

SCORE FUSION APP GUIDE TO FUSING MULTIBIOMETRIC SCORES.

User Manual

Version: 1.0

 $\begin{array}{c} Author \\ \text{Melissa R DALE} \end{array}$

Supervisor Arun Ross

Contents

1.1 Usin 2.1 2.2 2.3	Notes on Terminology ng the Score Fusion App Quick Start		 Hub) 	 													1 2 2 2 3
2.12.22.3	Quick Start	l	 Hub) 	· · · · · ·				•		 							$\frac{2}{2}$
2.2	2.1.1 Executable Download 2.1.2 Download from Source Input Files	l	 Hub) 	· · · · · ·				•		 							$\frac{2}{2}$
2.3	2.1.2 Download from Source Input Files	ce (GitH	Hub)	· · · · · · · · · · · · · · · · · · ·					 							2
2.3	Input Files					· · · ·											
2.3	2.2.1 Matrix Form	· · · · · · · · · · · · · · · · · · ·															2
	2.2.2 Column Form																O
	Missing Scores	 3						_							•	•	3
	2.3.1 Ignore Missing Scores2.3.2 Impute Missing Scores	8						-									4
2.4	2.3.2 Impute Missing Score																5
2.4	1 0	es															5
2.4	Train and Test Splitting																5
																	6
2.5	Normalization																6
	2.5.1 MinMax																6
	2.5.2 Z-Score																6
	2.5.3 Decimal																6
	2.5.4 Median & MAD																7
	2.5.5 Double Sigmoid																7
	2.5.6 Tanh Estimator																7
	2.5.7 None																7
2.6	Editing Detected Modalit	ies															7
2.7	Density Estimates																8
2.8	Fusion																9
	2.8.1 Sequential Fusion Ru	le															9
2.9	Experiments and Results																11
2.10	Saving Experiment Results t																12
Adv	vanced																17
		cture .															17
																	17
	_																18
3.2																	18
~· -	_																18
	3																19
																	19
	3.1	3.1 Internal Score Data Structure 3.1.1 Naming Conventions 3.1.2 Columns	3.1 Internal Score Data Structure . 3.1.1 Naming Conventions 3.1.2 Columns	3.1 Internal Score Data Structure 3.1.1 Naming Conventions	3.1 Internal Score Data Structure	3.1 Internal Score Data Structure	3.1 Internal Score Data Structure	3.1 Internal Score Data Structure	3.1 Internal Score Data Structure 3.1.1 Naming Conventions 3.1.2 Columns 3.2 Adding Fusion Rules 3.2.1 The Fuse Object 3.2.2 Updating the GUI	3.1 Internal Score Data Structure 3.1.1 Naming Conventions 3.1.2 Columns 3.2 Adding Fusion Rules 3.2.1 The Fuse Object 3.2.2 Updating the GUI	3.1 Internal Score Data Structure	3.1 Internal Score Data Structure 3.1.1 Naming Conventions 3.1.2 Columns 3.2 Adding Fusion Rules 3.2.1 The Fuse Object 3.2.2 Updating the GUI	3.1 Internal Score Data Structure	3.1 Internal Score Data Structure 3.1.1 Naming Conventions 3.1.2 Columns 3.2 Adding Fusion Rules 3.2.1 The Fuse Object 3.2.2 Updating the GUI			

4	Tro	uble Shooting	2 1
	4.1	Issue	21
	4.2	Known Issues	21
5	Fea	ture Requests, Support, and Future Directions	22
	5.1	Feature Requests and Support	22
	5.2	Future Directions	22

Introduction

The Score Fusion App was designed to facilitate experiments on multibiometric score data. This manual is intended to provide technical details on how to use and adapt this software application. To learn more about the science of multibiometrics, please consult [1].

