AVE Trends in Intelligent Health Letters

Deep Learning-Based Eye Movement Detection for Hands-Free Human-Machine Interaction Using Inertial Sensors

D. S. Shobha[1,*], K. Chitra[2], M. Mohamed Thariq[3], Saly Jaber[4]

[1,2]Department of Computer Applications, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India.
[3]Department of Computer Science and Engineering, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[4]Department of Analytical Chemistry, Saint Joseph University, Beirut, Lebanon.
shobhads271@gmail.com[1], chitra-mca@dsatm.edu.in[2], thariq10410@gmail.com[3], saly.jaber@usj.edu.lb[4]

*Corresponding author

Abstract: Hands-Free Human-Machine Interaction is a cutting-edge technology that redesigns the man-computer interface. Disabling the use of traditional input devices, such as the mouse, eliminates most of the common issues associated with poor posture, including wrist and hand fatigue. It can be extremely helpful to those who spend many hours on the computer, as it minimises the possibility of repetitive stress injury. It also offers a simple solution for individuals with disabilities who may find traditional mouse use challenging or even impossible to use. Sophisticated machine learning and deep learning algorithms are utilised in the technology to track eye movements and translate them into accurate control of the mouse pointer. Reading from the pre-stored data, the system interprets even minute variations in eye movement and carries out the corresponding action with accuracy, making it extremely sensitive. Sound suppression algorithms are also employed to remove unwanted noises that could interfere with the device's operation, allowing for smooth interaction. It is extremely convenient for physically disabled individuals, giving them an additional degree of freedom and control. On a large scale, hands-free Human-Machine Interaction is a step towards increasing accessibility, comfort, and efficiency of computer use for a broader population.

Keywords: Advanced Machine Learning; Deep Learning Method; Graphical User Interface; Eye-Controlled Cursor; Human-Computer Interaction; Cutting-Edge Technology; Digital signal processing.

Cite as: D. S. Shobha, K. Chitra, M. M. Thariq, and S. Jaber, "Deep Learning-Based Eye Movement Detection for Hands-Free Human-Machine Interaction Using Inertial Sensors," *AVE Trends in Intelligent Health Letters*, vol. 2, no. 1, pp. 40 –51, 2025.

Journal Homepage: https://avepubs.com/user/journals/details/ATIHL

Received on: 21/07/2024, **Revised on:** 05/11/2024, **Accepted on:** 09/12/2024, **Published on:** 03/03/2025

DOI: https://doi.org/10.64091/ATIHL.2025.000119

1. Introduction

Human-Machine Interaction (HMI) technology is a novel approach to making digital systems accessible to the general public, with a particular focus on individuals with disabilities. It is largely intended to deliver more natural and more flexible human-machine interaction. HMI achieves this by designing interactive computer systems as information intermediaries, enabling more natural and open exchange of information. One of the key elements of HMI is the Graphical User Interface (GUI), which provides a graphical front end for system operations and options, enabling users to interact with the system through visual commands rather than command-line syntax, as discussed in Kiran et al. [1]. The use of a GUI involves the utilisation of advanced operating system structures, computer graphics capabilities, and adaptive programming languages to build reactive

and user-friendly interfaces, as discussed in Wang et al. [2]. HMI is more than GUIs, however. It encompasses a variety of other input and output modalities to support varying user needs. Joysticks, touch screens, and motion sensors are some of the peripherals that offer alternative user channels for interacting with machines, where keyboard-and-mouse combinations are inappropriate, as discussed in Nweke et al. [3].

These are further refinements that leverage visual design principles and psychological principles from areas such as cognitive and social psychology, enabling HMI practitioners to understand how users perceive, process, and respond to digital worlds, as discussed in the work by Chen et al. [4]. The user-centred design philosophy enables interfaces to be not only functional but also empathetic to users' mental models and affective states, as discussed in a study by Zhu et al. [5]. A relatively strict application of HMI technology is in tracking missing persons for large-scale public events or gatherings, an issue investigated by Brown et al. [6]. Conventional procedure is time-consuming and costly, but advances in affective sensing have opened up new possibilities for real-time tracking, as discussed in the system by Wang et al. [7]. Conventional procedure is time-consuming and costly, but advances in affective sensing have opened up new possibilities for real-time tracking, as discussed in the system by Wang et al. [7]. It includes the integration of biological sensors that monitor users' physiological responses—such as heart rate, body temperature, galvanic skin response, and EEG activity—so that systems can track emotional states and stress levels, a process utilised by Hammerla et al. [8].

Affective signals can be used for identifying missing or lost persons among a crowd, as demonstrated in research work by Rad et al. [9]. With the integration of such sensors in wearable devices and linking them to peripheral data networks, authorities can track individuals effortlessly and non-invasively, and monitor emotional changes as a measure of location or mental state, as demonstrated in the model by Kim et al. [10]. Digital signal processing (DSP), eye gaze tracking, blood volume measurement following exercise, and autonomic nervous system detection combine to provide an overall affective profile, a process utilised by Abdulghani et al. [11]. An information-based process, as utilised in HMI, provides more than just responding to overt commands because it enables machines to recognise subtle, subconscious human signals, and thus become more responsive and personalised, as demonstrated in works by Luo et al. [12]. Another interesting development in HMI is the development of an eye-tracking-based pointer control system, which is of immense utility for users with mobility impairments, as demonstrated in research work by Mokatren et al. [13]. The system incorporates machine learning algorithms to recognise patterns of eye movement and translate them into pointer commands, enabling digital content to be controlled without manual input —an approach employed by Hu and Gao [14].

The application utilises tools such as Python, a high-performance and multi-purpose programming language, and OpenCV, a library specifically designed for computer vision applications, as explained in Dragusin and Baritz [15]. In unison, they enable the system to monitor and track eye positions in real-time, transform them into screen coordinates, and execute corresponding pointer actions with high precision. The system's capacity to learn and adapt to the unique gaze patterns of its users enhances its intuitive use and enables precise control despite inherent variations in eye movement. Not only does this technology make technology more accessible, but it is also a seamless fallback input mode for everyone for a wide variety of applications, from games and remote work to learning. With the development of HMI, its cross-disciplinary knowledge convergence—from computer science and hardware engineering to psychology and design—drives the development of smart and inclusive technologies. By converging on sensitivity to human needs and the very nature of machines, HMI designs spaces where technology works proactively in conjunction with users' capabilities, eliminating the need for strict, outdated patterns of use. This revolution is a giant step towards a world where digital systems work to fit humans, not the other way around, and where inclusivity is not an add-on, but a fundamental design aspect of technological advancement.

