USER SELF-TRIAGE MODEL USING TREE BASED METHODS

PAUL BUTLER

THE PROBLEM

- Most people are not medical professionals and lack the knowledge to accurately determine when to seek medical care.
- Combined with high medical care costs, this can create a disincentive to seek appropriate care.
- To solve this problem we need an easy way for people to answer two questions:
  - Do I need care?
  - How quickly should I seek care?

DATA SET

- Data for 4920 patients from a third world area.
- 133 symptoms:
  - Binary Presenting/ not presenting classifier
- Medical diagnosis
  - 41 categorical diagnoses

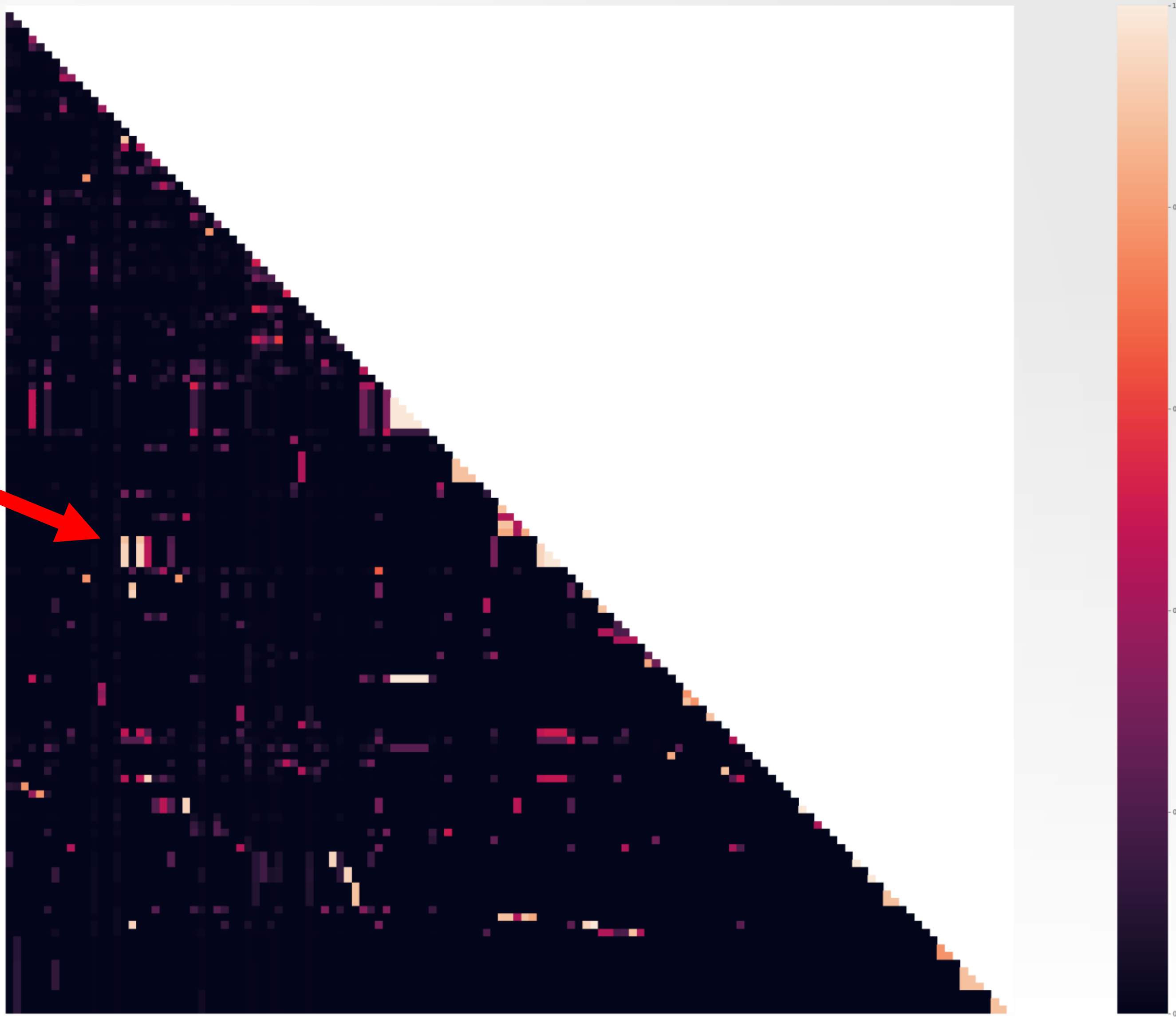
CLEANING HIGHLY RELATED SYMPTOMS

- Strong relationships between symptoms were detected using Cramer's V
- Any relationship > 0.4 was then cleaned by
  - Dropping the symptom(s) that non medically trained persons could not identify.
  - Dropping all but one symptom in highly related clusters of symptoms.
  - Dropping the symptom that non medically trained persons would be more likely to inaccurately identify.

CRAMER'S V RESULTS BEFORE CLEANING

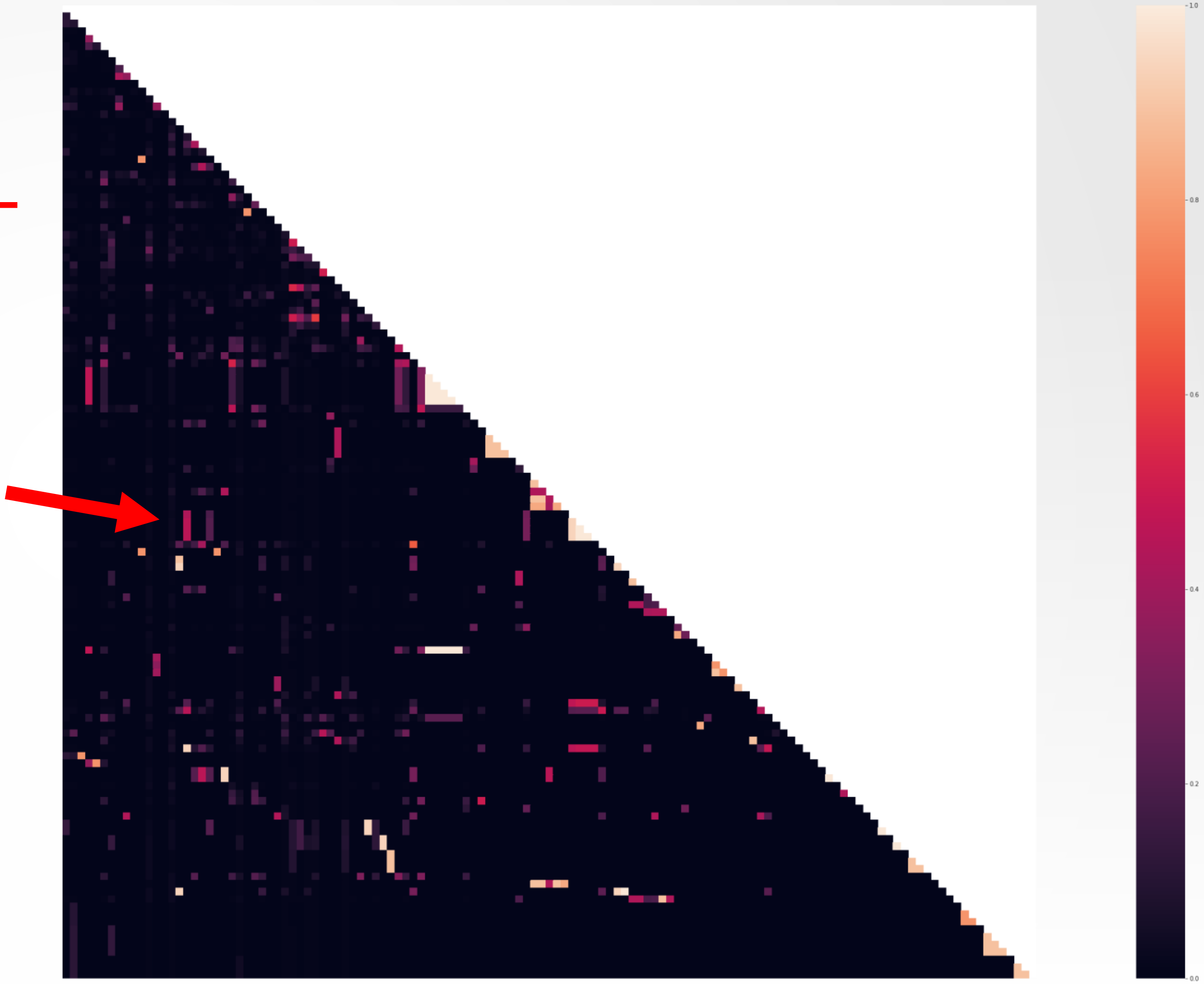
- Many hotspots of interrelated symptoms > 0.4 present.
- Example of two hotspots with high relationships:

