Correcting Eye Tracking Data Using Machine Learning

Luhang Sun, Deka Popov, Bret Miller, Sidney Liu

Background

- Existing research documents ten major algorithms and evaluates them using simulated and natural eye-tracking data
- Suggests that method based on dynamic time warping is a good option
- Some algorithms are better suited than others to particular types of drift phenomena and reading behavior

Carr, J.W., Pescuma, V.N., Furlan, M. et al. Algorithms for the automated correction of vertical drift in eye-tracking data. Behav Res 54, 287–310 (2022). https://doi.org/10.3758/s13428-021-01554-0

Motivation

- Reading Eye Tracking Data requires correction following trials
    - Maintaining proper camera calibration
- The most reliable method of correction: manual correction
    - Time intensive and not feasible for large datasets
- Synthetic data and real world data are drastically different
- Current automated correction options are not reliable and accurate with real data
    - Unreliable data means unreliable research

Research Question

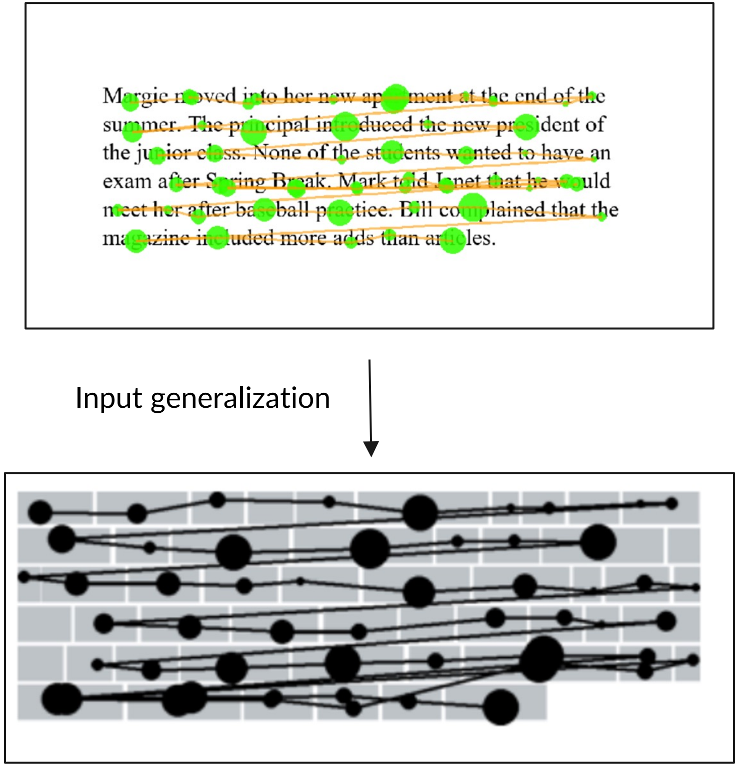
Does utilizing a classification layer between data collection and correction increase the accuracy of the corrections that are performed on reading eye tracking data?

Experiment Design

1. Create a synthetic dataset that simulates different forms of drift phenomena at different levels
2. Pick the best-performing algorithm for each type of error by correcting synthetic data with different levels of drift phenomena
3. Run this synthetic dataset through a Neural Network implementation to classify the most prominent type of error in the trial, then correct with appropriate algorithm
4. Compare the outcomes of the classification network corrections and non-classification network corrections
5. Manually correct a "Golden Dataset" of real world data that will be the ground truth
6. Use Golden Dataset to train and use in Neural Network
7. Compare accuracies of Golden Dataset without classification before and with classification before

Setting Up for Neural Networks

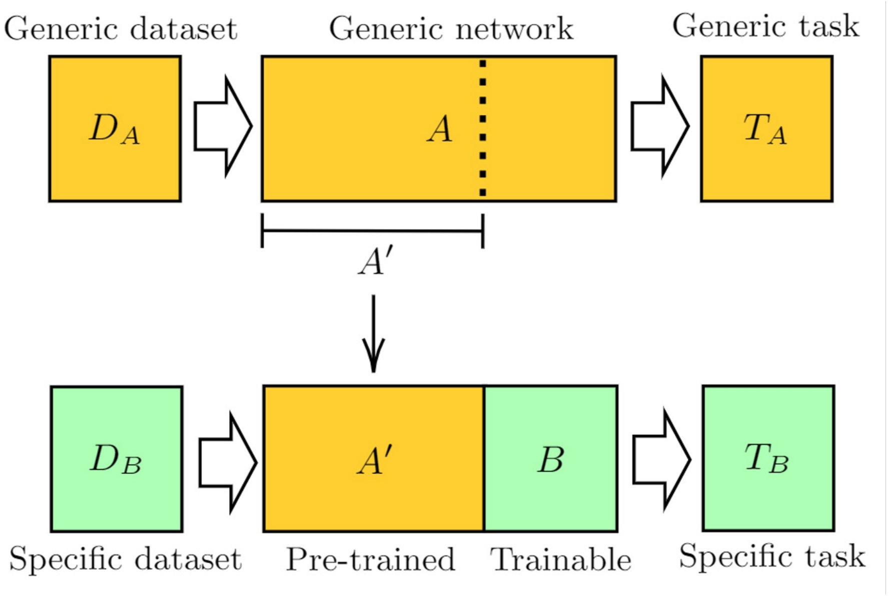
- “Trial as images” -> image classification problem
- Four major errors: noise, droop, shift, and offset
- Network comparison
  a. 1 Dense layer
  b. 1 Convolutional layer + 1 Dense layer
  c. Transfer learning and fine-tuning with MobileNet



