UNIVERSITAS
GADJAH MADA
**Perbandingan Metode Signal-to-Image Transformation untuk Klasifikasi Beban Kognitif Berdasarkan Data
Eye-Tracking Menggunakan Convolutional Neural Network**
Angelia Grace Cahyani, Dr. Eng. Sunu Wibirama, S.T., M.Eng., IPM. ; Noor Akhmad Setiawan, Ir., S.T., M.T., Ph.D.,
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Beban kognitif mencerminkan tuntutan mental yang dialami seseorang saat menjalankan suatu tugas dan menjadi indikator penting dalam pemantauan kapasitas mental di berbagai bidang aplikasi. Teknologi *eye-tracking* memungkinkan estimasi beban kognitif secara noninvasif melalui analisis pola pergerakan mata. Namun, meskipun pendekatan *deep learning* berbasis sinyal satu dimensi telah dikembangkan, metode tersebut masih memiliki keterbatasan dalam menangkap pola temporal nonlinier yang kompleks secara menyeluruh, karena konvolusi 1D hanya menelusuri dependensi lokal sepanjang sumbu waktu tanpa mampu merepresentasikan struktur rekurensi dan hubungan jarak jauh antar titik sinyal secara visual dan spasial. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi beban kognitif dengan mentransformasikan sinyal *gaze* ke dalam representasi citra dua dimensi menggunakan empat metode *signal-to-image transformation*, yaitu *Gramian Angular Summation Field* (GASF), *Gramian Angular Difference Field* (GADF), *Markov Transition Field* (MTF), dan *Recurrence Plot* (RP), sehingga pola temporal sinyal dapat diekstraksi secara otomatis oleh model *Convolutional Neural Network* (CNN).

Eksperimen dilakukan menggunakan dataset COLET (*COgnitive workLoad Estimation based on Eye-Tracking*) yang memuat rekaman *eye-tracking* partisipan saat mengerjakan tugas pencarian visual dengan tingkat kompleksitas yang bervariasi. Sinyal posisi *gaze* horizontal (X) dan vertikal (Y) ditransformasikan menjadi citra dua dimensi dan diklasifikasikan ke dalam tiga kelas beban kognitif, yaitu *Low* (C2), *Medium* (C1), dan *High* (C0), berdasarkan skor NASA-TLX. Arsitektur DenseNet121 digunakan sebagai model klasifikasi dengan strategi *fine-tuning* menyeluruh pada dua belas konfigurasi eksperimen (4 metode $\times$ 3 kanal sinyal, yaitu X, Y, dan kombinasi). Pelatihan menggunakan *weighted sampling*, *label smoothing*, *optimizer* AdamW, dan *early stopping* untuk menangani ketidakseimbangan kelas dan meningkatkan generalisasi. Evaluasi dilakukan menggunakan skema *5-fold stratified cross-validation* dengan metrik akurasi, presisi, *recall*, dan *F1-score*, serta uji statistik *Wilcoxon Signed-Rank* dan *Friedman* untuk memverifikasi signifikansi perbedaan performa.

Hasil penelitian menunjukkan bahwa seluruh metode transformasi menghasilkan akurasi yang jauh melampaui pendekatan *machine learning* dan *deep learning* konvensional pada dataset yang sama. RP secara konsisten mencapai performa tertinggi di semua konfigurasi kanal, dengan konfigurasi terbaik (RP pada sinyal vertikal) menghasilkan akurasi rata-rata 98,94% $\pm$ 0,09% dan *macro F1-score* 98,95% $\pm$ 0,09%. Uji *Friedman* mengonfirmasi perbedaan performa yang signifikan antar keempat metode ($\chi^2 = 40{,}760$, $p < 0{,}001$), dengan peringkat RP (3,87), GADF (3,13), GASF (2,00), dan MTF (1,00). Kanal sinyal vertikal (Y) secara konsisten menghasilkan akurasi sedikit lebih tinggi dari horizontal (X) pada metode RP dan GASF, dikonfirmasi secara statistik melalui uji *Wilcoxon*. Penelitian ini membuktikan bahwa RP merupakan metode transformasi optimal untuk klasifikasi beban kognitif berbasis sinyal *gaze*, dan membuka peluang pengembangan sistem pemantauan beban kognitif yang akurat dan noninvasif.

