Psychophysiological modelling of trust in technology:

Comparative analysis of algorithm ensemble methods

Ighoyota Ben A. et al

Background

- Artificial intelligence (AI) technologies emergence has resulted to technologies:
    - that operate as a conceptual extension of human beings [1-3].
    - that are radically transforming the way people live globally.
- For instance,
    - intelligent personal assistants [4].
    - autonomous vehicles (AV) [5].
    - expert diagnosis recommend systems [10].

Motivation

- Although AI based systems outperforms their human counterpart [11],
  - they are incapable of formulating ideas and constructively interacting with users [12].
  - the algorithms that controls these AI based systems are not infallible [13, 14], E.g. Teslar AV [15] and Uber AV [16] accidents.
- Consequently resulting to negative user experience (e.g. less users trusting it) that increases skepticism and lowers adoption [17 - 20]. For instance,
  - most users believe that autonomous vehicles are less safe than manual driving [21],
  - robot assisting surgeons are not as widely adopted as initially anticipated due to the delicate nature of surgical procedures, the expertise and precision required.
- Trust influences reliance during complex or uncertain situations [22]. E.G,
  - during users interactions with: recommender agents, and knowledge management systems [23 - 29].

Motivation

- Trust is a net cognitive state, and because cognitive states could be adapted to inform and correct system states [30],
  - researchers have developed cognitive models (i.e., adaptive interfaces) to enable measuring trust in real-time (see [31] for detailed review).
- Such cognitive trust models consists of
  - trained classifier model based on users psychophysiological signal data from interactions.
  - Psychophysiological signals data are [32-34] because it continuous
- As a result,
  - measurement of user’s trust in real-time to enable technologies adapt their operation to user trust state [32-34] has been investigated.
- However, assessing users trust in AI based systems continues to remain a challenge [35], especially in real-time,
  - This is due to the lack of a stable, accurate, and non-bias and variance sensitive trust level classifier models.

Related works and gaps

There has been substantial efforts towards developing a classifier model for assessing users trust in technology with psychophysiological signals.

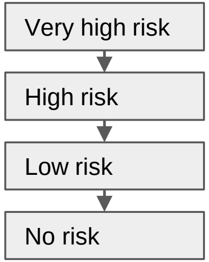
The use of single algorithms (KNN and Decision tree) resulted to models limited to the shortcoming of the individual algorithm. E.g In [36], and [38],

Though the combination of two more algorithms (ensemble) reduce classification error, improve variance sensitivity, and reduce bias [40], prior efforts utilized only voting ensemble method eg, [34, 32, 41, and 42] and are fairly accurate but unstable

Problem statement and goals

- As prior ensemble trust classifier models are unstable:
  - Only voting ensembling method has been utilized so far:
  - There are several ensembling methods
- Therefore the goal of this research article investigate is to investigate
  - what ensemble method should be used when developing a classifier model for assessing users trust with psychophysiological signals?
- Considering the
  - Four main algorithm ensemble methods (Boosting, Bagging, Voting, and Stacking).
  - Combination of four psychophysiological signals (EEG, EDA, ECG, and Facial EMG) that has been used in previous research [43 - 46].

