Application of Bayesian Networks for classifying Cognitive load signals

ỨNG DỤNG BAYESIAN NETWORK ĐỂ PHÂN LOẠI TÍN HIỆU NHẬN THỨC

Đỗ Minh Quân, Lê Thị Mỹ Hương, Trần Bùi Ty Ty

INTRODUCTION

Problems definition:

- **Input:** Physiological data from wearable devices
- **Output:** Bayesian network infers and classifies cognitive states

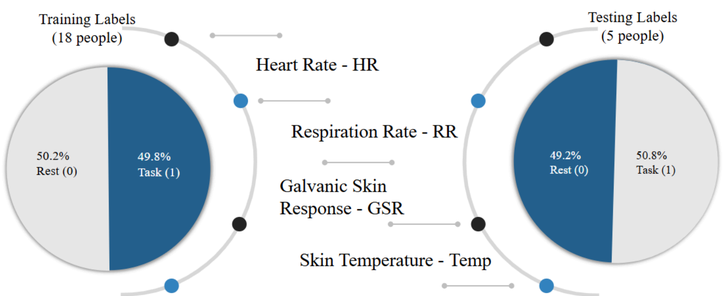
Challenge:

- Cognitive load is a latent state – it cannot be measured directly.
- Physiological signals (GSR, HR, Temp, RR) are indirect indicators.

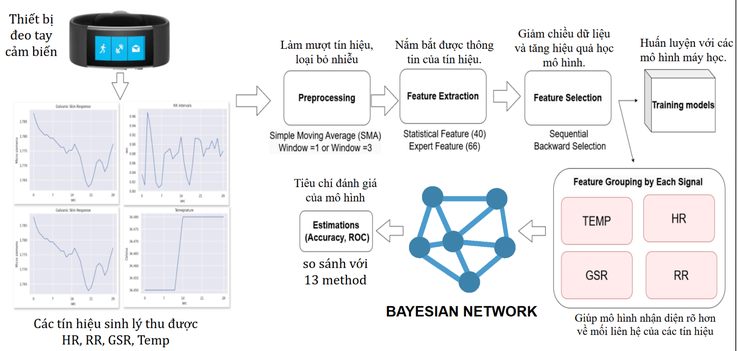
Thesis Goal: Analyze physiological signals and model their causal relationship with cognitive load using Bayesian Networks

DATASETS

- Quantity: 825 samples from 23 participants. 632 samples (training set) and 193 samples (test set).
- Data source: UbiComp2020
  https://colab.research.google.com/drive/1TcGrzhtlb6XdIE33l5cnedPeaPlyNrNE?usp=sharing



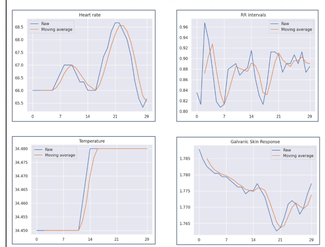
PROPOSED METHOD



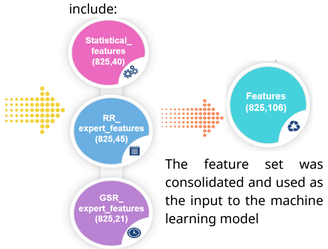
Overview of Cognitive Load Classification using Bayesian Network

1 Preprocessing & Feature Extraction

Raw (unsmoothed) physiological signals (blue) and smoothed physiological signals (orange)



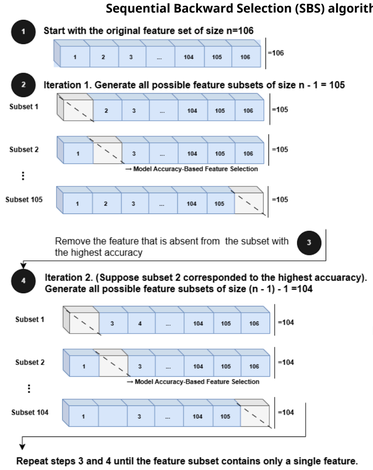
Feature groups extracted from physiological data include:



The feature set was consolidated and used as the input to the machine learning model

2 Feature Selection & Training Models

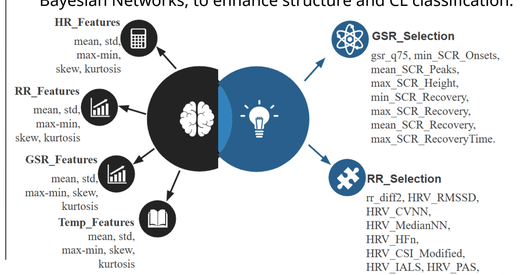
Sequential Backward Selection (SBS) algorithm



Using the SBS algorithm to select the optimal feature set that yields the highest model accuracy

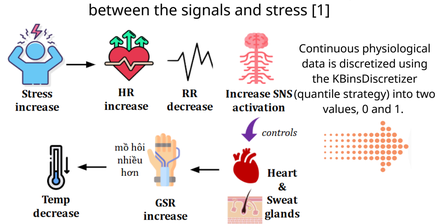
3 Feature Grouping By Each Signals

Combine statistical and selected features, organized by signal in Bayesian Networks, to enhance structure and CL classification.



