



### PROTEKSI ISI LAPORAN AKHIR PENELITIAN DOSEN PEMULA

Dilarang menyalin, menyimpan, memperbanyak sebagian atau seluruh isi proposal ini dalam bentuk apapun kecuali oleh pengusul dan pengelola administrasi pengabdian kepada masyarakat

#### LAPORAN AKHIR 2024

Rencana Pelaksanaan Penelitian Dosen Pemula: tahun 2024 s.d. tahun 2024

#### 1. JUDUL PENELITIAN

Pengembangan MindPro, Model Prediksi Beban Kerja Mental berbasis Sinyal Fisiologi dengan Hybrid Machine Learning dan Analisis Cross Dataset

Bidang Fokus	Tema	Topik (jika ada)	Prioritas Riset
Teknologi Informasi dan Komunikasi	Pengembangan sistem berbasis Kecerdasan buatan	Pengembangan aplikasi sistem cerdas	Digital Economy

Rumpun Ilmu Level 1	Rumpun Ilmu Level 2	Rumpun Ilmu Level 3
ILMU TEKNIK	TEKNIK ELEKTRO DAN INFORMATIKA	Teknik Perangkat Lunak

Skema Penelitian	Strata (Dasar/Terapan/Pengembangan)	Nilai SBK	Target Akhir TKT	Lama Kegiatan
Penelitian Dosen Pemula	Riset Dasar	50.000.000	3	1 Tahun

#### 2. IDENTITAS PENGUSUL

Nama, Peran	Jenis	Program Studi/Bagian	Bidang Tugas	ID Sinta
AUDITYA PURWANDINI SUTARTO 0011067801  Ketua Pengusul Universitas Qomaruddin	Dosen	Teknik Industri	Koordinator & penanggung jawab seluruh kegiatan operasional penelitian, pengembangan model, penyusunan laporan penelitian dan naskah publikasi	<a href="#">6705687</a>
ADE HENDI 0719107804  Anggota Universitas Qomaruddin	Dosen	Teknik Informatika	Pengembangan model, pengolahan data, membantu menyusun laporan	<a href="#">6779437</a>

#### 3. MITRA KERJASAMA PENELITIAN (Jika Ada)

Pelaksanaan penelitian dapat melibatkan mitra kerjasama yaitu mitra kerjasama dalam melaksanakan penelitian, mitra sebagai calon pengguna hasil penelitian, atau mitra investor

Mitra	Nama Mitra	Dana
-------	------------	------

#### 4. LUARAN DAN TARGET CAPAIAN

##### Luaran Wajib

Tahun Luaran	Kategori Luaran	Jenis Luaran	Status target capaian	Keterangan
1	Artikel di Jurnal	Artikel di Jurnal Bereputasi Internasional	Accepted/Published	<a href="https://ijai.iaescore.com/index.php/IJAI">https://ijai.iaescore.com/index.php/IJAI</a> , IAES International Journal of Artificial Intelligence (IJ-AI), Institute of Advanced Engineering and Science (IAES)

## 5. ANGGARAN

Rencana Anggaran Biaya penelitian mengacu pada PMK dan buku Panduan Penelitian dan Pengabdian kepada Masyarakat yang berlaku.

Total RAB 1 Tahun Rp47.010.000,00

Tahun 1 Total Rp47.010.000,00

Kelompok	Komponen	Item	Satuan	Vol.	Biaya Satuan	Total
Pengumpulan Data	Biaya konsumsi	Responden	OH	30	50.000	1.500.000
Pengumpulan Data	Biaya konsumsi	Peneliti dan Asisten	OH	60	50.000	3.000.000
Pengumpulan Data	Transport	Transport	OK (kali)	5	200.000	1.000.000
Sewa Peralatan	Peralatan penelitian	GSR	Unit	4	500.000	2.000.000
Analisis Data	Honorarium narasumber	Expert Machine Learning	OJ	3	900.000	2.700.000
Bahan	Bahan Penelitian (Habis Pakai)	chest strap polar	Unit	2	500.000	1.000.000
Bahan	ATK	ATK, FC, kertas	Paket	4	500.000	2.000.000
Bahan	ATK	Tinta Printer	Paket	1	350.000	350.000
Pengumpulan Data	HR Pembantu Peneliti	Asisten eksp	OJ	120	25.000	3.000.000
Pengumpulan Data	Uang Harian	Perjalanan Dinas dalam kota lebih dari 8 jam	OH	4	160.000	640.000
Pengumpulan Data	Uang Harian	Rapat 2 peneliti luar kantor	OH	4	140.000	560.000
Bahan	Bahan Penelitian (Habis Pakai)	Komite Etik	Unit	10	100.000	1.000.000
Sewa Peralatan	Obyek penelitian	Responden	Unit	30	150.000	4.500.000
Sewa Peralatan	Peralatan penelitian	Sensor polar	Unit	4	500.000	2.000.000
Analisis Data	HR Pengolah Data	paket	P (penelitian)	1	1.500.000	1.500.000
Analisis Data	Transport Lokal	transport	OK (kali)	4	200.000	800.000
Pelaporan Hasil Penelitian dan Luaran Wajib	Biaya Publikasi artikel di Jurnal Bereputasi Internasional	APC Jurnal	Paket	10	400.000	4.000.000
Pelaporan Hasil Penelitian dan Luaran Wajib	Biaya konsumsi rapat	Rapat	OH	8	50.000	400.000
Bahan	Barang Persediaan	Mindfield e-sense response device	Unit	1	5.000.000	5.000.000
Pelaporan Hasil Penelitian dan Luaran Wajib	Uang harian rapat di luar kantor	Rapat	OH	4	140.000	560.000
Analisis Data	Biaya analisis sampel	Analisis Data	Unit	300	25.000	7.500.000
Sewa Peralatan	Peralatan penelitian	Komputer	Unit	4	500.000	2.000.000

\*. KEMAJUAN PENELITIAN

## A. RINGKASAN

Perkembangan pesat teknologi dalam era digital ekonomi menuntut beban kerja mental tinggi bagi pekerja semua industri. Hal ini memicu kebutuhan sistem pemantauan real time yang dapat memprediksi mental workload demi menjaga produktivitas dan well-being, serta mengurangi kesalahan dan kecelakaan kerja. Model prediksi menggunakan sinyal fisiologi electrocardiography (ECG) dan electrodermal activity (EDA) dengan pendekatan machine learning (ML) berkembang pesat karena bersifat objektif, dapat digeneralisir untuk berbagai pekerjaan, dan mampu diukur kontinyu tanpa mengganggu aktivitas harian.

Penelitian ini merupakan kelanjutan penelitian kami yang menghasilkan akurasi 68% dengan satu jenis sinyal ECG. Adapun penelitian lain dengan akurasi di atas 80% belum menggunakan skenario simulasi mental workload di pekerjaan nyata. Penelitian ini bertujuan mengembangkan MindPro, suatu model hybrid ML untuk memprediksi beban kerja mental berdasarkan sinyal ECG dan EDA dengan simulasi pekerjaan sesungguhnya, guna mengatasi isu performa dan efisiensi pada model sebelumnya. Penelitian ini juga mengeksplorasi cross dataset untuk menguji generalisasi model.

Tahap pertama penelitian adalah eksperimen di laboratorium untuk mendapatkan sinyal ECG dan EDA. Eksperimen ini telah dilakukan pada 30 subjek mahasiswa yang menjalani empat sesi pengukuran, yaitu baseline 1, tugas mental workload 1 Iowa Gambling Test, tugas mental workload 2 arithmetic test, dan recovery. Output eksperimen sinyal ECG berupa parameter HRV dan sinyal EDA berupa parameter galvanic skin response (GSR).

Tahap kedua adalah pengembangan model ML berdasarkan input sinyal HRV dan GSR. Pada tahap ini disusun machine learning pipeline yang mencakup berbagai proses mulai dari data pre-processing, feature engineering, pemilihan fitur, pengembangan berbagai algoritma, hyperparameter tuning, dan evaluasi performansi model. Berbagai algoritma ML yang digunakan mencakup Support Vector Machine (SVM), Naïve Bayes NB), Extreme Gradient Boosting (XGB), Random Forest (RF), dan Multi-Layer Perceptron (MLP). Pengembangan ML sebagai bagian dari kecerdasan buatan ini diharapkan mampu menghasilkan model prediksi melalui klasifikasi tingkat kognitif yang lebih baik dibandingkan model statistik konvensional.

Adapun tahap ketiga yaitu pengujian algoritma pada dataset publik (cross-dataset analysis), di mana algoritma diuji menggunakan dua dataset publik: CogLoad dan WeSAD. Hasil analisis menunjukkan performa algoritma terbaik (MLP) lebih baik pada dataset WeSAD dibandingkan dengan dataset CogLoad. Kinerja yang kurang optimal pada dataset CogLoad disebabkan oleh jumlah fitur yang terbatas dan instances yang sedikit, sementara dataset WeSAD memiliki fitur lebih banyak dan instances yang sangat besar, yang mendukung performa model secara signifikan.

Penelitian ini diharapkan berkontribusi secara teoretis melalui pemahaman komprehensif respon ECG dan EDA terhadap beban kerja mental dan pengembangan model hybrid ML yang general, robust, dan optimal pada data baru. Adapun secara praktis, hasil penelitian mendukung pengembangan MindPro, sistem monitoring mental workload yang terintegrasi dengan wearable devices untuk meningkatkan awareness perubahan beban kerja mental dalam pekerjaan sehari-hari. Penelitian ini relevan dengan topik pengembangan aplikasi sistem cerdas dengan hasil tingkat kesiapan teknologi (TKT 3).

Saat ini hasil penelitian telah di-submit ke jurnal International Journal of Computing and Digital Systems (IJCDS) (<https://ijcnds.uob.edu.bh/>), e-ISSN 2210-142X (online) /) dengan status under-review.

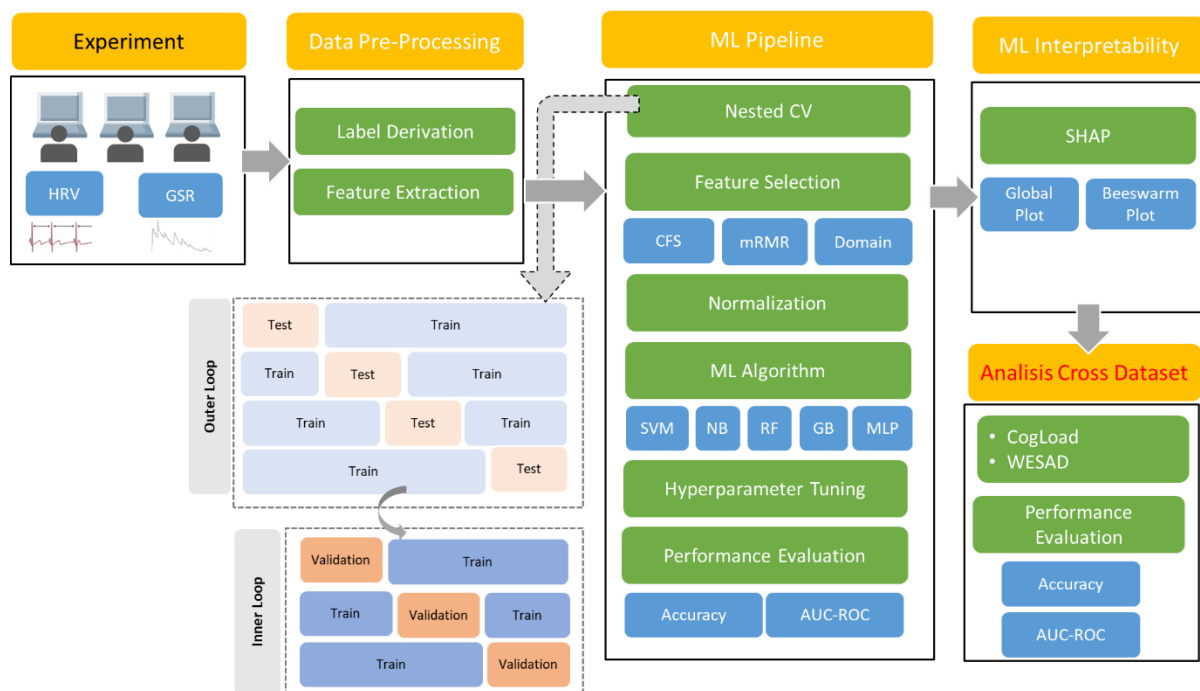
## B. KATA KUNCI

beban kerja mental; fisiologi; machine learning; monitoring; sistem cerdas

Pengisian poin C sampai dengan poin H mengikuti template berikut dan tidak dibatasi jumlah kata atau halaman namun disarankan ringkas mungkin. Dilarang menghapus/modifikasi template ataupun menghapus penjelasan di setiap poin.

**C. HASIL PELAKSANAAN PENELITIAN:** Tuliskan secara ringkas hasil pelaksanaan penelitian yang telah dicapai sesuai tahun pelaksanaan penelitian. Penyajian meliputi data, hasil analisis, dan capaian luaran (wajib dan atau tambahan). Seluruh hasil atau capaian yang dilaporkan harus berkaitan dengan tahapan pelaksanaan penelitian sebagaimana direncanakan pada proposal. Penyajian data dapat berupa gambar, tabel, grafik, dan sejenisnya, serta analisis didukung dengan sumber pustaka primer yang relevan dan terkini.