	weight_gain	Cold_hands_and_feets
puffy_face_and_eyes	0.89	0.89
enlarged_thyroid	0.94	0.94
brittle_nails	0.94	0.94
swollen_extremeties	0.94	0.94



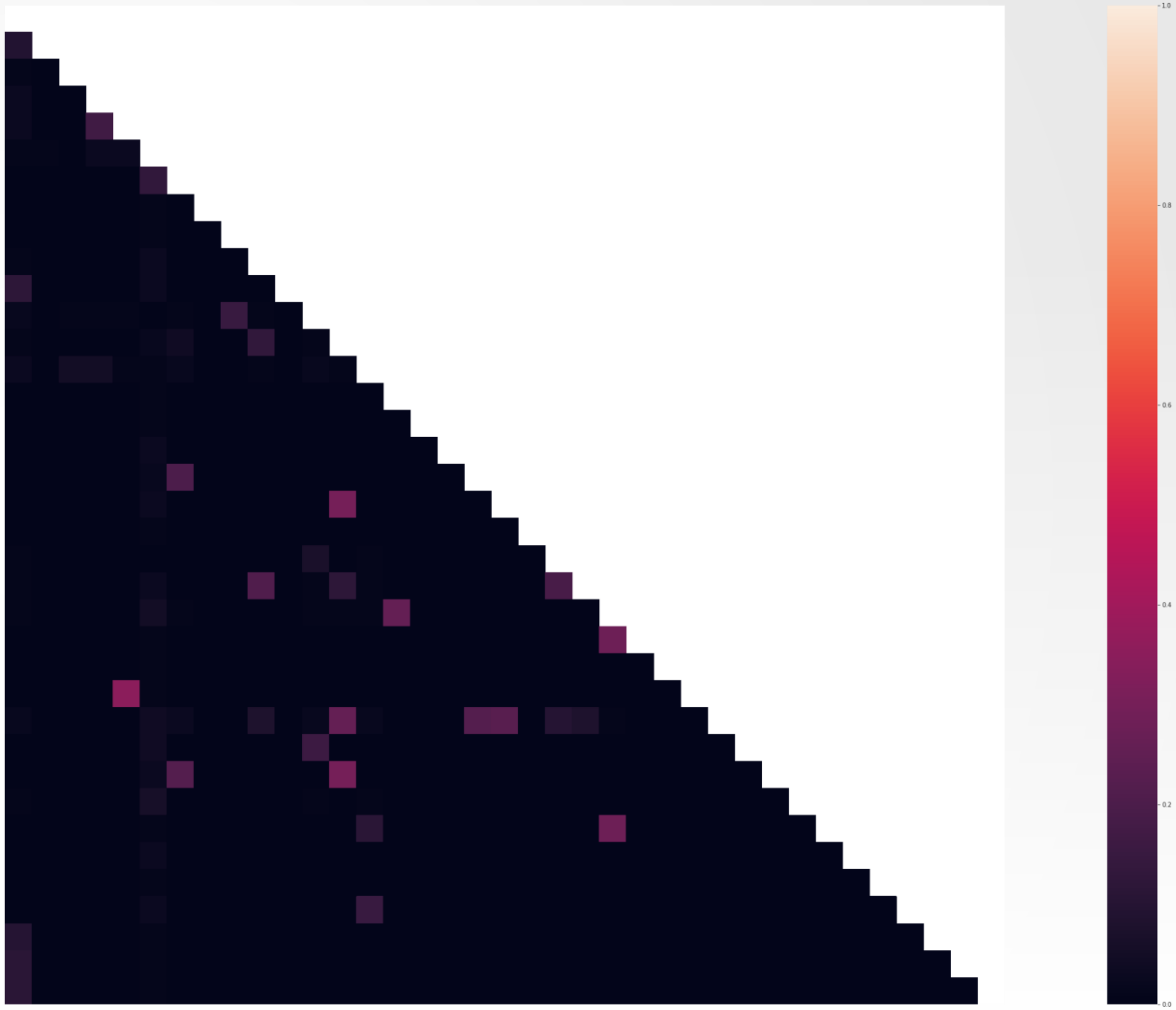
CRAMER'S V RESULTS CLEANING EXAMPLE

- Dropping "weight_gain" and "cold_hands_and_feets" eliminates the high relationships between these 2 variables and the 4 other variables



CRAMER'S V RESULTS AFTER CLEANING

- All hotspots > 0.4 eliminated from the dataset.
- 36 symptoms remaining



REMAINING FEATURES OVERVIEW

- 36 remaining features
- Features vary by severity and system affected
  - Severe
    - Example: weakness_of_one_body_side(neurological), muscle_weakness(mudculoskeletal)
  - Moderate
    - Example: stomach_pain(gastrointestinal) , skin_rash(dermatological)
  - Mild
    - stiff_neck(musculoskeletal), fatigue (multisystem)
- Features can be acute, chronic, or risk factors (for example history_of_alcoholism)

CREATING THE DECISION TREES MODEL

- Decision Trees Model
- Model overview:
  - Predictive features: 36 Symptoms
  - Target variable: Triage Classification (4 possible outcomes)
- Split data using stratified 80/20 train/test split.
- Random State = 42
- Technologies used
  - Scikit-learn for model building.