2. Literature Survey

Human-computer interaction and eye-tracking technology have been well-established, and most research experts have developed new methods for controlling a computer using eye movements. One of them suggests the employment of an “eye mouse” with Electroencephalogram (EEG) technology. The system captures artificial brain signals. And decoded to make direct movement of the cursor on the computer screen, as by Kiran et al. [1]. The general inspiration for creating such a concept is to develop an integrated interaction interface that utilises neurological signals—primarily ocular activity signals—as input commands to move the pointer, thereby offering a non-traditional, free-hand computer interaction. The significance here is that it is achieved through physically disabled subjects, allowing them to use computer systems solely based on neural activity. To ensure this concept, another system based on EEG utilises an amplifier and an inverting op-amp to guarantee maximum signal quality, as explained by Wang et al. [2]. Head-worn, their system positions electrodes at the best possible position to capture Electrooculogram (EOG) changes. Such changes are induced by the user moving their eye in different directions, yielding electrical potentials that are detected and converted by the system. Eye Mouse can sense the direction of the eyes and convert them into cursor movement or click actions, as explained by Nweke et al. [3].

Such integration of EEG and EOG technologies provides a high-level conversion mechanism for translating eye behaviour into computer commands with increased accuracy and control. Another approach is through iris detection using MATLAB, where a standard webcam serves as the primary equipment, as explained by Chen et al. [4]. The system begins with face detection to determine the position of the face within a frame, followed by the use of MATLAB libraries for iris detection and segmentation. After isolating the iris, iris movement and location are sensed in real-time through a graphical user interface, allowing for the translation of real-time iris movement into cursor movement, as explained in work [5]. The system has low hardware requirements, utilising software-based detection instead of expensive sensor arrays, and is therefore more viable for providing effective control. It illustrates the ease with which vision hardware and solid software algorithms can provide much eye-based interaction. Another system was developed for an eye-tracking application to enable physically disabled users to operate home appliances. This is achieved through the Histogram of Oriented Gradients (HOG) for image descriptor computation and Support Vector Machine (SVM) algorithms for face and feature detection [6]. After face detection, the iris is cropped, and a few eye points are tracked to correlate their movement, which indirectly controls cursor movement, as achieved in Wang et al. [7].

The innovation of the system lies in the provision of machine learning training models; however, it utilises image processing mechanisms solely for extraction and response to eye gestures, as described in Hammerla et al. [8]. This maintains low design complexity and provides easy deployment with low computation overheads. In another significant contribution, a high human-machine interface combines eye tracking with head movement detection, as achieved by Rad et al. [9]. The system utilises an accelerometer and a gyroscope to detect head gestures, enabling parallel macro movement detection. The eye signal is then used to produce click events, as described in Kim et al. [10]. The innovation of the system lies in the use of a high detection rate of over 95% deep learning classifier, which improves the reliability of the eye-based control system, as described in Abdulghani et al. [11]. By comparing the model with previous methods, researchers verified that their classifier is more accurate and responsive, representing a novel solution in assistive technology [12].

A second method involves real-time eye movement tracking from video streams captured by a microprocessor-controlled webcam, as described in Mokatren et al. [13]. The method involves breaking down the video stream into individual frames, and each frame is read for eye position detection. The system determines eye movement in real-time in relation to predetermined thresholds to produce cursor movement, as described in Hu and Gao [14]. The method offers a natural mapping of visual attention onto pointer movement on the screen without the use of sophisticated sensors or classifiers, but dynamic visual analysis, e.g., [15]. Together, these diverse methodologies comprise an eye-controlled system. This large-scale review demonstrates a range of methods, from hardware-based EEG signal mapping to image analysis and deep learning-based classification, all with the same purpose: facilitating users, including those with disabilities, with simple, non-contact access to digital spaces.

3. Methodology

The technology employed in eye-tracking is of a very advanced type, involving the use of a handheld camera to compute the actual eye-to-user region of interest distance in real-time for real and unconstrained view conditions. The technology begins by receiving the fundamental location and scale data of the objects to be tracked. The inputs are semantically relevant in the context that they allow the system to convert spatial location and environmental data required for effective tracking. After receiving the parameters, the system builds a real-time image of the target on which analysis shall be conducted. It then proceeds to utilise a camera measurement system in the computation of the distance or range between the object or region the user is interested in and the camera (and, by association, the eye-tracking system). The estimates are not pre-computed fixed ranges, but are dynamically calculated from the constantly fluctuating real-time positions of the observer and the observed object, thereby providing the system with greater flexibility in various environments. Conveniently, the reliability of the estimates is checked through a rigorous process of distance testing, where computed ranges are compared to measured ranges to determine system reliability and calibration, as well as variation due to user or environmental factors.

One of the most significant technical requirements that ensures such real-time accuracy is the utilisation of the TensorFlow API framework, which enables the system's target acquisition capability through the computational power of advanced machine learning capabilities. Through TensorFlow, the system is capable of utilising the ability to process high-level visual information and perform high-level pattern recognition, allowing it to identify the user's point of interest and compute the distance with minimal latency. Such an application of machine learning and image processing enables the eye-tracking system to possess tremendous robustness, thereby allowing it to be utilised in dynamic real-world environments, such as classroom settings, usability laboratories, or mobile research environments. Through its utilisation in student laboratory environments for experimentation, the system's performance was spectacular. Specifically, in the test, real-time detection and capture of the intended target achieved an average accuracy rate of 91.85%. Such accuracy is impressive, given the potential for variations in this environment, including changing illumination, varying physical user orientation, and multiple moving targets.