Penelitian ini dilaksanakan melalui dua tahapan utama: eksperimen dan pengembangan model machine learning (ML), sebagaimana ditunjukkan dalam Gambar 1. Data yang diperoleh dari eksperimen kemudian digunakan sebagai masukan untuk pengembangan model ML. Berikut adalah rincian dari masing-masing tahapan:



**Gambar 1.** Tahapan Penelitian

**TAHAP 1: EKSPERIMEN**

Pada tahap eksperimen, data primer berupa sinyal fisiologis HRV dan GSR dikumpulkan dalam beberapa kondisi. Protokol eksperimen disusun berdasarkan pedoman penelitian psikofisiologi dengan sinyal tubuh (1), yang mengharuskan perekaman sinyal dalam tiga kondisi berbeda: kondisi dasar (baseline), saat terpapar stimulus *mental workload* 1 dan 2 (test 1 dan test 2), dan saat pemulihan (recovery). Gambar 2 memperlihatkan suasana saat dilaksanakan eksperimen

**TAHAP 2: PENGEMBANGAN MODEL MACHINE LEARNING (ML)**

Pada tahap 2, sinyal HRV dan GSR dari hasil eksperimen akan diolah terlebih dahulu dengan software Kubios untuk mendapatkan berbagai parameter HRV domain seperti ditunjukkan dalam Tabel 1 (1,2). Parameter ini selanjutnya menjadi fitur yang menjadi masukan untuk membangun model ML. Adapun untuk sinyal GSR otomatis dihasilkan dari aplikasi E-Sense dalam bentuk: Time, Average of Session  $\mu$ S, Minimum  $\mu$ S, Maximum  $\mu$ S, Difference Min/Max  $\mu$ S, Time Increase, Time Decrease, Total SCR, % of SCR



**Gambar 2.** Contoh Suasana saat Eksperimen

**Table 1.** HRV parameters and their description

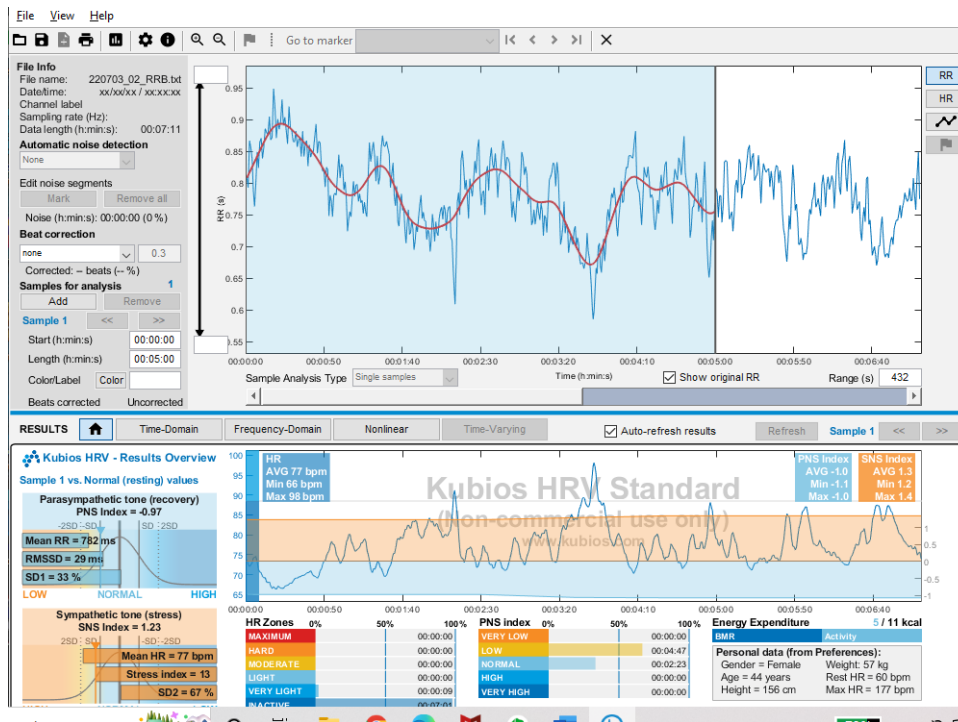
Parameters	Unit	Description
Time-domain		
Mean RR	ms	Mean of all RR intervals. The RR interval and heart rate (HR) are hyperbolically related ( $HR \times RR \text{ interval} = 60,000$ )
RMSSD	ms	Root mean square of successive differences between adjacent NN intervals (NN-intervals refer to interbeat intervals from which artifacts have been removed)
SDNN	ms	Standard deviation of NN intervals
pNN50	%	Percentage of successive NN intervals that differ by more than 50 m
Frequency Domain		
ULF power	ms <sup>2</sup>	Absolute power of the ultra-low-frequency band (< 0.0033 Hz)
VLF power	ms <sup>2</sup>	Absolute power of the very-low-frequency band (0.0033–0.04 Hz)
LF power	ms <sup>2</sup>	Absolute power of the low-frequency band (0.04–0.15 Hz)
LF power	nu	Relative power of the low-frequency band (0.04–0.15 Hz) in normalized units
HF power	ms <sup>2</sup>	Absolute power of the high-frequency band (0.15–0.4 Hz)
HF power	Nu	Relative power of the high-frequency band (0.15–0.4 Hz) in normalized units
LF/HF	%	Ratio of LF to HF power
Total power	ms <sup>2</sup>	Sum of the energy in the ULF, VLF, LF, and HF bands
Non-linear		
SD1	ms	Poincaré plot standard deviation perpendicular the line of identity (short-term variability)
SD2	ms	Poincaré plot standard deviation along the line of identity (long-term variability)
SD2/SD1	%	Ratio of SD1-to-SD2

Berikut adalah proses data preprocessing untuk HRV dan GSR

1. Ekstraksi Fitur (*Feature Extraction*)

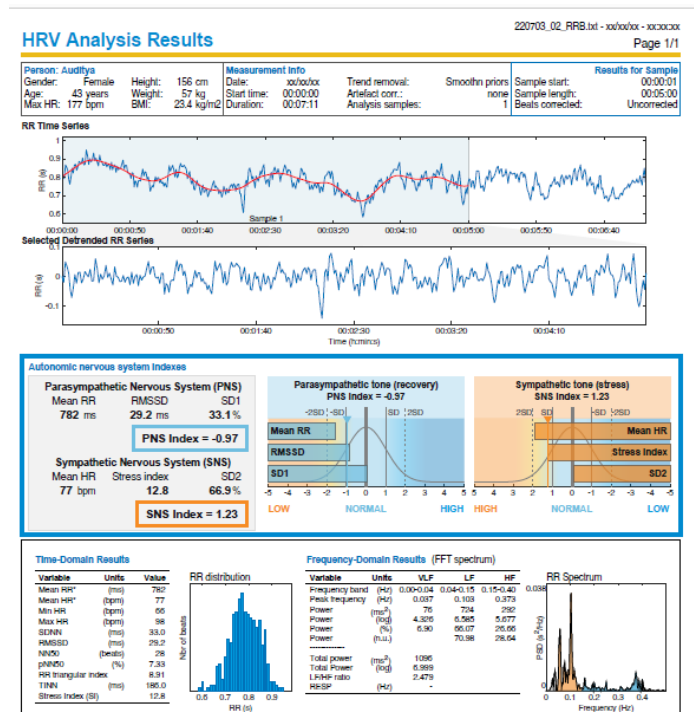
HRV

- a. Data rekaman sinyal berupa Interbeat Interval (IBI) atau RR yang diperoleh dari sensor Polar H10 diekspor melalui aplikasi **Elite HRV** ke email peneliti dalam format TXT
- b. Selanjutnya, data RR tersebut diolah menggunakan software Kubios versi 4.1.0 (3) untuk menghasilkan berbagai parameter HRV, sebagaimana ditunjukkan pada Gambar 3



Gambar 3. Tampilan layar proses analisis HRV dengan Kubios

- c. Jika ditemukan *noise (artifact)* artefak dalam sinyal, maka dilakukan koreksi pada tingkat medium
- d. Hasil analisis kemudian disimpan dalam format PDF dan Excel. Gambar 3 menampilkan contoh laporan hasil analisis menggunakan Kubios. Hasil analisis selanjutnya disimpan dalam format PDF dan Excel. Gambar 4 menunjukkan contoh laporan hasil analisis dengan Kubios



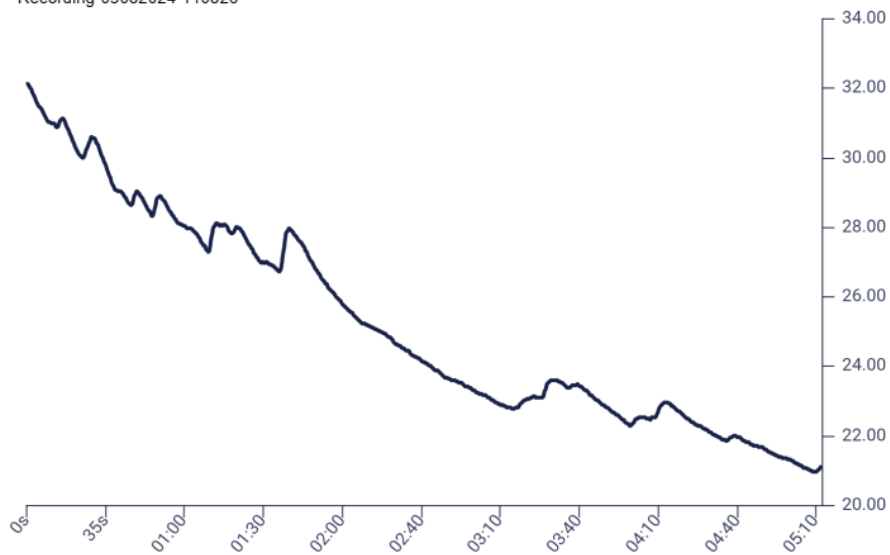
Gambar 4. Contoh hasil analisis HRV dengan Kubios

GSR

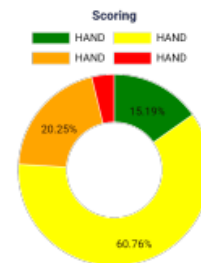
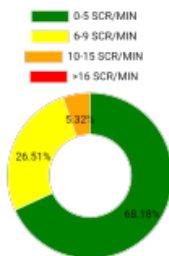
- a. Data rekaman sinyal diproses otomatis oleh aplikasi e-sense yang disediakan oleh pabrik pembuat alat ukur GSR (MindField, Germany) peneliti dalam format TXT

- b. Hasil pengolahan tersebut menghasilkan output dalam dua format.
- Format PDF berisi report untuk fitur Average of Session (diukur dalam micro Siemens atau  $\mu\text{S}$ ), minimum SCR, maximum SCR, difference between Min/Max, total SCR, dan percentage of SCR dilengkapi dengan visualisasi grafik
  - Format csv berisi fitur dalam PDF tanpa grafik namun dilengkapi data mentah sebanyak kurang lebih 1800 baris pengukuran SCR. Data mentah ini nantinya dapat diproses lebih lanjut menggunakan tools lain.
  - Contoh tampilan report kedua format ini ditunjukkan dalam Gambar 5

Recording-03082024-110826



Recording-03082024-110826



	A	B	C	D
1	Device: OPPO CPH2239	eSense app version: 6.9.7 android version: 11		
2	DESCRIPTION	VALUE		
3	RECORDING NAME	Recording-03082024-110012		
4	DATE (HH:MM:SS)	03.08.24 11:00:12		
5	TIME (HH:MM:SS)		00:06:18	
6	MAXIMUM	43.02 $\text{\AA}\mu\text{S}$		
7	MINIMUM	27.95 $\text{\AA}\mu\text{S}$		
8	DIFFERENCE MIN/MAX	15.06 $\text{\AA}\mu\text{S}$		
9	AVERAGE OF SESSION	33.42 $\text{\AA}\mu\text{S}$		
10	TIME INCREASE (HH:MM:SS)		00:02:12	
11	TIME STEADY (HH:MM:SS)		00:00:19	
12	TIME DECREASE (HH:MM:SS)		00:03:46	
13				
14	TOTAL SCR		57	
15	% of SCR OF SESSION		52.43	
16	Red hands		22	
17	Orange hands		53	
18	Yellow hands		19	
19	Green hands		1	
20	0-5 SCR/MIN (HH:MM:SS)		00:00:39	
21	6-9 SCR/MIN (HH:MM:SS)		00:03:33	
22	10-15 SCR/MIN (HH:MM:SS)		00:02:09	
23	>16 SCR/MIN (HH:MM:SS)		00:00:00	
24	STARS		0	
25	STARS WITH RIBBON		0	

**Gambar 5.** Contoh GSR Report yang Dihasilkan dari Aplikasi e-Sense dalam format PDF dan csv

## 2. Data Cleaning

### HRV

- Data dari proses Kubios dipindahkan dan direkapitulasi ke dalam satu file format excel.
- Setelah itu dilakukan pembersihan data yang mencakup pengecekan adakah data yang missing, mengandung outlier dan lain-lain.

### GSR

- Proses ini dilakukan otomatis oleh aplikasi e-sense

## 3. Feature Selection.

- Sebelum dilakukan pemilihan fitur terdapat 30 fitur yang terdiri atas 24 fitur HRV dan 6 fitur GSR. Fitur HRV mencakup RR, mean HR, min HR, max HR, SDNN, RMSSD, pNN50), total power, total power log, VLF absolute power, VLF log, peak VLF, LF absolute power, LF log, LF nu, peak LF, HF, HF log, HF nu, peak HF, and LF/HF), SD2, SD1, dan SD2/SD1. Sedangkan fitur GSR mencakup average SCR, minimum SCR, maximum SCR, difference between Min/Max, total SCR, dan percentage of SCR
- Dalam penelitian ini akan dibandingkan model yang dibangun berdasarkan tiga metode pemilihan fitur: 1) Correlation-based feature (CFS). variasi fitur atau parameter; 2) metode *minimum redundancy maximum relevancy* (mRMR) (4); 3) pemilihan parameter berdasarkan domain-expertise
- Tabel 3 menunjukkan daftar fitur terpilih berdasarkan metode *feature selection*.

**Tabel 3.** Daftar Fitur yang Digunakan untuk Mengembangkan ML berdasarkan Tiga Metode Fitur

Method	HRV	GSR
Domain-expert	hr, sdnn, rmssd, lfnu, hfnu, lf/hf	'scr', 'max', 'diff_minmax', 'pscr'
mRMR	'minhr', 'maxhr', 'peak_lf', 'hfnu', 'hr', 'rr', 'lfnu',	scr', 'min', 'pscr', '
CBS	'maxhr', 'peak_vlf', 'peak_lf', 'peak_hf', 'vlflog', 'hf', 'hfnu', 'lfhf', 'sd1', 'sd2', 'sd2sd1'	'scr', 'max', 'diff_minmax', 'pscr'

Singkatan. hr: heart rate, sdnn: standard deviation normal to normal, rmssd: root mean square standard deviation, lfnu: low frequency in normalized unit, hfnu: high frequency in normalized unit, lf/hf: ratio low frequency to high frequency, vlf: very low frequency, scr: skin conductance response, max: maximum amplitude, min: minimum amplitude of scr, diff\_minmax: differences between min and max amplitude of SCR, pscr: percentage of SCR.

#### 4. Data Transformation

- Transformasi data dilakukan baik untuk fitur HRV dan GSR untuk meningkatkan efisiensi, performansi, dan interpretability model ML (5). Dalam penelitian ini digunakan metode Min-Max Scaler.