Kata Kunci: *Eye-Tracking*, *Cognitive Workload*, *Signal-to-Image Transformation*, *Recurrence Plot*, *Convolutional Neural Network*

UNIVERSITAS
GADJAH MADA
**Perbandingan Metode Signal-to-Image Transformation untuk Klasifikasi Beban Kognitif Berdasarkan Data
Eye-Tracking Menggunakan Convolutional Neural Network**
Angelia Grace Cahyani, Dr. Eng. Sunu Wibirama, S.T., M.Eng., IPM. ; Noor Akhmad Setiawan, Ir., S.T., M.T., Ph.D.,
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Cognitive workload reflects the mental demands placed on an individual during the execution of a task and serves as a critical indicator for monitoring mental capacity in various real-world applications. Eye-tracking technology offers a noninvasive means of estimating cognitive workload by capturing eye movement patterns. However, although one-dimensional deep learning approaches have been developed for gaze signal analysis, they remain limited in capturing complex nonlinear temporal patterns comprehensively, as 1D convolutions only traverse local dependencies along the time axis without being able to represent recurrence structures and long-range relationships between signal points in a visual and spatial manner. This study proposes transforming gaze signals into two-dimensional image representations through four signal-to-image transformation methods, Gramian Angular Summation Field (GASF), Gramian Angular Difference Field (GADF), Markov Transition Field (MTF), and Recurrence Plot (RP), to enable automatic and effective feature extraction via Convolutional Neural Networks (CNNs).

Experiments were conducted using the COLET (Cognitive Workload Estimation Based on Eye-Tracking) dataset, which contains eye-tracking recordings of participants completing visual search tasks with varying levels of complexity. Horizontal (X) and vertical (Y) gaze position signals were independently transformed into two-dimensional images and classified into three cognitive workload levels — Low (C2), Medium (C1), and High (C0) — based on NASA-TLX scores. DenseNet121 was used as the classification architecture through a full fine-tuning strategy, trained across twelve experimental configurations (4 transformation methods $\times$ 3 signal channels: X, Y, and combined). Model training employed weighted sampling, label smoothing, AdamW optimization, and early stopping to address class imbalance and improve generalization. Evaluation was performed using 5-fold stratified cross-validation, assessed via accuracy, macro precision, recall, and F1-score. Statistical hypothesis testing using the Wilcoxon Signed-Rank test and Friedman test was conducted to verify the significance of observed performance differences.

The results demonstrate that all four transformation methods substantially outperform conventional machine learning approaches — including Gaussian Naive Bayes, Random Forest, and Logistic Regression — as well as conventional deep learning methods applied on the COLET dataset. RP consistently achieved the highest classification performance across all signal channel configurations, with the best configuration (RP on vertical gaze, RP_Y) reaching an average test accuracy of 98.94% $\pm$ 0.09% and macro F1-score of 98.95% $\pm$ 0.09%. The Friedman test confirmed statistically significant performance differences among the four methods ($\chi^2 = 40.760$, $p < 0.001$), with mean ranks of RP (3.87), GADF (3.13), GASF (2.00), and MTF (1.00). The vertical gaze channel (Y) produced slightly higher accuracy than horizontal (X) for RP and GASF, a difference confirmed statistically by the Wilcoxon Signed-Rank test. Combined channel configurations did not consistently outperform single-channel inputs, with performance being highly dependent on the transformation method used. These findings establish RP as the optimal signal-to-image transformation for cognitive workload classification from gaze data, and demonstrate the effectiveness of DenseNet121 with fine-tuning for this domain.

Keywords: *Eye-Tracking, Cognitive Workload, Signal-to-Image Transformation, Recurrence Plot, Convolutional Neural Network*