DECISION TREES MODEL ACCURACY

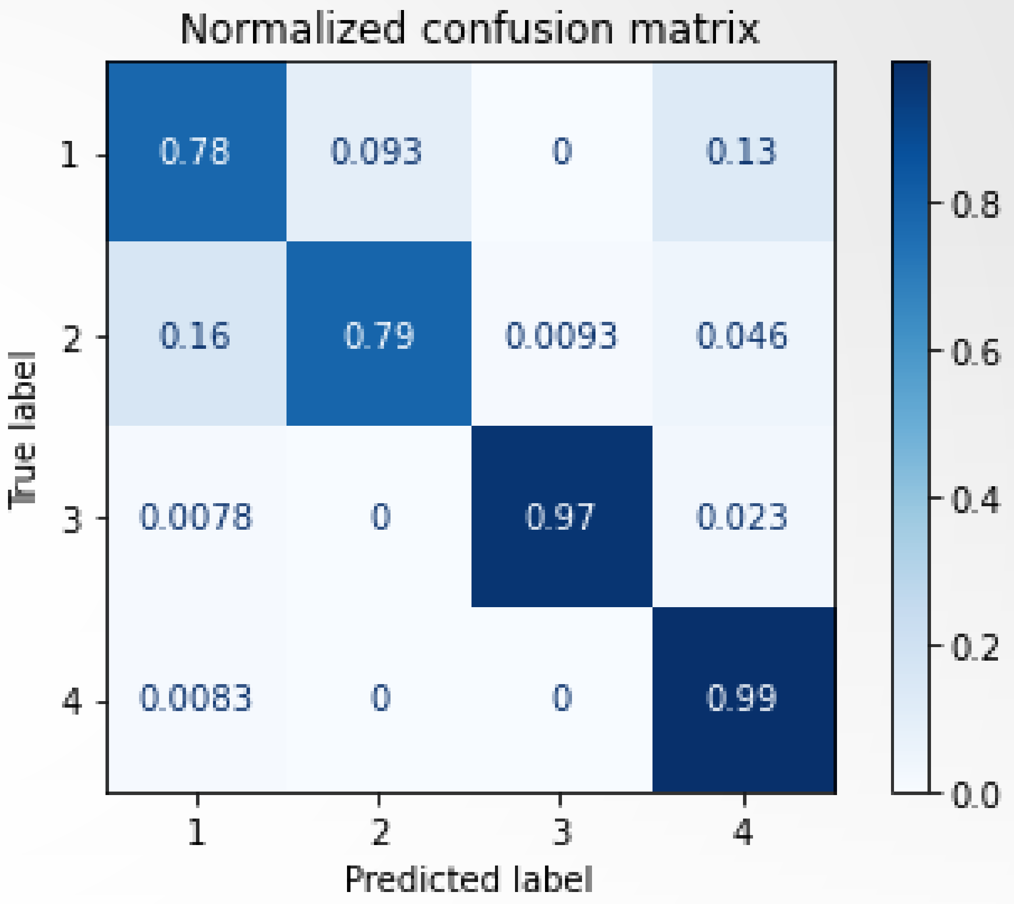
- Max Depth 30
  - Training Accuracy: 90%
  - Testing Accuracy: 88%
  - This indicates that the model is very accurate overall, but is slightly overfitted
  - Overall F1 Score: 0.89
  - Too complex to be explainable

Max Depth 3

  - Training Accuracy: 48%
  - Testing Accuracy: 48%
  - Not Useful

DECISION TREES ACCURACY CONT.

- Significant differences between classes in confusion matrix
- F1 scores for class 1 & 2 significantly lower than class 3 & 4.
- Large difference between precisions and recall for class 4



CAN WE DO BETTER?

Let's try again with a Random Forests Model

CREATING THE RANDOM FORESTS MODEL

- Random forest model implemented
- Model overview:
  - Predictive features: 36 Symptoms
  - Target variable: Triage Classification (4 possible outcomes)
- Technologies used
  - Scikit-learn for model building.
- Split data using stratified 80/20 train/test split.
- 10 estimators and entropy criterion
- Random State = 42