#### 5. Pengembangan model

- Model ML dikembangkan menggunakan tiga jenis classifier utama: 1) single classifier Support Vector Machine (SVM) dan Naïve Bayes (NB); 2) ensemble learning Gradient Boosting dan AdaBoost; 3) neural network Multi Layer Perceptron
- Model dikembangkan dengan bahasa pemrograman Python dan library *sci-kit*.
- Validasi model dilakukan secara hybrid menggabungkan teknik nested cross-validation
- Setelah proses validasi selesai selanjutnya menilai performa model menggunakan kriteria *accuracy* dan AUC yang hasilnya ditunjukkan dalam tabel 3.

**Tabel 3.** Nilai Akurasi dan AUC Model ML.

	CFS		mRMR		Domain Expertise	
	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC
SVM	0.6591	0.7070	<b>0.6818</b>	0.6149	0.6364	<b>0.7412</b>
Naïve	0.5909	0.6128	0.5682	0.6377	0.6591	0.7019
GB	0.6818	0.6812	0.6364	<b>0.7598</b>	<b>0.6818</b>	0.6708
RF	0.5909	0.6563	0.6136	0.6749	0.6818	0.6584
MLP	<b>0.7500</b>	<b>0.8219</b>	0.5682	0.6853	0.5909	0.6439

Catatan. CFS: Correlation-Based Feature Selection, mRMR: minimum redundancy maximum relevancy. Performansi terbaik dinyatakan dalam huruf tebal.

### TAHAP 3: ANALISIS CROSS DATASET

Tahapan terakhir dalam penelitian ini adalah melakukan analisis cross-dataset, di mana algoritma yang dikembangkan diuji menggunakan dataset publik. Dalam penelitian ini, pengujian dilakukan pada dua dataset publik berikut:

#### 1. CogLoad (6)

Dataset CogLoad mencakup data dari 23 partisipan yang menyelesaikan berbagai tugas kognitif dengan tingkat kesulitan berbeda, termasuk tugas N-back dan enam tugas kognitif dasar (Elementary Cognitive Tasks/ECT) dengan jeda istirahat di antaranya. Respon fisiologis yang diukur meliputi detak jantung (HR), interval R-to-R (RR), respons galvanik kulit (GSR), suhu kulit (ST), dan akselerasi 3-sumbu, yang direkam menggunakan perangkat Microsoft Band.

#### 2. WeSAD (7)

Dataset WeSAD adalah dataset multimodal yang diperoleh dari 15 partisipan dalam skenario terkontrol yang dirancang untuk membangkitkan stres, hiburan (amusement), dan relaksasi. Dataset ini berfokus pada deteksi stres dan emosi menggunakan perangkat wearable dan mencakup data fisiologis serta gerakan, seperti aktivitas elektrodermal (EDA), detak jantung (HR), suhu tubuh, data akselerometer, dan lainnya. WeSAD banyak digunakan dalam penelitian untuk mendukung pengembangan teknologi wearable dalam pemantauan kesehatan dan komputasi afektif.

Hasil analisis menunjukkan bahwa performa model terbaik untuk setiap dataset bervariasi, sebagaimana ditunjukkan dalam Tabel 4. Pada dataset CogLoad, performa algoritma terbaik (MLP) relatif kurang baik, sementara performa pada dataset WeSAD sangat memuaskan. Kinerja yang kurang optimal pada dataset CogLoad disebabkan oleh beberapa faktor, termasuk jumlah fitur yang terbatas (hanya RR, HR, TEMP, dan GSR) serta jumlah *instances* yang relatif sedikit. Sebaliknya, pada dataset WeSAD, meskipun hanya menggunakan sinyal

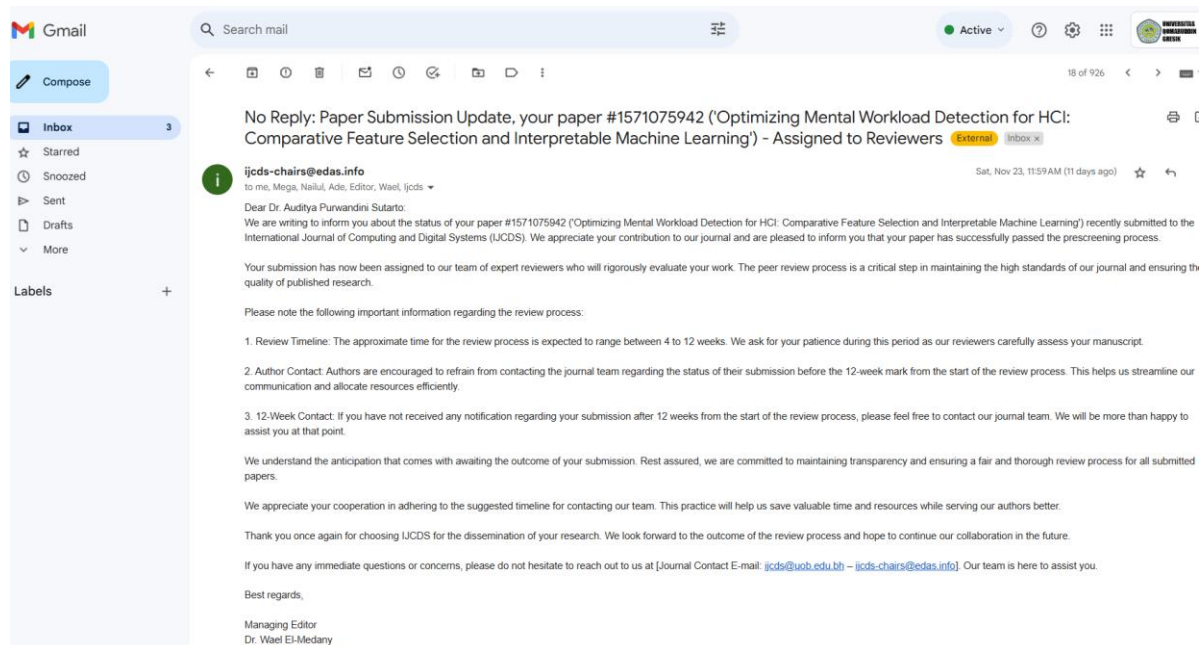
HRV, terdapat 23 fitur dengan jumlah *instances* yang sangat besar, mencapai lebih dari 100.000, sehingga mendukung performa algoritma dengan lebih baik.

**Tabel 4.** Perbandingan Nilai Akurasi dan AUC Model ML terbaik antara Data Eksperimen dengan Dataset Public.

Dataset	subjek	sinyal	# final features	# instances	akurasi	AUC
Eksperimen sendiri	34	HRV, GSR	15	144	0.750	0.822
cogload	23	RR, HR, TEMP, GSR	4	825	0.550	0.530
wesad	15	HRV	23	135,650	0.79	0.850

**D. STATUS LUARAN:** Tuliskan jenis, identitas dan status ketercapaian setiap luaran wajib dan luaran tambahan (jika ada) yang dijanjikan. Jenis luaran dapat berupa publikasi, perolehan kekayaan intelektual, hasil pengujian atau luaran lainnya yang telah dijanjikan pada proposal. Uraian status luaran harus didukung dengan bukti kemajuan ketercapaian luaran sesuai dengan luaran yang dijanjikan. Lengkapi isian jenis luaran yang dijanjikan serta mengunggah bukti dokumen ketercapaian luaran wajib dan luaran tambahan melalui BIMA.

Luaran wajib dalam penelitian ini adalah jurnal internasional bereputasi. Saat ini luaran wajib dalam penelitian ini telah di- submit ke jurnal *International Journal of Computing and Digital Systems (IJCDS)* (<https://ijcds.uob.edu.bh/>), e-ISSN 2210-142X (online) /) dengan status *under-review* sebagaimana ditunjukkan dalam Gambar 3.



**Gambar 3.** Tangkapan Layar Notifikasi Status Naskah

**E. PERAN MITRA:** Tuliskan realisasi kerjasama dan kontribusi Mitra baik *in-kind* maupun *in-cash* (untuk Penelitian Terapan, Penelitian Pengembangan, PTUPT, PPUPT serta KRUPPT). Bukti pendukung realisasi kerjasama dan realisasi kontribusi mitra dilaporkan sesuai dengan kondisi yang sebenarnya. Bukti dokumen realisasi kerjasama dengan Mitra diunggah melalui BIMA.

Penelitian ini tidak melibatkan mitra

**F. KENDALA PELAKSANAAN PENELITIAN:** Tuliskan kesulitan atau hambatan yang dihadapi selama melakukan penelitian dan mencapai luaran yang dijanjikan, termasuk penjelasan jika pelaksanaan penelitian dan luaran penelitian tidak sesuai dengan yang direncanakan atau dijanjikan.

Beberapa kendala yang dijumpai dalam penelitian ini

1. Smartphone yang digunakan lambat dan terkadang tidak dapat menangkap sinyal GSR. Solusinya adalah menggunakan tablet untuk instal aplikasi e-sense yang dapat menangkap sinyal GSR lebih stabil
2. Karena keterbatasan waktu dan sumber daya, pengolahan data GSR menggunakan data mentah belum dapat dilakukan. Solusinya adalah menggunakan output yang dihasilkan oleh aplikasi e-sense. Hal ini menjadikan output kurang dapat dibandingkan dengan rekomendasi dari peneliti lain (8). Namun demikian, saat ini sedang diupayakan mengolah menggunakan piranti dan tool lain yang lebih *powerful*.
3. Cross-data set analysis belum dilakukan karena fokus pada pengolahan data eksperimen dan pengembangan model ML awal. Solusinya adalah menguji algoritma ML yang diperoleh pada dataset publik lain setelah laporan kemajuan ini dikumpulkan

**G. RENCANA TAHAPAN SELANJUTNYA:** Tuliskan dan uraikan rencana penelitian di tahun berikutnya berdasarkan indikator luaran yang telah dicapai, rencana realisasi luaran wajib yang dijanjikan dan tambahan (jika ada) di tahun berikutnya serta *roadmap* penelitian keseluruhan. Pada bagian ini diperbolehkan untuk melengkapi penjelasan dari setiap tahapan dalam metoda yang akan direncanakan termasuk jadwal berkaitan dengan strategi untuk mencapai luaran seperti yang telah dijanjikan dalam proposal. Jika diperlukan, penjelasan dapat juga dilengkapi dengan gambar, tabel, diagram, serta pustaka yang relevan. Pada bagian ini dapat dituliskan rencana penyelesaian target yang belum tercapai.

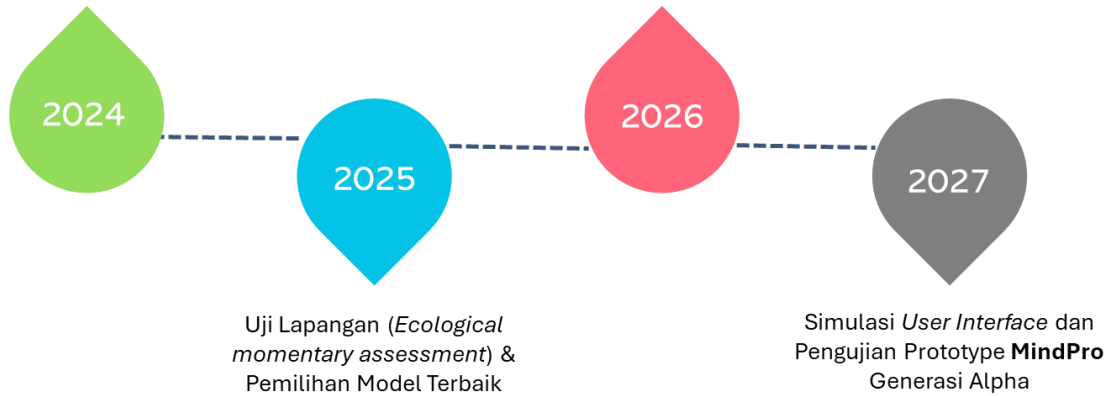
Rencana penelitian untuk tahapan selanjutnya berdasarkan penelitian ini adalah

1. Mengembangkan model ML yang lebih baik melalui
  - Menggunakan pendekatan model ML berbasis *artificial neural network* untuk meningkatkan performansi model.
  - Mengembangkan model *multiclass* sehingga dapat diperoleh model deteksi mental workload yang tidak hanya membedakan antara kondisi load dan non-load namun juga berdasarkan tingkat pembebanan seperti *low*, *medium*, dan *high*.
  - Kedua usulan di atas memerlukan data lebih banyak, maka dalam penelitian selanjutnya, data HRV dan GSR akan dipecah menjadi segmen pengukuran (*time windows*) yang lebih sempit. Jika pada penelitian ini menggunakan *time windows* sesuai lama kondisi (minimal lima menit), penelitian selanjutnya dapat menerapkan segmen 30-60 detik (9).
2. Melakukan eksperimen pada pekerja dengan simulasi atau aktivitas yang melibatkan kerja mental sesungguhnya untuk meningkatkan generalisasi dan mengatasi isu validasi eksternal
3. Merancang prototype MindPro untuk memonitor tingkat beban kerja mental melalui tahapan berikut:
  - **Analisis.** Pada tahap ini, dilakukan analisis kebutuhan sistem yang meliputi hardware, software, serta data yang diperlukan.
  - **Desain.** Semua aktivitas yang akan dilakukan oleh sistem digambarkan secara rinci pada tahap ini
  - **Pembuatan Prototipe.** Dalam tahap ini, prototipe akan dibangun dan diuji coba oleh pengguna.
  - **Pengembangan.** Umpan balik dari pengguna dievaluasi dan dijadikan dasar untuk perbaikan sistem hingga siap digunakan sesuai kebutuhan

# Peta Jalan Penelitian

Pengembangan Hybrid Model  
ML Prediksi *Mental Workload*

Pengembangan *Prototype MindPro*  
Integrasi Prediksi *Mental Workload*  
dalam *Wearable Devices*



**H. DAFTAR PUSTAKA:** Penyusunan Daftar Pustaka berdasarkan sistem nomor sesuai dengan urutan pengutipan. Hanya pustaka yang disitasi pada laporan akhir yang dicantumkan dalam Daftar Pustaka.

1. Malik M. Heart rate variability Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*. 1996;17(3):354–81.
2. Laborde S, Mosley E, Thayer JF. Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*. 2017;8(213).
3. Kubios. User's Guide Kubios HRV Scientific. Kubios Oy; 2023.
4. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*. 2017 Dec;18(1):9.
5. Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. Sebastopol, California: O'Reilly Media, Inc; 2019.
6. Gjoreski M, Kolenik T, Knez T, Luštrek M, Gams M, Gjoreski H, et al. Datasets for cognitive load inference using wearable sensors and psychological traits. *Applied Sciences (Switzerland)*. 2020;10(11).
7. Schmidt P, Reiss A, Duerichen R, Van Laerhoven K. Introducing WeSAD, a multimodal dataset for wearable stress and affect detection. *ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction*. 2018;400–8.
8. Chen F, Zhou J, Wang Y, Yu K, Arshad SZ, Khawaji A, et al. Galvanic Skin Response-Based Measures. In: *Robust Multimodal Cognitive Load Measurement [Internet]*. Cham: Springer International Publishing; 2016 [cited 2024 Sep 1]. p. 87–99. (Human–Computer Interaction Series). Available from: [http://link.springer.com/10.1007/978-3-319-31700-7\\_5](http://link.springer.com/10.1007/978-3-319-31700-7_5)

9. Castaldo R, Montesinos L, Melillo P, James C, Pecchia L. Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making*. 2019;19(1):1–13.



