IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 7, 2020, accepted February 15, 2020, date of publication March 3, 2020, date of current version March 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978028

A Visual Analytic in Deep Learning Approach to Eye Movement for Human-Machine Interaction Based on Inertia Measurement

SHAHRIAR RAHMAN FAHIM[1], DRISTI DATTA[2], MD. RAFIQUL ISLAM SHEIKH[1], (Member, IEEE), SANJAY DEY[3], YEAHIA SARKER[4], SUBRATA K. SARKER[2], FAISAL R. BADAL[4], AND SAJAL K. DAS[4]

[1]Department of Electrical and Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[2]Department of Electrical and Electronic Engineering, Varendra University, Rajshahi 6204, Bangladesh
[3]Department of Computer Science and Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[4]Department of Mechatronics Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh

Corresponding author: Subrata K. Sarker (skshuvo138008@gmail.com)

ABSTRACT This paper proposes a hand free human-machine interaction (HMI) system to establish a novel way for communication between humans and computers. A regular interaction system based on the computer mouse puts the user's hand for too long in a pronation posture that increases inflammation in the wrist and hand. Additionally, the need for hand obstructs the use of computers for handicap people. In this paper, we develop a new pointing device for differently able people based on open and closed human eyes with inertia measurement that restrict to deal with carpal tunnel syndrome (CTS) for regular people and enables a novel way to interact with computers for the handicap people. The proposed system carries the human head gesture and eyes to perform the movement and clicking event of the mouse cursor. A combined three-axis accelerometer and gyroscope is used to detect the head gesture and translate it into the position of the mouse cursor on the computer monitor. To perform the left and right-clicking event, the user needs to shut down the left and right eye for a moment while opening another eye. This paper is also carried out the design of a deep learning approach to classify the individual openness and closeness of both human eyes with quite a high accuracy of 95.36% that ensures the comprehensive control over the clicking performance. The use of complementary filter removes the noise and drift from the obtained performance and confirms the smooth and accurate operation of the proposed device. An experimental validation is added to show the effectiveness of the proposed HMI system. The experimental details along with the performance evaluation prove that the proposed HMI system has extensive control over its performance for differently able people.

INDEX TERMS Accelerometer, gyroscope, complementary filter, deep learning approach, Fitts's law.

I. INTRODUCTION

Human-machine interaction (HMI) develops an assistive communication technology for differently able people with more flexibilities. The HMI creates a phenomenon to design, evaluate and implement interactive computing devices for the interaction. The development of this technology enables a medium of communication between machines and humans by using graphical user interfaces (GUIs). It includes the efficient design of the operating systems, computer graphics, and different programming languages to control the machine for the interaction. On the contrary, other forms of communication include joysticks or tactile screens, graphic design skills, and social and cognitive psychology that control the human factor during the interaction with machines/computers [1]–[4].

A computer interaction based on regular mouse requires a human hand and plane surface for its operation which is not adequate for fixed pointing on the computer screen. To address this problem, the idea of air mouse have been emerged in [5]. These modern computer mice are wireless and use the light sensor to detect its movement. The operation of these mice also depend on the plain and unobstructed surface to effectively poll the user movement. When the

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Mercaldo.

user spends too much time on the computer with air mouse, they feel numbness and pain in their wrist which leads to a heightened risk of the CTS [6]. An ergonomic mouse is designed to solve the CTS problem by reducing the ulnar deviation [7]. The increased pressure in the carpal tunnel area limits the application of ergonomic mouse. Additionally, the need for hand limits the use of above mouses for the handicap people [8]. However, a hand-free operation of the pointing device or mouse is needed to make the computer-access for the disable people as well as to remove the strain in finger and cuts down the joint pain.

Recently, people started to use emerging technology for human-computer interaction (HCI) based on visual information, human voice recognition, brain-computer interfaces (BCI) and head-operated joysticks [9]. A large number of people with hand disabilities pay much attention to this kind of technology for computer interaction because no hand or GUI is required to control the computer mouse. Acoustic based pointing system is one widely used system where the human voice is required for the interaction [10]–[12]. The control unit of this system receives the sound through the microphone and converts it into an electrical signal to perform the desired tasks. Although the use of this technique can relief the use of hands for computer operation, the effect of the external noise complex the proper selection of voice that suffers the accurate pointing.

Bionic technology is proposed to overcome the effect of external noise [13]. It is the combination of biology, robotics and computer science, and plays an important role in the pointing system by using the biological signal from the human organs. Brain-computer interfacing (BCI) is one of the widely used bionic techniques that takes the signal from the neuron of the human brain through a number of electrodes. The amplitude of these signals is quite low that requires an external amplification system. Electrooculography (EOG) is another interaction way to detect and track the eye movement using the computer camera where no amplification system is required to make the interaction with computer. The peoples who are unable to use above interfacing devices for HCI, they can use this technique [14], [15]. This technique requires an additional eye-tracking algorithm to perform the interaction.

Number of eye movement tracking algorithm has already been developed in the state-of-art. One of widely used technique based on the pattern of eye movement have been proposed in [16]. This method is designed based on the minimum redundancy with maximum relevance feature selection of eye. In this method, the pattern of eye movement is achieved by introducing saccades, blinks and fixations characteristics. This algorithm is able to increase the tracking precision up to 76.1% by using the support vector machine. Web browsing and detecting the single identifiable activity is the limitation of this algorithm.

Experts research on the field of eye tracking is divided into signal tracking-based and vision-based method. A signal based eye tracking method is proposed in [17] where neural network with wavelet transform is used to find the potential differences between the cornea and retina to identify the eye movement. This system is able to reduce the tracking error less than 2°. The need of electrodes and fixation problems are the main drawbacks of this method.

A feature-based approach to detect the eye-blink artifact has been proposed for HCI based on EOG [18], [19]. The features of the eye-blink artifact may have maximum absolute value [20], entropy [21], second-order transform [18] and teager-kaiser energy [22]. An artifact detection method by using distance-based approach is presented in [23] for HCI. This is developed by reconstructing a template of the artifacts. The distance between the templates are measured by using dynamic time warping [24] and support vector machine [25]. A constant threshold value used in these technique produces the detection error.

Vision-based gesture recognition technique has been applied to obtain high accuracy for mouse pointing [26]. This system requires the engagement of the human hand to control a pointing device by using 2D and 3D hand gestures. The advantage of this technique is its capability to obtain information from the captured image at long distance. This method has high precision but requires lots of time for computing. The training based methods by using active appearance models (AAM) [27], or histogram of oriented gradient (HOG) based SVM (HOG-SVM) [28] requires less data processing time. However, it is complicated to find an ideal feature dimension that results in produce lower accuracy.

Electroencephalogram (EEG) technique has been developed to overcome the limitation of the feature and distance-based approach [29]. This algorithm combines with automatic thresholding algorithm where the extracted features are processed by using the digital filters. The individual threshold value is able to minimize the detection error. The advantage of this method is rapid data acquisition. Moreover, the noise due to the user movement affects the signal that may develop the lower signal to noise ratio and decreases the accuracy level.

To address the above problems, in this paper, we propose a novel hand free human-machine or computer interaction system for communication between the machine or computer and people with quite high level of accuracy. The main aim of this paper is to develop a new pointing device that able to make interaction with computers for the differently able people.

The main contributions of this paper are as follows:

(i) Development of a novel hand free human-machine interaction system to carry on the combination of left and right eye movement classification for mouse pointing and head gesture for mouse movement on the computer screen.

(ii) Design of a deep learning technique for enhancing the classification performance of the open and closed eye images obtained from a general-purpose webcam.

(iii) Integrating a complementary filter (combination of high and low pass filter) with an inertial measurement unit (IMU) to remove the noise and drift from the obtained

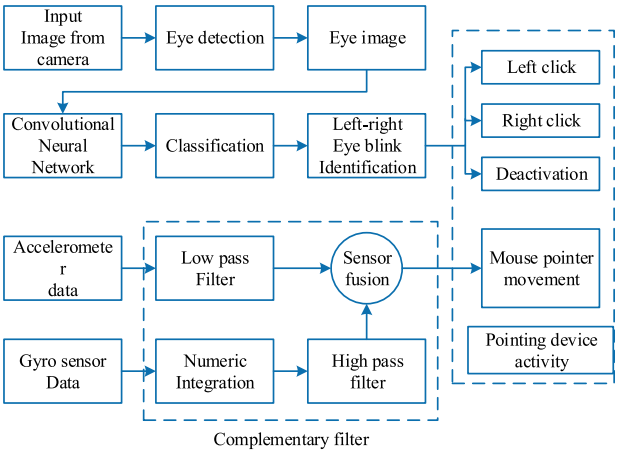

FIGURE 1. Basic block diagram of the proposed system.

performance that results in accurate pointing on the computer screen.