RANDOM FORESTS MODEL ACCURACY

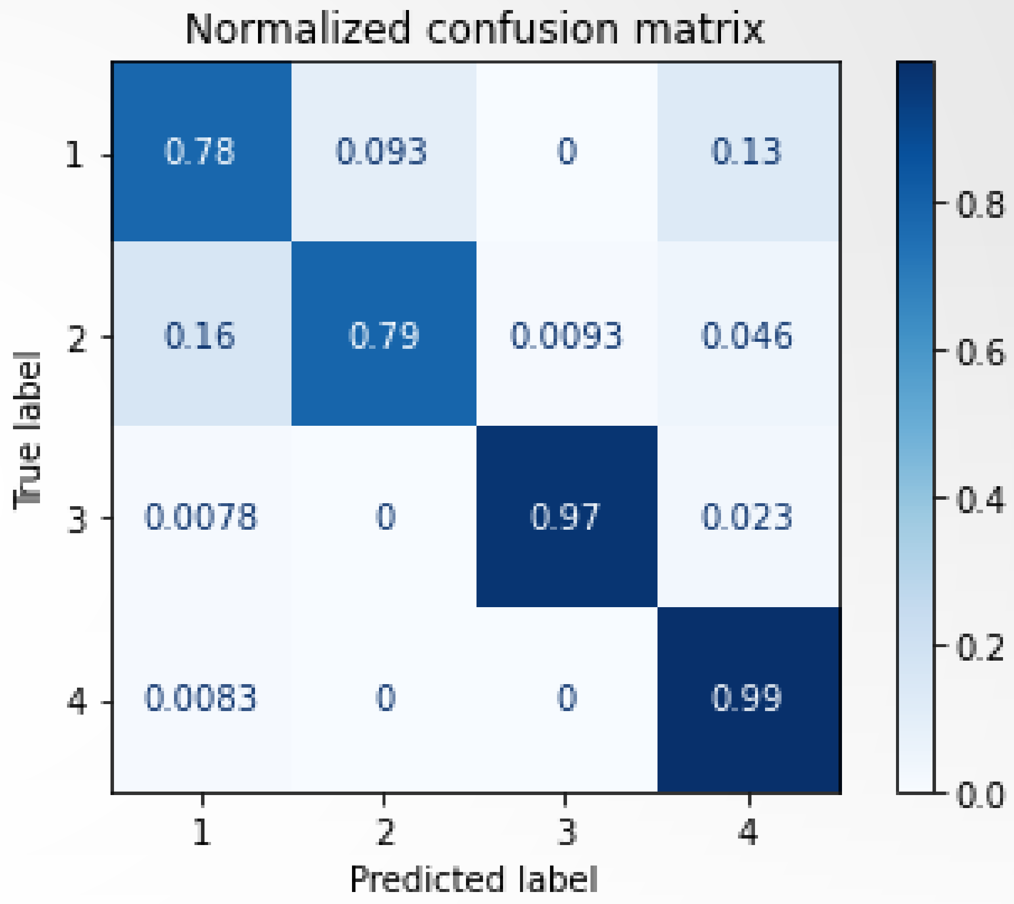
- No Max Depth
  - Same Results as Decision Tree Model
  - Training Accuracy: 90%
  - Testing Accuracy: 88%
  - This indicates that the model is very accurate overall, but is slightly overfitted
  - Overall F1 Score: 0.89

Max Depth 3

  - Training Accuracy: 64%
  - Testing Accuracy: 63%
  - Better result than using decision trees at same depth but still not Useful

RANDOM FORESTS ACCURACY CONT.

- Same Results as Decision Tree Model
- Significant differences between classes in confusion matrix
- F1 scores for class 1 & 2 significantly lower than class 3 & 4.
- Large difference between precisions and recall for class 4



FURTHER RANDOM FORREST MODEL EXPLANATIONS

- Recursive Feature Dropping
  - 8 Features not contributing to model
- SHapley Additive exPlanations (SHAP)
  - All Features contribute to each class, but in varying degrees