Gmail

Compose

Inbox 3

- Starred
- Snoozed
- Sent
- Drafts
- More

Labels +

No Reply: Paper Submission Update. your paper #1571075942 ('Optimizing Mental Workload Detection for HCI: Comparative Feature Selection and Interpretable Machine Learning') - Assigned to Reviewers

ijcds-chairs@edas.info  
to me, Mega, Nailul, Ade, Editor, Wael, ijcds

Sat, Nov 23, 11:59AM (11 days ago)

Dear Dr. Auidiya Purwandini Sutarto:

We are writing to inform you about the status of your paper #1571075942 ('Optimizing Mental Workload Detection for HCI: Comparative Feature Selection and Interpretable Machine Learning') recently submitted to the International Journal of Computing and Digital Systems (IJCDs). We appreciate your contribution to our journal and are pleased to inform you that your paper has successfully passed the prescreening process.

Your submission has now been assigned to our team of expert reviewers who will rigorously evaluate your work. The peer review process is a critical step in maintaining the high standards of our journal and ensuring the quality of published research.

Please note the following important information regarding the review process:

- Review Timeline:** The approximate time for the review process is expected to range between 4 to 12 weeks. We ask for your patience during this period as our reviewers carefully assess your manuscript.
- Author Contact:** Authors are encouraged to refrain from contacting the journal team regarding the status of their submission before the 12-week mark from the start of the review process. This helps us streamline our communication and allocate resources efficiently.
- 12-Week Contact:** If you have not received any notification regarding your submission after 12 weeks from the start of the review process, please feel free to contact our journal team. We will be more than happy to assist you at that point.

We understand the anticipation that comes with awaiting the outcome of your submission. Rest assured, we are committed to maintaining transparency and ensuring a fair and thorough review process for all submitted papers.

We appreciate your cooperation in adhering to the suggested timeline for contacting our team. This practice will help us save valuable time and resources while serving our authors better.

Thank you once again for choosing IJCDs for the dissemination of your research. We look forward to the outcome of the review process and hope to continue our collaboration in the future.

If you have any immediate questions or concerns, please do not hesitate to reach out to us at [Journal Contact E-mail: [ijcds@uob.edu.bh](mailto:ijcds@uob.edu.bh) – [ijcds-chairs@edas.info](mailto:ijcds-chairs@edas.info)]. Our team is here to assist you.

Best regards,

Managing Editor  
Dr. Wael El-Medany



# Comparative Feature Selection for Physiological Signal-Based Mental Workload Detection: Optimizing Human-Computer Interaction with Interpretable Machine Learning

Auditya Purwandini Sutarto<sup>1\*</sup>, Mega Bagus Herlambang<sup>2</sup>, Nailul Izzah<sup>3</sup>, and Ade Hendi<sup>4</sup> Author 1

<sup>1</sup> Department of Industrial Engineering, Universitas Qomaruddin, Gresik, Indonesia

<sup>2</sup> Department of Industrial Engineering, Institut Teknologi Indonesia, Tangerang, Indonesia

<sup>3</sup> Department of Industrial Engineering, Universitas Qomaruddin, Gresik, Indonesia

<sup>4</sup> Department of Informatics Engineering, Universitas Qomaruddin, Gresik, Indonesia

E-mail address: [auditya@uqgresik.ac.id](mailto:auditya@uqgresik.ac.id), [mega.herlambang@iti.ac.id](mailto:mega.herlambang@iti.ac.id), [nailulizzah@uqgresik.ac.id](mailto:nailulizzah@uqgresik.ac.id), [hendi.ade@gmail.com](mailto:hendi.ade@gmail.com)

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

**Abstract:** Detecting mental workload (MWL) is essential for optimizing task performance, preventing cognitive overload, and enhancing well-being and safety, especially in contexts involving human-computer interaction (HCI) and complex task environments. This study introduces a novel dataset for MWL detection using accessible physiological signals, specifically heart rate variability (HRV) and galvanic skin response (GSR), collected from 36 participants engaged in switcher and arithmetic tasks designed to induce varying levels of mental workload. To classify binary MWL states, we trained various machine learning (ML) algorithms, including Multi-Layer Perceptron (MLP), Gradient Boosting (GB), and Support Vector Machine (SVM). This approach provided significant insights for optimizing human-computer interaction (HCI) systems. To improve model performance, we employed three different feature selection techniques: correlation-based, minimum redundancy maximum relevancy (mRMR), and domain-expert selection. We optimized hyperparameters using grid search cross-validation (CV) and validated the results through nested CV. Among the models, the MLP with correlation-based feature selection demonstrated the highest performance, reaching an area under the curve (AUC) of 0.822. The GB and SVM models also performed well with mRMR and domain-expert feature selection (AUC = 0.760 and 0.741, respectively). To provide interpretability and better understand feature importance in HCI contexts, SHapley Additive exPlanations (SHAP) identified several GSR features and heart rate as key predictors of mental workload, offering critical insights for designing adaptive HCI systems that account for cognitive load.

**Keywords:** Human-Computer Interaction, Feature Selection, Heart Rate Variability, SHAP, Mental Workload

## 1. INTRODUCTION

The workplace has undergone significant changes since the mid-20th century, with the rise of knowledge-based work and the increasing need for flexible work [1]. These changes highlight the importance of detecting mental workload (MWL) for preventing cognitive fatigue, decision-making errors, and productivity declines. Although no single definition of mental workload is universally agreed upon, a commonly accepted description by Young and Stanton's defines it as the amount of attentional resources needed to fulfill performance requirements, affected by task complexity, available support, and previous experience [1]. MWL is often confused with cognitive load, but the two are

distinct; cognitive load focuses on instructional design, while MWL emphasizes task demands and resource allocation [2]. MWL is commonly discussed in ergonomics and human-machine interaction and is categorized into overload, underload, and optimal load states [3]. For this study, both terms will be used interchangeably, as in affective computing, the term "cognitive load" is more widely used.

Advances in sensor technology and machine learning (ML) have enabled automatic MWL detection in real-world environments, supporting self-monitoring and adaptive human-machine interaction (HMI) systems [4]. These systems adjust workload to prevent overload or under-stimulation [5]. Physiological signals, like heart



rate variability (HRV) and Galvanic Skin Response (GSR), offer a reliable means of detecting cognitive load, as they reflect changes in the body's autonomic nervous system [6], [7]. Wearable technologies now enable continuous, non-intrusive monitoring of these signals, making them invaluable for real-time MWL assessment.

While affective computing has been extensively studied, fewer publicly available datasets focus on mental workload compared to stress or emotion (for review, see [8], [9], [10], [11], [12]). Seitz et al. [13] reviewed cognitive load datasets, but few capture physiological signals from non-intrusive devices. Therefore, new open datasets are essential for advancing research, reproducing results, and validating findings [14]. Since physiological signal correlations are not necessarily strong, ML models are crucial for reliably detecting workload, as they can account for individual differences in latent mental workload. Additionally, creating datasets tailored to specific populations, such as Indonesia, is important because physiological responses like HRV can vary by ethnicity [15].

On the other hand, enhancing the interpretation of features used in building ML models is critical for better understanding predictions and promoting fairness, trust, and transparency [16], [17], [18]. The importance of explainable ML (XML) is particularly emphasized in healthcare, where clinicians must be confident in AI systems to provide the best care for patients [19]. Healthcare poses unique ethical, legal, and regulatory challenges, as decisions can affect well-being or life immediately. However, these challenges also apply to fields like Human Factors and Ergonomics. AI systems in these domains, which are used to optimize workspaces, assess mental workload, or improve ergonomics, should also be transparent so practitioners can evaluate their impact on worker safety. This transparency builds trust among engineers and safety managers and ensures compliance with safety regulations. While ergonomic risks may not be as immediate as in healthcare, poor ergonomic designs can still have long-term effects on physical and mental health, highlighting the need for transparent, data-driven AI interventions to minimize potential harm.

This transparency can be achieved by combining feature selection with explainable ML (XML) methods. When used together, feature selection and XML techniques provide a clearer understanding of how each feature contributes to predictions while reducing model complexity by removing irrelevant or redundant features. However, little research has been comparing domain expertise-driven and data-driven feature selection methods, particularly in HRV. Features should be chosen based on clinical or physiological motivations to enhance the contextual interpretation of model performance [16],

[20]. Since HRV research offers well-established guidelines for key parameters [21], [22], comparing ML models using both selection methods is essential for detecting mental workload, and these should be evaluated alongside XML techniques.

Thus, this study aims to establish a new dataset for detecting mental workload using multimodal physiological signals and evaluate the interpretability of ML models developed from this dataset. It improves our prior study [23], by incorporating GSR data alongside HRV to enhance model performance. The dataset focuses on mental workload elicitation tailored to an Indonesian context. It also compares feature selection approaches—domain-expert and data-driven—in detecting MWL. To further enhance model interpretability, we performed SHapley Additive exPlanations (SHAP) to clarify the contribution of features and interactions differentiating MWL from non-load states [24]. Key contributions of this study include:

1. Introducing a novel dataset focused on the underexplored Indonesian population, improving classifier performance for adaptive HMI.
2. Comparing domain-expert and data-driven feature selection in XML models, enhancing accountability, fairness, and transparency.
3. Developing an automatic mental workload detection system using accessible physiological signals from wearable devices, which offers potential for real-world mental workload monitoring.

## 2. LITERATURE REVIEW

### A. Mental Workload Assessment from Physiological Data

Measuring MWL can be challenging, given its multifaceted nature [1], [25]. Traditional methods have often relied on subjective self-reports or task performance metrics, but these approaches can be limited by their retrospective nature and the potential for bias. There has been a growing shift towards physiological measures, which offer objective and real-time insights into a user's cognitive load [25]. Recent advancements in sensor technologies have made physiological measurements more accessible, portable, and less intrusive. Tools now allow the collection of data such as HRV, electrodermal activity (EDA), respiration, electromyography (EMG), and photoplethysmography (PPG) without interrupting the primary task [25], [26]. These technologies enable researchers to gather precise and continuous data, which correlate directly with the user's mental state, offering a valuable alternative to traditional MWL assessments.

HRV and GSR, also known as EDA, are both valuable physiological indicators of mental workload, as they



reflect the autonomic nervous system's (ANS) activity. The ANS regulates the body's involuntary physiological responses and consists of two branches: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS is responsible for the "fight or flight" response during stress or high cognitive demand, while the PNS manages the "rest and digest" state, promoting relaxation.

HRV, the variation in time between heartbeats, is influenced by the balance between SNS and PNS activity [22]. Under increased mental workload, SNS activity rises, leading to an elevated heart rate and reduced HRV as the body becomes more alert and engaged with the task. Conversely, in more relaxed states, PNS activity dominates, resulting in lower heart rates and higher HRV, signaling a more adaptive and rested state. HRV can be derived from electrocardiography (ECG) and photoplethysmography (PPG), offering flexibility in physiological measurement.

Similarly, skin conductance response (SCR) is a direct measure of SNS activity. When cognitive demands or emotional arousal increase, the SNS activates sweat glands, especially in the palms, increasing the skin's electrical conductivity. Since SCR is primarily driven by SNS activity without significant influence from the PNS, it is an effective marker of arousal and stress.

HRV analysis can be performed in the time, frequency, and non-linear indices [21], [22], [27]. Time-domain measures assess the variability in the intervals between normal heartbeats (normal-to-normal intervals), frequency-domain measures quantify power distribution across different frequency bands using spectral analysis, and non-linear indices capture the complexity and unpredictability of heart rate time series. Table 1 summarizes the most commonly used HRV indices in psychophysiological research.

TABLE I. HRV PARAMETERS AND THEIR DESCRIPTION

Parameter	Unit	Description
Time-domain		
Mean RR	ms	Mean of all RR intervals. The RR interval and heart rate (HR) are hyperbolically related ( $HR \times RR \text{ interval} = 60,000$ )
RMSSD	ms	Root mean square of successive differences between adjacent NN intervals (NN-intervals refer to interbeat intervals from which artifacts have been removed)
SDNN	ms	Standard deviation of NN intervals
pNN50	%	Percentage of successive NN intervals that differ by more than 50 m
Frequency Domain		
ULF power	ms <sup>2</sup>	Absolute power of the ultra-low-frequency band (< 0.0033 Hz)
VLF power	ms <sup>2</sup>	Absolute power of the very-low-frequency band (0.0033–0.04 Hz)
LF power	ms <sup>2</sup>	Absolute power of the low-frequency band (0.04–0.15 Hz)

LF power	nu	Relative power of the low-frequency band (0.04–0.15 Hz) in normalized units
HF power	ms <sup>2</sup>	Absolute power of the high-frequency band (0.15–0.4 Hz)
HF power	Nu	Relative power of the high-frequency band (0.15–0.4 Hz) in normalized units
LF/HF	%	Ratio of LF to HF power
Total power	ms <sup>2</sup>	Sum of the energy in the ULF, VLF, LF, and HF bands
Non-linear		
SD1	ms	Poincaré plot standard deviation perpendicular the line of identity (short-term variability)
SD2	ms	Poincaré plot standard deviation along the line of identity (long-term variability)
SD2/SD1	%	Ratio of SD1-to-SD2

B. Related Work

Over the past decade, affective computing research has primarily focused on stress and emotional states, with relatively fewer studies dedicated to cognitive load inference. Understanding cognitive load is critical for developing adaptive HCI and human-machine interaction (HMI) systems that reduce distractions, minimize errors during demanding tasks, and enhance user performance. In the past five years, significant advancements have been made in detecting cognitive load using physiological signals and machine learning, as summarized in Table 2. Commonly used signals for workload detection include electrodermal activity (EDA), electrocardiography (ECG), and accelerometer (ACC) data.