The residue of the paper is ordered as follows: The method of filter implementation and mouse pointer control is described in section II. The implementation of proposed deep learning approach for eye status detection is manifested in section III. Section IV describes the evaluation of usability performance using the computer. A concluding part of the paper is depicted in section V.

II. MOUSE POINTER CONTROL

The basic block diagram of the proposed system is illustrated in Fig.1. The system performs the fundamental mouse event, i.e. left click, right click and the mouse pointer movement on the screen without the intervention of the human hand. A radio frequency (RF) transmitter transfers the head motion activity to the computer to locate the cursor on the screen. The movement of the cursor is followed by the user's head motion activity which is detected by the IMU (combination of accelerometer and gyroscope sensor). The obtained information from IMU is employed with a sensor fusion technique to acquire minimal uncertainty that may affect the system performance. Hence, a complementary filter based sensor fusion technique is used to remove the problems appeared in IMU. The advantages of the proposed system are its capability to reduce the burden of holding the air mouse for its operation, decrease the wrist pain and ensure the reliable communication between the computer and differently able people.

In this paper, the control of the mouse pointer's position is presented with respect to the movement of human head. When the user tilt his head on the left-right or forward-backward direction, the system performs the mouse pointer movement action. Accordingly, for calling the pointer to move in horizontal direction, the condition is to tilt the head to right (for right moving) or to left (for left moving). Similarly, for moving the pointer in vertical direction, the user needs to tilt his head in forward (for moving up) or to the backward (for moving down). An embedded control is used in the proposed design which takes the signal from IMU and performs the required action of mouse pointer. The precious reading from IMU is essential for better placement of the cursor of the mouse in the computer screen. The performance of IMU may be affected due to the presence of external noise which initiates the design of proper filtering system to overcome the noise. The detail design of complementary filter is discussed in the following section.

A. FILTER IMPLEMENTATION

The proposed system determines the current head tilt angle of a user using the combination of gyro sensor and accelerometer. The signals generated from the gyro sensor and accelerometer have to be interpreted to the pixel information in the computer monitor. The gyroscope sensor is used to measure the angular velocity of the head tilt. The output of this sensor is linearly proportional to the rate of change of rotation in degrees per second along the corresponding axis. The angular velocity from the gyroscope reading along with x-axis and y-axis can be represented as,

$$\begin{aligned} x_{\text{gyro}} &= x_{\text{gyro\_ADC}} - x_{\text{gyro\_offset}} * x_{\text{gyro\_scale}} \\ y_{gyro} &= y_{\text{gyro\_ADC}} - y_{\text{gyro\_offset}} * y_{\text{gyro\_scale}} \end{aligned} \tag{1}$$

where,

x_{gyro} and y_{gyro} = Angular velocity along x- and y-axis

$x_{\text{gyro\_ADC}}$ and $y_{\text{gyro\_ADC}}$ = Gyroscope raw data along x- and y-axis

$x_{\text{axis\_offset}}$ and $y_{\text{axis\_offset}}$ = Gyroscope reading when lying stationary with x- and y-axis

$x_{\text{axis\_scale}}$ and $y_{\text{axis\_scale}}$ = Conversion factor along x- and y-axis.

The desired angle of the head tilt is measured by integrating of eq. (1) that represents the rotation of head along with x- and y-axis respectively. The resultant integration of the gyroscope sensor reading can be represented as,

$$\begin{aligned} x_{\text{gyroangle}} &= x_{\text{gyro\_angle}} + x_{\text{gyro}} * dt \\ y_{\text{gyroangle}} &= y_{\text{gyro\_angle}} + y_{\text{gyro}} * dt \end{aligned} \tag{2}$$

Here, $x_{\text{gyro\_angle}}$ and $y_{\text{gyro\_angle}}$ presents the initial angle of the gyroscope. To attain the change in angle, the gyroscope records the previous angle and adds the change in angle with the starting point. The measured angle from the gyroscope data is shown in the Fig.2(a). From the Fig.2(a), it is seen that the measured angle using gyroscope has an aptitude to drift due to the noise in the device and the inherent imperfection.

An accelerometer have been used in the proposed design to determine the small movement of mouse, where the user needs to tilt his head less than 30 degrees in left-right and forward-backward direction. The small angle approximation method is adopted to get the desired angle from the accelerometer output data. The measurement of angle along the x- and y-axis can be given as,

$$\begin{aligned} x_{\text{axis}} &= \frac{(x_{\text{axis\_ADC}} - x_{\text{axis\_offset}}) * x_{\text{axis\_scale}} * 180}{\pi} \\ y_{\text{axis}} &= \frac{(y_{\text{axis\_ADC}} - y_{\text{axis\_offset}}) * y_{\text{axis\_scale}} * 180}{\pi} \end{aligned} \tag{3}$$

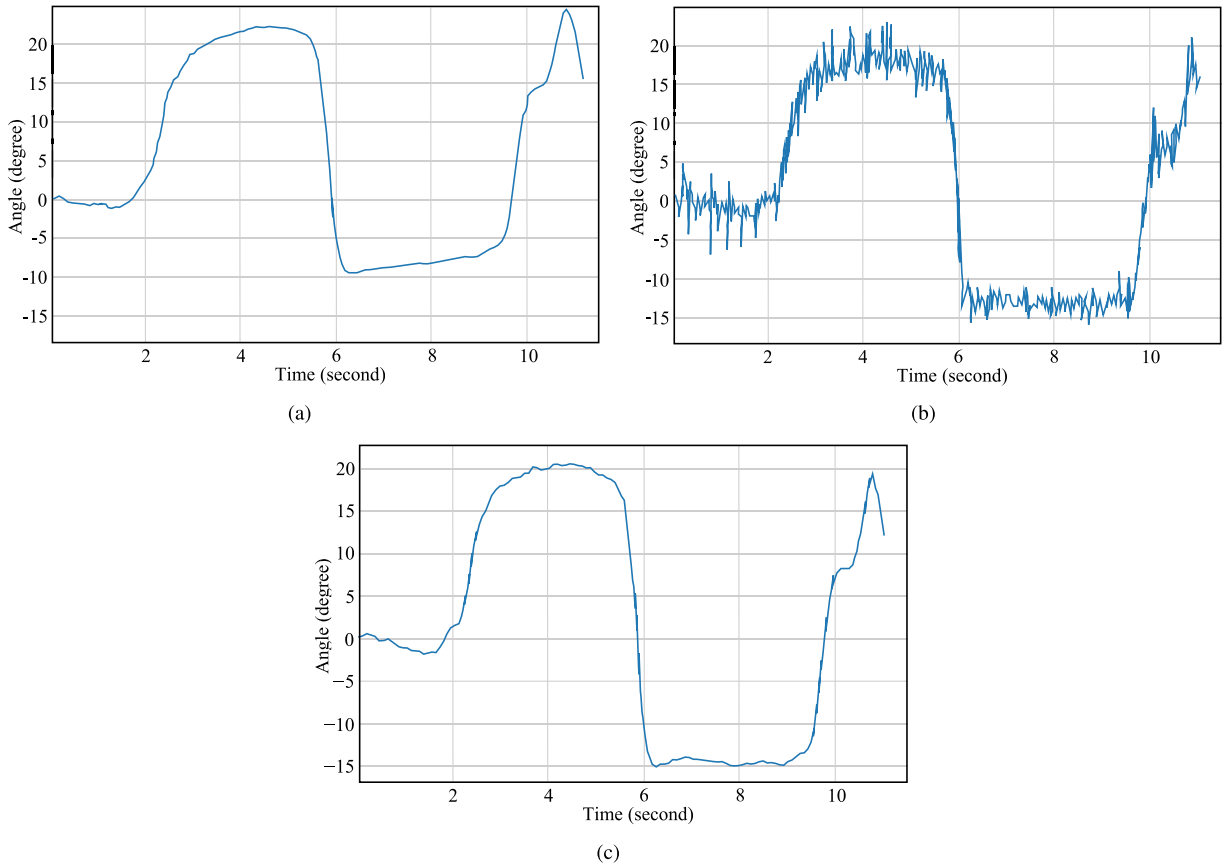

FIGURE 2. Measurement of angle for mouse cursor movement using the (a) gyroscope data only; (b) accelerometer data only and (c) combination of filtered accelerometer and gyroscope data.

where, x_{axis} and y_{axis} presents the measured x- and y-axis angle using accelerometer, $x_{\text{axis\_ADC}}$ and $y_{\text{axis\_ADC}}$ represents the raw accelerometer reading along the x- and y-axis reading. The term $x_{\text{axis\_offset}}$ and $y_{\text{axis\_offset}}$ measures the accelerometer reading in lying flat conditions. The accelerometer sensor measures all the forces playing on it. However, the prone is the disturbance in measurement resulting the small forces. An accelerometer sensor in x-axis angle estimation without filtering is shown in Fig. 2(b).