The application loads biometric score data in a variety of file formats, provides data grooming, visualizations, and metric descriptors of the loaded scores, in addition to providing a variety of fusion rules to analyze. The following chapters of this manual describe how to acquire the app, what score files should look like, and how to customize fusion experiments. In the Advanced chapter, the software itself is described in detail to facilitate the addition of new fusion rules and customization options.

1.1 Notes on Terminology

Biometric Match Scores A typical biometric system compares the *probe* data with the identities stored in a *gallery* database of user templates [1]. To accomplish this, the feature set extracted from the probe is compared against each gallery template resulting in a match **score**. These match scores are then sorted to determine those gallery identities with the highest degree of similarity to the probe.

The Genuine and Imposter Class. The application uses the terms Genuine and imposter throughout the app and the user manual. A genuine score indicates a score that

Verification vs **Identification**. *Verification* tasks aim to verify a claimed idea. The output of the system is binary: Genuine or Imposter. *Identification* tasks attempt to identity a user from the system's enrolled users. The output of an identification task is the identity of the probe.

Using the Score Fusion App

2.1 Quick Start

2.1.1 Executable Download

Executable versions of the Score Fusion App can be downloaded from the releases directory (https://github.com/melissadale/ScoreFusionApp/tree/master/Releases). Work is ongoing to increase stability of the portable executable. If it does not currently work on the user's machine, please consider compiling from source.

2.1.2 Download from Source (GitHub)

If you would like to adapt the application, or if the packaged executable does not work for you, you can set up the application by following this section.

Prerequisites

- Conda: Conda creates a virtual environment which facilitates acquiring the correct python packages and the correct versions of those packages.
- git: Git allows you to collect the code from GitHub.

Once the prerequisites are installed, you can run the .sh script provided in the Score Fusion App's GitHub page to check out the code as well as install necessary packages to a Conda environment.

If you are unable to run the .sh script, you can manually run each of the following commands from your terminal or powershell, and navigate to the directory you'd like the score fusion app code to reside).

```
conda create --name scorefusion_environment python=3.8.5
conda activate scorefusion_environment
git clone https://github.com/melissadale/ScoreFusionApp.git
pip install -r $PWD/ScoreFusionApp/Utilities/requirements.txt
sudo apt-get install libmtdev1
```

You can now edit the score fusion app code or run the application through your favorite editor, like PyCharm after configuring the python interpreter to be the scorefusion environment conda environment.

If assistance is needed, please consult the Feature Requests and Support chapter.

2.2 Input Files

This application accepts score data in 2 formats, **matrix** format or **column** formats. A matrix format assumes that each modality's scores are saved in a separate file (either csv or txt file), with the genuine scores along the diagonal. A column format contains all the score data in a singular file. The details and differences of these formats are explored in the following subsections, and summarized in Figure 2.1.

The user does not need to specify which format the data is in, the application will automatically detect this information.

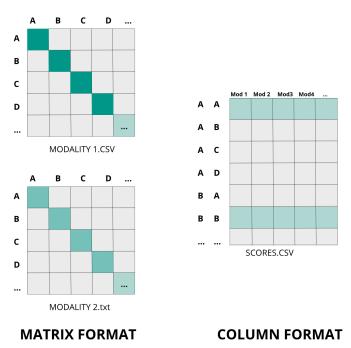


Figure 2.1: Expected input file format: Matrix Form (Left) and Column Form (Right)

Things that are true, regardless of input file format:

- Subject IDs: The application does its best to identify what might be a subject id. IF IDs are strictly numeric, that makes detection more challenging. If Subject IDs are included as column or row headers, ensure they are not strictly numeric.
- Test / Train detection: The application checks filenames for "test" and "train" to determine if the data is already split into training and testing samples. Please be sure to have train and test anywhere in the filename where appropriate.

2.2.1 Matrix Form

This score format contains subjects along the rows and columns such that genuine scores are along the diagonal, and the impostor scores are off diagonal.