As presented in Table 2, multimodal approaches generally achieve higher accuracy rates. In classifying participants' expertise during a pedagogy simulation scenario, the accuracy of single-modality signals was 53.32% for ECG and 79.35% for GSR [28]. However, combining these signals using the kNN algorithm increased accuracy to 83.0%. A similar trend was observed by [18], who, although not listed in the table due to a focus on stress detection devices, found that incorporating a second biosignal (EDA) significantly enhanced performance, further supporting the effectiveness of multisensor wearables. Furthermore, most studies utilized specific feature selection techniques, such as sequential backward floating selection (SBFS) and minimum redundancy maximum relevancy (mRMR), alongside various k-fold cross-validation methods. However, none of the studies explicitly reported using hyperparameter optimization strategies.

Furthermore, there are public datasets that classify mental or cognitive load. Since we focused on long-term MW monitoring scenarios, we limited our summary to easily accessible physiological signals. Consequently, we do not consider brain-activity-related databases but instead concentrate on MWL public datasets that include easily acquired signals such as ECG, PPG, and GSR.



TABLE II. OVERVIEW OF RELATED WORKS AND PUBLIC DATASET ON AUTOMATED MENTAL (COGNITIVE) WORKLOAD DETECTION IN CHRONOLOGICAL ORDERS

Study & Year	N Subject	Scenarios	Signals	Class	Feature Engineering	Ground truth	Best Acc
[23], 2023	34	d2 attention, Switcher	ECG	Binary: MWL	FS: mRMR, domain expert, SFBS NS Hype LOOCV	Activity Type	67.89% SVM
[4], 2020	18	Arithmetic test	EMG, ECG, EDA	Multiclass: easy	FS: ANOVA NS Hype 10-fold	Activity Type	78.30% BPNN
[29], 2020	23	Maastricht Acute Stress Test	EOG, ECG	Binary: task	FS: SFFS NS Hype 8-fold CV	Activity Type	94.0% XGB
[30], 2019	24	Social exposure, event recall, cognitive load, stressful videos	ECG	Binary: mental stress	mRMR NS Hyp. 10-fold CV	Activity Type	84.4% SVM
[31], 2019	16	Psychomotor vigilance task (PVT), <i>N</i> -back, visual search	ECG, EDA	Multiclass: types of task	NS FS NS Hyp LOOCV	Activity Type	66% kNN
[28], 2019	10	Trauma Simulation	ECG, GSR	Binary: novice and experts	LASSO NS Hyp. 5-fold CV	Expert level	83.0% kNN

Notes. FS = Feature Selection, Hype = Hyperparameter Tuning, Acc = Accuracy, ECG= Electrocardiograph, EDA= Electrodermal Activity, EOG= Electrooculography, GSR= Galvanic Skin Response, NS= Not specified, FS= feature selection method, Hyp= Hyperparameter Tuning, LASSO= Least Absolute Shrinkage and Selection Operator, SFBS= Sequential Backward Floating Selection, SFFS= Sequential Forward Floating Selection, mRMR= minimum redundancy maximum Relevancy, SVM= Support Vector Machine, XGB= Extreme Gradient Boosting, BPNN: Back Propagation Neural Network.

The datasets listed in Table 3 employ various strategies to induce various MW conditions, with the choice of stimuli being crucial. Different stimuli can evoke significantly different physiological responses or in some cases, no response. Thus, the selection of appropriate stimuli directly impacts their accuracy. For example, CLAS [32] used a series of interactive tasks like solving math problems, logic puzzles, and the Stroop test to generate different levels of MW. Participants were required to complete these tasks quickly to assess momentary cognitive load. In contrast, MMOD-COG [33] induced MW through various complex arithmetic tasks, with 4-operand arithmetic representing high complexity and 2-operand arithmetic representing low complexity. Similarly, CogLoad [34] and MAUS dataset [35] used the 2-back and 3-back conditions, where higher *N*-back levels induced greater MW.

Signals recorded in these datasets include ECG, breathing rate, fingertip and wrist PPG, 3-axis ACC, GSR,

and skin temperature. The datasets use different instruments for recording signals, resulting in varied sampling rates. In contrast, the CLAS and CogLoad datasets collected signals using wearable devices, and the MMOD-COG and MAUS datasets employed clinical-grade physiological monitoring systems. Furthermore, differences in MW stimuli and experimental setups lead to varying signal durations across the datasets. For example, the CLAS dataset provides 30-minute recordings per participant, while MMOD-COG offers 15-minute recordings. SVM was the most commonly used machine learning model in these public datasets, with accuracy rates ranging from 68.2% to 78.2%. Two datasets (MMOD-COG and CogLoad) used context (activity type) as ground truth, while CLAS relied on participant self-assessments, and MAUS employed NASA-TLX, a widely used self-report measure of mental workload [36].

TABLE III. PUBLICLY AVAILABLE DATASETS FROM RELATED WORK

Dataset (year)	N Subject	Scenario	Signals	Devices	Class (target)	Ground Truth	Feature Engineering	Best Acc
CLAS, 2019, [32]	59	Math & Logic Problem, Stroop Test, Neutral Music Video Clip,	ECG, GSR, PPG	Shimmer 3	High and Low	Self-assessment	NS FS Hyp: C & Gamma  NS Validation	SVM 78.2%
MMOD-COG, 2019, [33]	40	Arithmetic Task	ECG, GSR, Speech	BIOPAC	Low and High  Pre- and High	Context	NS FS Hyp: GridSearch Nested LOOCV	SVM 76.66%
CogLoad 2020, [34]	23	N-back Task	GSR, TEMP, ACC, ECG (PPG)	Microsoft Band	Rest and Task	Context	FS: Ranking method based on mutual information NS Hyp LOOCV	DT 68.2%
MAUS, 2021, [35]	22	N-back Task	ECG, PPG (Fingertip & Wrist), GSR, TEMP	Procomp Infiniti	High and Low Level	NASA-TLX	NS FS NS Hyp. 5-fold CV	SVM (71.6%)
Proposed dataset	36	Switcher test, Arithmetic task	ECG, EDA (GSR)	Polar H10 & E-sense	MWL and not	Context	Correlation, mRMR, Domain-expert Hyp: GridSearch Nested CV	MLP (5.0%)

Notes. ECG= Electrocardiograph, EDA: Electrodermal Activity, GSR= Galvanic Skin Response, PPG= Photoplethysmograph, TEMP= Skin Temperature, NS= Not specified, FS= feature selection method, Hyp= Hyperparameter Tuning, mRMR= minimum redundancy maximum Relevancy, SVM= Support Vector Machine, MLP = Multi-Layer Perceptron, DT = Decision Tree, NASA-TLX= NASA Task Load Index

### 3. METHOD

#### A. Participants

A total of 36 healthy participants (16 males, 20 females, age =  $21.9 \pm 1.38$  years, BMI =  $23.22 \pm 4.79$ ) were recruited through advertisements on our university campus. Participants were required to avoid consuming caffeinated beverages for at least two hours before the experiment [37] and had a regular sleep routine the night before [38]. All participants were right-handed. To be eligible, they needed to meet the following criteria: 1) no history of neurological, cardiac, or psychiatric disorders; 2) no long-term medical treatments; and 3) no allergies to adhesives or rubbing alcohol. Participation was voluntary, and written informed consent was obtained from each participant before the experiment. The protocol of the study's methodology adheres with the 1964 Declaration of Helsinki [36] and obtained approval from the Health Research Ethics Committee (1497/KEP-UNISA/VII/2022).

#### B. Instruments and Devices

The sensors utilized in this study included the Polar H10 Heart Rate Sensor and the Mindfield eSense Skin Response Sensor for collecting physiological data. The Polar H10, a validated ECG chest strap, measures heart rate at 130 Hz and transmits RR intervals to a smartphone for real-time recording [12], [18]. GSR data were collected using the eSense Skin Response Sensor, which detects sweat gland activity through electrodes placed on the index and middle fingers. The skin conductance level

(SCL), measured in microSiemens ( $\mu S$ ), was recorded as a time series on a smartphone at 5 Hz. While this study used the sensor's built-in output, future research could explore custom models for raw electrodermal data analysis.

#### C. Experimental Procedure

The experimental procedure began with an introduction where participants were informed about the data collection process, including its purpose and voluntary nature, with the option to withdraw at any time. After agreeing to participate, signing consent forms, and completing a demographics questionnaire, participants underwent the same experimental session, visiting the laboratory once.

Next, sensors were attached: the Polar H-10 electrocardiogram (ECG) on the chest and the Mindfield eSense GSR sensor on two fingers (index and middle) of the non-dominant hand. Baseline stress levels were recorded for five minutes while participants sat upright and avoided movement.

Participants then completed two cognitive tasks: a five-minute switcher task using PEBL software and a five-minute arithmetic task. The switcher task was implemented using the PEBL software. Participants were required to identify pairs of figure elements based on a switching feature rule among ten figures, each a unique combination of five colors, five shapes, and five letters displayed on a black screen (see Figure 1). Participants practiced briefly before the task, during which recording was paused. Following practice, participants completed nine blocks, each lasting until twelve correct answers

were recorded. The task lasted approximately five minutes, and reaction time (RT) and accuracy were recorded for each trial [39], [40] (see Fig. 1). The arithmetic task, part of the Trier Social Stress Test (TSST), required participants to count backward from 2043 in steps of seven under time pressure. All tasks followed well-documented and scripted protocols to ensure high data quality. Both the switcher and arithmetic tasks are widely accepted in the scientific community to induce autonomic nervous system related to mental workload [39], [40], [41].

After completing both tasks, participants rested for five minutes (recovery). Next, the sensors were removed, and participants completed the NASA-TLX survey on perceived workload (not used in this analysis). The total session duration was about 30 minutes.

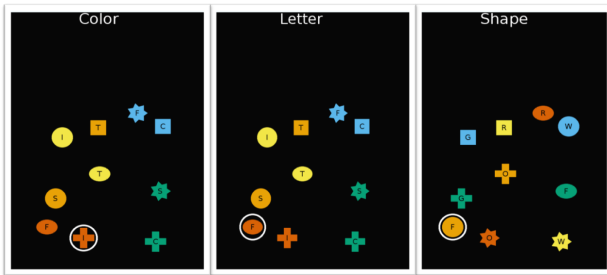


Figure 1. Switcher Featuring Task

#### D. Machine Learning Pipeline

An illustration of the machine learning pipeline, including data pre-processing, feature selection, model selection, and evaluation, is shown in Figure 2; begins with signal and raw data pre-processing, moves through the ML pipeline, and concludes with explainable ML (XML) and performance metrics. All ML algorithms in this study were implemented using the Scikit-Learn Python package.

We implemented binary classification in this study. Given that the MWL—Switcher and Arithmetic tasks—were administered uniformly to all participants in a controlled laboratory setting, the known activity type (context) served as the data labels. This approach helps minimize potential bias that could arise from subjective self-reporting. Two of the authors were present throughout all data collection sessions, carefully noting the exact start and end times of each session and the

corresponding activity type. The experiment was structured to differentiate between various levels of mental workload and non-MWL conditions, including baseline and recovery ("no load") phases, as well as the Switcher and Arithmetic tasks ("load").

##### 1) Data Preprocessing

Data pre-processing followed guidelines from the North American Society for Pacing and Electrophysiology and the Task Force of the European Society of Cardiology [21], [22]. Since there are no formal guidelines for GSR recording duration, we matched it to the HRV window. Using session timestamps, raw heart rate data was divided into four phases: Baseline, MWL 1, MWL 2, and Recovery.

A 5-minute window was applied, following the recommendation [21], [22]. The Elite HRV smartphone app extracted R-R intervals from the raw ECG data produced by the Polar H-10 [43]. Then, Kubios HRV (version 4.1.0), a scientifically validated software for HRV analysis, was used for artifact removal and HRV feature extraction [42]. It results in 24 HRV features across three domains: time (7 features), frequency (14 features), and non-linear (3 features). The final dataset contained 144.

For GSR, we quantified skin conductance, which consists of two components: tonic (static level) and phasic (fluctuating responses), typically referred to as Skin Conductance Responses (SCR). A higher number of SCRs per minute generally indicates increased stress. Consistent with the scientific literature, a relaxed state is indicated by 0 to 5 SCRs per minute [44]. The device recorded several parameters, including the number of SCRs per minute, maximum and minimum values, the difference between min/max, total SCR, and the percentage of SCR per session.

During the pre-processing stage, we applied the MinMaxScaler from Scikit-learn to perform min-max normalization. This technique adjusts each value by subtracting the feature's minimum and dividing the result by the feature's range (i.e., the difference between its maximum and minimum values). This process preserves the original distribution of the data without altering the underlying information in the features. By default, the MinMaxScaler scales values to a range between zero and one.

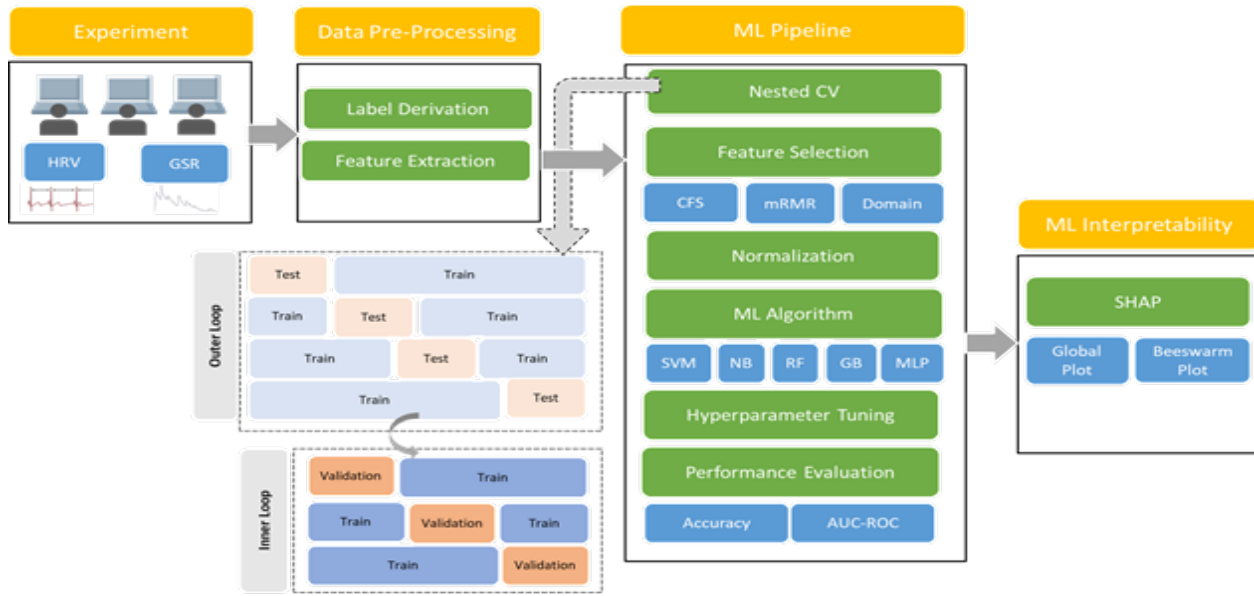


Figure 2. Steps of Our Methodology for the Mental Workload Detection with Physiological Signals and ML models

## 2) Feature Selection

Feature selection is a process that aims to identify the optimal subset of features from a more extensive set that best differentiates between classes [45]. Reducing dimensionality through feature selection can improve classification accuracy. The performance of classifiers often depends on the features used. Feature selection methods are typically classified into three categories: filters, wrappers, and embedded methods [29], [46]. In this study, compare three methods of feature selections as follows:

- **Domain Expertise.** From the 24 HRV features extracted using Kubios HRV software (v4.1.0), we selected six based on recommendations from prior cognitive and HRV research [21], [22]. Chosen features, both time and frequency domains, followed prior HRV and stress studies [47]. Key features include mean heart rate, SDNN, RMSSD, HF in normalized unit (HF nu), LF nu, and ratio LF to HF [22]. As to our knowledge, there is no recommendation from the profession or agreement on GSR; we relied on features provided by the internal system of the sensor manufacturer, so the total features used are six, including average SCR per minute, maximum value, minimum value, difference min/max, total SCR, and percentage of SCR per session.
- **Correlation-based Feature Selection (CFS).** This method evaluates feature subsets by considering their relevance to the target variable and their intercorrelation. Highly correlated features

(Pearson correlation coefficient  $> 0.8$ ) are considered redundant and are removed to retain the most relevant features. CFS is particularly effective when working with many features, as it helps identify the most important ones for the task.