Both the accelerometer and the gyro sensor are sensible to the noise at a certain frequency. The accelerometer needs a low pass filtering system as it is sensible to the high-frequency noise originated by the system vibration. The gyroscope is also a high-frequency noise sensitive device which becomes a low-frequency drift after integration and causing the data to increase linearly over the time. Thus, a high pass filtering system is necessitated. The filter design for gyroscope sensor can be represented as,

$$\begin{aligned}\bar{X}_{\text{gyro-angle}}[n] &= \alpha * [X_{\text{gyro-angle}}[n-1] + X_{\text{angle}}[n] \\ &\quad + X_{\text{angle}}[n-1]] \\ \bar{Y}_{\text{gyro-angle}}[n] &= \alpha * [Y_{\text{gyro-angle}}[n-1] + Y_{\text{angle}}[n] \\ &\quad + Y_{\text{angle}}[n-1]]\end{aligned}$$

and the design of filter for the accelerometer can be given as,

$$\begin{aligned}\bar{X}_{\text{acc-angle}}[n] &= \beta * x_{\text{axis}} + X_{\text{acc-angle}}[n-1] \\ \bar{Y}_{\text{acc-angle}}[n] &= \beta * y_{\text{axis}} + Y_{\text{acc-angle}}[n-1]\end{aligned} \tag{4}$$

where,
$\bar{X}_{\text{gyro-angle}}$ and $\bar{Y}_{\text{gyro-angle}}$ = Filtered gyro angle along the x- and y-axis;
$\bar{X}_{\text{acc-angle}}$ and $\bar{Y}_{\text{acc-angle}}$ = Filtered accelerometer angle along the x- and y-axis;
and α and β are the smoothing coefficient. The value of α and β are selected as 0.98 and 0.02 respectively for removing the drift from the signal obtained from IMU. The resultant filter design algorithm for the system can be represented for both x- and y-axis as,

$$\begin{aligned}X_{final-angle} &= \bar{X}_{\text{gyro-angle}} + \bar{Y}_{\text{acc-angle}} \\ Y_{final-angle} &= \bar{y}_{\text{gyro-angle}} + \bar{y}_{\text{acc-angle}}\end{aligned} \tag{5}$$

The resulted filter angle for X axis is shown in Fig.2(c). It is observed that the measured angle using the complementary filter is more accurate and provides the lower amount of error as compared to the only gyroscope and accelerometer angle. It is also seen that the final output has no signal coupling. The filtered angle for Y-axis introduces the similar output as like as X-axis.

Algorithm 1: Cursor position on computer screen.

Input: Raw IMU sensors data obtained from user's head movement.
Output: Cursor position(x_{cursor} , y_{cursor}) on screen.

1. x_{gyro} and y_{gyro} ← get angular velocity from the raw gyroscope data using (1).
2. Measure angle along x and y axis by integrating x_{gyro} and y_{gyro} over time.
3. x_{axis} and y_{axis} ← get angle along x and y axis from raw accelerometer data using (3).
4. Implement low pass filter on x_{axis} and y_{axis}. } Complementary filter
5. Implement high pass filter on x_{gyro} and y_{gyro}. } Complementary filter
6. $final_{x\text{-}angle}$ and $final_{y\text{-}angle}$ ← get final x and y axis rotational angle using (5). } Complementary filter
7. Map x and y axis rotational angle range within $P_{h\text{-}min}$ to $P_{h\text{-}max}$ and $P_{v\text{-}min}$ to $P_{v\text{-}max}$.
8. ∇x and ∇y ← get the difference between current and previous pixel.
9. **for** each ∇x
   move cursor by one pixel along ∇x direction.
   until ∇x==0
10. **for** each ∇y
    move cursor by one pixel along ∇y direction.
    until ∇y==0
11. **end**

FIGURE 3. Algorithm for moving the mouse pointer on the computer screen.

B. MOUSE CURSOR MOVEMENT ON THE SCREEN

The proposed HMI system adopts the relative coordinate system where the current locus of the mouse cursor is always at (0,0). For a display of 1280*1024 resolution, the mouse cursor position always ranges from -1280 to 1280 and -1024 to 1024 along the x and y-axis. The general purpose mouse uses dots per inch (DPI) as a measure of the sensitivity while the controlled head movement mouse uses the term dots per degree (DPD) for the measurement of sensitivity. It means how far the mouse pointer will move on the screen for a change of one degree. As the human head has a limited bending angle, the selection of DPD affects much in the performance. The selected DPD for the system is chosen 45° to obtain the promising performance of the system.

The head gesture involves both of the left-right and up-down movement of the user's head. The up-down head movement is responsible to move the mouse cursor along the y-axis, while the left-right movement is for moving it along the x-axis. The head tilt angle along both axes are appropriately mapped to point the cursor coordinates on the computer screen. When the user's head is in the rest position, the noise due to the tiny movement makes the mouse pointer unstable. For this reason, a neutral zone is selected which ranges 5° in each side from the normal to head along both axes where the head movement is neglected. Fig. 3 shows the algorithm for moving the mouse pointer on the screen. At the beginning stage, the system receives the raw gyroscope data and integrate it over the time. Thereafter, the system receives the accelerometer data. As both the accelerometer and the gyroscope are prone to noise, the final head tilt angle is calculated by applying the complementary filtering technique as illustrated in *step:4 to 6*. The rotational information is mapped within the minimum horizontal and vertical pixel

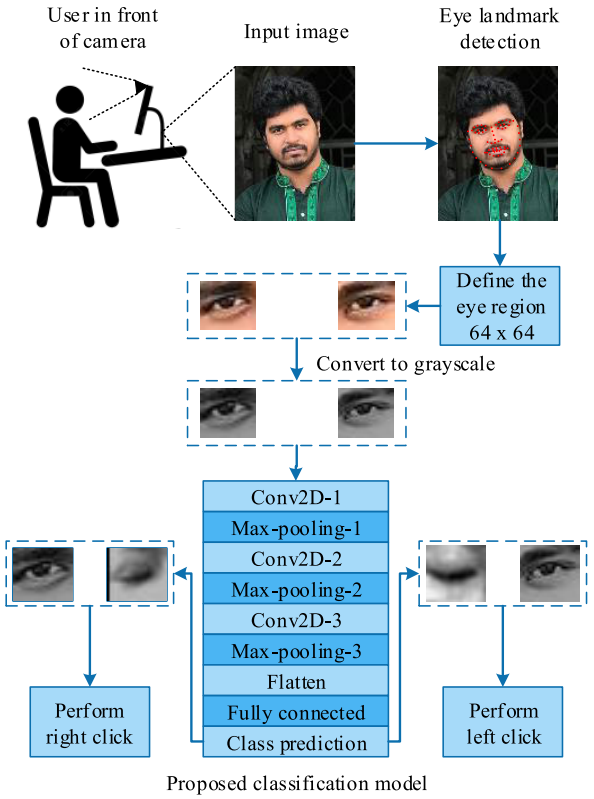

FIGURE 4. Flowchart for performing the mouse clicking event.

(P_{h-min}, P_{v-min}) to the maximum horizontal and vertical pixel (P_{h-max}, P_{v-max}) of display. If there is a difference between the previous and current pixel value, ∇x is detected, the pointer moves in the ∇x direction (+ve axis for $+\nabla x$ and -ve axis for $-\nabla x$) with $|\ \nabla x\ |$ pixels.

Algorithm: Performing clicking event

Input: User's image from general purpose webcam
Output: Perform left and right clicking event

1. Capture image with webcam installed in front of the user.
2. Detecting both left and right eye (E_r and E_l) using pre trained model.
3. Cropping both eyes to 64 x 64 dimension.
4. Feed the image to the proposed model.
Eye state classifier
5. Applying Convolutional Filter
6. Adding non-linearity utilizing rectified non-linear unit
7. Applying max-pooling operation to Converting the data to 1-dimensional array
8. Feeding the 1-D array to the fully connected layer
9. Classifying the image using sigmoid function
10. Predicting the eye state.
11. **if** left eye close and right eye open
while (true)
{start timer
repeat 1 to 10
if left eye close and right eye open for a time $T_{threshold}$
do left click
else
break}
12. **else if** right eye close and left eye open
while (true)
{start timer
repeat 1 to 10
if right eye close and left eye open for a time $T_{threshold}$
do right click
else
break}
13. **else**
repeat 1

FIGURE 5. Algorithm for performing the eye state controlled mouse clicking event.