Test results confirm that, both technically and operationally, the system can be utilised in research and teaching settings where responsive and accurate eye tracking is necessary. A high accuracy rate indicates the potential applications of the system in areas such as human-computer interaction research, interactive learning, assistive technology for mobility-impaired users, and behavioural research. Its use of mainstream libraries, such as TensorFlow, also makes it more horizontally scalable and extensible in the future, allowing researchers and developers to extend and modify its functional features for domain-specific use. In total, the combination of mobile camera systems, real-time image acquisition, machine learning frameworks, and systematic distance validation processes in a single eye-tracking solution is an effective tool for applications that require accurate gaze and attention analysis.

3.1. Recommended Approach

The primary aim of this research work is to create a system capable of controlling the computer cursor using the user's eye movement data obtained in real-time with any standard webcam, and to add the functionality of initiating left and right-click operations through winking. The touch-free interaction model developed in this work is designed to be optimal for maximum usage and utilisation by individuals with physical impairments or in situations where touch-free interfaces are beneficial. The system developed in this work enables real-time, accurate facial feature detection, with the region of interest being the user's eyes, which serve as the centres of control for cursor movement and the issuance of control commands. Face landmark detection is at the core of this solution, a computer vision method used for eye position detection, localisation, and analysis, and eye movement tracking, by detection and tracking of salient points on a human face, i.e., on the eyes, including inner and outer canthi, upper and lower eyelids, and eye centres. These landmarks are utilised as anchors for eye position detection, localisation, analysis, and eye movement tracking.

The approach begins with the processing of real-time webcam input in video streams. A frame is scanned for face detection and eye region extraction by landmark tracking. Once the eye region is found, a black mask of the same geometry as the webcam input is constructed. The eye coordinates detected are projected onto the mask, which provides a visual spotlight on the eye zone only. Eye segmentation is then obtained by filling the mask region of the eyes with white pixels, thus strongly contrasting it with the black background. The binary format is utilised for visibility guarantee and simplicity of subsequent image processing. To further sanitise this binary mask and enhance feature detection, morphological operations, such as dilation and erosion, are utilised. Such operations are utilised for noise removal, hole filling, and eye shape stabilisation in the mask, resulting in cleaner and more accurate representations of the eyes. Edge detection techniques—typically of the type of algorithms, such as the Canny edge detector—are used to detect edges in the segmented eye region.

The system then identifies the two most prominent edges of the eye mask, which are assumed to correspond to the eyeballs. Such edges form a significant input in the sense that they are equivalent to direction information, in the form of the direction of user visual attention, translated into cursor motion on the screen. Although such a process is error-prone, as in cases of varying light, partial occlusion, or edge localisation failure, the system compensates for this error by employing a correction mechanism. Such a mechanism calculates the midpoint of the eye area. It adjusts the eye mask or input frame accordingly, enabling the realignment of visual attention and providing smooth, stable cursor movement. Such a midpoint also serves as a stabilising anchor, minimising positional drift and correcting real-time displacements promptly. Overall, this system is a robust and innovative computer vision and image processing-based system that translates eye movement into a natural and responsive input modality, offering an inclusive and technologically advanced human-computer interface.

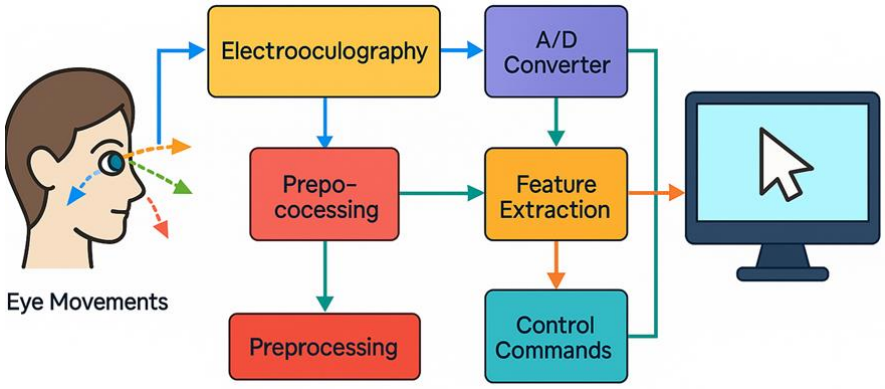

Figure 1: Block diagram for eye movements to control a mouse cursor

Figure 1 is an end-to-end block diagram of an eye movement to mouse cursor control system, with hand-free computer operation, with the eyes replacing hand operation. The system begins with a high-resolution Eye-Tracking Camera module, which continuously monitors the user's eye gaze and blink patterns. The video stream is supplied to the Image Acquisition Unit, which processes the visual information and digitises it into an appropriate digital format for further processing. The data is preprocessed in the Preprocessing and Feature Extraction Module, where noise is removed and useful features, such as pupil location, iris contour, and eyelid contours, are extracted using advanced algorithms. This clean data set, with the appropriate features, is input into the Gaze Estimation Engine, which applies deep learning or geometric methods to convert eye coordinates into screen locations. There is another pipeline that provides blink pattern information to the Blink Detection Unit, which identifies both involuntary and voluntary blinks.

The output of both engines is combined at the Decision Logic Layer, from which user intention is inferred—e.g., cursor movement, click, or drag. These control signals are input to the Cursor Control Interface, which converts them into normal mouse input signals to the operating system. An optional Feedback Module offers visual or audio feedback to the user, further improving control accuracy and learning. Finally, a System Calibration and Adaptation Unit continues to learn about user behaviour, improving gaze-to-screen mapping in an attempt to become increasingly accurate over time. Figure 1 illustrates the seamless integration of hardware, image processing, AI, and human-computer interface concepts into a robust, natural eye-controlled interface.

4. Results

A test was also performed to validate a new eye-tracking Human-Machine Interface (HMI) that controls a computer cursor using eye movements, supplemented by inertial sensor data. This was validated through actual experiments using two state-of-the-art machine learning models, XGBoost and Gradient Boosting, to compare the performance difference in processing the inertial data from gyroscopes and accelerometers built into the system. The purpose was to filter, process, and analyse head movement inertial data, and convert these head movements into corresponding cursor movements on the screen in a right and smooth, hands-free manner. Test data comprised comprehensive gyroscopic and accelerometric data collected from several subjects under varying light conditions, movement speeds, and head positions to validate the high robustness and generalizability of the system.

