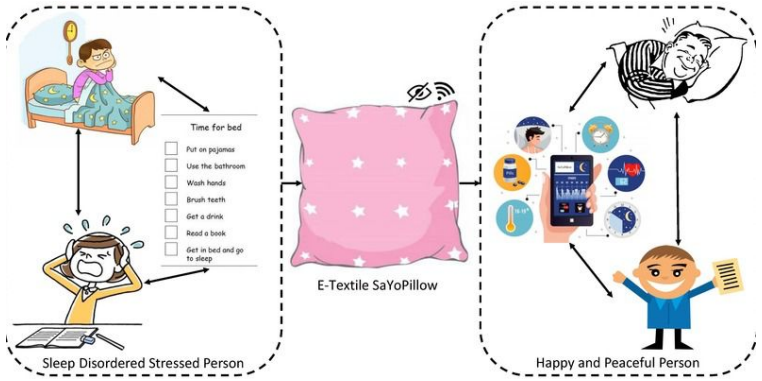
Human Stress Detection In and Through Sleep

Semestre Project
Presentation Duration : 15 min.



Makbule Ozge Ozler
23501039

Content

1. **Problem Definition**
2. **Motivation**
3. **Literature Review**
4. **Contribution to Literature**
5. **Methods Used**
6. **Methodology**
7. **Experimental Results**
8. **Conclusion**

1. Problem Definition

Stress level detection (0- low/normal, 1 – medium low, 2- medium, 3-medium high, 4 -high) after the awakening based on changes in parameters - snoring range of the user, respiration rate, body temperature, limb movement rate, blood oxygen levels, eye movement, number of hours of sleep, heart rate- during sleep.

Data has been generated from Literature Review, therefore using small and balanced dataset for prediction of stress level in and through sleep.

2. Motivation

★ To complete suggestions proposed in previous assignment.
★ To fill the gap in the literature on stress level detection after the awakening.
★ To experience ensemble learning for the first time, and to apply CGAN for the first time.
★ To attend a national or international conference.
★ To pass the course Blm5134 at Ytu in term fall'24.

3. Literature Review

As it is stated that the state of the art today does not address the relationship between stress variations during the day and sleeping behaviors at night [2], the connected papers on these topics [5] seem scarce and need to be focused on in the first place.