III. PROPOSED DEEP LEARNING ARCHITECTURE FOR MOUSE CLICKING EVENT

This section presents the detail information of making the proposed approach for left and right clicking event of the proposed pointing device. The flow chart of the mouse clicking event with the proposed eye state classification approach is shown in Fig. 4. The process begins from the clicking of user's image with a general purpose webcam installed on the computer. The user's eye image of size 64 × 64 is then extracted on basis of facial landmark detection as shown in Fig.4. The extracted grayscaling eye image is fed into the proposed eye state classification model. Based on the prediction of proposed eye state classification model, the system performs left or right clicking event.

Fig. 5 shows the algorithm to describe the details of performing the eye state controlled mouse clicking event. In this algorithm, the *step:5 to 9* outlines the procedure of the proposed eye state classification model. Once the eye state is classified by performing *step:1 to 10*, the system waits for a time $T_{threshold}$ which can be adjusted by the user. This time delay separates the regular eye blink from the command blink. If the user closes the left or right eye greater than the threshold value, the necessary control action will be performed. The proposed eye state classification approach comprises repeated sets of neurons that are applied over the space of the user's eye image. This sets of neuron apply over and over upon all the patches of the image and alluded as 2-D convolutional kernels.

The term kernels is used as a matrix in machine learning to extract the most significant features from every subspace of an image. The working process of proposed approach is similar to the traditional convolution neural network. The proposed architecture classify the image by using the sequential operation of convolution and pooling layer and transfers the activation values in one volume to another volume by means of a differentiable function. The first layer consists of convolutional kernels that provides an output using the rectified linear unit (ReLu). On the contrary, the polling layer reduces the amount of computation time and optimize the parameters to acquire better accuracy. The proposed structure uses the spatial invariance to avoid the over fitting of the obtained result that differentiates the traditional CNN and proposed architecture. The model used in the system has three convolution layer accompanied by three max-pooling layers

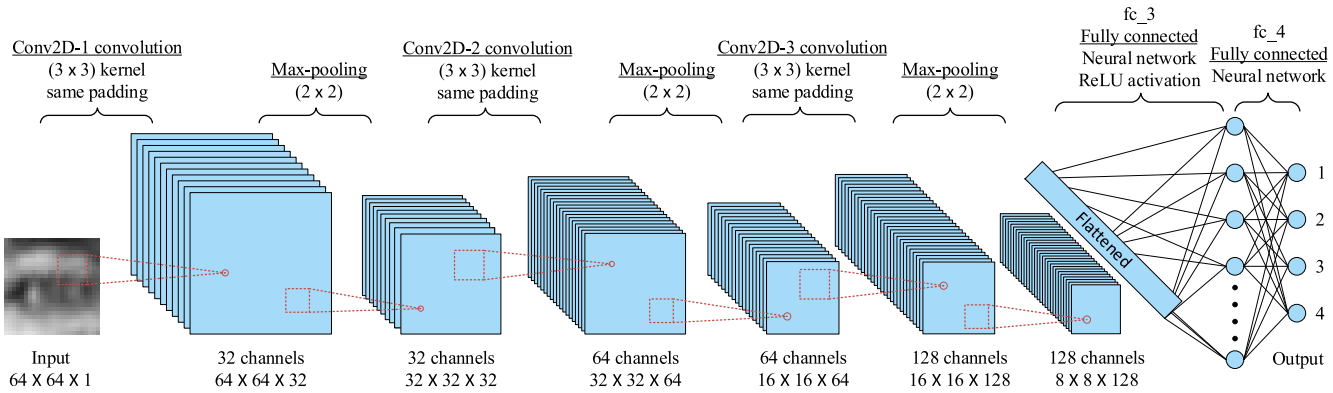

FIGURE 6. Structure of eye state classifier model.

TABLE 1. Numbers and sizes of filters in the CNN network used in the system.

Layer Name	Output shape	parameter
Conv2D	$64 \times 64 \times 32$	320
Activation	$64 \times 64 \times 32$	0
MaxPooling2D	$32 \times 32 \times 32$	0
Conv2D	$32 \times 32 \times 64$	8256
Activation	$32 \times 32 \times 64$	0
MaxPooling2D	$16 \times 16 \times 64$	0
Conv2D	$16 \times 16 \times 128$	32896
Activation	$16 \times 16 \times 128$	0
MaxPooling2D	$8 \times 8 \times 128$	0
Flatten	8192	0
dense	512	4194816
Activation	512	0
Dense	256	131328
Activation	256	0
dense	4	514
Activation	4	0
classifier		Softmax

of size 2×2. The number and the size of the filter in the convolution and pooling layers are shown in Table 1.

A. STRUCTURE OF PROPOSED DEEP LEARNING ARCHITECTURE

The proposed structure of the deep learning algorithm for the eye state classification scheme is presented in Fig. 6. This architectural element with regards to the convolutional and max-pooling layers, filters, filter sizes, and nodes are similar to that of a CNN model like AlexNet [30] except in fully connected (FC) layer. In order to achieve the less delay time during the HCI, the fine tuning of the neural network output is not done.

The proposed deep learning model takes an image of size 64×64 pixel as an input and perform classification of open or closed left or right eye. The generated image after resizing is used as the final input for this model. In Table 1, the conv2D-1 to conv2D-3 are the convolutional layer. Additionally, the max-pooling layers or the sub sampling layers are the layer that selects the maximum value in one of the feature abstraction stages. The information is passed through the 3 convolutional layers and the 3 max-pooling layers. Additionally, the three FC layers after the convolutional and max-pooling layer stack performs the similar computations when applying it after using the inner products in a neural network. By progressing through each of the aforementioned layers, the proposed architectures extracts the features of left/right and open/closed eyes. The generated features are passed through the Softmax layer to classify the left and right eye open/closed.

B. FEATURE EXTRACTION VIA CONVOLUTIONAL LAYER

The process of feature extraction using convolutional layer is discussed in this section. In the convolutional layers, the feature extraction procedure is staged by applying a 2D convolutional operation to the input eye image. The application of convolutional operation to the input image changes the stride number for the horizontal and vertical direction, filter exploration movement range and the dimension size of the resulting image. Therefore, the filter size, number of filter, padding operation and stride values are the main factors need to be considered in the convolutional layer.

From Table 1, the conv2D-1 layer has 32 filters of size 64×64 and strides by 1 pixel unit in the horizontal and vertical directions. One pixel unit padding is applied in the horizontal and vertical direction. A max-pooling filter of size 32×32 with a stride of 1 pixel unit explores in the horizontal and vertical direction. The remaining conv2d-2 to conv2d-3 layer performs the operation with 64 and 128 filters of size 32×32 and 16×16 respectively with a padding of 1 pixel and striding by 1 pixel unit. The pool size of the second and third max pooling layer is selected as 16×16 and 8×8 with a padding of 1 pixel unit. The term Relu and softmax activation function is used in convolutional layer and output layer respectively. The function of Relu activation function is to convert the nonlinear separable data to linear data which

is fed to output layer. The proposed architecture successively apply the pooling operations and convolution filters to input data and creates a hierarchy of layers. The output of those layers are increasingly complex feature vectors than the input data which is necessary to simplify the data for obtaining high accuracy.

In the proposed methodology, every pixel is given as input of $n_1 \times 1$ for the input layer where n_1 defines the number of bands in eye image. The hidden convolution layer filters the $n_1 \times 1$ input vector through t kernels with the size of $k_1 \times 1$. The number of nodes in convolution layer can be defined as $t \times n_2 \times 1$, where $n_2 = n_1 - k_1$. The activation map of the convolutional layer can be represented as,

$$\tilde{y}^i(r) = max\left(0, \tilde{b}^{j(r)} + \sum_j \tilde{k}^{ij(r)} * \tilde{x}^{i(r)}\right) \tag{6}$$

where, $\tilde{x}^{ir} = i$th input activation map.
$\tilde{y}^{jr} = j$th output activation map.
$\tilde{b}^{j(r)} =$ bias of the jth output map.
$\tilde{k}^{ij(r)} =$ convolution kernel between the ith input map and the jth output map.