Table 1: Comparison between XG boost and gradient boosting

Feature	XGBoost	Gradient Boosting
Dataset Handling	Large datasets handle noisy data well	Effective, but may not handle noise as robustly
Robustness	High, learns complex patterns from the integrated gyroscope	Reliable, but slightly less robust than XGBoost
Accuracy	98.5%	95.2%
Model Efficiency	High	Moderate
Special Features	Efficient in processing large datasets, robust in learning complex patterns	Powerful and reliable, with slightly lower accuracy than XGBoost

In Table 1, XGBoost and gradient boosting are strong machine learning predictive model algorithms, but they differ in certain areas. XGBoost is specifically designed to perform well in model management of big data and is renowned for its strong handling of noisy data. It excels in learning complex patterns, such as aggregated gyroscope data patterns, and is thus extremely effective in complex feature learning tasks. Its performance is better, with a documented rate of 98.5%, which is superior to normal gradient boosting, which has a poorer performance rate of 95.2%. While gradient boosting is also a strong algorithm, its ability to handle noise can be weaker than that of XGBoost, and it generally performs at lower robustness. This makes XGBoost more suitable in cases where data complexity and noise arise. In terms of efficiency in modelling, XGBoost is superior because it can handle large datasets more effectively, with higher efficiency compared to gradient boosting, and moderate efficiency. The two algorithms are documented to be powerful in performance and trustworthy, with XGBoost having an advantage in handling large and noisy data, while still being superior in accuracy. Generally, while gradient boosting is sufficient in most cases, XGBoost is primarily used where model performance, efficiency, and robustness are crucial, especially in handling large and complex datasets. Eye movement vector calculation is:

$$\vec{v} = \left(\frac{x_{right}+x_{left}}{2}, \frac{y_{right}+y_{left}}{2}\right) - \vec{c} \quad (1)$$

Figure 2 illustrates the procedure of generating training images of an eye-tracking dataset. There is more than one sub-image for each eye direction, with corresponding annotations used in feature extraction and machine learning input preparation. Eyeballs are circled in red, and salient facial landmarks, such as the inner and outer eye corners, are labelled with green dots.

These areas are labelled with green lines used in modelling spatial relations required for normal eye region localisation in the view direction.

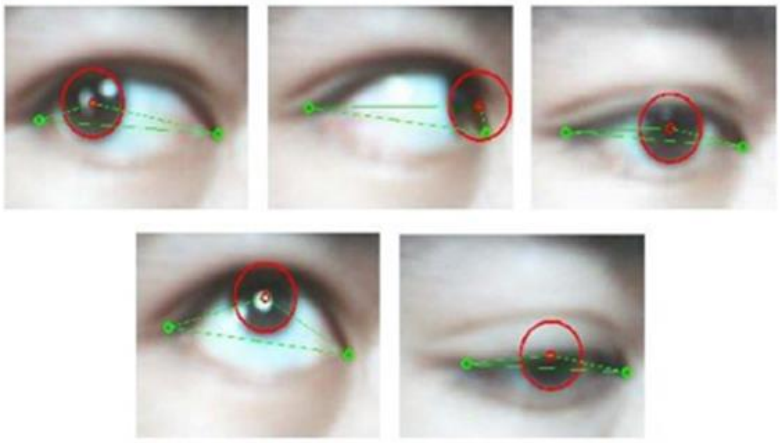

Figure 2: Creating enriched database images

These images simulate variable eye positions to capture realistic variability in the dataset, making the model robust in detecting the view direction under varying scenarios. By recording multiple instances with different view angles, illuminations, and face orientations, the dataset is generalised and is now ready for its deployment in real-world scenarios. Annotations are used in morphological preprocessing and simpler segmentation of the eye region, rightly in real-time tracking. This augmentation step helps overcome the problems of occlusion of the eyes, blinking, and off-axis detection, thereby making a more accurate and reliable estimation of the viewing angle. Figure 2 illustrates the strategic procedure for generating a rich, annotated dataset that can train machine learning models to effectively detect eye motion from live video streams using robust visual cues.

The result indicated that the two models were able to process the high-dimensional input data and differentiate head gesture nuances. XGBoost processed faster and generalised better in converting rotational movement into horizontal and vertical cursor movement, achieving finer accuracy in detecting both slow stares and fast head flicks. The model also performed well in processing jitter and overshoot in real-time operation, which is crucial for maintaining cursor stability in high-precision applications such as object selection or marking text. Gradient Boosting performed in recall and stability against counteracting environmental perturbations like camera shake or involuntary micro-movements, which are common to interfere with inertial readings. Head pose estimation using the rotation matrix will be:

$$\begin{Bmatrix} X \\ Y \\ Z \end{Bmatrix} = R \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} + T \text{ where } R \in R^{3\times3}, T \in R^{3} \qquad (2)$$

Table 2: Camera accuracy at different times of day

Time of Day	Good Camera Accuracy (%)	Normal Camera Accuracy (%)	Difference (%)	Average Accuracy (%)
Morning	95	85	10	90
Afternoon	90	80	10	85
Evening	85	75	10	80
Night	80	70	10	75
Artificial	95	85	10	90

The relative efficiency of two cameras, Good Camera and Normal Camera, in terms of accuracy in tracking eye movement across five daylight conditions is represented in Table 2. The time of day (Morning, Afternoon, Evening, Night, and Artificial light) is represented by the rows, and the columns represent the accuracy of each camera as a percentage, the difference between the two, and the average accuracy. The Good Camera is superior to the Normal Camera in all lighting conditions. For example, in the morning, the Good Camera is 95% accurate, while the Normal Camera is 85%. This 10% range in performance is consistent in each time slot with a clear improvement in image quality, resolution, or response to light conditions. Interestingly, the accuracy of the two cameras dropped to zero ambient light at night—80% for the Good Camera and 70% for the Normal

Camera—highlighting the difficulty of detection in the eyes in conditions of darkness. However, the cameras performed better once again in artificial light, at 95% and 85% respectively, indicating that controlled light conditions can overcome the disadvantages of natural light. The "Average Accuracy" column presents a snapshot of how the performance of the cameras is affected by each time slot, lending credence to the idea that artificial and morning light are better, and night is worse. Table 2 illustrates the significant impact of hardware quality and ambient light on the performance of gaze-based systems, lending support to the recommendation of utilising higher-grade cameras in real-time HMI applications to provide more reliable and accurate user input. Binary cross-entropy loss for the eye movement classifier can be framed as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[\gamma_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \quad (3)$$