Methodology



Very high risk

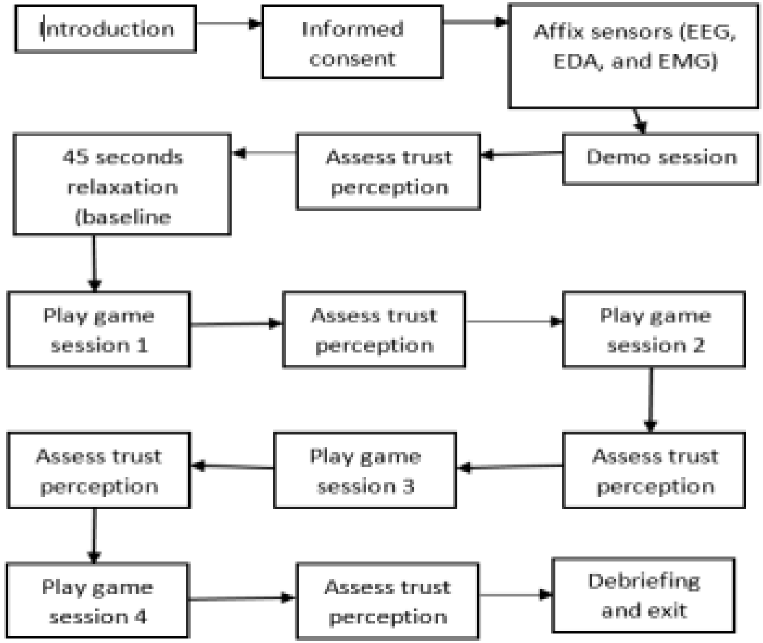
High risk

Low risk

No risk

- A within subject four condition experiment was implemented
- Apparatus used includes
    - Autonomous vehicle game (see [51] for detailed description).
    - An MSI core i7 high performance gaming computer.
    - A 30inch LCD monitor.
    - A joystick (Logitech 3D).
    - Lab-stream Layer software
    - Google hangout.
- 31 participants that are
    - >=18yrs,
    - have prior driving experience,
    - symmetric personality traits,
    - no known medical condition and
    - right handed participated

Methodology



Introduction

Informed consent

Affix sensors (EEG, EDA, and EMG)

Demo session

Assess trust perception

45 seconds relaxation (baseline

Play game session 1

Assess trust perception

Play game session 2

Assess trust perception

Play game session 3

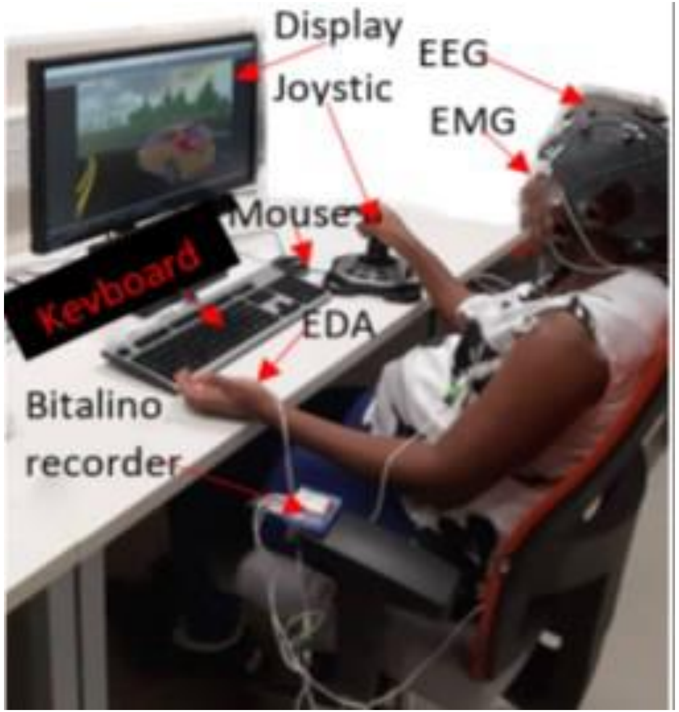
Assess trust perception

Play game session 4

Assess trust perception

Debriefing and exit

Experiment procedure

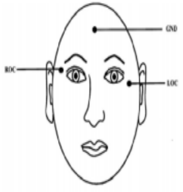


Participant during experiment

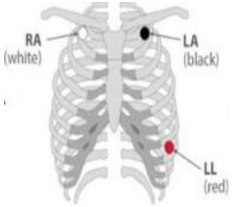
Methodology

- Data collection and processing.
  - The Continuous EEG data were recorded at a
    - sampling rate of 250Hz with
    - impedance kept at <20kohm
    - The ground reference electrode was placed on the right earlobe
    - 75% metabolic spirit fluid to wipe the right ear lobe before affixing the ground electrode
    - Electrolyte gel was applied to each electrode
    - Low pass filter of 120hz, remove sharp spikes
    - high pass filter of 0.10Hz low-frequency drift noise
    - notch filter of 50hz high-frequency sinusoidal power line noise respectively.
  - The continuous ECG, EDA and Facial-EMG signals were recorded at a
    - sampling rate of 1000hz.
    - the EDA signals down-sampled to 50hz [55].
    - ECG signals were down-sampled to 50hz[56].