Good things to know:

- labels are not necessary, scores along the diagonal are *genuine*, off diagonal scores are *impostor*
- Each modality should be its own file. That file name will be what the modality is referred to as in the application.

A pictorial example of this format for the NIST Biometric Score Dataset (Face x Finger) at NIST BSSR1, with subject IDs as row and column headers.

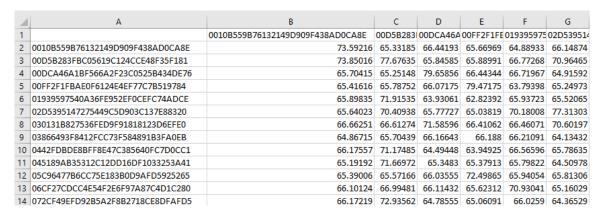


Figure 2.2: Score Matrix with Headers.

Assumptions:

- The order of subjects is consistent along the column and rows
- There are the same number of rows and columns
- IF there are row or column headers identifying the subject id, that id is not in the same format as the scores (i.e. subject ids should not be floats or pure integers)

2.2.2 Column Form

Column formatted data are assumed to be represented as a collection of columns, which contain $at\ least$:

- Subject ID
- Gallery ID
- Score Column(s)

N: M Score Data

If there are more than 1:N scores for subject to gallery ids, please ensure that the score data is formatted such that the following column headers exist:

- probe subject id
- probe file id
- candidate subject id

- candidate file id
- genuine_flag

Important things to know:

This format is based along the concept that there are modalities along the columns, with the last column containing the label(0 impostor, 1 genuine) OR the first 2 columns must be subject ids - in which case, genuine impostor labels may be determined by checking if the first two column ids are equal (genuine) or not (imposter).

- Each column must have a header. This is how the application knows modality titles.
- The label for this last column should be titled label or class (capitalization does not matter)

2.3 Missing Scores

The Score Fusion App provides multiple options on how to deal with missing scores. Many multimodal biometric datasets are complete, meaning that each probe contains a score for each gallery identity. In this case, there is no need to address the issue of missing scores (e.g. ignore missing scores). However, in many real world use cases, there are many situations where a probe may not be able to return the scores for each identity in the gallery. This can lead to a sparse dataframe.

In these circumstances, it is necessary to either drop incomplete score vectors, or fill empty spaces with a proxy value. The following two subsections define the options for handling missing score values.

2.3.1 Ignore Missing Scores

The default option is to ignore missing scores. If this option is selected, incomplete score vectors are simple dropped. If there are a small number of missing scores, or no missing scores, use this option.

Select Ignore if:

- The score data is complete and does not contain any missing scores.
- The number of missing scores is small

2.3.2 Impute Missing Scores

If the data contains more than a few missing scores, you can choose to impute missing score values. The app supports a variety of imputation methods:

- Mean and Median: These two options are univariate and straightforward. A modality's missing scores are filled with that modality's mean or median score.
- Multiple Imputation by Chained Equations (MICE): MICE approach to imputations uses a machine learning model to iteratively update missing values by initially filling missing values with the mean or median score value described above,

and then iteratively updating the placeholder values. The following approaches are available in the app:

- Bayesian Ridge Regression (Linear)
- Decision Tree Regression (Non-Linear)
- K Nearest Neighbors Regression Note, this option often returns good results, however can also be prohibitively time expensive to run.

2.4 Train and Test Splitting

2.5 Normalization

Normalization puts all the scores into a common domain. Consider a matcher's whose scores are between -1 and 1, while another matcher's scores are between 100 and 500. In order to fuse the scores from these two matchers, the scores must be normalized into a common range.

The following are the normalization techniques supported by the Score Fusion App.