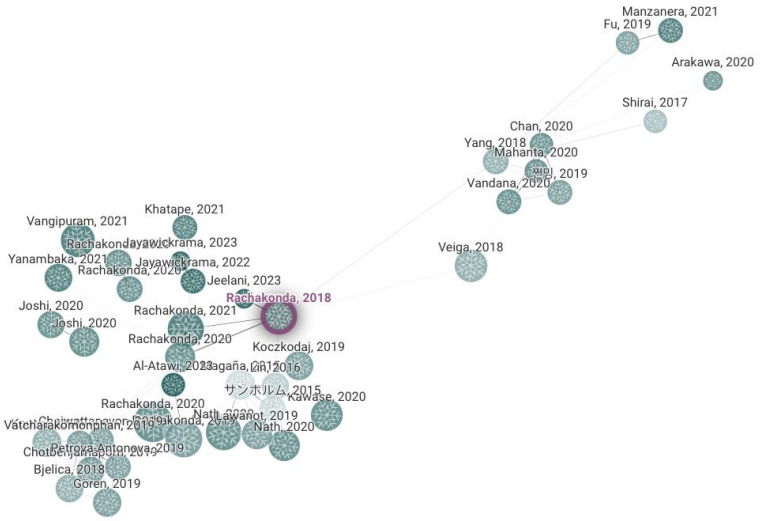


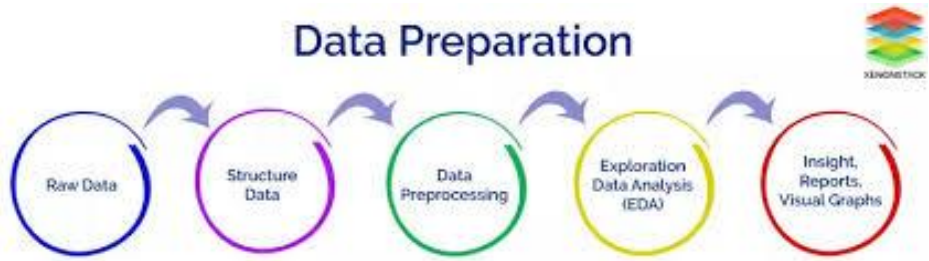
Table 1 Literature Review

	Stressors Used	*ML Algorithm Performed*	*Features Extracted*	*Stress Levels Classified*	*Activities Considered*	*Accuracy Obtained (%)*	*Provided Security?*	*Performed Stress Controlled?*
[1]	Snoring Rate, Respiration Rate, Body Temperature, Limb Movement Rate, Blood Oxygen, Rem Rate, Heart Rate	Multilayer Perception (MLP), Random Forest, (SVM), Decision Trees, Naïve Bayes and Logistic Regression	7	2	Sleeping Habits	91.27%	No	-
[2]	Heart Rate, Respiration Rate, Blood Oxygen range, REM period, Limb movement, Body Temperature, Snoring Range, Sleep Duration	FCNN	8	5	Sleeping Habits	96%	Yes	Yes
[3]	Body Temperature, Heart Rate, Snoring range and Sleep Duration	Fuzzy Logic	4	5	Sleeping Habits	Up-to 65	Yes	No

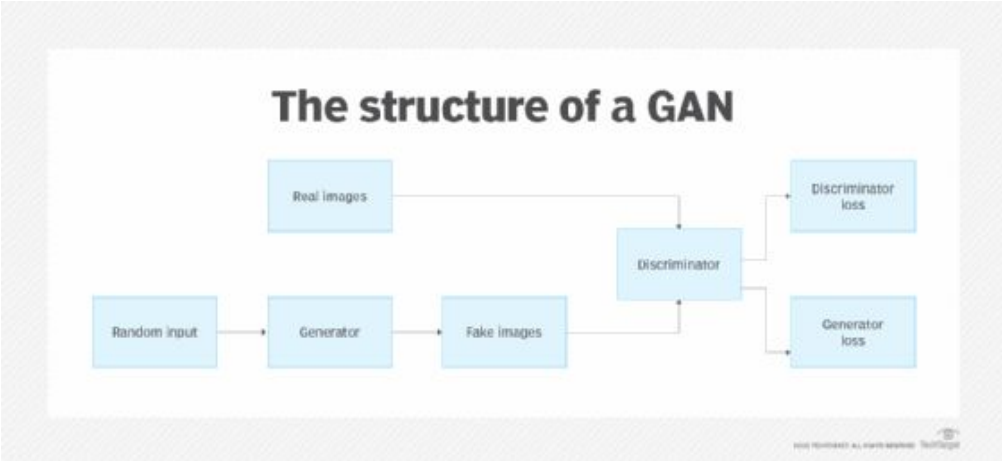
4. Contribution to Literature

- Using augmentation techniques and
- Applying ensemble learnings for such data.

5. Methods Used



1. Data Preprocessing: Data has already balanced and has no missing value.
2. Data Augmentation:



1- Split df as X and y

2- Split X and y as train and test (0.999 train, 0.001 test)

3- Define GAN (discriminator, generator, train_gan, build_gan, generate_synthetic_data)

4- Build and train GAN

5- Train a RandomForestClassifier on the original data (X_train and y_train)

6- Generate synthetic feature data using the trained generator (with X_train and y_train)

7- Use the trained RandomForestClassifier to predict y values (synthetic data)

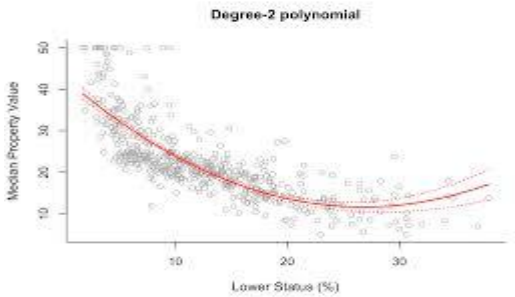
8- Combine original and synthetic data of X and y seperately and save as csv. (X+X_synthetic and y+ y_synthetic)

9- Train a RandomForestClassifier on the combined data (combined_X and combined_y)

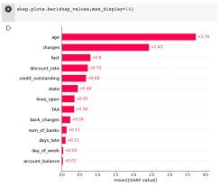
10- Split data again for train and test

11-Performance metrics displayed

5. Methods Used



$$X_{new} = \frac{X_i - \min(X)}{\max(x) - \min(X)}$$

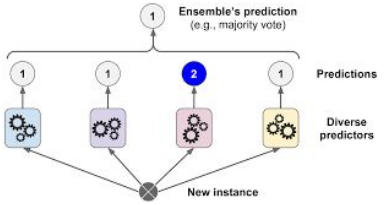


3. Feature Selection/Elimination:

   a. 2-degree Polynomialization for all features
   b. MinMaxScaler to normalize the data before GAN
   c. Feature Importance with RandomForestClassifier
   d. Create a DataFrame for feature importances , Sort the DataFrame by importance, Select the first 5 features.

