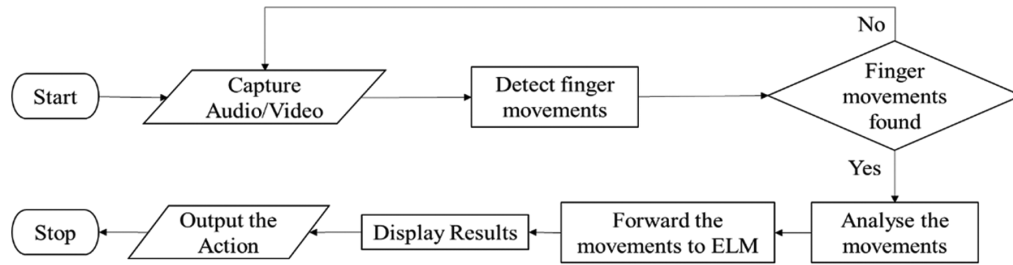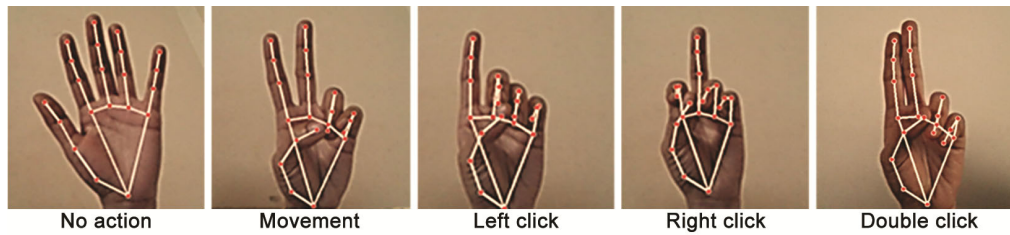
Fig. 2 — Gesture recognition process

Fig. 3 — Hand gestures being recognised as mouse operations

Experimentation

The proposed Hand gesture detection model is implemented to simulate mouse actions virtually. It presents the notion of increasing human-computer interaction using computer vision and hence be called as virtual mouse. Mouse operations are expected to reflect the action in no time as the user cannot wait for a few seconds even. Moreover, the virtual mouses face this challenge in a sceptical manner as the end-user of these virtual systems demand high level human computer interaction. Hence, the extreme learning is followed and implemented to develop the virtual mouse operations and simulated the mouse actions. However, comparing the end-to-end performance of the virtual mouse systems or simulations were however difficult due to the restricted number of datasets available. For tracking and detection, hand motions and fingertip detection have also been evaluated in a variety of complex backgrounds, as well as at extreme distances from the webcam.

Virtual Mouse Operations

Depending on which hand is used and which fingers are up, the virtual mouse executes all click functions (left, right, double), drag, and scrolling. It just requires the in-built webcam of the system and doesn't require additional physical equipment to integrate with the system. As mentioned in earlier sections, the captured hand gestures would be processed and identifies the operation defined on them. The details on the finger gestures and their definitions as the mouse operations are as shown in Fig. 3.

The experiment is carried out with various sets, with each set containing 100 instances in which each mouse action is detected based on the displayed finger gestures. These 100 instances include both male and female gendered hands, as well as hands with differing skin tones. The interpreted results are addressed in the next section.

Results & Discussion

We conclude that the suggested extreme learning based gesture recognition is 99% accurate while AI based virtual mouse operation methods read 98%. The experimental results shown in the Table 2 are yielding with an accuracy of around 99%. However, the table also depicting the granularity for "Right Click" is inadequate since it is the most difficult move for the laptop to comprehend. Because the motion used to conduct the precise mouse function is more difficult, the precision of right click on is low. Furthermore, the precision is both suitable and excessive for all the various movements. In comparison with other existing methods for simulating virtual mouse, this model performed admirably, with 99% accuracy. In the table, the description of the Finger Index TF refers to Thumb, IF is Index finger, MF is Middle finger, RF is Ring finger, and LF is Little finger.