Margie moved into her new apartment at the end of the
summer. The principal introduced the new president of
the junior class. None of the students wanted to have an
exam after Spring Break. Mark told Janet that he would
meet her after baseball practice. Bill complained that the
magazine included more adds than articles.

Input generalization

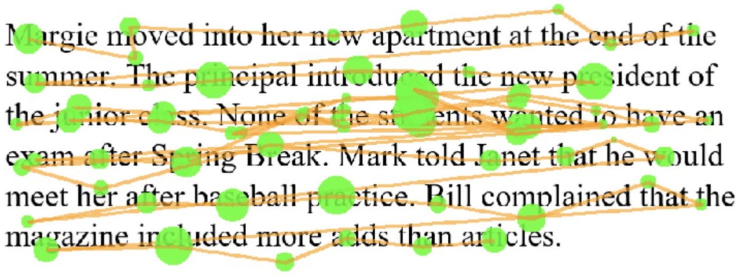
Transfer Learning



Synthetic Dataset

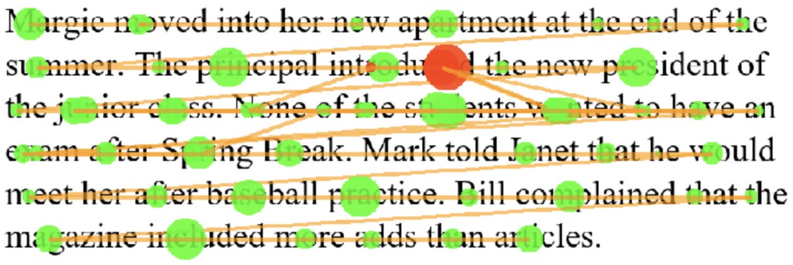
```
error_type = [“noise”, “shift”, “droop”, “offset”, “no error”]

FOR each error_type:
    FOR error_magnitude = 1 to 10
        FOR trial = 1 to 100:
            Generate fixations with random regression and skipping
            Add an error phonomena with specified magnitude to fixations
```

Synthetic Data Correction: Noise (“chain” algorithm)



Margie moved into her new apartment at the end of the summer. The principal introduced the new president of the junior class. None of the students wanted to have an exam after Spring Break. Mark told Janet that he would meet her after baseball practice. Bill complained that the magazine included more adds than articles.

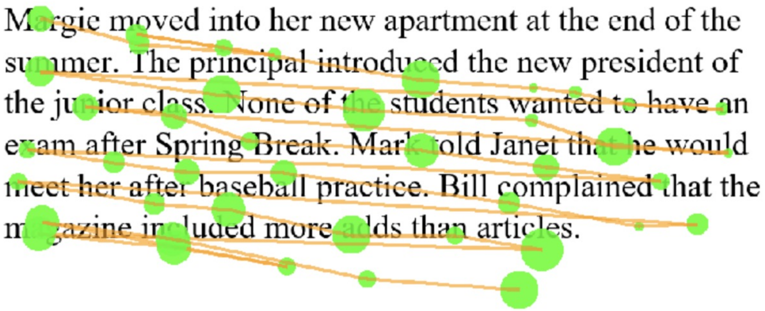
Uncorrected



Margie moved into her new apartment at the end of the summer. The principal introduced the new president of the junior class. None of the students wanted to have an exam after Spring Break. Mark told Janet that he would meet her after baseball practice. Bill complained that the magazine included more adds than articles.

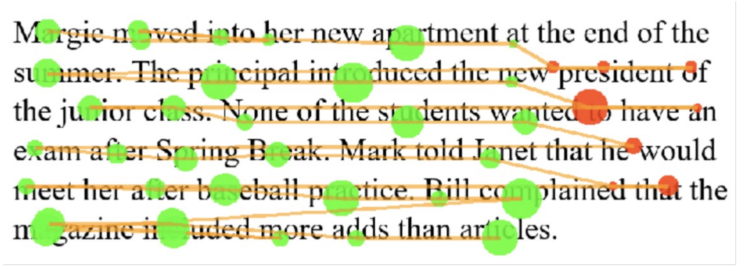
Corrected

Synthetic Data Correction: Droop (“regress” algorithm)



Margie moved into her new apartment at the end of the summer. The principal introduced the new president of the junior class. None of the students wanted to have an exam after Spring Break. Mark told Janet that he would meet her after baseball practice. Bill complained that the magazine included more adds than articles.

