Model Card - Multiple Sclerosis Disease Predictors

Jeffrey Kong, Noah Lee, Jacob Ventura

Model Details

- Created by Jeffrey Kong, Noah Lee, and Jacob Ventura with methods taught by Adam Loy in STATTAT 270 at Carleton College.
- Employed K-nearest neighbors, Naive Bayes, and a Random Forest model.
- Finalized the model on 11/25/2024.

Intended Use

- The model was designed as a capstone project for STAT 270 at Carleton College.
- Goal is to accurately predict if a patient diagnosed with CIS will develop Multiple Sclerosis.
- Intended for educational purposes, and should not be used as the sole basis for conclusions.

Factors

- Demographic: Age, Gender
- Medical History: Initial_Symtpoms. Mono_or_Polysymptomatic, Breastfeeding, Varicella
- Instrumentation: Periventricular_MRI, Cortical_MRI, Infratentorial_MRI, Spinal_Cord_MRI
- Environments: Schooling
- Potential Future Variable: Neurological Imaging

Metrics

- Accuracy is the primary metric used to compare models, while sensitivity, specificity, and ROC-AUC were also considered.
- The validation dataset was used to test performance metrics.

Training Data

- 70% of the data was randomly split and assigned as the training data, stratified by group.

Evaluation Data

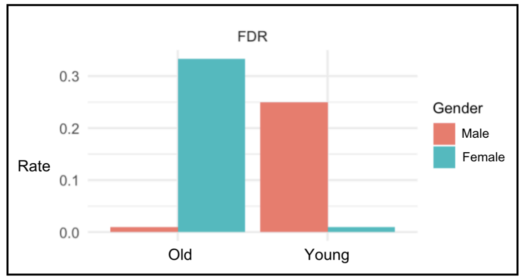
- 15% of the data was randomly assigned to testing data, and the remaining 15% was used as the validation data, also stratified by group.
- Data was pre-processed to omit NA or unknown values.

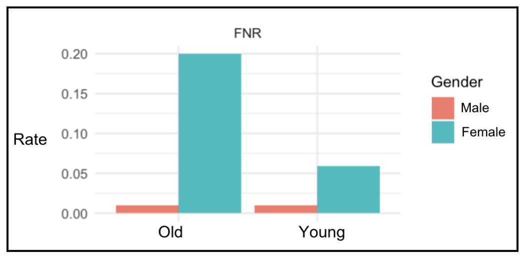
Ethical Considerations

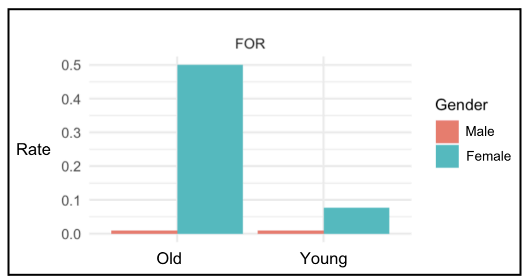
- Data was collected through clinical evaluations, diagnostic tests, and follow-ups of CIS patients at the National Institute of Neurology and Neurosurgery in Mexico City. Individuals presumably agreed to their anonymous information being collected.

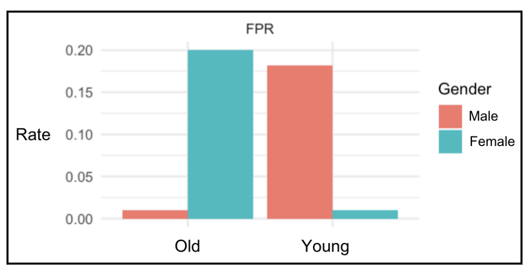
Quantitative Analysis

Gender	Age Group	False Positive Rate	False Negative Rate	False Discovery Rate	False Omission Rate
Male	Old	0	0	0	0
Male	Young	0.18	0	0.25	0
Female	Old	0.20	0.20	0.33	0.50
Female	Young	0	0.06	0	0.08









Model Card - Smiling Detection in Images

Model Details

- Developed by researchers at Google and the University of Toronto, 2018, v1.
- Convolutional Neural Net.
- Pretrained for face recognition then fine-tuned with cross-entropy loss for binary smiling classification.