2.5.1 MinMax

MinMax normalization scales score values between 0 and 1, and it is a common normalization approach. This is the Score Fusion App's default normalization. MinMax normalization is sensitive to outliers.

$$s' = \frac{s - M_{min}}{M_{max} - M_{min}} \tag{2.1}$$

2.5.2 **Z-Score**

Z-Score normalization is more robust to outliers by using scores' means (μ) and standard deviations (σ) . Note that if the unnormalized data has a large standard deviation, the normalized score values will be near 0.

$$s' = \frac{s - M_{\mu}}{M_{\sigma}} \tag{2.2}$$

2.5.3 Decimal

Decimal scaling puts scores into the range of 0 to 1. For instance, consider scores in the range 100 to 9,000. These are scaled to the range 0.01 to 0.9 by dividing score s by 10^4 . Equation 2.3 formally defines decimal scaling. In the above example, j=4, as the largest score value is 9,000 (10^4).

$$s' = \frac{s}{10^j} \tag{2.3}$$

2.5.4 Median & MAD

$$s' = \frac{s - median(s)}{MAD(s)} \tag{2.4}$$

2.5.5 Double Sigmoid

$$s' = \begin{cases} 1/(1 + (exp(-2((s-t)/r_1))), & \text{if } s < t \\ 1/(1 + (exp(-2((s-t)/r_2))), & \text{if } s \ge t \end{cases}$$
 (2.5)

2.5.6 Tanh Estimator

$$s' = \frac{s - median(s)}{MAD(s)} \tag{2.6}$$

2.5.7 None

You can choose to not normalize the scores to a common domain. This option is not recommended unless you are certain that all modalities already are in a consistent range.

$$s' = s \tag{2.7}$$

2.6 Editing Detected Modalities

Modalities are automatically detected by either the file names or column headers, and by default are set as similarity scores. Users can edit these by clicking on the pencil icon next to the "Detected Modalities" pane. This will bring up the "Edit Modalities" popup as seen in Figure 2.3.

This popup allows users to change the names of the modalities, select a subsection of available modalities for fusion, and specify if a modality's scores are dissimilarity scores. Note, the Update button must be clicked for these changes to take effect.

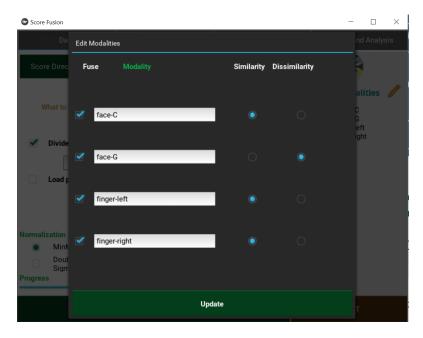


Figure 2.3: Edit detected modalities screen.

2.7 Density Estimates

The Density Estimates Tab features descriptions of the score data. This includes estimated Probability Density Functions (PDFs) and histograms of the genuine and imposter class scores for each of the detected modalities. These visualizations, as well as metrics such as the numbers of genuine and imposter users in the entire dataset, the training dataset, and the testing dataset for each modality.

Figure 2.4 provides a screenshot of this panel.

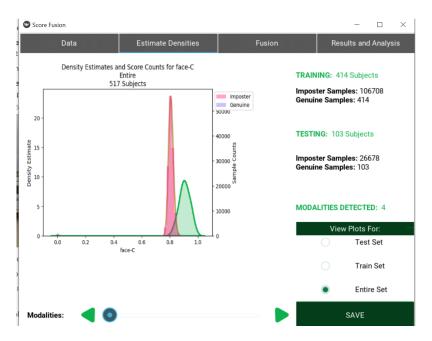


Figure 2.4: Estimate Densities tab, which features score distribution plots and test/train metrics for modalities.

2.8 Fusion

The Fusion Panel is where the user defines the experiments they wish to run by checking which fusion rules to apply, give the experiment names, and specify the type of task (verification and/or identification).

This panel is shown in Figure 2.5.

2.8.1 Sequential Fusion Rule

If the Sequential Fusion is selected, a popup will present the user with a screen to define the baseline modality, as well as an option to automatically detect or manually define score ranges that indicate contentious scores.