C. MAX POOLING LAYER

Pooling layer decreases the size of the convolutional layers' output. When passing through multiple pooling layers, a large image will be scaled down but keeps the features required for recognition. In max pooling layer the maximum value is stored. The max-pooling layer can be defined as,

$$\tilde{y}^i_{jk} = max\left(\tilde{x}^i_{js+m,ks+n}\right) \tag{7}$$

Here, $\tilde{y}^i_{jk}$ denotes a neuron in the ith output activation map, which is computed over an non-overlapping local region of size $(s \times s)$ in the ith input map.

D. FULLY CONNECTED LAYER

In the fully connected layers, every neuron in layer l is fully connected to the outputs of all neurons in layer $l-1$. Each of the connection is necessitated for the calculation of the weighted sum. The output $y^{(i)}j$ of neuron j in a fully connected layer l is dependent on,

$$y^i_{jk} = \phi(x)\left(\sum_{i=1}^{n^{(l-1)}} y^{(l-1)}(i).w^{(l)}(i,j) + b^{(l)}(j)\right) \tag{8}$$

where $N^{(l-1)}$ defines the number of neurons in the previous layer $(l-1)$, $w^{(l)}(i,j)$ is the weight for the connection from neuron i in layer $(l-1)$ to neuron j in layer l, $b^{(l)}j$ is the bias of neuron j in layer l and $\phi(x)$ is the activation function.

E. SOFTMAX CLASSIFIER

Softmax classifier is applied to handle multi-class classification in the fully connected condition of the system. Assume that there are K classes and n labeled training sample. For each test input, the softmax classifier creates a K-dimensional vector whose elements sum to 1. Each element of output vector represents the estimated probability of each class label as follows,

$$P\left(y_i = m \mid x_i; W\right) = a_i = \frac{e^{w_i^T x_i}}{\sum_{j=1}^{K} e^{w_j^T x_j}} \tag{9}$$

where, $W = w_1, w_2, w_3, \ldots, w_k$ are the parameters which is learned by the back-propagation algorithm. The cross entropy loss function is used as the cost function for the Softmax classifier and it can be calculated as,

$$J(W) = -\sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij} \log\left(\frac{e^{w_j^T x_i}}{\sum_{m=1}^{K} e^{w_m^T x_i}}\right) \tag{10}$$

With N is the number of data points in the training set. Then, the gradient descent method is applied to solve the minimum of the $J(W)$ as,

$$\nabla_W J\left(W\right) = \sum_{i=1}^{N} x_i \left(a_i - y_1\right)^T \tag{11}$$

Finally, the parameters are updated as,

$$W^n ew = W^o ld - \eta \nabla_W J\left(W\right) \tag{12}$$

where, η is the learning rate.

F. TRAINING OF THE PROPOSED DEEP LEARNING ARCHITECTURE

The neural network imitates the application of neurons presented in the human brain. The term weight is used as the strength of electrical signal while the neurons are connected. The weight of the specific neuron is multiplied with the specific value given as input to the neuron. The summation of all weighted values associated to the next layer applies as a input for the activation function. In this kinds of networks, the interaction between the neurons affect the output neurons when a new data is inserted to the network. The inclusion of new data needs to optimize to comply with previous input. The back-propagation algorithm is applied to optimize the neuron weights in the proposed neural network architecture. The popular adaptive moment estimation (Adam) [31] optimizer is adopted to learn the proposed architecture that works based on the extension to the conventional stochastic gradient descent method (SGD) [32]. The conventional SGD computes on a random selection of data examples which is inefficient. On the other hand, the Adam optimizer computes the individual adaptive learning rates for the different parameters.

At the beginning of the training procedure, the input data is labeled with the correct output class in advance. The labeled data sample passes through the neural network. The actual and the generated output from the neural network can either be different or same. If there is a difference between the outputs, the learning rate is multiplied with a weighted parameter that reflects the differences. The obtained result is applied when the new weight values are updated. The gaussian distribution with a standard deviation of 0.001 and

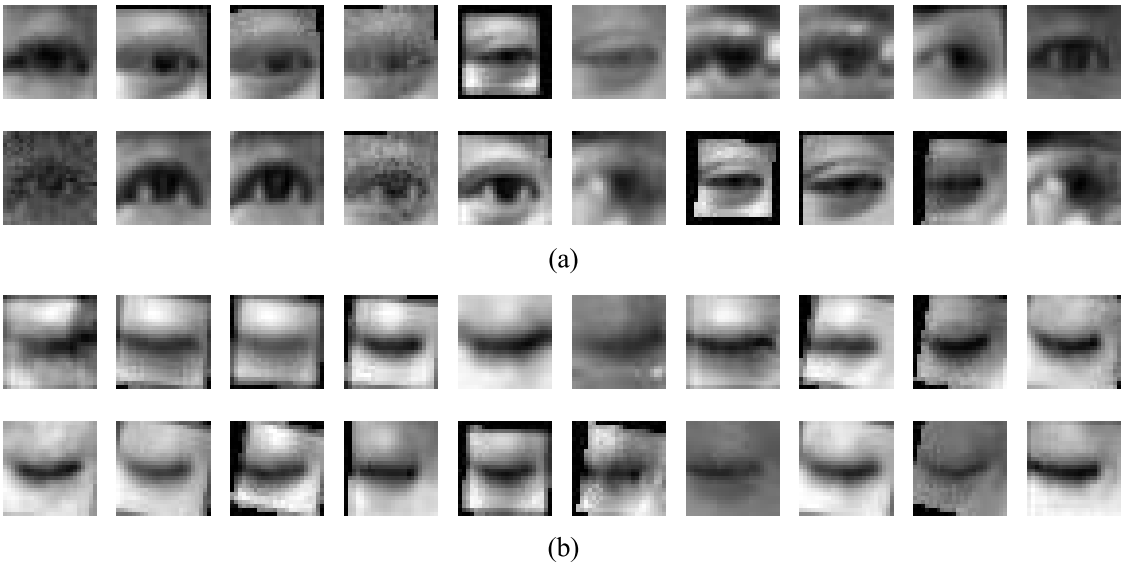

FIGURE 7. Examples of (a) open; and (b) closed eye image from data base 1.

a mean value of 0 randomly initializes the weights used in th FC layer. The optimal parameters for the Adam optimizer are determined by conducting several experiments to obtain the lowest loss value and the highest model accuracy.

G. TESTING RESULTS OF THE PROPOSED DEEP LEARNING ARCHITECTURE

For validating the opened and closed eye classification using the proposed approach, we uses the combination of an open database DB1 [36] and a purpose build database DB2. The image resolution in DB1 is 24 × 24 which is then resized to 64 × 64 pixels that contains a total of 2423 images where the value of 1192 are closed eye images and 1231 are open eye images. The examples of open and closed eye image from DB1 is shown in Fig.7. With the purpose of checking the robustness of the proposed approach, the system uses a combined database for testing and training the proposed model. Due to the small amount of opened and closed eye images in DB1, an abundant of data is needed to ascertain the optimal values for copious coefficients. The use of the abundant data produces an over fitting result of the training set in a traditional CNN structure. This is happened during the training and testing the model with a diminutive amount of data. However, the proposed architecture uses a data augmentation technique to make the purpose build database DB2. Using the image rotation, horizontal flip, image scaling and image translation, the DB2 is created as a number of 542,752 images of 64 × 64 pixels. All images are rotated by 10 degrees for 3 times in both clockwise and anti-clockwise direction to get a data multiplication factor of 7 and the image translation and scaling technique is applied to 16 times. All the translated and rotated images are flipped horizontally that creates a total of 224 images from a single image. The data augmentation techniques with the original image is presented in Fig. 8. The use of data augmentation technique avoids to require the additional data in training data set and removes the chances to create the over fitting result in the training set.

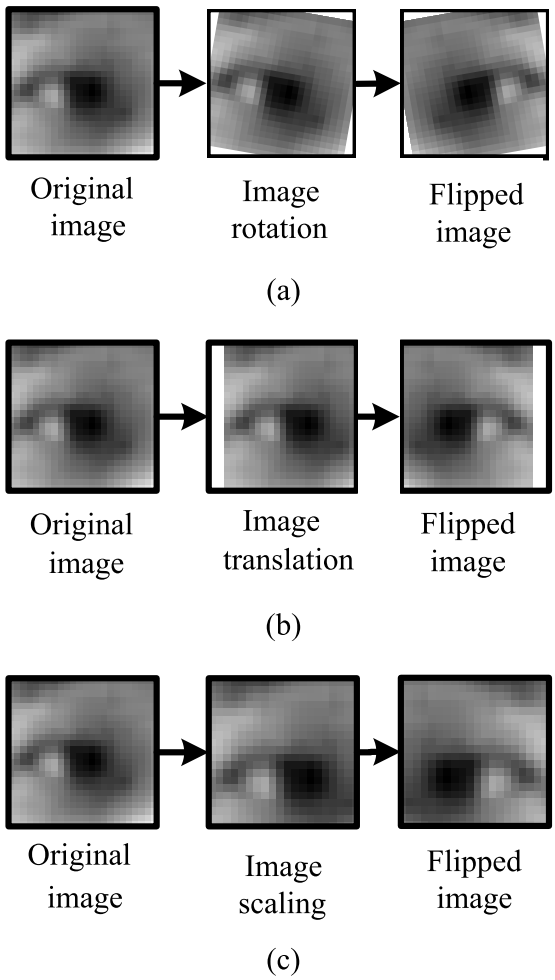

FIGURE 8. Creating image for augmented database.