Intended Use

- Intended to be used for fun applications, such as creating cartoon smiles on real images; augmentative applications, such as providing details for people who are blind; or assisting applications such as automatically finding smiling photos.
- Particularly intended for younger audiences.
- Not suitable for emotion detection or determining affect; smiles were annotated based on physical appearance, and not underlying emotions.

Factors

- Based on known problems with computer vision face technology, potential relevant factors include groups for gender, age, race, and Fitzpatrick skin type; hardware factors of camera type and lens type; and environmental factors of lighting and humidity.
- Evaluation factors are gender and age group, as annotated in the publicly available dataset CelebA [36]. Further possible factors not currently available in a public smiling dataset. Gender and age determined by third-party annotators based on visual presentation, following a set of examples of male/female gender and young/old age. Further details available in [36].

Metrics

- Evaluation metrics include **False Positive Rate** and **False Negative Rate** to measure disproportionate model performance errors across subgroups. **False Discovery Rate** and **False Omission Rate**, which measure the fraction of negative (not smiling) and positive (smiling) predictions that are incorrectly predicted to be positive and negative, respectively, are also reported. [48]
- Together, these four metrics provide values for different errors that can be calculated from the confusion matrix for binary classification systems.
- These also correspond to metrics in recent definitions of "fairness" in machine learning (cf. [6, 26]), where parity across subgroups for different metrics correspond to different fairness criteria.
- 95% confidence intervals calculated with bootstrap resampling.
- All metrics reported at the .5 decision threshold, where all error types (FPR, FNR, FDR, FOR) are within the same range (0.04 - 0.14).

Training Data

- CelebA [36], training data split.

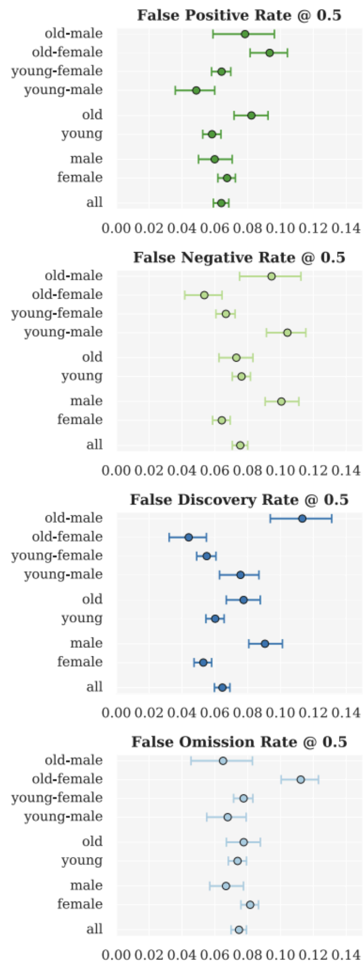
Evaluation Data

- CelebA [36], test data split.
- Chosen as a basic proof-of-concept.

Ethical Considerations

- Faces and annotations based on public figures (celebrities). No new information is inferred or annotated.

Quantitative Analyses



Caveats and Recommendations

- Does not capture race or skin type, which has been reported as a source of disproportionate errors [5].
- Given gender classes are binary (male/not male), which we include as male/female. Further work needed to evaluate across a spectrum of genders.
- An ideal evaluation dataset would additionally include annotations for Fitzpatrick skin type, camera details, and environment (lighting/humidity) details.

Figure 2: Example Model Card for a smile detector trained and evaluated on the CelebA dataset.

- Example

Gender	Age Group	False Positive Rate	False Negative Rate	False Discovery Rate	False Omission Rate
Male	Old	0	0	0	0
Male	Young	0.18	0	0.25	0
Female	Old	0.20	0.20	0.33	0.50
Female	Young	0	0.06	0	0.08