This popup is shown in Figure 2.6.

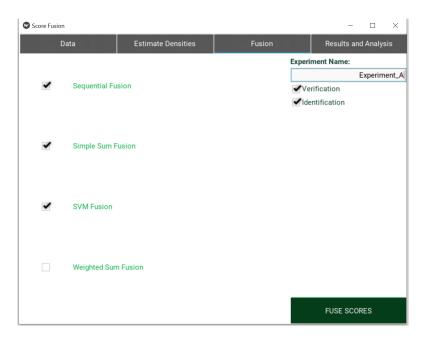
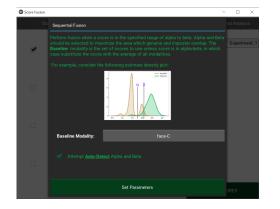


Figure 2.5: Screenshot of Fusion panel, where fusion rules and experiments are defined.



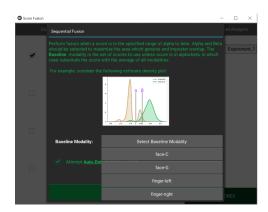


Figure 2.6: Sequential Fusion Popup

2.9 Experiments and Results

The Experiment and Results panel display the experiment's results, including plots and numeric metrics. On the bottom of the panel is a slider, that allows the user to traverse named experiments.

The verification task experiments show ROC plots, AUC accuracy, Equal Error Rate, and Estimated TMR values for a user defined Fixed FMR value. The ROC plots are click through. The first plot contains the ROC curves for all modalities and fusion rules, each fusion rule compared to the modalities, and only the modalities.

The Identification task experiments show CMC plots, Rank 1 and Rank 2 accuracies, as well as the option for a user to select Rank K, where K<20. The CMC plots are click through. The first plot contains the CMC curves for all modalities and fusion rules, the baseline modalities, and the fusion rules.

The identification and verification task panels are shown in Figure 2.7.

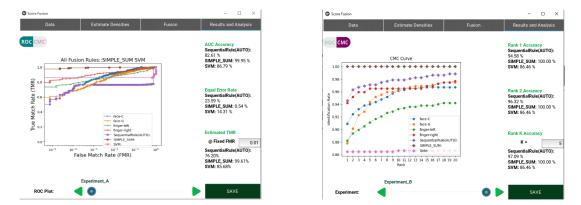


Figure 2.7: Identification Task Results Panel (Left) and Verification Task Results Panel (Right)

2.10 Saving Experiment Results to Report

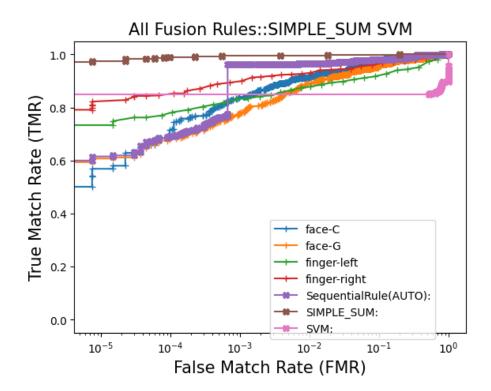
The Score Fusion App can save the data displayed in the results panel to a PDF report. The following pages provide examples of the reports generated for verification and identification tasks.

Generated on: 2021-08-25

Experiment:

Fusion Rules Applied: SequentialRule(AUTO), SIMPLE_SUM, SVM

Modalities: face-C, face-G, finger-left, finger-right



Fusion Rule	AUC	EER	TMR @0.01FMR
face-C	0.98898	0.04449	0.91489
face-G	0.98288	0.05803	0.89168
finger-left	0.96265	0.08393	0.87427
finger-right	0.98233	0.04865	0.9265
SequentialRule(AUTO):	0.99266	0.02901	0.96518