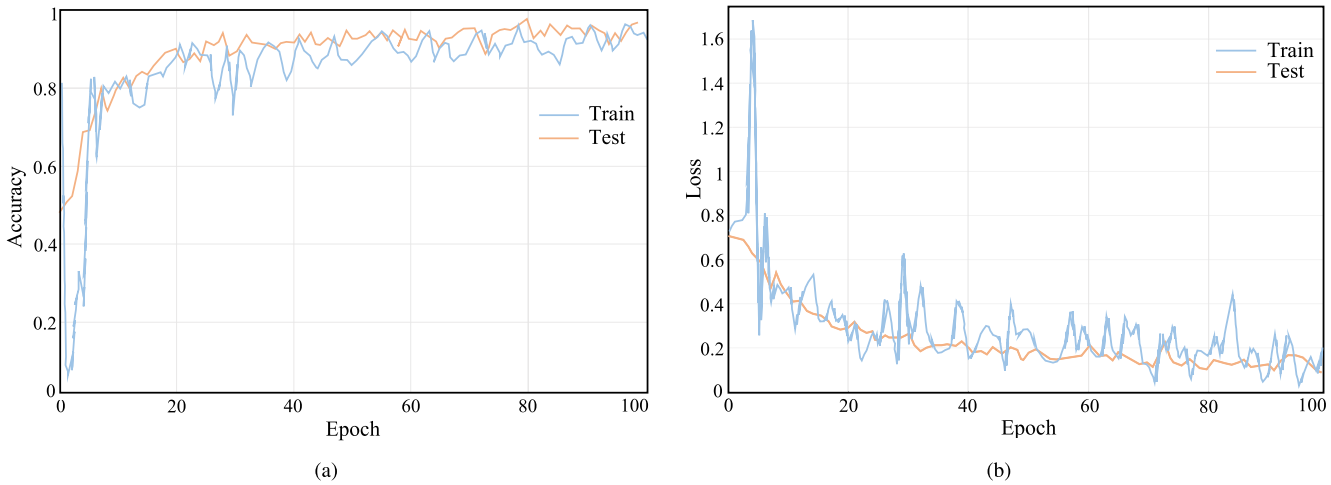

FIGURE 9. (a) Accuracy curve for training and test; (b) Loss curve for training and test.

The proposed model is trained for the 100 epochs with Adam optimization algorithm where the categorical cross-entropy cost function is used to obtain optimum result. After the training phase, the proposed model appears to be promising performances with the augmented dataset. The overall classification accuracy and loss curve conferred in Fig. 9(a) and Fig. 9(b) shows that the accuracy of the proposed approach is 95.36% and the loss is 0.2.

H. COMPARATIVE ANALYSIS WITH THE PROPOSED METHOD AND OTHERS

In this study, the testing accuracies with the proposed classification model to the other models are compared. For the comparison, The combined source separation (SS) and pattern recognition algorithm (PRA), vertical projection, and HOG-SVM technique are taken under consideration. In the case of HOG-SVM based on eye state classification, the HOG extracts the features from the input image and SVM performs the classification of open and closed eye using a radial basis function (RBF). On the other hand, the SS and PRA based technique uses EEG signal data extracted from the eye region. The overall accuracy for the system is noted as 89%. The quantitative measurements regarding the overall accuracy and the input data type are presented in Table 2. It is observed that proposed architecture is capable to provide higher accuracy as compared to other methods. A comparative advantages between the proposed and other existing method is reported in Table 3. In the case of classification, the proposed architecture provides more reliable performance as compared to other existing networks.

IV. EXPERIMENTAL RESULTS

This section exhibits the experimental results of proposed architecture applying in the proposed pointing device to compare between the usability of the traditional computer mouse and pointing devices performance. An experimental setup for the performance measurement of proposed HMI system

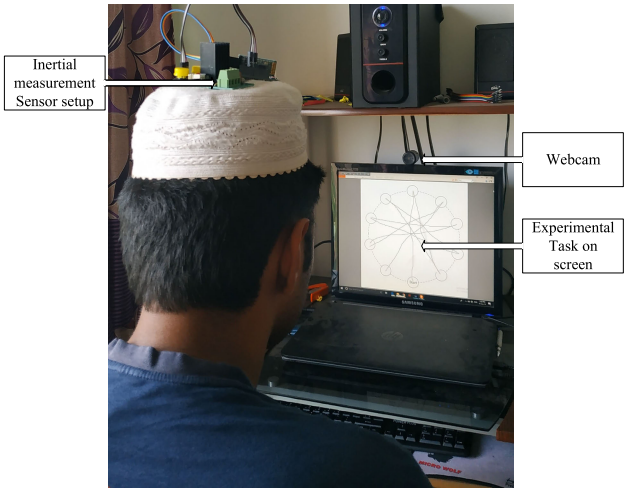

FIGURE 10. An experimental setup for performance evaluation.

TABLE 2. Comparison of classification accuracy using different methods.

No	Reference	Method used	Input data type	Overall accuracy
1	[33]	SS and PRA	EEG signal data	89%
2	[34]	vertical projection	Eye image	89.5%
3	[35]	HOG-SVM	Eye image	85.62%
4	- -	Proposed	Eye image	95.36%

is shown in Fig.10. The system consists of a user, general purpose camera, IMU and multi-directional position selection channel. The function of camera is to take an user image which is passed through the proposed architecture. The output from proposed structure defines the state of the user image and performs the desired task based on multi-directional position channel. The detail performances of the proposed system are discussed in the following section.

TABLE 3. Comparison of advantages between existing and proposed category, and architecture.

Category	Method	Advantages	Disadvantages
Signal based	(i) EOG, (ii) EMG, (iii) EEG	(i) Fast data acquisition speed.	(i) Inconvenient because sensors have to be attached to a user (ii) Low accuracy (iii) Lack of flexibility
Video based	(i) eyelid and facial movements, (ii) Pixel based	(i) High accuracy because it uses information from several images in a videos	(i) Longer processing time due to working with several images.
Non Training image based	(i) Iris detection, (ii) Template matching, (iii) oriented edges	(i) Classifying open/closed eyes from a image without any additional training process	(i) Lower accuracy because information about open/closed eyes is extracted from a single image.
Training based (other than CNN)	(i) SVM-based, (ii) HOG-SVM, (iii) AAM-based	(i) Shorter processing time (ii) High accuracy as compared to non Training image based method	(i) Difficult to find an optimal feature
Training based	**Proposed method**	(i) Automatically extract the optimal feature from the training data without pre or post-processing, (ii) Provide higher classification Accuracy (iii) Require less processing time to process the large time (iv) Able to provide "Condition-change-enduring" classification performance because a large-capacity DB learning under different conditions	(i) Difficult to use for the blind people

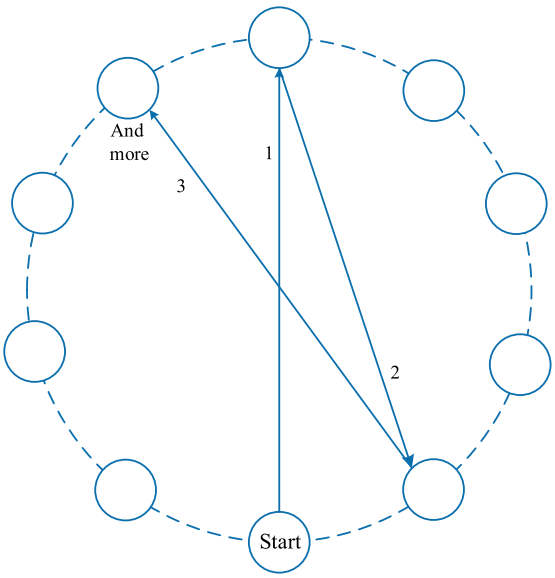

FIGURE 11. Multi-directional position selection task.