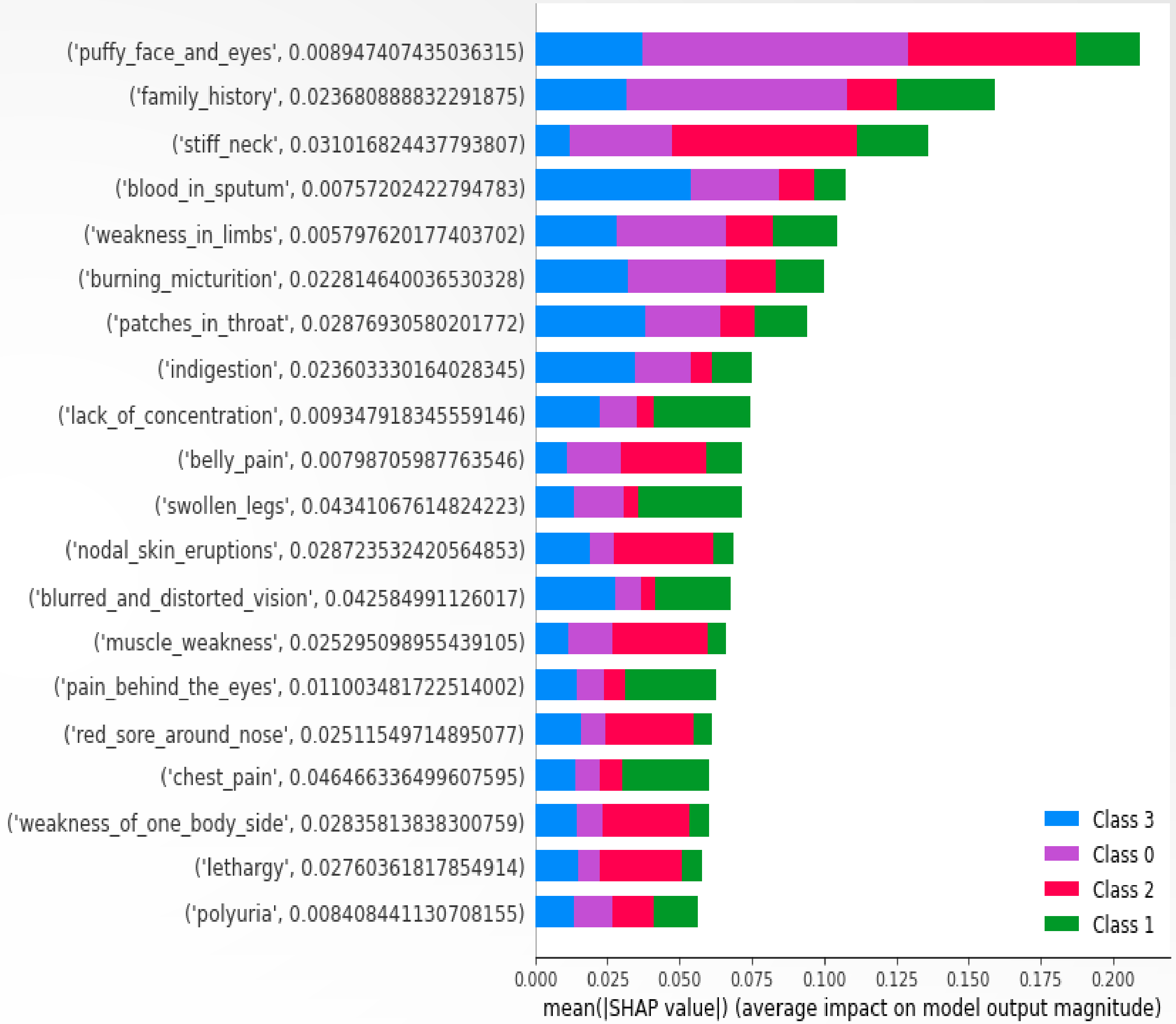
RECURSIVE FEATURE DROPPING

- 8 Features did not contribute to model
- 3 Negative Features
- 25 Positive Features
- This was not the same finding as SHAP

Feature	Impact
indigestion	0
blurred_and_distorted_vision	0
puffy_face_and_eyes	0
slurred_speech	0
muscle_weakness	0
lack_of_concentration	0
blood_in_sputum	0
silver_like_dusting	0
weakness_in_limbs	-0.00081
loss_of_balance	-0.00081
pus_filled_pimples	-0.00081

SHAPLEY ADDITIVE EXPLANATIONS (SHAP)

- All features contribute to each class, but in varying degrees
- Some feature contribute significantly to one class, and then less to others.
- Some features moderately contribute to two classes, and then less to the other two



CONCLUSIONS

- This model is a good first step
- Random Forests does not have superior performance to decision trees when depth limit is not implemented but is superior when depths are limited.
  - As dataset gets larger in future versions, this pattern may not hold.
- Test production pilot in representative area
  - Gather more data for future versions
- Future versions
  - Expand number of diseases triaged
  - Use more symptoms as predictors

OTHER TREE MODEL ATTEMPTS

- Decision Trees Max Depth 10
  - Training Accuracy: 76%
  - Testing Accuracy: 77%
- Random Forests Max Depth 10
  - Training Accuracy: 86%
  - Testing Accuracy: 85%