Generated on: 2021-08-25

SIMPLE_SUM: 0.99957 0.00387 0.99613

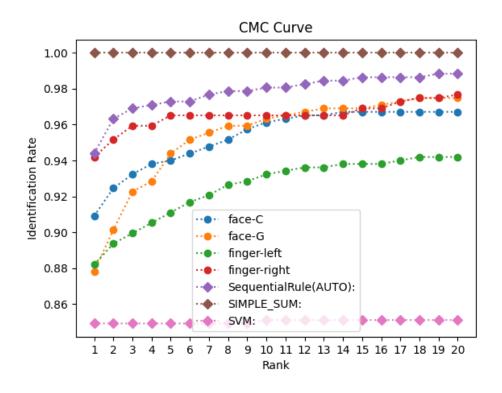
SVM: 0.86136 0.15087 0.84913

Generated on: 2021-08-25

Experiment:

Fusion Rules Applied: SequentialRule(AUTO), SIMPLE_SUM, SVM

Modalities: face-C, face-G, finger-left, finger-right



Fusion Rule	Rank 1	Rank 2	Rank 5
face-C	0.90909	0.92456	0.94004
face-G	0.87814	0.90135	0.94391
finger-left	0.88201	0.89362	0.91103
finger-right	0.94197	0.95164	0.96518
SequentialRule(AUTO):	0.94391	0.96325	0.97292
SIMPLE_SUM:	1.0	1.0	1.0

Generated on: 2021-08-25

SVM: 0.84913 0.84913 0.84913

Advanced

This chapter is for the advanced user of Score Fusion App, and is intended to help others adapt the code for their own purposes. This section focuses on the software design and naming conventions.

3.1 Internal Score Data Structure

No matter how the input files are structured, this app creates a consistent score matrix representation to be used throughout interactions with the application. This score matrix dataframe contains the following columns: Train_Test category, Probe_ID, Gallery_ID, Label, Original Scores (multiple columns), Normalized Scores (multiple columns), and Fused Scores (one or more multiple columns).

For Example, Figure 3.1 is a screenshot of the internal score matrix generated for the NIST BSSR1 multimodal dataset.



Figure 3.1: Screenshot of internal Score matrix derived from the NIST BSSR1 dataset.

3.1.1 Naming Conventions

There are a number of ways this app relies upon consistent naming conventions. We describe these instances below.

• Original, Unedited Modalities: "_ORIGINAL". Original scores read in from the data files are saved in separate columns, so that modality edits or different normalization techniques can be applied without restarting the app.

• Fusion Rules: ":" When analyzing the performance of fusion experiments, it is easy to identify the fused score representations by collecting columns which contain ":". This means that modalities must not use ":" in the name, and that fused modalities must be identified by using ":" in their names.

3.1.2 Columns

The following describe the columns within the internal score matrix.

Train Test

This column specifies whether a probe should belong to the TRAIN or TEST set. This ensures that when performing fusion experiments, results are comparable.

Probe ID, Gallery ID

These columns provide the identifier of the probe and the gallery template compared to that probe to produce the match scores.

Original Scores

Original scores are retained in the score matrix so that the user can choose a different normalization technique or edit modalities without having to restart the app.

Normalized Scores

All fusion techniques are applied to the modalities' normalized scores. Each modality contains a column of normalized scores. Normalized scores serve as the baseline modalities.

Fused Scores

There is a fused column for each fusion rule applied. These columns can be easily identified as all fused score modalities contain ":" in the name. This steps allows the separation of baseline modalities and fused modalities.

3.2 Adding Fusion Rules

There are 2 sections of code that need to be updated to add a new fusion rule: the fuse object and the GUI itself.

3.2.1 The Fuse Object

The Fuse object is defined in Fuse.py (https://github.com/melissadale/ScoreFusionApp/blob/master/App/Analytics/Fuse.py). This file can be updated with a function for your new fusion rule, keeping the following in mind:

• It is important to use only training data for any parameter setting, in order to be consistently compared to other fusion rules.