- **Minimum redundancy maximum relevancy (mRMR).** This method selects features by considering both their relevance to the target variable and their redundancy with other selected features. The goal is to maximize relevance to the target while minimizing redundancy among features. Relevance is typically measured using F-statistics or mutual information, and redundancy is evaluated using Pearson correlation or mutual information. Features are ranked based on these criteria, and the most informative ones are selected iteratively [46].

Table 4 shows the final of selected features.

TABLE IV. LIST OF THE FEATURES SELECTED

Method	HRV	GSR
Domain-expert	hr, sdnn, rmssd, lfnu, hfnu, lf/hf	'scr', 'max', 'diff_minmax', 'pscr'
mRMR	'minhr', 'maxhr', 'peak_lf', 'hfnu', 'hr', 'rr', 'lfnu',	'scr', 'min', 'pscr',
CBS	'maxhr', 'peak_vlf', 'peak_lf', 'peak_hf', 'vlflog', 'hf', 'hfnu', 'lfhf', 'sd1', 'sd2', 'sd2sd1'	'scr', 'max', 'diff_minmax', 'pscr'

*Notes.* hr= heart rate, sdnn= standard deviation normal to normal, rmssd= root mean square standard deviation, lfnu= low frequency in normalized unit, hfnu= high frequency in normalized unit, lf/hf= ratio low frequency to high frequency, vlf= very low frequency, scr= skin conductance response, max= maximum amplitude, min= minimum amplitude of scr, diff\_minmax= differences between min and max amplitude of SCR, pscr= percentage of SCR



### 3) Cross Validation

In the ML approach, cross-validation helps evaluate model performance on unseen data. Given the relatively small size of our dataset and the participant-dependent nature of the experimental sessions, we employed leave-one-out cross-validation (LOOCV) across all ML models. This CV type was also used in prior ML models for cognitive load and stress detection[9], [34], [48], [49]. Both hyperparameter optimization and model selection can be conducted through this procedure. However, using the exact data for hyperparameter tuning and model evaluation can introduce data leakage, leading to overfitting and overly optimistic results. To avoid this, we employed a nested cross-validation strategy, ensuring the reliability and fairness of the model performance assessment. The procedure was as follows:

1. Outer Loop (Train-Test Split). Split the dataset into two subsets: 70% for training and 30% for testing. The test set is held out and will only be used for the final evaluation.
2. Inner Loop (Leave-One-Out Cross-Validation on Training Set)
  - a. Within the training set (70% of the data), employ leave-one-out cross-validation (LOOCV).
  - b. In each iteration of LOOCV, one sample is held out as validation, while the remaining samples are used to train the model.
  - c. During this step, hyperparameter tuning is performed using methods like GridSearch to find the optimal parameters for the model.
3. Hyperparameter Optimization
  - a. GridSearchCV is used within the inner loop to search for the best combination of hyperparameters across multiple iterations of LOOCV. (see Table 5)
  - b. After LOOCV is completed, the best model configuration (with optimized hyperparameters) is selected based on the performance in the inner loop.
4. Model Evaluation with the Outer Loop
  - a. Once the best hyperparameters are selected, the model is retrained on the entire training set (70%) using those parameters.
  - b. The model is then evaluated on the outer loop's held-out test set (30%) to assess its generalization performance on unseen data.

TABLE V. PREDEFINED HYPERPARAMETERS FOR THE GRIDSEARCH CV

Algorithm	Hyperparameter	Value
Decision Tree	min_samples_split	[2, 3, 4]
	min_samples_leaf	[1, 2, 3, 4]
	criterion	['gini', 'entropy', 'log_loss']
	max_depth	[1, 2, 3, 4]
Support Vector Machine	C (regularization parameter)	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 1]
	gamma (kernel coefficient)	[1, 0.1, 0.01, 0.001, 0.0001]
	degree (for poly kernels)	[1, 2, 3, 4, 5]
	kernel	['linear', 'poly', 'rbf', 'sigmoid']
Random Forest	n_estimators	[200, 250, 300]
	criterion	['gini', 'entropy', 'log_loss']
	max_depth	[1, 2, 3, 4]
	min_samples_split	[2, 3]
GradientBoost	min_samples_leaf	[1, 2, 3]
	loss	['log_loss', 'exponential']
	learning_rate	[0.1, 0.3, 0.4, 0.5, 1, 1.1]
Multi-Layer Perceptron	n_estimators	[10, 20, 30, 50, 70]
	hidden_layer_sizes	[(100,), (200,), (300,), (400,)]
	activation	['identity', 'logistic', 'tanh', 'relu']
	solver	['sgd', 'adam']

### 4) Classifiers (ML Algorithms)

The ML algorithms in this study were selected to compare classical techniques (SVM, kNN), ensemble methods (RF, GB), and Artificial Neural Networks (MLP) for detecting mental workload (MWL) based on our experimental data. These classifiers were chosen due to their frequent use and strong performance in HRV-based ML studies (for review, see [9], [10], [12]).

1. Support vector machines (SVM) are discriminative models designed to find an optimal hyperplane that separates data into different classes, particularly in high-dimensional spaces[50].
2. Naïve Bayes is a family of probabilistic classifiers based on Bayes' theorem. It operates with strong independence assumptions between features and is computationally efficient due to its simple algorithmic structure. Naïve Bayes assumes that the value of each feature is independent of others within the same class [51].
3. Random Forest (RF) is an ensemble learning method that combines multiple decision trees for accurate predictions [52]. It creates diverse trees by randomly sampling the dataset with replacements and makes final predictions by averaging or majority voting from these trees. RF is effective for classification and regression tasks, handling large

datasets and reducing overfitting. Therefore, they are widely used in ML applications.

4. Gradient Boosting (GB) is a popular ensemble machine learning technique that addresses the limitations of weak learners by iteratively improving their performance using gradient descent [53]. The loss function, which represents the difference between true and predicted values, is minimized with each step. By adding predictors sequentially, each new predictor corrects the errors of the previous one, thereby strengthening the overall model. GB is particularly effective in reducing noise, variance, and bias in the data, resulting in a more robust predictive model.
5. Multi-layer perceptron (MLP) is a type of neural network that processes information in a forward direction, from input to output [54]. It consists of three types of layers: the Input Layer, Hidden Layer(s), and Output Layer. In MLP, each node, except in the Input Layer, represents a neuron that applies a non-linear activation function to transform the weighted sum of its inputs into an output. The Input Layer receives the input signal, while the desired regression or classification task is performed in the Output Layer.

#### 5) Performance Metrics

In this study, machine learning performance was assessed using two primary metrics: accuracy and area under the curve (AUC). Accuracy, a basic metric, represents the ratio of correct predictions (true positives and true negatives) to the total number of predictions and is suitable for datasets with balanced class distributions, as in our case. However, for imbalanced datasets, accuracy can be misleading. To address this, we also utilized AUC-ROC (Area Under the Receiver Operating Characteristic Curve) as a complementary measure. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) for a binary classifier, and AUC quantifies the area under this curve. A perfect classifier has an AUC of 1.0, while an AUC of 0.5 indicates random guessing [54].

#### 6) Estimation of SHAP Value

In this study, SHapley Additive exPlanations (SHAP) are utilized to enhance the interpretability of our machine learning models. SHAP is an open-source, game-theoretic method that explains model predictions by calculating the contribution of each feature to the final output. It relies on Shapley values from game theory, representing each feature's average marginal contribution in a coalition. SHAP generates explanations by simulating all possible combinations of features and measuring their individual contributions to the model's predictions. This approach quantifies each feature's positive or negative impact on the

model's outcome, both on a global scale (across the entire dataset) and locally (for specific observations), thereby improving transparency in model interpretation. [24].

The critical advantage of SHAP is its ability to explain any machine learning model with a set of feature contributions, offering more interpretable insights than standard feature importance scores. In this study, SHAP was applied to our best-performing models, and two types of SHAP plots were generated—a global bar plot and a bee swarm plot—to provide deeper insights into the models' decision-making processes.

## 4. RESULTS

### A. Evaluation of ML Models

Table 6 summarizes the performance of five machine learning algorithms (SVM, NB, GB, RF, and MLP) evaluated using three feature selection methods: CFS, domain expertise, and mRMR. The results reveal different patterns in how these models interact with the chosen feature selection methods. SVM performed best when features were selected using domain expertise, achieving the highest AUC (0.7412), which indicates that manually selected, relevant features improved the model's ability to classify effectively. GB performed consistently across all methods. Random Forest, known for handling correlated features well, performed best with domain expertise, which helped reduce noise. However, its performance remained stable across all methods, with only slight improvements under mRMR. MLP excelled with CFS, achieving the highest accuracy and AUC across all models and methods. This result suggests that MLP benefits from correlated features to create complex patterns. However, domain expertise and mRMR negatively impacted its performance, indicating that MLP requires more diverse features to perform optimally. Overall, MLP and Gradient Boosting were the strongest performers, though their success depended heavily on the feature selection method. CFS benefited more complex models like MLP, while mRMR enhanced AUC for models such as Gradient Boosting by reducing feature redundancy.

TABLE VI. MACHINE LEARNING EVALUATION RESULTS WITH BINARY CLASSIFICATION ACCURACY



	CFS		mRMR		Domain Expertise	
	Acc	AUC	Acc	AUC	Acc	AUC
SVM	0.6591	0.7070	<b>0.6818</b>	0.6149	0.6364	<b>0.7412</b>
Naïve	0.5909	0.6128	0.5682	0.6377	0.6591	0.7019
GB	0.6818	0.6812	0.6364	<b>0.7598</b>	<b>0.6818</b>	0.6708
RF	0.5909	0.6563	0.6136	0.6749	0.6818	0.6584
MLP	<b>0.7500</b>	<b>0.8219</b>	0.5682	0.6853	0.5909	0.6439

Notes. CFS= Correlation Feature Selection, mRMR= minimum redundancy maximum relevancy. Best performance based on accuracy and AUC are marked in Bold.

### B. Model Interpretation with SHAP Values

Figure 3 visualizes feature contributions from each feature selection method using the SHAP bar and beeswarm plots for the best-performing tuned ML algorithms. For CFS, the plots are based on MLP; for domain expertise, on SVM; and for mRMR, on GB. The SHAP bar plot displays the mean absolute SHAP values, highlighting the relative importance of each feature in the model's predictions, with longer bars indicating more influential features.

For Correlation-based feature selection, maximum GSR amplitude has the highest SHAP value (0.18), indicating its most influential feature, followed by 'sd1' (non-linear HRV feature, see Table 1) and 'maxhr' (maximum heart rate). Features like 'sd2sd1' and 'LFHF' have less influence.

The SHAP beeswarm plot offers a comprehensive visualization of how each feature impacts the model's predictions across all data points. Features are ranked by importance along the y-axis, while the x-axis represents SHAP values, indicating the extent to which each feature increases or decreases the prediction. The color of each dot corresponds to the feature's value, with red signifying higher values and blue representing lower values. When multiple dots have the same SHAP value (x-axis position), they are stacked to represent density. The color scale on the right-hand side reflects the feature values, allowing for a quick visual assessment of how feature magnitude affects predictions. The plot highlights how feature values (represented by color) influence predictions, providing both global and local interpretability. For example, higher maximum GSR amplitude values are associated with positive predictions, while lower values have minimal features like 'diff\_minmax', which show more scattered effects, indicating a less consistent impact. The plot highlights that 'max', 'sd1', 'maxhr', and 'sd2' are the most influential features, driving predictions upwards when their values are high. The beeswarm plot clarifies each

feature's role and provides global and local insights into the model's decision-making.

The SHAP bar plot derived from the mRMR illustrates that 'percentage of skin conductance response' ('pscr') is the most influential feature with a SHAP value of 0.19, followed by an average of SCR ('scr') and heart rate ('hr') with values of 0.15 and 0.12, respectively. Lower-ranked features like maximum and minimum GSR amplitudes contribute less, with values around 0.03.

The SHAP beeswarm plot shows that 'pscr' and 'scr' appear as the most influential features, with higher values significantly increasing predictions, as shown by red dots to the right of zero. Lower-ranked features, such as 'minhr', 'rr', and 'hr', exhibit mixed effects with less consistent contributions. In summary, the SHAP bar and beeswarm plots emphasize the critical roles of 'pscr' and 'scr' in driving predictions, with other features like heart rate and HRV contributing variably to the model's decision-making process.

The SHAP bar plot, based on the domain expertise feature selection method, ranks 'pscr' (percentage of skin conductance response) as the most influential feature with a SHAP value of 0.19, followed by 'scr' and 'hr', with values of 0.15 and 0.12, respectively. Lower-ranked features, such as 'max' and 'min' (GSR amplitudes), contribute less with SHAP values around 0.03. Skin conductance and heart rate measures, particularly 'pscr', play the most critical roles in driving predictions.