A. METHOD

A total number of 10 participants with different abilities, of that 8 are male and 2 are female take part in this experiment. All of the participants are from the computer science and engineering department and their ages range from 22 to 24. For evaluating the performance, most widely used ISO 9241-9 [37] standard multi-directional position selection task is adopted as in Fig.11.

In the task, the user selects the target arranged in a circular pattern and the sequence of the selection is shown in Fig.11. No time limits are given for completion of the task and no penalties for miss-selection. The evaluation process is divided into three sets to appraise whether fatigue is present and the sets are further broken into rounds to find the increasing or decreasing the performances. For each round, each target selection time (MT), target distance (D), target width (W) and the number of total target hits are recorded for calculating the index of difficulty and the index of performance (IP) as a measure of the HCI performance. The performance of proposed system is measured by using the following equations,

$$ID = log_2\left(\frac{2 \times D}{W}\right)$$

$$IP = \left(\frac{ID}{MT}\right)$$

B. RESULTS

The result evaluation is performed by comparing the usability of the proposed pointing device with an A4 Tech OP-620D modeled mouse with USB Optical cable. The IP accomplished by each participant for both pointing devices are shown in Fig. 12(a). From the result, it is observed that the participants with traditional computer mouse perform better as compare to the gesture and eye-controlled mouse. Since all the participants are regular mouse user, the gap in the performance is probably the familiarity with a pointing device which influences the person's ability to efficiently use it.

In the case of successful target hitting, the successful attempts are lower for the proposed pointing device than the regular mouse. The number of unsuccessful selection per 100 targets is illustrated in Fig. 12(b) with a box and whisker chart. The unsuccessful attempts of regular computer mouse

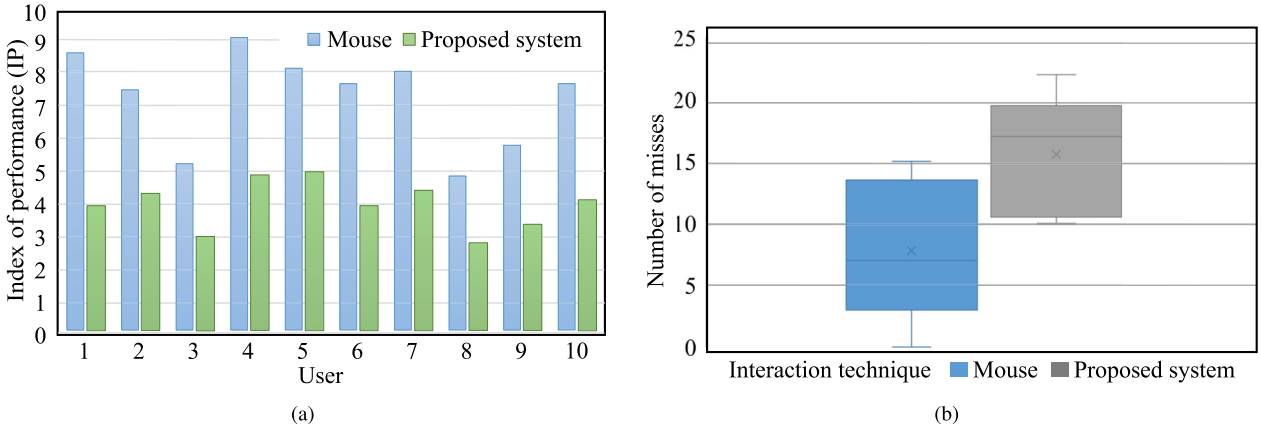

FIGURE 12. (a) Participants' IP for pointing device; (b) Number of misses for 100 target selected.

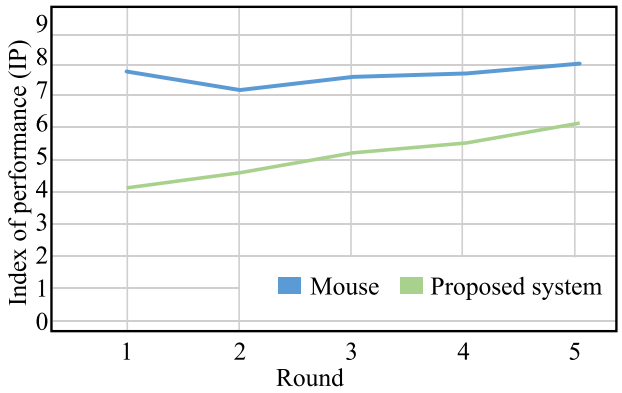

FIGURE 13. Participant IP over different round.

are recorded with a mean value of 7.77 where the highest unsuccessful attempt for the same mouse is recorded as 15. On the other hand, the mean value of unsuccessful attempts for the proposed pointing device is recorded as 15.55 while in lowest unsuccessful attempts are recorded as 10 which is lower than the upper quarterly of the regular mouse.

Another result in Fig. 13 shows that the participants' performance is improved over the successive rounds while the performance for the regular computer mouse almost remains constant. It also clearly indicates that the participants are getting intimated with the new pointing device on the computer screen. The increment of user performance ensure that the participants with proposed pointing device perform better with the traditional mouse. The adaptability of the pointing device depends upon the user's familiarity and convenience. The evaluation of user's performance can be improved by training the user. However, this system provides access to the computer for the differently people. The general user familiar with the other pointing device may find difficulty at the beginning but the continuous using of the device make it comfortable to use.

V. CONCLUSION

This paper presnts a novel human-machine interaction system for controlling a pointer on the computer screen based on open and closed eyes. An accelerometer and a gyroscope sensor tracks the users head gesture and relocate the cursor on the screen. The clicking event functions are performed by detecting the user's eye movement. The detection of both eye movement is accomplished by using a deep learning classifier which is able to provide 95.36% accuracy in the clicking event on the computer screen. The obtained results in this paper is compared with other methods that ensure the comprehensive control over its classification performance. The validation of classifier performance is studied by using the HCI where the output of the classifier provides the input to control the clicking task of the proposed device. Also, this paper carries an experimental validation for the usability of the pointing device while comparing with the obtained performance for a regular computer mouse following the ISO standard. The comparison results show that the proposed system is promising to perform better with a trained user and over the using time. The future work of this research is to enhance the proposed system with more functionality like scrolling or performing double clicking.

REFERENCES

[1] J. Cannan and H. Hu, *Human-Machine Interaction (HMI): A Survey*. Colchester, U.K.: Univ. Essex, 2011.

[2] D. Gorecky, M. Schmitt, M. Loskyll, and D. Zuhlke, "Human-machine-interaction in the industry 4.0 era," in *Proc. 12th IEEE Int. Conf. Ind. Informat. (INDIN)*, Jul. 2014, pp. 289–294.

[3] X. Tong, D. Ding, and W. Li, "Dynamic gesture based short-range human-machine interaction," U.S. Patent 9 164 589, Oct. 20, 2015.

[4] S. Combefis, D. Giannakopoulou, C. Pecheur, and M. Feary, "A formal framework for design and analysis of human-machine interaction," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2011, pp. 1801–1808.

[5] S. Seneviratne, Y. Hu, T. Nguyen, G. Lan, S. Khalifa, K. Thilakarathna, M. Hassan, and A. Seneviratne, "A survey of wearable devices and challenges," *IEEE Commun. Surveys Tutr.*, vol. 19, no. 4, pp. 2573–2620, 4th Quart., 2017.

[6] K. M. Ali and B. W. C. Sathiyasekaran, "Computer professionals and carpal tunnel syndrome (CTS)," *Int. J. Occupational Saf. Ergonom.*, vol. 12, no. 3, pp. 319–325, Jan. 2015.

[7] A. B. Schmid, P. A. Kubler, V. Johnston, and M. W. Coppieters, "A vertical mouse and ergonomic mouse pads alter wrist position but do not reduce carpal tunnel pressure in patients with carpal tunnel syndrome," *Appl. Ergonom.*, vol. 47, pp. 151–156, Mar. 2015.