- Use a ':' in the rule name. Much of the app determines the fused scores by checking for a ':'. For instance, the simple sum fusion rule is titled 'SIMPLE SUM:'.
- It is a good idea to save the important parameters of the new rule so that it may be saved as an experiment model. The following line should be included in some form a the end of your function:

3.2.2 Updating the GUI

Currently, the Fusion panel code resides in the master styles.kv class. This is a god class and will be broken up into its own kv class in the future. In the meantime, it is possible to get to the relevant code by searching for *text: 'Fusion'*.

From here, the new rule will need to have an additional check box by adding something like:

```
CheckBox:
    id: chk_newrule
    size_hint_x: 0.2
    color: hex('#000000')

Label:
    text: 'New Fusion Rule'
    size_hint_x: 0.8
    halign: 'left'
    valign: 'middle'
    text_size: self.width, None
    color: hex('#0DB14B')
```

Next add the chk_newrule to the header of the <Main> object (located towards the top of the Styles.kv file), and to the fuse function inside of the GUI.py class.

3.3 Software UML

The following image is the layout of the Score Fusion App code development 3.2.

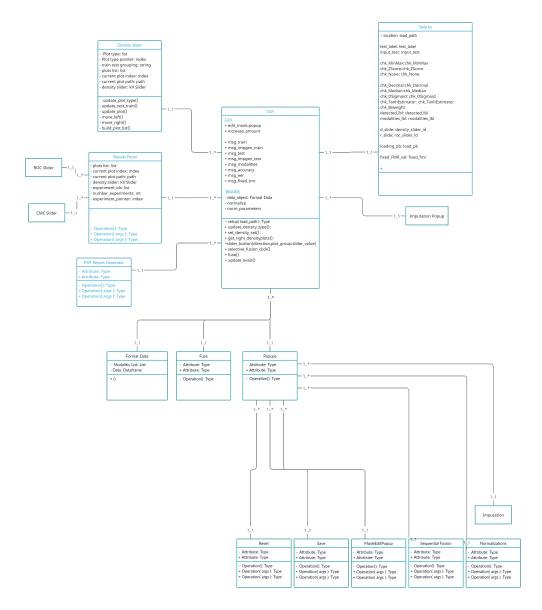


Figure 3.2: UML of Score Fusion App Design

Trouble Shooting

4.1 Issue

Only test or training data is being loaded. Ensure that the word "train" or "test" is not in any of the parent directories. If either of these words appear in the path, the app will believe data is already training or testing data.

I click on a button, but nothing is happening. If nothing is happening, check that you have performed necessary previous steps. For instance, attempting to apply sequential fusion without choosing a baseline modality.

4.2 Known Issues

If running the Score Fusion App on a very large dataset or on a machine with few resources, the following issues are known to occur.

- Image or GIF turns to black square. This can occur when computational resources are limited, it does not impact the performance of the Score Fusion App.
- App says Application not Responding. This warning appears when performing a computationally expensive task, such as using KNN imputation on a very large dataset. If allowed to continue running, the score fusion app will often complete the task and return to normal.

Feature Requests, Support, and Future Directions

5.1 Feature Requests and Support

Feature Requests or Bugs can be submitted at: http://www.github.com/melissadale/ ScoreFusionApp/issues/new

For help or to inquire about this app, contact the author directly at: dalemeli@msu.edu

5.2 Future Directions

This project contains a number of god classes, which are actively being refactored. In addition, the author acknowledges a number of spaghetti code instances as a result of scope creep in addition to the process of learning app development with the Kivy language. A priority moving forward is to repay the technical debt that has been incurred over the course of this project.

Bibliography

[1] Arun A Ross, Karthik Nandakumar, and Anil K Jain. *Handbook of multibiometrics*. Vol. 6. Springer Science & Business Media, 2006 (page 1).