Data collection



Facial EMG



ECG

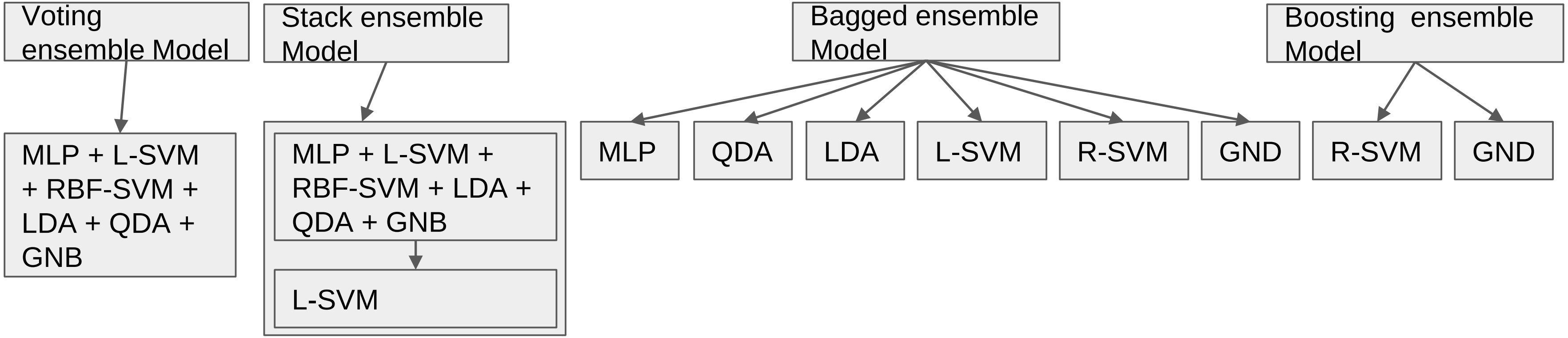
EDA

Methodology

- Feature extraction
  - The continuous EEG, ECG, EDA, and facial EMG data were first
    - divided into epochs, each epoch
      - is 4s long.
      - was labeled as
        - high trust (coded as 2, if the joystick was not used during a trial) or
        - low trust (coded as 1 if the joystick was used during a trial).
  - Using customized python script implementing python libraries from
    - Python MNE [57] and MNE-feature extraction [70],
    - Mathlab and Ledlab software [55],
      - we extracted
        - total of 172 feature set from both time and frequency domain of the psychophysiological signals (EEG, EDA, ECG and Facial EMG).

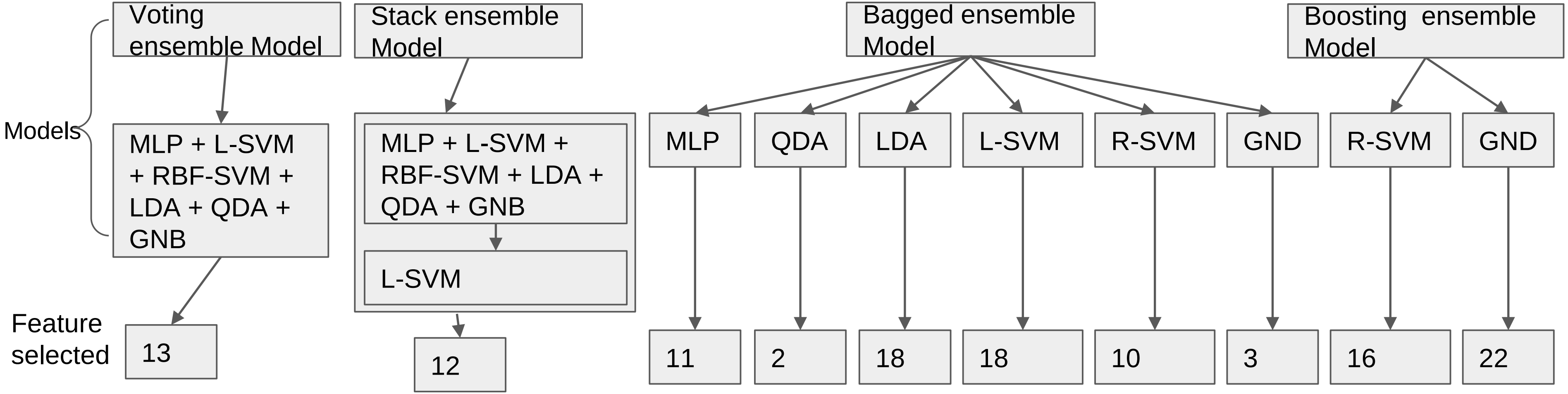
Methodology

- Ensemble trust classifier models were developed based on the following algorithms
  - Multi-Layer Perceptron (MLP),
  - Linear support vector machine (L-SVM),
  - Regularized support vector machine (RBF-SVM),
  - Linear discriminant analysis (LDA),
  - Quadratic discriminant analysis (QDA),
  - Gaussian Naive Bayes (GNB))



Methodology

- Feature selection
  - Applying hybrid feature selection method on subset of the training data(80%) as follows:
    - First filter feature selection method resulted to
      - model independent top fourty(40) features
    - Secondly, for each model, using subset of the training data(80%) with only the top 40 features we
      - applied sequential forward floating feature selection method(SFFFS)[63].