Also, the Table 2 explains that the implemented virtual mouse operations executed is exceptionally well with respect to the accuracy factor. The recommended mannequin is unusual in that it can do most mouse tasks via fingertip detection, including left and right clicks, scrolling up and down, and mouse pointer motions, as well as managing the PC in virtual mode like a hardware mouse. The Fig. 4

Table 2 — Experimental results of the virtual mouse operations, based on the proposed gesture recognition model

| Gesture	Virtual mouse operation	No. of times succeeded	No. of times failed	Accuracy (%)
All the five fingers (TF, IF, MF, RF and LF) are open & pointing to top.	No action	100	0	100
IF or IF & MF are open & pointing to top.	Mouse movement	100	0	100
TF & IF are open with the finger-gap less than or equal to 30mm.	Left button clicks (single click)	99	1	99
TF & IF are open with the finger-gap less than or equal to 40mm.	Left button clicks (double click)	96	4	96
IF & MF are open with the finger-gap less than or equal to 30mm.	Right button clicks	95	5	95
IF & MF are open with the finger-gap more than 40 and moved towards top of the page.	Scroll-up	100	0	100
IF & MF are open with the finger-gap more than 40 and moved towards down the page.	Scroll-down	100	0	100
Total		594	6	99
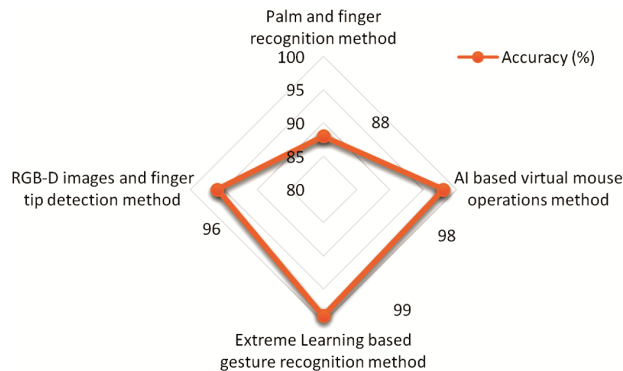
Fig. 4 — Comparative insight on the virtual mouse methods

provides a brief comparative insight on the virtual mouses devised based on existing gesture recognition methods and the proposed method (which attains 99%).

As presented, the implications of ELM show a large impact on the performance of certain learning tasks and solving time series problems.[29,30] Thus, ELM is frequently recommended[31–34] in a multitude of sectors, over other learning methods because of its robust pace, strong applicability, and simple execution.

Conclusions

As extreme learning is quick in training by virtue, the ultimate goal of the virtual mouse devices is achieved by implementing this gesture recognition model. From the model's findings, we can determine that the proposed method helped virtual mouse device to function brilliantly and with greater accuracy than present systems. Also, such effective human computer interaction is highly required in the current world of computing and can be utilised in many real-world applications like security, privacy, integrity, etc., as well as decreases the spread of COVID-19 or any such pandemics. However, the virtual mouse with the

proposed approach faces certain challenges, including a modest reduction in the precision of the appropriate click mouse feature, as well as the mannequin has trouble in picking, clicking and dragging the text. These are some of the challenges that the suggested hand gesture recognition system has, which we will work to overcome in the future. Obviously, the proposed technology might be modified to handle with both keyboard and mouse input. Another aspect is mouse capability, which has the potential to expand the realm of HCI.

References
1   Yang J & Ismail A W, A review: deep learning for 3D reconstruction of human motion detection, *Int J Innov Comput*, **12(1)** (2022) 65–71, https://doi.org/10.11113/ ijic.v12n1.353.
2   Lai Z, Wang M, Wang L & Zhou Y, Intelligent running posture detection based on artificial intelligence combined with sensor, *J Sensors*, (2022), https://doi.org/10.1155/ 2022/6746260.
3   Panwar M, Hand gesture recognition based on shape parameters, *Int Conf Comput, Commun Appl* (IEEE) 2012, 1–6, https://doi.org/10.1109/ICCCA.2012.6179213.
4   Sriraman G, Sountharrajan S & Suganya E, Agile and touchless automation in the software industry, *8th Int Conf Adv Comput Commun Syst* (IEEE) 2022, 1492–1498, https:// 10.1109/ICACCS54159.2022.9785260.
5   Harshitha R, Syed I A & Srivasthava S, HCI using hand gesture recognition for digital sand model, *IEEE 2nd Int ConfImage Info Proc* (IEEE) 2013, 453–457, https://doi.org/ 10.1109/ICIIP.2013.6707633.
6   Subba T & Chingtham T S, A review on types of machine learning techniques for biosignal evaluation for human computer interaction, Advanced Computational Paradigms and Hybrid Intelligent Computing, *Adv Intell Syst Comput*, **1373** (2022), 457–466, https://doi.org/10.1007/978-981-16-4369-9_45.
7   Qi J, Jiang G, Li G, Sun Y & Tao B, Intelligent human-computer interaction based on surface EMG gesture recognition, *IEEE Access*, **7** (2019) 61378–61387, https://doi.org/10.1109/ ACCESS.2019.2914728.