4 Analyze the relationship between physiological signals

A 2021 research analyzed the relationship between the signals and stress [1]



Continuous physiological data is discretized using the KBinsDiscretizer (quantile strategy) into two values, 0 and 1.

Rest
1 HR↓
0 RR↑
1 GSR↓
0 Temp↑

Task
0 HR↑
1 RR↓
0 GSR↑
1 Temp↓

EXPERIMENTS

Experiment 1: Identify the model with the highest performance. Evaluate and compare models to find the one with the optimal feature set based on accuracy.

Experiment 2: Identify the nodes and edges. Group features by signal type and assess the impact of adding feature groups to the model.

Experiment 3: Examine the influence of observed variables on the target variable. Identify which physiological signals have the strongest effect on cognitive load

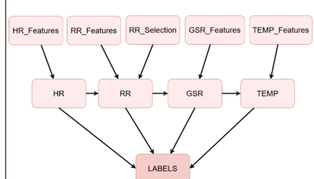
Experiment 1: Identify the model with the highest performance

Experimental results of 10 models with SMA = 0

Model	Features	Accuracy	ROC
SVM	52	0.6788	0.69
XGB	85	0.7098	0.71
LR	11	0.7098	0.73
RF	**21**	**0.7565**	**0.73**
MLP(SKLearn)	25	0.6943	0.71
MLP(Keras)	11	0.5751	0.50
E7GB	24	0.7306	0.75
WGLR	31	0.7306	0.74
ESVM	16	0.7047	0.68
CNN	106	0.6010	0.61
RNN	106	0.6528	0.67

A total of 21 important features were selected using the Random Forest model. The features are categorized based on signal type as follows:

- GSR: 8 features
- RR: 11 features
- Temp: 2 features
- HR: 0 features



The Bayesian Network model uses groups of input signals to predict Task/Rest state

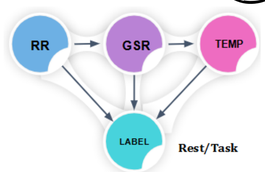
Experiment 2: Identify the nodes and edges

Experiments using different combinations of physiological features

Signal groups (number of features)	Accuracy
• HR(5): HR_Features • RR(5): RR_Features • GSR(5): GSR_Features • Temp(5): Temp_Features	0.6269
• HR(5): HR_Features • RR(5): RR_Features • GSR(13): GSR_Features, GSR_Selection • Temp(5): Temp_Features	0.6269
• **HR(5): HR_Features** • **RR(16): RR_Features, RR_Selection** • **GSR(5): GSR_Features** • **Temp(5): Temp_Features**	**0.6425**
• HR(5): HR_Features • RR(16): RR_Features, RR_Selection • GSR(13): GSR_Features, GSR_Selection • Temp(5): Temp_Features	0.5907

Experiment 3: Examine the influence of observed variables on the target variable

Signal impact table (exhaustive search) in Bayesian Task/Rest model

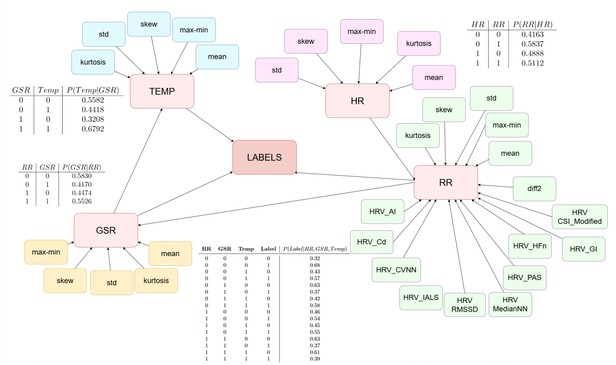


RR (HRV), GSR, and Skin Temperature all have clear physiological associations with stress

Case	Signals Influencing The Label	Accuracy
1	HR	0.5544
2	RR	0.6010
3	GSR	0.5803
4	Temp	0.4819
5	HR-RR	0.6010
6	HR-GSR	0.5699
7	HR-Temp	0.5544
8	RR-GSR	0.6010
9	RR-Temp	0.6010
10	GSR-Temp	0.5803
11	HR-RR-GSR	0.6062
12	HR-RR-Temp	0.5803
13	HR-GSR-Temp	0.5544
14	**RR-GSR-Temp**	**0.6580**
15	HR-RR-GSR-Temp	0.6425

RESULTS

COGNITIVE LOAD BAYESIAN NETWORK



SOTA results of participating teams in the UBiComp 2020 challenge

Method	Proposed Classifier	Accuracy
I	Ensemble of 7 Gradient boosting decision trees	0.694
II	Support vector machine	0.679
III	Ensemble of support vector machine	0.674
IV	**Logistic regression**	**0.663**
V	**Random Forest**	**0.653**
VI	Weighted sum of individualized and global logistic regression models	0.653
VII	Multilayer perceptron	0.648
VIII	Logistic regression	0.648
IX	Support vector machine	0.627
X	XGBoost Classifier	0.580
XI	Convolutional neural network...	0.560
XII	Recurrent neural network	0.554
XIII	Logistic regression	0.503
Our team	**Bayesian Networks**	**0.658**

Submitting paper "A Multiscale Approach for Inferring Cognitive Load from the CogLoad Physiological Dataset" ICCIES2025 - *International Conference On Computational Intelligence In Engineering Science 2025*