The SHAP beeswarm plot complements this by showing how individual data points contribute to the model's predictions. Features such as 'pscr', 'scr', and 'hr' again emerge as the most important, with higher values (in red) pushing predictions upward. HRV metrics (like 'lfnu' and 'hfnu') and 'diff\_minmax' have more variable and inconsistent impacts, while features like 'max' and 'min' show weaker effects. In summary, 'pscr', 'scr', and 'hr' are important, while other physiological metrics contribute less consistently to the model's predictions.

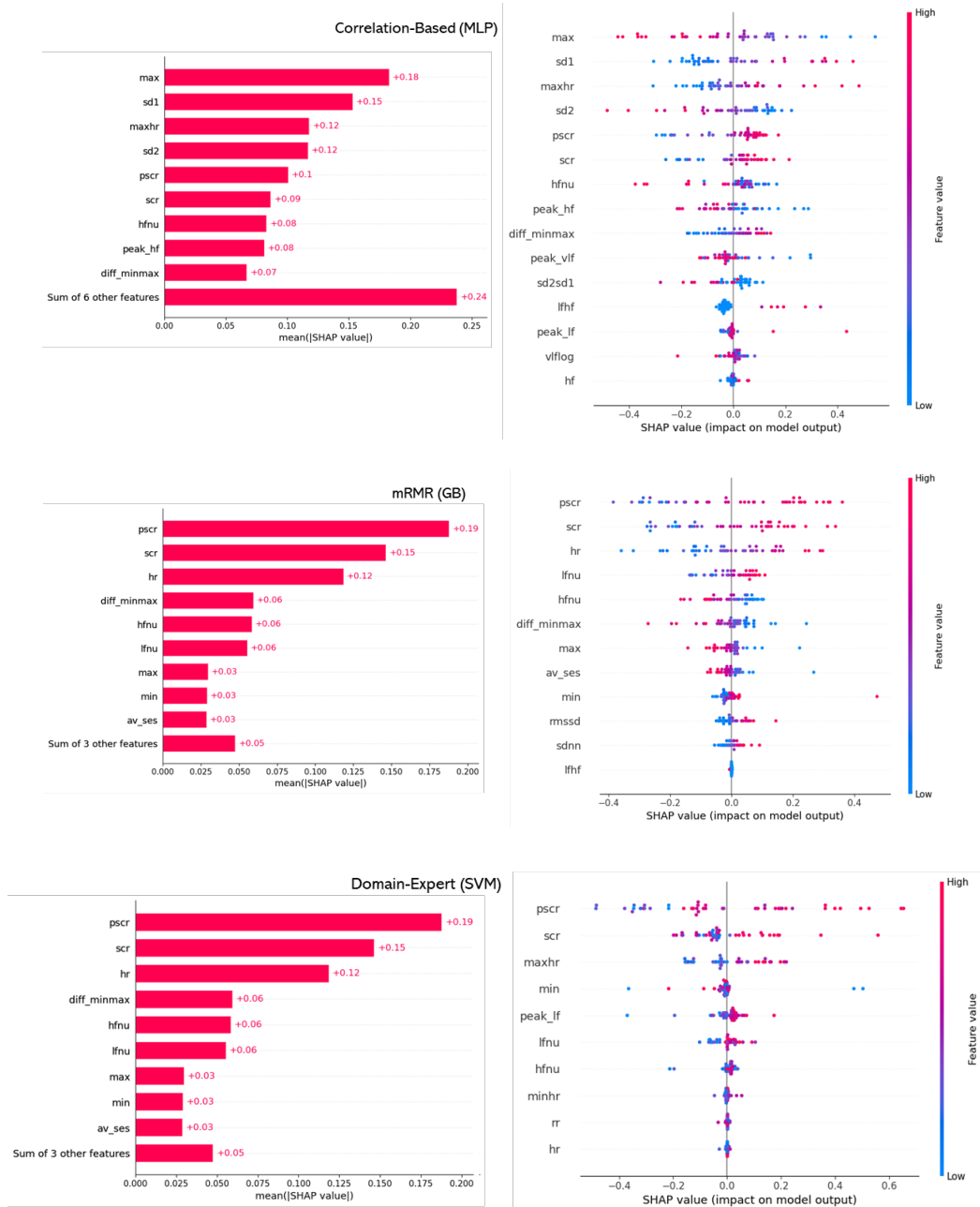


Figure 3. Global SHAP Plot (Left) and Beeswarm SHAP Plot (Right) for the Best Model from Each Feature Selection Method

### 5. DISCUSSION

In this study, we present a new dataset and develop a machine-learning pipeline to detect mental workload using physiological data from ECG (HRV) and GSR,

collected in a laboratory experiment with 36 participants across two types of MWL tasks. The pipeline integrates various algorithms, including SVM, kNN, RF, GB, and MLP. It also incorporates correlation, mRMR, and domain expert-based feature selection, nested CV,



hyperparameter tuning, and SHAP value computation for model interpretability.

Our findings show that the MLP model, trained on HRV and GSR signals, outperformed the other models in predicting mental workload, achieving the highest accuracy and AUC scores. GB performed best with the mRMR method, while SVM yielded the best results with domain-expertise-based feature selection.

#### A. Comparison with Other Public Dataset

Directly comparing our study with previous works is challenging due to substantial differences in experimental setups, including variations in datasets, sensors, pre-processing techniques, machine learning methods, classification tasks, and evaluation procedures. Moreover, discrepancies in the generated features and performance metrics make such comparisons even more difficult. While some prior studies report high performance in detecting cognitive or mental load, these results may be overly optimistic due to methodological limitations (Bencheikroun et al., 2023; Naegelin et al., 2023). Despite these constraints, we provide a comparison with selected notable studies, particularly those listed in Table 3, to offer context for our findings.

Regarding the experimental protocol used to generate training data, the work by [34] may be the most comparable to ours. Their experiment also aimed to induce cognitive load, with 23 participants completing 2-back and 3-back tasks. For ML model development, the authors utilized a feature selection ranking method based on mutual information and LOOCV. Their study employed various algorithms, RF, NB, KNN, Logistic Regression (LR), Decision Trees (DT), GB, XGB, and MLP, to distinguish the cognitive load from non-load from GSR and HRV features. Their best performance was achieved with DT (68.2%), while other algorithms, such as RF (67.9%), NB (60.8%), and MLP (64.3%), showed results similar to ours.

Moreover, as shown in Table 3, CLAS showed the highest accuracy of ML models for the public dataset, which resulted in 78.2% using SVM to detect high and low cognitive load using ECG and GSR signals [32]. However, this data employed various stimuli (mathematics & logic problems, Stroop test, neutral music, and video clip), administered in less than five minutes. These might not be enough to induce a mental workload.

#### B. Comparison across Feature Selection and Algorithms

Our results indicate that the MLP, a neural network-based ML algorithm, outperformed other methods in a binary classification of mental workload data using correlation-based feature selection (CFS), hyperparameter optimization, and nested cross-validation. Despite the

relatively small dataset, this strong performance can be attributed to CFS, which reduces dimensionality by selecting only the most relevant features, helping to prevent overfitting. By removing irrelevant or redundant features, MLP can focus on the most critical patterns in the data. MLP's capacity to model complex, non-linear relationships through its hidden layers and regularization techniques, such as dropout, further enhances its generalization ability. The superiority of MLP was also demonstrated in prior studies using a larger number of instances [18], [34], [55]

With regard to domain expertise feature selection, the superior performance of SVM in our study might be attributed to the linear nature of the selected features, especially those related to HRV. These features exhibit linear properties, which align well with SVM's strength in handling linearly separable data. Additionally, SVM is well-suited for small datasets like our datasets, as it maximizes the margin between classes. Our HRV data appears to have a distinct margin of separation, making SVM efficient and accurate in this context.

However, SVM tends to scale poorly with larger datasets, potentially becoming computationally expensive in future studies involving more subjects or additional physiological signals and cognitive states. This observation supports a work by Choi (2022), which noted that SVM often demonstrates superior generalization ability, mainly when training data is limited and across diverse samples [28], [29], [34].

#### C. Interpretation of Results with SHAP Values

The SHAP analysis across the three feature selection methods (CFS, mRMR, and domain expertise) consistently highlights the significance of specific physiological features. However, their relative influence varies depending on the selection method. Across all methods, 'pscr' (percentage of SCR) and 'scr' (average skin conductance response) consistently emerge as the most influential features, particularly concerning skin conductance, followed by heart rate. In contrast, features related to heart rate variability and GSR amplitudes play a less prominent and more inconsistent role in model predictions. This pattern emphasizes the critical importance of skin conductance and heart rate in driving the models' decision-making processes.

#### D. Implications

The findings of this study have several implications for mental workload detection and workplace applications. The strong performance of the MLP model, particularly when using HRV and GSR signals, suggests that neural networks have significant potential for real-time workload monitoring through wearable devices. In particular, GSR features, such as the percentage of SCR

and average SCR, consistently emerged as critical predictors, highlighting their importance in future stress and workload detection systems.

The study also shows that the choice of feature selection method plays a critical role in optimizing model performance. The variations observed, with MLP performing best with correlation-based selection, Gradient Boosting excelling with mRMR, and SVM performing well with domain-expertise-based selection, underline the need for tailored feature selection strategies to enhance model accuracy.

SHAP analysis also improves model interpretability by identifying the most influential features for MWL detection. This analysis added transparency, which makes the ML models more interpretable and trustworthy, which is crucial in workplace environments where transparency is necessary for user acceptance.

Lastly, while the laboratory-based results are promising, field studies are essential to validate these models in real-world conditions. Testing the models in actual work environments, such as office tasks or monitoring jobs, would increase their external validity and refine the models for practical use. Ultimately, this could lead to the development of personalized, continuous workload detection systems that adapt to individual employee needs, improving both productivity and overall health in the workplace.

#### E. Limitations

While this research shows the potential to predict mental workload (MW) using easily acquired physiological signals, several limitations should be noted. First, the small, homogeneous sample of young university students limits generalizability. Future studies should involve larger, more diverse samples. Second, although we used the recommended five-minute window segments [21], [22], shorter windows (30–60 seconds), commonly used in HRV-based studies [55], could increase data points, enable multiclass MWL levels classification and the use of deep learning models. Third, incorporating simulated tasks that resemble real-life scenarios would improve external validity. Including subjective self-reports, as done in other studies [4], [16], [34], would also offer insights into participants' perceived stress or relaxation levels, providing a fuller understanding of MW as both an input and output [25].

Multimodal data integration, such as eye movements, facial expressions, and task performance, is recommended for a more comprehensive analysis. Lastly, while we used two separate devices (Polar H10 and eSense) to collect signals, future studies could benefit from a single wearable device like the wrist-worn Empatica E4, which, though less precise, would be more practical and less intrusive for field research.

## 6. CONCLUSION

Operators' cognitive or mental workload is crucial for optimizing human resource allocation, improving task performance, and ensuring well-being and safety. This study introduces a new dataset that predicts mental workload from easily acquired physiological signals, specifically HRV and GSR. The data were collected in a laboratory setting where participants completed switcher and arithmetic tasks designed to induce mental workload. The study also explores various feature selection techniques, extensive hyperparameter tuning, and machine learning algorithms to classify resting versus workload states based on HRV and GSR features.

Our results show that the MLP algorithm achieved the best performance for MWL detection with feature selection, while Gradient Boosting performed best with mRMR and SVM, which excelled using domain knowledge. Additionally, using SHAP values as an explainable ML method highlighted several HRV and GSR features strongly associated with mental workload level.

Future research should include field studies to validate these findings in real-world conditions, such as office work or monitoring tasks with actual employees. Such studies would enhance the external validity of the research and contribute to the development of adaptive, personalized systems for continuous mental workload detection in workplace environments.

## ACKNOWLEDGMENT

This work was supported by the Directorate of Indonesian Higher Education under Research Grant for Junior Lectures [Grant number: 068/SP2H/PT/LL7/2024]

## REFERENCES

- [1] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, Jan. 2015, doi: 10.1080/00140139.2014.956151.
- [2] T. Luong, A. Lecuyer, N. Martin, and F. Argelaguet, "A Survey on Affective and Cognitive VR," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 12, pp. 5154–5171, Dec. 2022, doi: 10.1109/TVCG.2021.3110459.
- [3] R. McKendrick, B. Feest, A. Harwood, and B. Falcone, "Theories and Methods for Labeling Cognitive Workload: Classification and Transfer Learning," *Frontiers in Human Neuroscience*, vol.