8   Jiang D, Li G, Sun Y, Kong J & Tao B, Gesture recognition based on skeletonization algorithm and CNN with ASL database, *multimed Tools Appl*, **78** (2019) 29953–29970, https://doi.org/10.1007/s11042-018-6748-0.

9   Matlani R, Dadlani R, Dumbre S, Mishra S & Tewari A, Virtual mouse using hand gestures, *Int Conf Tech Adv Innov* (IEEE) 2021, 340–345, https://doi.org/10.1109/ICTAI53825.2021.9673251.

10 Li G, Jiang D, Zhou Y, Jiang G, Kong J & Manogaran G, Human lesion detection method based on image information and brain signal, *IEEE Access*,**7** (2019) 11533–11542, https://doi.org/10.1109/ACCESS.2019.2891749.

11 Peng F, Chen C, Lv D, Zhang N, Wang X, Zhang X & Wang Z, Gesture recognition by ensemble extreme learning machine based on surface electromyography signals, *Front Hum Neurosci*, **16** (2022), https://doi.org/10.3389/ fnhum.2022.911204.

12 Buń P, Husár J & Kaščak J, Hand tracking in extended reality educational applications, *Advances in Manufacturing III: Volume 2-Production Engineering: Research and Technology Innovations, Industry 4.0 (Cham: Springer International Publishing)*, (2022) 317–325, https://doi.org/10.1007/978-3-030-99310-8_25.

13 Sodhro A H, Sennersten C & Ahmad A, Towards cognitive authentication for smart healthcare applications, *Sensors*, **22(6)** (2022) 2101, https://doi.org/10.3390/s22062101.

14 Alnaim N, *Hand Gesture Recognition using Deep Learning Neural Networks*, Ph.D. Thesis, Brunel University, London, UK, 2020, http://bura.brunel.ac.uk/handle/2438/20923.

15 Ouda M, Al-NajiA & Chahl J, Elderly care based on hand gestures using kinect sensor, *Computers*, **10(5)** (2021) 5, https://doi.org/10.3390/computers10010005.

16 Zhang Y, Liu B & Liu Z, Recognizing hand gestures with pressure sensor based motion sensing, *IEEE Trans Biomed Circuits Syst*, **13** (2019) 1425–1436, https://doi.org/10.1109/TBCAS.2019.2940030.

17 Mujahid A, Awan M J, Yasin A, Mohammed M A, Damaševičius R, Maskeliūnas R & Abdulkareem, K H, Real-Time Hand gesture recognition based on deep learning YOLOv3, *Model Appl Sci*, **11(9)** (2021) 4164, https://doi.org/10.3390/app11094164.

18 Min Y, Zhang Y, Chai X & Chen X, An efficient point LSTM for point cloud based gesture recognition, *Proc the IEEE/CVF Conf Comput Vis Patt Recog* (IEEE) 2020, 5761–5770, https://doi.org/10.1109/CVPR42600. 2020.00580.

19 Neethu P, Suguna R & Sathish D, An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks, *Soft Comput*, **24** (2020) 15239–15248, https://doi.org/10.1007/s00500-020-04860-5.

