

## ABSTRACT

Eye tracking technology plays a critical role in assistive applications by enabling gaze-based interaction for individuals with motor impairments. It allows users to control digital interfaces using only their eye movements, promoting hands-free interaction and digital accessibility. Traditional gaze interaction systems typically rely on fixation-based mechanisms, which require frequent recalibration to maintain accuracy. This process can be cumbersome and fatiguing for users, especially those with limited motor control or visual instability.

Smooth pursuit offers a compelling alternative for gaze-based input. By allowing users to follow moving targets naturally and continuously, it enables calibration-free and intuitive interaction. This modality is particularly well-suited for users with motor impairments, as it reduces the reliance on sustained gaze fixation and aligns with natural ocular behavior. However, smooth pursuit rarely occurs in isolation, as eye movement data in practical settings often include fixations, saccades, and blinks. As a result, accurate eye movement classification is necessary to reliably detect smooth pursuit and support robust gaze-based interaction.

This study utilizes GazeCom and HMR datasets to address limitations in smooth pursuit classification. Recent hybrid deep learning models report F1 scores of up to 82.89% on the GazeCom dataset but require long temporal input windows of 1500 milliseconds, increasing model complexity and limiting their applicability in real-time or embedded systems.

To address these limitations, this work proposes a lightweight hybrid deep learning model that integrates one-dimensional convolutional neural networks (1D-CNN) with multi-head self-attention. The architecture is specifically adapted for online applications and does not require access to future time steps, allowing low-latency inference in real-time scenarios. It operates on a compact temporal window of 20 milliseconds—corresponding to five samples in GazeCom sampled at 250 Hz and four samples in HMR sampled at 200 Hz—thereby significantly reducing memory requirements and computational load.

The model is evaluated using stratified 5-fold cross-validation. It achieves F1 scores of 91.89% and 95.51% for smooth pursuit classification on GazeCom and HMR, respectively, while also maintaining strong performance across other event types such as fixation, saccade, and blink detection.

These findings demonstrate that accurate and efficient eye movement classification can be achieved using minimal temporal input. The proposed architecture offers a promising solution for real-time, resource-constrained gaze-based interaction, particularly supporting the development of practical applications for individuals with motor impairments in assistive technology and human-computer interaction.

**Keywords**—Eye tracking, 1D-CNN, multi-head self-attention, hybrid deep learning model

## INTISARI

Teknologi *eye tracking* memainkan peran penting dalam aplikasi assistif dengan memungkinkan interaksi berbasis pandangan mata untuk individu dengan gangguan motorik. Teknologi ini memungkinkan pengguna mengontrol antarmuka digital hanya menggunakan gerakan mata mereka, mendorong interaksi bebas tangan dan aksesibilitas digital. Sistem interaksi pandangan mata tradisional biasanya mengandalkan mekanisme berbasis *fixation*, yang memerlukan recalibrasi berkala untuk mempertahankan akurasi. Proses ini dapat merepotkan dan melelahkan bagi pengguna, terutama mereka yang memiliki kontrol motorik terbatas atau ketidakstabilan visual.

*Smooth pursuit* menawarkan alternatif yang menarik untuk input berbasis pandangan mata. Dengan memungkinkan pengguna mengikuti target bergerak secara alami dan berkelanjutan, teknologi ini memungkinkan interaksi yang bebas kalibrasi dan intuitif. Modalitas ini sangat cocok untuk pengguna dengan gangguan motorik, karena mengurangi ketergantungan pada *fixation* yang berkelanjutan dan selaras dengan perilaku okular alami. Namun, *smooth pursuit* jarang terjadi secara terisolasi, karena data *eye movement* dalam pengaturan praktis sering mencakup *fixation*, *saccade*, dan *blink*. Akibatnya, klasifikasi *eye movement* yang akurat diperlukan untuk mendeteksi *smooth pursuit* secara andal dan mendukung interaksi berbasis pandangan mata yang robust.

Penelitian ini menggunakan dataset GazeCom dan HMR untuk mengatasi keterbatasan dalam klasifikasi *smooth pursuit*. Model *hybrid deep learning* terbaru melaporkan *F1 score* hingga 82,89% pada dataset GazeCom namun memerlukan jendela input temporal yang panjang yaitu 1500 milidetik, meningkatkan kompleksitas model dan membatasi aplikabilitasnya dalam sistem *real-time* atau *embedded*.

Untuk mengatasi keterbatasan ini, penelitian ini mengusulkan model *hybrid deep learning* yang ringan yang mengintegrasikan *one-dimensional convolutional neural networks* (1D-CNN) dengan *multi-head self-attention*. Arsitektur ini secara khusus diadaptasi untuk aplikasi *online* dan tidak memerlukan akses ke langkah waktu masa depan, memungkinkan inferensi latensi rendah dalam skenario *real-time*. Model ini beroperasi pada jendela temporal yang kompak yaitu 20 milidetik—yang sesuai dengan lima sampel dalam GazeCom yang di-*sampling* pada 250 Hz dan empat sampel dalam HMR yang di-*sampling* pada 200 Hz—sehingga secara signifikan mengurangi kebutuhan memori dan beban komputasi.

Model dievaluasi menggunakan *stratified 5-fold cross-validation*. Model mencapai *F1 score* 91,89% dan 95,51% untuk klasifikasi *smooth pursuit* pada GazeCom dan HMR secara berturut-turut, sambil juga mempertahankan kinerja yang kuat di seluruh jenis *event* lainnya seperti deteksi *fixation*, *saccade*, dan *blink*.

Temuan ini menunjukkan bahwa klasifikasi *eye movement* yang akurat dan efisien dapat dicapai menggunakan input temporal minimal. Arsitektur yang diusulkan menawarkan solusi yang menjanjikan untuk interaksi berbasis pandangan mata *real-time* dengan sumber daya terbatas, khususnya mendukung pengembangan aplikasi praktis untuk individu dengan gangguan motorik dalam teknologi assistif dan interaksi manusia-komputer.

**Kata kunci**—*Eye tracking*, 1D-CNN, *multi-head self-attention*, *hybrid deep learning model*