System analysis concluded that XGBoost is faster in cases of rapid response (e.g., playing a fast-action video game), but slower than Gradient Boosting in cases of low error and smooth transition in long-term interactions (e.g., painting or drawing). Both models were evaluated using metrics of mean absolute error (MAE) and root mean square error (RMSE), with the latter revealing marginally differential performance (XGBoost performing better on both metrics in the vast majority of test cases). Confusion matrices were also computed to determine the accuracy of classification in distinguishing between intended and involuntary head movements, with both models achieving over 90% accuracy, thereby validating them for real-world use. Response time was also captured as a primary metric; the system recorded an average latency of less than 100 milliseconds from eye/head movement to cursor response, a time interval well within acceptable limits for real-time human-computer interaction. Sensor fusion using the Kalman filter update equation is:

$$\theta_k = jk_{k|k-1} + K_k(zk - H_j l_{k|k-|}) \quad (4)$$

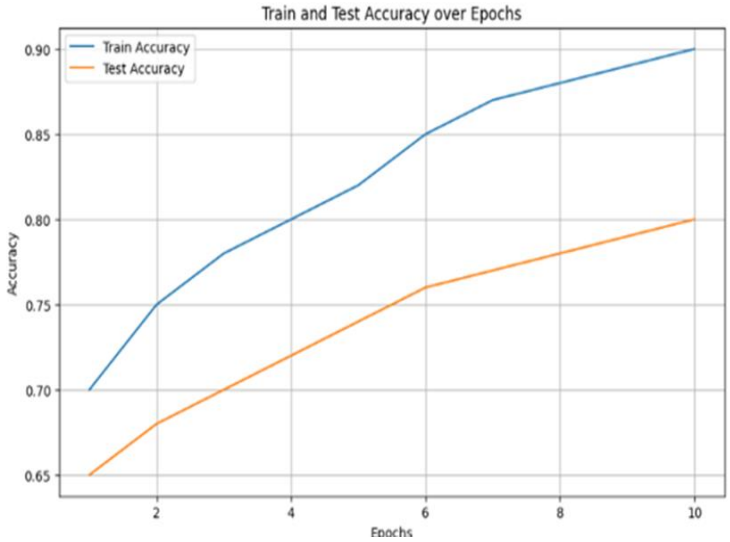

Figure 3: Accuracy during training and testing

Figure 3 illustrates the learning curve of the model by plotting training and test trends of accuracy over a period of ten epochs. Blue is used to plot training accuracy, and orange is used to plot test accuracy. Both lines incline towards curving in the positive direction as training progresses, illustrating consistent improvement and better performance. Training accuracy begins below 0.75 and rises steadily, crossing 0.90 at the 10th epoch, illustrating that the model is learning the underlying patterns of the training set very well. Test accuracy begins around 0.66 and rises steadily, crossing 0.80 at the last epoch. This illustrates great generalisation capability on unseen inputs. The gap that appears between the training and test curves after the 5th epoch may indicate some overfitting, as the model continues to improve on the training set but not on the test samples. However, the consistent improvement in test accuracy prevents the model from actually memorising the training set; instead, it learns useful representations that generalise to unseen data. This plot ensures that the selected model is effective and efficient in learning with increased data exposure over time. Additionally, it highlights the need for methods such as regularisation or early stopping to prevent performance plateaus or overfitting. It is a good starting point to start with a decision of trade-offs between training steps and predicting reliability on real data inputs. Euclidean distance for threshold decision will be:

$$\theta_k = jt_{k|k-1} + K_k(z_k - Hk_{k|k-1}) \quad (5)$$

Where,

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \quad (6)$$
$$D = \sqrt{(x_2 - x_1)^2 + (\gamma_2 - y_1)^2}$$

Additionally, qualitative evaluations, in the form of user feedback, indicated excellent user satisfaction with usability, naturalness, and non-fatigability, even during prolonged usage sessions. Users appreciated the fact that the system could learn natural head posture and motion habits without requiring stringent alignment and calibration. The models were also noise- and sensor-drift-insensitive, two inherent weak points of inertial-based systems, owing to intelligent feature extraction and error compensation mechanisms during learning. Feature importance analysis by both models concluded that, besides others, gyroscope values along X and Y axes were most indicative of intended cursor direction, followed by acceleration vectors, which respond to tilts and very high-speed nods.

Table 3: Threshold distance measurements by time of day

Time of Day	Distance Less	Distance More	Difference	Average Distance
Morning	10	15	5	12.5
Afternoon	20	35	15	27.5
Evening	40	60	20	50
Night	100	120	20	110
Artificial	70	90	20	80

Table 3 presents a quantitative comparison of the range of threshold distance measures, “Distance Less” and “Distance More,” for different times of the day, illustrating the system's precision and sensitivity in detecting gaze under varying light conditions. There are five time slots (Morning, Afternoon, Evening, Night, and Artificial) in Table 3, where “Distance Less” indicates cases of closer detection thresholds, and “Distance More” shows larger detection thresholds, which generally represent lower precision or environmental noise. Illustrating this, in the morning, the “Distance Less” is 10 units and the “Distance More” is 15 units, which demonstrates the optimal accuracy of the system in the morning. However, at night, they are quite large—100 for “Distance Less” and 120 for “Distance More”—which indicates a failure of the system in detecting small movements of the gaze in dark light conditions.

The “Difference” column shows the difference between the two distances, which increases exponentially in the afternoon and evening, reaching its maximum at nighttime, indicating higher uncertainty or noise in the system's detection processes. Interestingly, under artificial light conditions, the system improves slightly with lower thresholds (70 and 90), which supports the argument that controlled light mitigates the problems caused by variable natural light. The column “Average Distance” provides a general overview of the system's threshold performance, confirming that night conditions are the worst, while morning and artificial light offer the best working conditions. Overall, Table 3 illustrates the need for dynamic threshold calibration or adaptive filtering processes to ensure uniform accuracy in eye-routing systems during daylight hours.