20 Asadi-Aghbolaghi M, Clapes A, Bellantonio M, Escalante H J, Ponce-López V, Baró X, Guyon I, Kasaei S & Escalera S, A survey on deep learning-based approaches for action and gesture recognition in image sequences, *Proc 12th IEEE Int Conf Auto Face & Gesture Recog* (IEEE)*, **30** (2017) 476–483, https://doi.org/10.1109/FG.2017.150.

21 Cao C, Zhang Y, Wu Y, Lu H & Cheng J, Egocentric gesture recognition using recurrent 3d convolutional neural networks with spatiotemporal transformer modules, *Proc IEEE Int Conf Comput Vis* (IEEE) 2017, 3763–3771, https://doi.org/ 10.1109/ICCV.2017.406.

22 JohnV, Boyali A, Mita S, Imanishi M & Sanma N, Deep learning-based fast hand gesture recognition using representative frames, *Proc Int Conf Digi Imag Comput Technol Appl* (IEEE) 2016, 1–8, https://doi.org/10.1109/DICTA.2016.7797030.

23 Wang S, Song J, Lien J, Poupyrev I & Hilliges O, Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum, *Proc 29th Ann Sym User Interf Soft Tech,* (2016) 851–860, https://doi.org/10.1145/2984511.2984565.

24 Funke I, Bodenstedt S, Oehme F, Von Bechtolsheim F, Weitz J & Speidel S, Using 3D convolutional neural networks to learn spatiotemporal features for automatic surgical gesture recognition in video, *Proc Int Conf Med Imag Compu Compu-Assis Interv* (Springer International Publishing) 2019, 467–475, https://doi.org/10.48550/arXiv.1907.11454.

25 Huang, Guang-Bin, Qin-Yu Z & Chee-Kheong S, Extreme learning machine: Theory and applications, *Neurocomput*, **70(1-3)** (2006) 489–501, https://doi.org/10.1016/j.neucom.2005.12.126.

26 Wang J, Lu S, Wang S H & Zhang Y D, A review on extreme learning machine, *Multimed Tools Appl*, **81(29)** (2022) 41611–41660, https://doi.org/10.1007/s11042-021-11007-7.

27 Chen Z H, Kim J T, Liang J, Zhang J & Yuan, Y B, Real-time hand gesture recognition using finger segmentation, *The Scientific World Journal,* (2014), https://doi.org/10.1155/2014/267872.

28 Shriram S, Nagaraj B, Jaya J, Shankar S, Ajay P, Deep learning-based real-time ai virtual mouse system using computer vision to avoid COVID-19 spread, *J Healthc Eng,* (2021), https://doi.org/10.1155/2021/8133076.

29 Weng F, Chen Y, Wang Z, Hou M, Luo J & Tian Z, Gold price forecasting research based on an improved online extreme learning machine algorithm, *J Ambient Intell Humaniz Comput,* **11** (2020) 4101–4111, https://doi.org/10.1007/s12652-020-01682-z.

30 Prakhar K, Sountharrajan S, Suganya E, Karthiga M & Kumar S, Effective stock price prediction using time series forecasting, *6th Int Conf Trends Electron Info* (IEEE) 2020, 1636–1640, https://10.1109/ICOEI53556.2022.9776830.

31 Moccia S, Solbiati S, Khornegah M, Bossi F F & Caiani E G, Automated classification of hand gestures using a wristband and machine learning for possible application in pill intake monitoring, *Comput Methods Programs Biomed,* **219** (2020) 106753, https://doi.org/10.1016/j.cmpb.2022.106753.

32 Varma A, Pawaskar S, More S & Raorane A, Computer control using vision-based hand motion recognition system, *ITM Web of Conf EDP Sci*, **44** (2022) 03069–03074, https://doi.org/10.1051/itmconf/20224403069.

33 Jain R, Jain M, Jain R & Madan S, Human computer interaction – Hand gesture recognition, *Adv J Grad Res*, **11(1)** (2021) 1–9, https://doi.org/10.21467/ajgr.11.1.1-9.

34 Sairam U, Gowra D K & Kopparapu S C, Virtual mouse using machine learning and GUI automation, *28th Int Conf Adv Comput Commun Syst* (IEEE), (2022) 1112–1117, https://doi.org/10.1109/ICACCS54159.2022.9784972.