- 13, no. September, pp. 1–20, 2019, doi: 10.3389/fnhum.2019.00295.
- [4] Y. Ding, Y. Cao, V. G. Duffy, Y. Wang, and X. Zhang, “Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning,” *Ergonomics*, vol. 63, no. 7, pp. 896–908, Jul. 2020, doi: 10.1080/00140139.2020.1759699.
- [5] V. Borisov, E. Kasneci, and G. Kasneci, “Robust cognitive load detection from wrist-band sensors,” *Computers in Human Behavior Reports*, vol. 4, no. June, p. 100116, 2021, doi: 10.1016/j.chbr.2021.100116.
- [6] F. Chen *et al.*, “Galvanic Skin Response-Based Measures,” in *Robust Multimodal Cognitive Load Measurement*, in Human–Computer Interaction Series. , Cham: Springer International Publishing, 2016, pp. 87–99. doi: 10.1007/978-3-319-31700-7\_5.
- [7] T. Pereira, P. R. Almeida, J. P. S. Cunha, and A. Aguiar, “Heart rate variability metrics for fine-grained stress level assessment,” *Computer Methods and Programs in Biomedicine*, vol. 148, pp. 71–80, 2017, doi: 10.1016/j.cmpb.2017.06.018.
- [8] P. J. Bota, C. Wang, A. L. N. Fred, and H. Placido Da Silva, “A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals,” *IEEE Access*, vol. 7, pp. 140990–141020, 2019, doi: 10.1109/ACCESS.2019.2944001.
- [9] S. Gedam and S. Paul, “A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques,” *IEEE Access*, vol. 9, pp. 84045–84066, 2021, doi: 10.1109/ACCESS.2021.3085502.
- [10] S. S. Panicker and P. Gayathri, “A survey of machine learning techniques in physiology based mental stress detection systems,” *Biocybernetics and Biomedical Engineering*, vol. 39, no. 2, pp. 444–469, 2019, doi: 10.1016/j.bbe.2019.01.004.
- [11] P. Schmidt, A. Reiss, R. Dürichen, and K. V. Laerhoven, “Wearable-based affect recognition—a review,” *Sensors (Switzerland)*, vol. 19, no. 19, pp. 1–42, 2019, doi: 10.3390/s19194079.
- [12] G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, “Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review,” *International Journal of Medical Informatics*, vol. 173, no. February, p. 105026, 2023, doi: 10.1016/j.ijmedinf.2023.105026.
- [13] J. Seitz and A. Maedche, “Biosignal-Based Recognition of Cognitive Load: A Systematic Review of Public Datasets and Classifiers,” in *Information Systems and Neuroscience*, vol. 58, F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, A. B. Randolph, and G. R. Müller-Putz, Eds., in Lecture Notes in Information Systems and Organisation, vol. 58. , Cham: Springer International Publishing, 2022, pp. 35–52. doi: 10.1007/978-3-031-13064-9\_4.
- [14] L. Longo, C. D. Wickens, G. Hancock, and P. A. Hancock, “Human Mental Workload: A Survey and a Novel Inclusive Definition,” *Front. Psychol.*, vol. 13, p. 883321, Jun. 2022, doi: 10.3389/fpsyg.2022.883321.
- [15] L. K. Hill *et al.*, “Ethnic differences in resting heart rate variability: A systematic review and meta-analysis,” *Psychosomatic Medicine*, vol. 77, no. 1, pp. 16–25, 2015, doi: 10.1097/PSY.000000000000133.
- [16] M. Bahameish, T. Stockman, and J. Requena Carrión, “Strategies for Reliable Stress Recognition: A Machine Learning Approach Using Heart Rate Variability Features,” *Sensors*, vol. 24, no. 10, p. 3210, May 2024, doi: 10.3390/s24103210.
- [17] J. S. Banerjee, M. Mahmud, and D. Brown, “Heart Rate Variability-Based Mental Stress Detection: An Explainable Machine Learning Approach,” *SN COMPUT. SCI.*, vol. 4, no. 2, p. 176, Jan. 2023, doi: 10.1007/s42979-022-01605-z.
- [18] N. Chalabianloo, Y. S. Can, M. Umair, C. Sas, and C. Ersoy, “Application level performance evaluation of wearable devices for stress classification with explainable AI,” *Pervasive and Mobile Computing*, vol. 87, p. 101703, Dec. 2022, doi: 10.1016/j.pmcj.2022.101703.
- [19] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, “The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies,” *Journal of Biomedical Informatics*, vol. 113, p. 103655, Jan. 2021, doi: 10.1016/j.jbi.2020.103655.
- [20] B. Remeseiro and V. Bolon-Canedo, “A review of feature selection methods in medical applications,” *Computers in Biology and Medicine*, vol. 112, p. 103375, Sep. 2019, doi: 10.1016/j.compbiomed.2019.103375.
- [21] S. Laborde, E. Mosley, and J. F. Thayer, “Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting,” *Frontiers in Psychology*, vol. 8, no. 213, 2017, doi: 10.3389/fpsyg.2017.00213.
- [22] M. Malik, “Heart rate variability Standards of measurement, physiological interpretation, and clinical use,” *European Heart Journal*, vol. 17, no.

- 3, pp. 354–381, 1996, doi: 10.4324/9781315372921.
- [23] N. Izzah, A. P. Sutarto, A. Hendi, M. Ainiyah, and M. N. B. Abdul Wahab, “Physiological Signals as Predictors of Mental Workload: Evaluating Single Classifier and Ensemble Learning Models,” *J. Optimasi Sist. Ind.*, vol. 22, no. 2, pp. 81–98, Dec. 2023, doi: 10.25077/josi.v22.n2.p81-98.2023.
- [24] S. M. Lundberg *et al.*, “From local explanations to global understanding with explainable AI for trees,” *Nat Mach Intell*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.
- [25] R. L. Charles and J. Nixon, “Measuring mental workload using physiological measures: A systematic review,” *Applied Ergonomics*, vol. 74, no. September 2016, pp. 221–232, 2019, doi: 10.1016/j.apergo.2018.08.028.
- [26] D. Tao, H. Tan, H. Wang, X. Zhang, X. Qu, and T. Zhang, “A systematic review of physiological measures of mental workload,” *International Journal of Environmental Research and Public Health*, vol. 16, no. 15, pp. 1–23, 2019, doi: 10.3390/ijerph16152716.
- [27] M. Mohanty, P. Sekhara Rath, and A. G. Mohapatra, “IoMT-based Heart Rate Variability Analysis with Passive FBGSensors for Improved Health Monitoring,” *IJCDS*, vol. 15, no. 1, pp. 1135–1147, Mar. 2024, doi: 10.12785/ijcds/150180.
- [28] K. Ross *et al.*, “Toward Dynamically Adaptive Simulation: Multimodal Classification of User Expertise Using Wearable Devices,” *Sensors*, vol. 19, no. 19, p. 4270, Oct. 2019, doi: 10.3390/s19194270.
- [29] K. Pettersson, J. Tervonen, J. Narvainen, P. Henttonen, I. Maattanen, and J. Mantyjarvi, “Selecting Feature Sets and Comparing Classification Methods for Cognitive State Estimation,” *Proceedings - IEEE 20th International Conference on Bioinformatics and Bioengineering, BIBE 2020*, pp. 683–690, 2020, doi: 10.1109/BIBE50027.2020.00115.
- [30] G. Giannakakis, K. Marias, and M. Tsiknakis, “A stress recognition system using HRV parameters and machine learning techniques,” *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2019*, pp. 269–272, 2019, doi: 10.1109/ACIIW.2019.8925142.
- [31] H. F. Posada-Quintero and J. B. Bolkhovsky, “Machine learning models for the identification of cognitive tasks using autonomic reactions from heart rate variability and electrodermal activity,” *Behavioral Sciences*, vol. 9, no. 4, 2019, doi: 10.3390/bs9040045.
- [32] V. Markova, T. Ganchev, and K. Kalinkov, “CLAS: A Database for Cognitive Load, Affect and Stress Recognition,” in *2019 International Conference on Biomedical Innovations and Applications (BIA)*, Varna, Bulgaria: IEEE, Nov. 2019, pp. 1–4. doi: 10.1109/BIA48344.2019.8967457.
- [33] I. Mijic, M. Sarlija, and D. Petrinovic, “MMOD-COG: A Database for Multimodal Cognitive Load Classification,” in *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Dubrovnik, Croatia: IEEE, Sep. 2019, pp. 15–20. doi: 10.1109/ISPA.2019.8868678.
- [34] M. Gjoreski *et al.*, “Datasets for cognitive load inference using wearable sensors and psychological traits,” *Applied Sciences (Switzerland)*, vol. 10, no. 11, 2020, doi: 10.3390/app10113843.
- [35] W.-K. Beh and Y.-H. Wu, “MAUS: A Dataset for Mental Workload Assessment on N-back Task Using Wearable Sensor,” *Arxiv*, 2021.
- [36] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” *Advances in Psychology*, vol. 52, no. C, pp. 139–183, 1988, doi: 10.1016/S0166-4115(08)62386-9.
- [37] F. Zimmermann-Viehoff, J. Thayer, J. Koenig, C. Herrmann, C. S. Weber, and H.-C. Deter, “Short-term effects of espresso coffee on heart rate variability and blood pressure in habitual and non-habitual coffee consumers – A randomized crossover study,” *Nutritional Neuroscience*, vol. 19, no. 4, pp. 169–175, Apr. 2016, doi: 10.1179/1476830515Y.0000000018.
- [38] P. K. Stein and Y. Pu, “Heart rate variability, sleep and sleep disorders,” *Sleep Medicine Reviews*, vol. 16, no. 1, pp. 47–66, Feb. 2012, doi: 10.1016/j.smr.2011.02.005.
- [39] R. Langner, F. Scharnowski, S. Ionta, C. E. G. Salmon, B. J. Piper, and G. S. P. Pamplona, “Evaluation of the reliability and validity of computerized tests of attention,” *PLoS ONE*, vol. 18, no. 1, p. e0281196, Jan. 2023, doi: 10.1371/journal.pone.0281196.
- [40] S. T. Mueller and B. J. Piper, “The Psychology Experiment Building Language (PEBL) and PEBL Test Battery,” *Journal of Neuroscience Methods*, vol. 222, pp. 250–259, Jan. 2014, doi: 10.1016/j.jneumeth.2013.10.024.
- [41] N. F. Narvaez Linares, V. Charron, A. J. Ouimet, P. R. Labelle, and H. Plamondon, “A systematic review of the Trier Social Stress Test methodology: Issues in promoting study comparison and replicable research,” *Neurobiology*



- of Stress*, vol. 13, p. 100235, Nov. 2020, doi: 10.1016/j.yinstr.2020.100235.
- [42] Kubios, "User's Guide Kubios HRV Scientific." Kubios Oy, 2023.
- [43] A. S. Perrotta, A. T. Jeklin, B. A. Hives, L. E. Meanwell, and D. E. R. Warburton, "Validity of the Elite HRV Smartphone Application for Examining Heart Rate Variability in a Field-Based Setting," *Journal of Strength and Conditioning Research*, vol. 31, no. 8, pp. 2296–2302, Aug. 2017, doi: 10.1519/JSC.0000000000001841.
- [44] "Mindfield eSense Skin Response - GSR sensor for iPhone & Android." Accessed: Sep. 05, 2024. [Online]. Available: <https://bio-medical.com/mindfield-esense-skin-response-gsr-sensor-for-iphone-andriod.html>
- [45] H. Alshaher, "Studying the Effects of Feature Scaling in Machine Learning," Dissertation, North Carolina Agricultural and Technical State University, 2021.
- [46] M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic, "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data," *BMC Bioinformatics*, vol. 18, no. 1, p. 9, Dec. 2017, doi: 10.1186/s12859-016-1423-9.
- [47] C. Rudin and J. Radin, "Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition," *Harvard Data Science Review*, vol. 1, no. 2, Nov. 2019, doi: 10.1162/99608f92.5a8a3a3d.
- [48] M. Choi and J. J. Jeong, "Comparison of Selection Criteria for Model Selection of Support Vector Machine on Physiological Data with Inter-Subject Variance," *Applied Sciences (Switzerland)*, vol. 12, no. 3, 2022, doi: 10.3390/app12031749.
- [49] W. Kraaij *et al.*, "Personalized support for well-being at work: an overview of the SWELL project," *User Modeling and User-Adapted Interaction*, vol. 30, no. 3, pp. 413–446, 2020, doi: 10.1007/s11257-019-09238-3.
- [50] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/j.neucom.2019.10.118.
- [51] B. Alam, "Naive Bayes Classifier Python Tutorial 2023." Accessed: Aug. 11, 2023. [Online]. Available: <https://hands-on.cloud/naive-bayes-classifier-python-tutorial/>
- [52] L. Breiman, "Random Forest," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [53] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Lecture Notes in Computer Science*, vol. 1857, no. 2, Berlin, Heidelberg, 2000, pp. 1–15. doi: 10.1007/3-540-45014-9\_1.
- [54] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. Sebastopol, California: O'Reilly Media, Inc, 2019.
- [55] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–13, 2019, doi: 10.1186/s12911-019-0742-y.



**Auditya Purwandini Sutarto** is a faculty member in the Department of Industrial Engineering at Universitas Qomaruddin Gresik, Indonesia. She holds a Bachelor's degree in Industrial Engineering from Institut Teknologi Bandung, a Master's degree in Mathematics from Universitas Gadjah Mada, and a PhD in Technology Management from Universiti Malaysia

Pahang. Her research interests lie at the intersection of psychophysiology, ergonomics, and artificial intelligence, demonstrating a strong dedication to interdisciplinary research.



**Mega Bagus Herlambang** holds a PhD in Human Factors from the University of Groningen, Netherlands. He currently serves as the Head of the Department of Industrial Engineering at Institut Teknologi Indonesia (ITI) in South Tangerang, Indonesia, where he has been a faculty member since 2014. In addition to his academic role, he is an independent researcher, a

professional data scientist, and a consultant specializing in human development and small-medium enterprises. He also shares his expertise through his YouTube channel on data science



**Nailul Izzah** is a lecturer in the Department of Industrial Engineering at Universitas Qomaruddin Gresik, Indonesia. She holds a Bachelor's degree in Mathematics Education from Universitas Islam Malang and a Master's degree in Health from Universitas Airlangga, Surabaya. With more than 15 years of

experience in teaching mathematics and statistics, her research focuses on clustering and applied statistics



**Ade Hendi** earned both his Bachelor's and Master's degrees in Computer Science from Institut Sains dan Teknologi Palapa, Malang, and Institut Teknologi Surabaya, respectively. He is currently a faculty member in the Department of Informatics Engineering at Universitas Qomaruddin, Gresik, Indonesia where he teaches system programming and mobile computing. His research interests span across the fields of the Internet of Things and mobile computing.

SURAT PERNYATAAN TANGGUNG JAWAB BELANJA

Yang bertanda tangan di bawah ini :

Nama : AUDITYA PURWANDINI SUTARTO S.T, M.Si, PhD

Alamat : Jln Nias 17 GKB Kebomas

berdasarkan Surat Keputusan Nomor 109/E5/PG.02.00.PL/2024 dan Perjanjian / Kontrak Nomor 068/SP2H/PT/LL7/2024, 474.06/A.2/UQ.7/MoU/VI/2024 mendapatkan Anggaran Penelitian Pengembangan MindPro, Model Prediksi Beban Kerja Mental berbasis Sinyal Fisiologi dengan Hybrid Machine Learning dan Analisis Cross Dataset Sebesar Rp.47.010.000

Dengan ini menyatakan bahwa :

1. Biaya kegiatan Penelitian di bawah ini meliputi :

No	Uraian	RAB 100%	Realisasi
1	Bahan ATK Bahan Penelitian Habis Pakai Barang persediaan	Rp.9.350.000	Rp.9.350.000
2	Pengumpulan Data HR Pembantu Peneliti Uang Harian Biaya konsumsi Transport	Rp.9.700.000	Rp.9.700.000
3	Analisis Data Transport Lokal HR Pengolah Data Honorarium Narasumber Biaya Analisis Sampel	Rp.12.500.000	Rp.12.500.000
4	Sewa Peralatan Obyek Penelitian Peralatan Penelitian	Rp.10.500.000	Rp.10.500.000
5	Pelaporan Luaran Wajib Uang Harian Rapat diluar Kantor Biaya Konsumsi Biaya Publikasi Artikel di Jurnal Bereputasi Internasional	Rp.4.960.000	Rp.4.960.000
6	Lain-lain	Rp.0	Rp.0
Realisasi ( 100 % )			Rp.47.010.000

2. Jumlah uang tersebut pada angka 1, benar-benar dikeluarkan untuk pelaksanaan kegiatan Penelitian dimaksud.

Demikian surat pernyataan ini dibuat dengan sebenarnya.

Gresik, 20-12-2024, Ketua



AUDITYA PURWANDINI SUTARTO S.T, M.Si, PhD

NIP/NIPK 0011067801