[8] J. Hammel, S. Magasi, A. Heinemann, D. B. Gray, S. Stark, P. Kisala, N. E. Carlozzi, D. Tulsky, S. F. Garcia, and E. A. Hahn, "Environmental barriers and supports to everyday participation: A qualitative insider perspective from people with disabilities," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 4, pp. 578–588, Apr. 2015.
[9] P. Tsarouchi, S. Makris, and G. Chryssolouris, "Human–robot interaction review and challenges on task planning and programming," *Int. J. Comput. Integr. Manuf.*, vol. 29, no. 8, pp. 916–931, 2016.
[10] L. Kessous, G. Castellano, and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis," *J. Multimodal User Inter.*, vol. 3, nos. 1–2, pp. 33–48, Dec. 2009.
[11] J. Rennies, S. Goetze, and J.-E. Appell, "Personalized acoustic interfaces for human-computer interaction," in *Human-Centered Design of E-Health Technologies: Concepts, Methods and Applications*. Hershey, PA, USA: IGI Global, 2011, pp. 180–207.
[12] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: A review," in *Proc. IEEE 7th Int. Colloq. Signal Process. Appl.*, Mar. 2011, pp. 410–415.
[13] S. Deeny, C. Chicoine, L. Hargrove, T. Parrish, and A. Jayaraman, "A simple ERP method for quantitative analysis of cognitive workload in myoelectric prosthesis control and human-machine interaction," *PLoS ONE*, vol. 9, no. 11, Nov. 2014, Art. no. e112091.
[14] J. M. Hahne, S. Dahne, H.-J. Hwang, K.-R. Muller, and L. C. Parra, "Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 618–627, Jul. 2015.
[15] D. Tan, T. S. Saponas, D. Morris, and J. Turner, "Wearable electromyography-based human-computer interface," U.S. Patent 9 037 530, May 19, 2015.
[16] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 741–753, Apr. 2011.
[17] R. Barea, L. Boquete, S. Ortega, E. López, and J. M. Rodríguez-Ascariz, "EOG-based eye movements codification for human computer interaction," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2677–2683, Feb. 2012.
[18] A. Klein and W. Skrandies, "A reliable statistical method to detect eyeblink-artefacts from electroencephalogram data only," *Brain Topography*, vol. 26, no. 4, pp. 558–568, Mar. 2013.
[19] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, Jan. 2011.
[20] H. Nolan, R. Whelan, and R. B. Reilly, "FASTER: Fully automated statistical thresholding for EEG artifact rejection," *J. Neurosci. Methods*, vol. 192, no. 1, pp. 152–162, Sep. 2010.
[21] G. Barbati, C. Porcaro, F. Zappasodi, P. M. Rossini, and F. Tecchio, "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals," *Clin. Neurophysiol.*, vol. 115, no. 5, pp. 1220–1232, May 2004.
[22] L. Breuer, J. Dammers, T. P. L. Roberts, and N. J. Shah, "Ocular and cardiac artifact rejection for real-time analysis in MEG," *J. Neurosci. Methods*, vol. 233, pp. 105–114, Aug. 2014.
[23] A. Aarabi, K. Kazemi, R. Grebe, H. A. Moghaddam, and F. Wallois, "Detection of EEG transients in neonates and older children using a system based on dynamic time-warping template matching and spatial dipole clustering," *NeuroImage*, vol. 48, no. 1, pp. 50–62, Oct. 2009.
[24] W.-D. Chang and C.-H. Im, "Enhanced template matching using dynamic positional warping for identification of specific patterns in electroencephalogram," *J. Appl. Math.*, vol. 2014, Apr. 2014, Art. no. 528071.
[25] S.-Y. Shao, K.-Q. Shen, C. Jin Ong, E. Wilder-Smith, and X.-P. Li, "Automatic EEG artifact removal: A weighted support vector machine approach with error correction," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 336–344, Feb. 2009.
[26] A. A. Argyros and M. I. Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 40–51, doi: 10.1007/11754336_5.
[27] I. Bacivarov, M. Ionita, and P. Corcoran, "Statistical models of appearance for eye tracking and eye-blink detection and measurement," *IEEE Trans. Consum. Electron.*, vol. 54, no. 3, pp. 1312–1320, Aug. 2008.
[28] L. Pauly and D. Sankar, "Detection of drowsiness based on HOG features and SVM classifiers," in *Proc. IEEE Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN)*, Nov. 2015, pp. 181–186.
[29] W.-D. Chang, J.-H. Lim, and C.-H. Im, "An unsupervised eye blink artifact detection method for real-time electroencephalogram processing," *Physiol. Meas.*, vol. 37, no. 3, pp. 401–417, Feb. 2016.
[30] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360
[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[32] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: http://arxiv.org/abs/1609.04747
[33] R. N. Roy, S. Charbonnier, and S. Bonnet, "Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms," *Biomed. Signal Process. Control*, vol. 14, pp. 256–264, Nov. 2014.
[34] M. Dehnavi and M. Eshghi, "Design and implementation of a real time and train less eye state recognition system," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 30, Feb. 2012.
[35] L. Pauly and D. Sankar, "Non intrusive eye blink detection from low resolution images using HOG-SVM classifier," *Int. J. Image, Graph. Signal Process.*, vol. 8, no. 10, pp. 11–18, Oct. 2016.
[36] F. Song, X. Tan, X. Liu, and S. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," *Pattern Recognit.*, vol. 47, no. 9, pp. 2825–2838, Sep. 2014.
[37] *Ergonomic Requirements for OfficeWork With Visual Display Terminals (VDTs): Requirements for Non-Keyboard Input Devices*. International Organization for Standardization, Geneva, Switzerland, 2000.

SHAHRIAR RAHMAN FAHIM was born in Bangladesh, in 1997. He received the B.Sc. degree in electrical and electronic engineering from the Rajshahi University of Engineering & Technology (RUET), Rajshahi, Bangladesh. His research interests include artificial intelligence and applications, renewable energy systems, robotics, mechatronics systems, and power system control.

DRISTI DATTA was born in October 1992. He received the B.Sc. and M.Sc. degrees from the Rajshahi University of Engineering and Technology, in 2014 and 2019, respectively. He is currently working as an Assistant Professor with the Department of EEE, Varendra University, Rajshahi, Bangladesh. His research interest fields are power system control and stability, smart-grid, micro-grid, and the IoT-based power plant.

MD. RAFIQUL ISLAM SHEIKH (Member, IEEE) was born in Sirajgonj, Bangladesh, in October 1967. He received the B.Sc. (Eng.) and M.Sc. (Eng.) degrees from the Rajshahi University of Engineering & Technology (RUET), Bangladesh, in 1992 and 2003, respectively, and the Ph.D. degree from the Kitami Institute of Technology, Hokkaido, Kitami, Japan, in 2010, all in electrical and electronic engineering (EEE). He joined as a Lecturer with the EEE Department, RUET, in 1994, where he is currently a Professor. He is also working as a Vice Chancellor with RUET. His research interests are, power system stability enhancement by using FACTs devices, renewable energy technologies, smart grid, and load frequency control of multi-area power systems. He has published many technical journal and conference papers, and authored or coauthored three books, and three book chapters.

Dr. Sheikh is the Fellow IEB and the member of BCS Bangladesh.

SANJAY DEY is currently pursuing the degree with the Computer Science and Engineering Department, Rajshahi University of Engineering & Technology. His research interest is mainly on AI systems, machine learning, deep learning, data mining, and computer vision.

YEAHIA SARKER is currently pursuing the Department of Mechatronics engineering, Rajshahi University of Engineering & Technology (RUET), Rajshahi, Bangladesh. He is passionate about innovative research in the field of data science, human–computer interaction, and data processing to solve challenging problems. His current research focuses on hyperspectral image classification with deep learning methods.

SUBRATA K. SARKER was born in Bangladesh in 1996. He received the B.Sc. degree in mechatronics engineering from the Rajshahi University of Engineering & Technology (RUET), Rajshahi, Bangladesh. He is currently working as a Lecturer with the Electrical and Electronic Engineering Department, Varendra University, Rajshahi. His research interests include control theory and applications, robust control of electro-mechanical systems, robotics, mechatronics systems, and power system control.

FAISAL R. BADAL was born in Bangladesh. He received the B.Sc. degree in mechatronics engineering from the Rajshahi University of Engineering & Technology (RUET), Rajshahi, Bangladesh. He is currently working as a Lecturer at RUET. His research interests include control theory and applications, and power system control.

SAJAL K. DAS received the Doctor of Philosophy (Ph.D.) degree in electrical engineering from the University of New South Wales, Australia, in 2014. In May 2014, he was appointed as a Research Engineer with the National University of Singapore (NUS), Singapore. In January 2015, he joined the Department of Electrical and Electronic Engineering, AIUB as an Assistant Professor. He continued his work at AIUB until he joined the Department of Mechatronics Engineering, Rajshahi University of Engineering & Technology (RUET), in September 2015 as a Lecturer. He is currently working as an Assistant Professor with RUET. His research interests include control theory and applications, mechatronics system control, robotics, and power system control.