Model Card- Thyroid Disease Detection

March 10, 2021

Model Details

- Person developing the model: Alyssa Hajj Assaf
- Model date: February 13, 2021
- Model Type: Parzen Window Classifier with Gaussian Kernel.
- Feedback on the model: For any questions and/or comments on the model, please email alyssa.hajj.assaf@umontreal.ca

Intended Use

- Primary intended uses: This classifier was developed to aid and support the diagnosis of patients with hypothyroid, hyperthyroid, or normal thyroid function.
- Primary intended users: Healthcare professionals practicing in hospital or in clinics.
- Out-of-scope uses: The classifier should not be used in the assistance of diagnosis other than hypothyroid, as the parameters of the classifier has been chosen from dataset conveying thyroid diseases information.

Factors

- Based on other thyroid data set information, potentially relevant factor includes gender, age, thyroid surgery, other sickness, thyroid medication and tumor. Note that these information about the patients were not disclosed by the dataset used to train this classifier.
- Evaluation factors are 1- normal, 2-hyperthyroid, and 3-hypothyroid. Further factors were not available for this dataset.
- Further factors affecting the model might include laboratory test accuracy that are fed as input to the classifier.

Metrics

- Hyperparameter for this classifier includes Euclidean distance in the prediction procedure and number of folds for the k-fold cross validation

- Parameter sigma of the Gaussian kernel was chosen among a range of value and selected as the one producing the smaller rate of classification error over 3-fold cross validation on training data split.
- Evaluation metrics includes the rate of correct classifications (e.g. normal-normal) and rate of incorrect classifications for all erroneous combinations (e.g. hypothyroid-normal) calculated from a confusion matrix.

Training Data

- Thyroid dataset donated by Stefan Aeberhard, training data split
- Preprocessing: Thyroid data set was shuffled as the original file was ordered by labels
- This data set was chosen due to its completeness. Other thyroid data set contained missing values.

Evaluation Data

- Thyroid dataset donated by Stefan Aeberhard, test data split

Ethical Consideration

- The thyroid dataset was sample from a population where thyroid diseases are rare.
- This classifier should be used to aid or support the healthcare professional diagnosis of hypothyroid, hyperthyroid, or normal thyroid function of a patient referred to the clinic or hospital. It should not be used as a primary evidence for the diagnosis, nor should it be used to weight against the recommendations of the healthcare professional.

Caveats and Recommendations

- An ideal evaluation data set would include gender, age, race, thyroid medication and presence of tumor.

Source of the dataset

- https://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/

Quantitative Analyses

Figure 1: Confusion Matrix.