Models

Feature selected

Voting ensemble Model

MLP + L-SVM + RBF-SVM + LDA + QDA + GNB

13

Stack ensemble Model

MLP + L-SVM + RBF-SVM + LDA + QDA + GNB

L-SVM

12

Bagged ensemble Model

MLP

QDA

LDA

L-SVM

R-SVM

GND

11

2

18

18

10

3

Boosting ensemble Model

R-SVM

GND

16

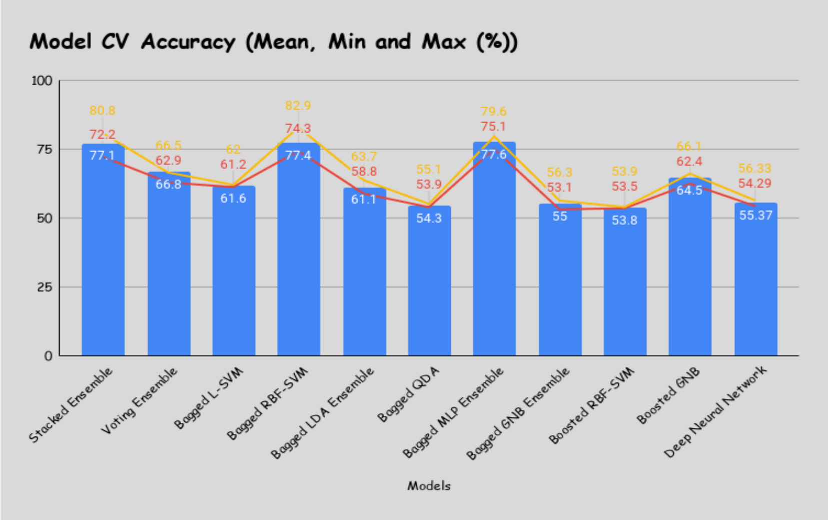
22

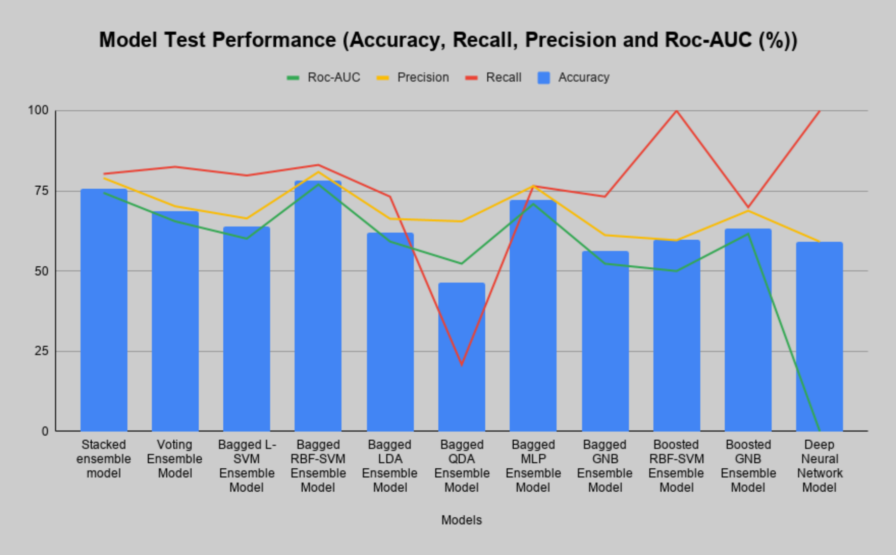
Result

Model

Training and Validation (Accuracy %) Performance

Model Testing





Conclusion

All models irrespective of ensemble method Is stable due to hybrid feature selection method employed. [64].

Though the bagged ensemble methods and its resulting models (RBF-SVM and MLP algorithms) had first and third high performance, this also suggest that the bagged ensemble method performance varies between algorithms.

stack ensemble trust classifier model is most superior with regards to generalizability and performance.

Future work

Future research could explore unsupervised deep neural network trust classifier models and reinforcement learning approach.

what algorithms are most suitable for assessing users trust with psychophysiological signals.

what are the most suitable psychophysiological signals for assessing users trust with psychophysiological signals.