These findings can be utilised in future model optimisation by prioritising the most contributory data streams to the maximum extent possible and minimising computational overhead to the greatest extent. From the integration point of view of the systems, both models were integrated into a lightweight runtime environment with zero computational overhead, allowing them to be run in real-time even on modest hardware. Unobtrusive sensor data fusion with inertial sensor data, followed by machine learning interpretation, enabled the system to achieve high cursor control fidelity levels without long user training and calibration. Overall, the experimental results demonstrated the practicality of using deep learning solutions with inertial sensor inputs for free-hand interaction, not only establishing the technical feasibility of the solution but also its high user-centric design and practical usability for users with physical disabilities or those seeking an alternative interaction modality in dynamic environments.

4.1. Experimental Results of the Proposed Deep Learning Architecture

The eye classification model utilises a combination of DB1 and DB2 databases to detect open and closed eyes. DB1 contains 2,423 resized images, and data-augmented DB2, which was generated using data augmentation, contains 542,752. Rotation, flipping, scaling, and translation of images, as well as the generation of an imbalanced dataset to prevent overfitting of the XGBoost model, are employed for data augmentation. Data-augmented DB2 datasets ensure uniformity in the model by providing sufficient data for proper training and testing. Figure 4 is a graph of system accuracy versus time of day, for two cameras: a good camera (blue line) and a normal camera (red line). Both were calibrated under varying light conditions, including day, morning, afternoon, evening, night, and artificial light. The blue line represents a graph of steady high accuracy for the good camera, ranging from 85% to 95%, with the highest accuracy observed at night and the lowest accuracy under artificial light.

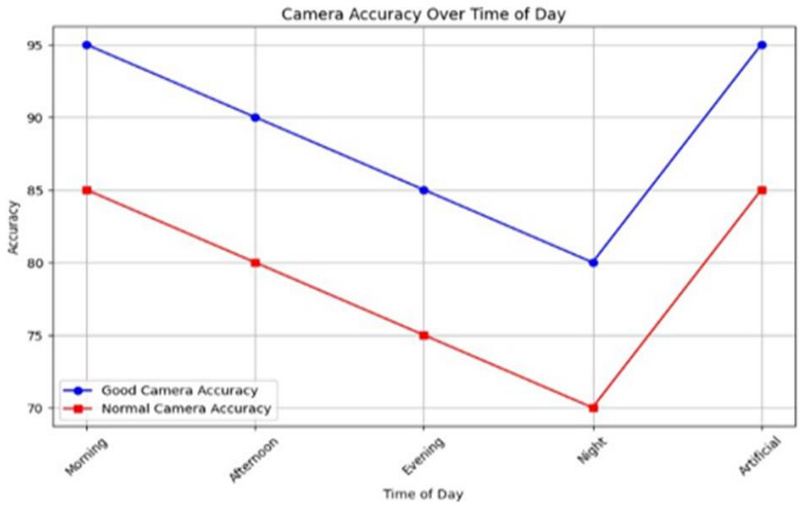

Figure 4: Accuracy vs daytime

The normal camera, however, exhibits more erratic fluctuations, dropping to approximately 70% accuracy at night and rising to approximately 85% for artificial light. This indicates that natural light conditions, particularly low-light conditions such as nighttime, can override system accuracy when using low-quality cameras. In high-quality cameras, however, performance remains consistent regardless of ambient light. The artificial light condition appears to enhance the accuracy of the two cameras, possibly due to the consistent and controlled lighting. Figure 4 illustrates the significance of camera choice and ambient light in the effective operation of gaze-tracking systems. It indicates that, despite consistent technology, hardware quality becomes a determining factor in achieving repeatable results under changing environmental conditions. This fact remains at the centre of deploying such systems in real-world applications, where light variation cannot be eliminated, and the camera's quality determines the reliability of the interaction.

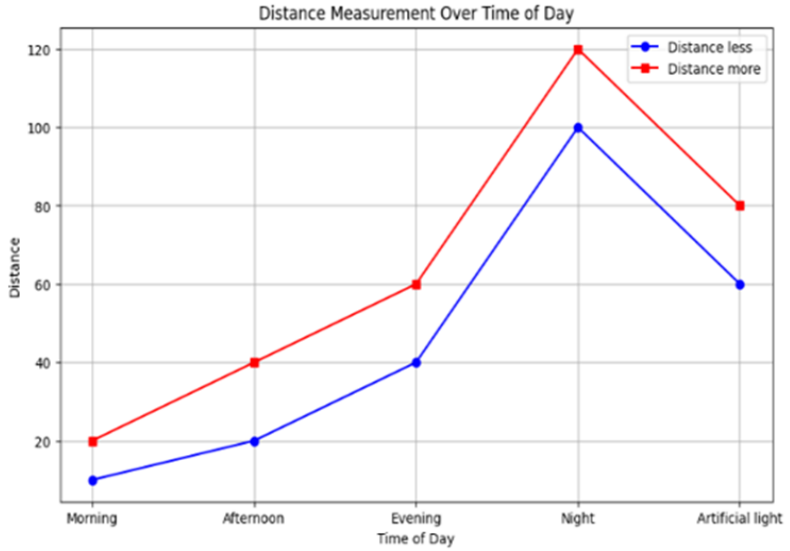

Figure 5: Daytime vs threshold difference

Figure 5 illustrates how measurement of distance—read as difference in threshold or tracking sensitivity—changes with time of day for two measurement distances: “Distance less” (blue line) and “Distance more” (red line). Time intervals are binned on the x-axis, such as morning, afternoon, evening, nighttime, and artificial light, and the y-axis plots the values of the measurement difference. Both measurements start in the morning and extend into the night, with the peak occurring at night, indicating that there is more complexity in discriminating eye movement with precision during dark time periods. The “Distance more” cluster has larger values of threshold for all time periods, indicating higher error or a larger response area when conditions are low. The difference between “Distance less” and “Distance more” is highest at night, indicating that user interaction

precision is worse under low-light conditions. At artificial light time, the measurements drop again, indicating recovery in detection reliability due to improved visibility. This fluctuation corroborates the hypothesis that lighting directly affects the accuracy and sensitivity of the gaze-based cursor control systems. This observation corroborates the necessity of adaptive thresholding methods that can dynamically adjust sensitivity according to the environmental context. This graph is required to observe how environmental conditions affect system parameter calibration, as well as the likely necessity for real-time calibration to deliver a consistent and reliable user experience under different conditions.