4. Ensemble Learning:

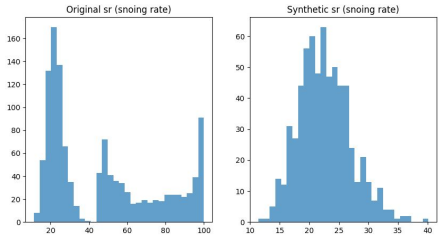
   (i) Selection of ML algorithms based on literature. LogisticRegression, RandomForestClassifier, GradientBoostingClassifier, SVC, GaussianNB
   Split the data into training and test sets
   (ii) Train each algorithm with train data.
   Create the ensemble model 1 : Averaging
   Create the ensemble model 2 : Weighted Averaging
   Create the ensemble model 3 : Voting

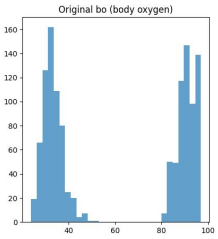


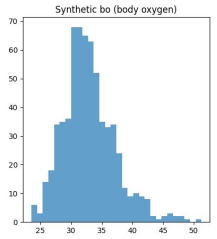
5. Voting (Majority or Weighted) : Both majority and Weighted
6. Performance Evaluation
   a. ensemble model 1 : Averaging has better score in accuracy wrt other ensemble methods.
7. Understanding the impact of each feature on the model's predictions
   a. rr (respiration rate)^2 > sr (snoring rate) lm (limb movement rate) > rr (respiration rate) *lm (limb movement rate) > sr (snoring rate) rr (respiration rate) > rr (respiration rate)
8. Hyperparameter tuning and adjusting ensemble composition.

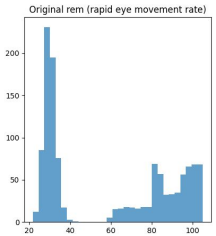
7. Experimental Results

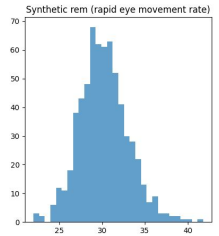
Even though features generated by GAN vs original features was quite different,











Ensemble learning methods (1,2,3) coped with this.

Ensemble accuracy 1: 0.99%
Ensemble Precision: 0.99%
Ensemble Recall: 0.99%
Ensemble F1 Score: 0.99%

Ensemble Confusion Matrix:
[[147 0 0 0 0]
[1 20 0 0 0]
[0 0 26 0 0]
[0 0 2 33 0]
[0 0 0 0 27]]

Ensemble accuracy 2: 0.98%
Ensemble Precision: 0.98%
Ensemble Recall: 0.980%
Ensemble F1 Score: 0.98%

Ensemble Confusion Matrix:
[[145 2 0 0 0]
[0 21 0 0 0]
[0 0 26 0 0]
[0 0 3 32 0]
[0 0 0 0 27]]

Ensemble accuracy 3: 0.94%
Ensemble Precision: 0.94%
Ensemble Recall: 0.94%
Ensemble F1 Score: 0.94%

Ensemble Confusion Matrix:
[[147 0 0 0 0]
[4 17 0 0 0]
[0 5 21 0 0]
[0 0 6 28 1]
[0 0 0 0 27]]

For hyperparameter tuning, values obtained:

Best parameters: {'gbc__n_estimators': 50, 'gnb__var_smoothing': 1e-09, 'lr__C': 0.1, 'rf__n_estimators': 50, 'svm__C': 10.0}

Best score: 0.994

8. Conclusion

Recommendation:

- Features were main area, but still other features such as Electrodermal Activity (EDA) which measures the skin's electrical conductance or ambient sound can be included.
- The mood of tomorrow might be added to the scope of the study according to the situation as sometimes, person’s stress level can be examined in another behaviour such as eating frequency.
- An urge need for real dataset about stress level detection in and through sleep is there.

Self-evaluation:

- Due to its complexity, CGAN could not applied, GAN applied for the study.
- Among important features, rapid eye movement was not selected for the first five which was expected by the experimenter, at least in a interaction with another feature.
- I strengthen my muscle in the data area after a long time.
- Experimenting such small and balanced dataset was tough.