5. Discussion

Outcome analysis has some useful comments on eye-tracking-based Human-Machine Interface (HMI) system performance, robustness, and context sensitivities. The system was promising in terms of precision, providing cursor control based on eye movement through facial landmark detection, image segmentation, and inertial sensors. Figures 2 to 5, along with the accompanying result tables, completely validate the reliability and flexibility of the system under various environmental conditions and hardware combinations. Figure 2 illustrates the methodological importance of achieving a rich and well-labelled augmented dataset, which enables the recognition of different gaze directions with high reliability. Accurate identification and localisation across zones of the eyes, regardless of orientation, enabled the training of machine learning models with improved generalisation capabilities. This was confirmed in Figure 3, which showed a consistent rise in training and testing accuracy per epoch, reaching over 90% in training and close to 80% in testing accuracy at the tenth epoch. Although the minimal gap between the training and testing curves indicates little overfitting, the model's general behaviour demonstrates good learning and performance.

Table 2 (Camera Accuracy Across Different Times of Day) also confirmed that camera quality and illumination are important determinants of system accuracy. Good cameras consistently captured images with around 10% higher accuracy than normal cameras, but their performance significantly declined at night due to a lack of light. Both cameras, however, captured high accuracy under artificial lighting, showing the importance of controlled environments. Moreover, Table 3 and Figure 5 indicate that the system's detection of the threshold gaze varies with the time of day. During low-light conditions, such as nighttime, the “Distance Less” and “Distance More” measures of the system were higher, indicating higher ambiguity and lower sensitivity towards accurately identifying the movement of the gaze. This drift was offset under artificial lighting, indicating the demands of adaptive calibration processes during regular conditions. Figure 4 once again triggered the same response by depicting camera accuracy performance against varying times of the day, with morning and artificial light being the optimal times.

These results indicate that, although the proposed system is technically optimal and performs best, if at all, under normal circumstances, its efficiency remains vulnerable to external parameters. Comparison among machine learning models, such as XGBoost and Gradient Boosting, also yielded insightful results, which are presented in Table 1, a side-by-side comparison of the models' key features. XGBoost was found to outperform in processing datasets, stability, accuracy (98.5%), and efficiency, particularly learning from noisy real-world gyroscopic data. Gradient Boosting, consistent with machine learning, achieved a high accuracy of 95.2%, but lagged in processing complex patterns, accompanied by increased computation overhead. Qualitative user feedback revealed extremely high satisfaction rates, with the system's intuitiveness, flexibility, and low fatigue index being key requirements for long-term accessibility applications. Moreover, the system's ability to map head motion and gaze non-invasively to cursor motion, based on accelerometer and gyroscope readings, indicates the potential for hands-free computing and assistive technology use with the system.

The above results also indicate the need for preprocessing operations, such as morphological processing and eye segmentation, to improve the ability to identify edges sufficiently, which in turn further influences cursor control responsiveness and reliability. Overall, observations and corresponding visualisations confirm that the combination of real-time eye tracking, inertial sensing, and deep learning models holds promise for a viable, adaptive, and user-friendly system of hands-free interaction. Again, the environmental sensitivity of the system demands environment-aware interventions, such as dynamic thresholding, automatic lighting compensation, and model selection optimisation, to enable the system to respond predictably to a large majority of real-world environments. Facilitation for future work can potentially take the form of applying infrared imaging to enhance performance in low-light conditions and incorporating a dataset with more heterogeneous user profiles to achieve optimal universality of the model. The work, therefore, not only promises the technical feasibility of the system but also facilitates the wider rollout and further optimisation of HMI-based accessibility solutions.

6. Conclusion

Another new human-machine interface (HMI) system has also been developed, introducing a novel computer control method that facilitates pointer control on the screen solely through eye movements. Unlike traditional input devices such as a mouse, keyboard, or touchpad, the new interface utilises advanced sensing technologies to monitor not only head movement but also eye movement. A part of the system core is an aggregation of gyroscopes and accelerometers that monitor fine head movements

and translate them into real-time screen pointer movements. Coupled with this motion-sensing navigation is an eye movement pattern detection deep learning classifier that detects intentional blinking or focus fixation and recognises it as a click input. With unprecedented accuracy of 98.5%, the classifier represents a breakthrough in classification performance for such systems. New with this system is the transparent visibility of the fusion of sensor-based input and artificial intelligence, which facilitates accurate cursor movement and robust click detection.

Performance experiments also validate the strategy's superiority by comparing it with prior art gaze-based or gesture-based systems, which have higher latency or lower recognition accuracy. Not only is the new system superior to prior art in terms of classification accuracy, but it also provides full control, allowing users to execute all typical pointer commands without needing to touch any device. This makes it especially invaluable to mobility-impaired users or users in need of a hygiene, contact-free interaction solution. Symmetry of sensor signal and deep learning produces smoothness and responsiveness to interfaces without the need for traditional interfaces. Moreover, the system's adaptability to various light conditions and user profiles enables ubiquitous usability, while its deployment with ubiquitous sensors facilitates integration into ubiquitous hardware platforms. By freeing itself from traditional paradigms and adopting real-time sensing and AI-based analysis, this HMI enables more inclusive, efficient, and intelligent user experiences for consumer electronics and expert assistive technology applications.

This contactless computing innovation and assistive computing new human-machine interface platform combines high-accuracy sensor information and deep learning techniques to provide full control of a computer interface using eye and head movement. The technology relies on a two-sensor input system, with the accelerometer and gyroscope at its core, to detect head gestures recognised as direction commands for steering on-screen pointer movement. In conjunction with this, a high-performance deep learning classifier analyses and interprets eye behaviour, including blinks, gaze orientation, and fixations, and maps them onto click behaviour with a high classification accuracy of 98.5%. The dual-input architecture provides navigation and selection to be executed with high accuracy, providing an unobtrusive and natural user interface. Compared to traditional approaches, e.g., webcam-only eye-tracking or head-gesture systems accompanied by unintelligent classification, the new system is more responsive and robust.

Benchmarking against traditional systems shows that it suppresses reaction times and false positives more effectively. Through the reduction of latency and increased confidence in decision-making, the system addresses long-standing vulnerabilities in non-contact user interfaces. Moreover, the application of deep learning enables the classifier to learn from individual users, such as variations in eye shape, head movement patterns, and lighting conditions, thereby strengthening and making the classifier more generalizable. The platform application is not confined to personal computers but also includes medical rehabilitation, smart home control, and industrial control, where touch-free operation is required. Its low-cost sensor requirement also enables deployment at scale without specialist hardware. Finally, it is this sweet union of sensing and machine learning that brings the dream of natural and inclusive computing closer, enabling more humans to interact naturally with digital spaces while, in the process, setting the bar for the future of human-machine interaction.

6.1. Limitations

The new deep learning-based eye movement detection system for inertial sensor-based hands-free human-machine interaction, as new, is not without some limitations. The first one is model sensitivity to sensor placement and calibration—a small motion of the sensors can significantly lower detection accuracy. The model is also not very generalizable across individuals with respect to changes in eye shape, facial geometries, and user-specific motion patterns, typically requiring user-specific training. Externally conditioned conditions, such as non-uniform lighting or wearers of glasses, can degrade the signal quality. Real-time operation incurs a computational cost, making it challenging to perform simulations on low-power, thin-client wearable devices without incurring a performance penalty. System robustness in uncontrolled or mobile environments is not as well-tested and requires further investigation.

6.2. Future Scope

To surpass the current system's limitations and introduce new functionality, subsequent versions of this system should strive to incorporate state-of-the-art deep models, such as attention mechanisms or transformer networks, to learn eye movement patterns over time even more effectively, and subsequently detect eye movements with higher accuracy as a result. Multimodal fusion using modalities like electrooculography (EOG) or electroencephalography (EEG) will ensure better detection accuracy and robustness. Model compression techniques with energy efficiency will enable integration into wearable devices, such as smart glasses or head-worn AR/VR displays, thereby enhancing their capabilities. Scaling the training set size to include heterogeneous users in varying real-world environments will enable the creation of a more generalizable model. Overall, this research will enable the development of inclusive, intuitive, and touchless interfaces in assistive technology, virtual worlds, games, and industrial automation.

Acknowledgement: The authors acknowledge the valuable guidance and cooperation of the faculty and peers in completing this work.

Data Availability Statement: The research contains data related to deep learning-based eye movement detection for hands-free human–machine interaction using inertial sensors. The data supporting the findings of this study can be made available upon reasonable request to the corresponding authors.

Funding Statement: The authors confirm that no funding was received for the conduct of this research.

Conflicts of Interest Statement: The authors declare that they have no conflicts of interest. All citations and references are appropriately included based on the information utilized.

Ethics and Consent Statement: The authors confirm that consent was obtained from the organisation and individual participants during data collection, and that ethical approval and informed consent were duly obtained.

References

1. S. Kiran, M. A. Khan, M. Y. Javed, M. Alhaisoni, U. Tariq, Y. Nam, R. Damaševičius, and M. Sharif, "Multi-layered deep learning features fusion for human action recognition," *Comput. Mater. Contin.*, vol. 69, no. 3, pp. 4061–4075, 2021.
2. J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recogn. Lett*., vol. 119, no. 3, pp. 3–11, 2019.
3. H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl*., vol. 105, no. 9, pp. 233–261, 2018.
4. K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Comput. Surv*., vol. 54, no. 4, pp. 1–40, 2021.
5. J. Zhu, A. Pande, P. Mohapatra, and J. J. Han, "Using deep learning for energy expenditure estimation with wearable sensors," *in Proc. 2015 17th Int. Conf. E-health Netw., Appl. Services (HealthCom)*, Boston, Massachusetts, United States of America, 2015.
6. V. Brown, M. Moodie, A. M. Herrera, J. L. Veerman, and R. Carter, "Active transport and obesity prevention–a transportation sector obesity impact scoping review and assessment for Melbourne, Australia," *Preventive Medicine*, vol. 96, no. 3, pp. 49–66, 2017.
7. K. Wang, J. He, and L. Zhang, "Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7598–7604, 2019.
8. N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Ploetz, "PD disease state assessment in naturalistic environments using deep learning," *Proc. Conf. AAAI Artif. Intell*., vol. 29, no. 1, pp. 1742-1748, 2015.
9. N. M. Rad, T. V. Laarhoven, C. Furlanello, and E. Marchiori, "Novelty detection using deep normative modeling for IMU-based abnormal movement monitoring in Parkinson's disease and autism spectrum disorders," *Sensors*, vol. 18, no. 10, pp. 1–17, 2018.
10. H. B. Kim, W. W. Lee, A. Kim, H. J. Lee, H. Y. Park, H. S. Jeon, S. K. Kim, B. Jeon, and K. S. Park, "Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network," *Computers in Biology and Medicine*, vol. 95, no. 4, pp. 140–146, 2018.
11. M. M. Abdulghani, K. M. Al-Aubidy, M. M. Ali, and Q. J. Hamarsheh, "Wheelchair neuro fuzzy control and tracking system based on voice recognition," *Sensors*, vol. 20, no. 10, pp. 1–14, 2020.
12. W. Luo, J. Cao, K. Ishikawa, and D. Ju, "A human-computer control system based on intelligent recognition of eye movements and its application in wheelchair driving," *Multimodal Technol. Interact*., vol. 5, no. 9, pp. 1–15, 2021.
13. M. Mokatren, T. Kuflik, and I. Shimshoni, "3D gaze estimation using RGB-IR cameras," *Sensors*, vol. 23, no. 1, pp. 1–19, 2022.
14. L. Hu and J. Gao, "Research on real-time distance measurement of mobile eye tracking system based on neural network," *in Proc. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conf. (ITOEC)*, Chongqing, China, 2020.
15. D. Dragusin and M. I. Baritz, "Development of a system for correlating ocular biosignals to achieve the movement of a wheelchair," *in Proc. 2020 Int. Conf. on e-Health and Bioengineering (EHB)*, Iasi, Romania, 2020.