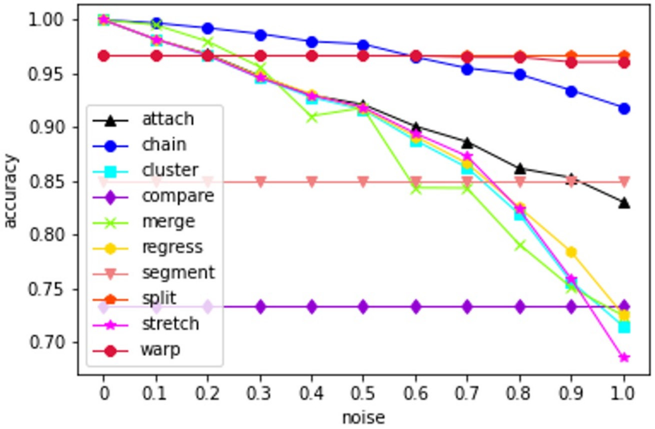
Uncorrected

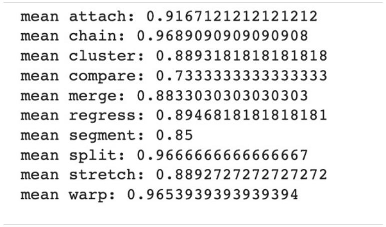


Margie moved into her new apartment at the end of the summer. The principal introduced the new president of the junior class. None of the students wanted to have an exam after Spring Break. Mark told Janet that he would meet her after baseball practice. Bill complained that the magazine included more adds than articles.

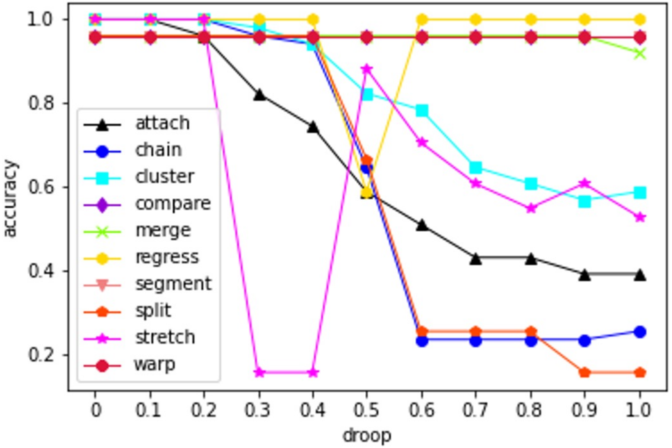
Corrected

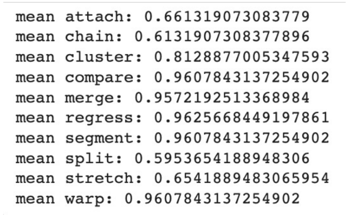
Algorithm Accuracy on Synthetic Data (noise, droop)





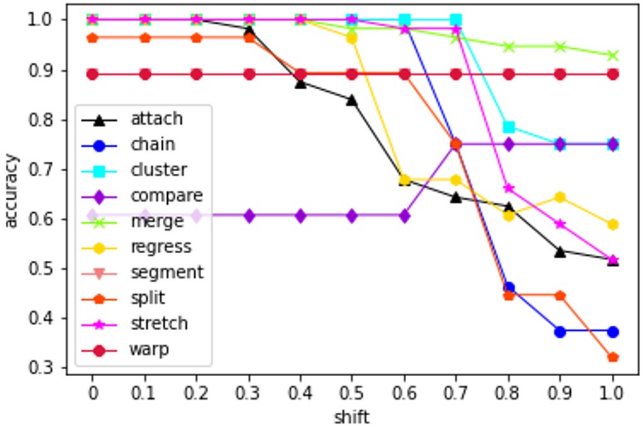
```
mean attach: 0.9167121212121212
mean chain: 0.9689090909090908
mean cluster: 0.8893181818181818
mean compare: 0.7333333333333333
mean merge: 0.8833030303030303
mean regress: 0.8946818181818181
mean segment: 0.85
mean split: 0.9666666666666667
mean stretch: 0.8892727272727272
mean warp: 0.9653939393939394
```

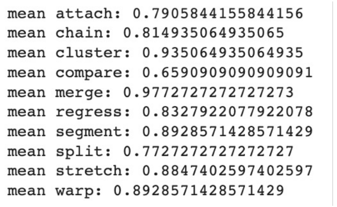




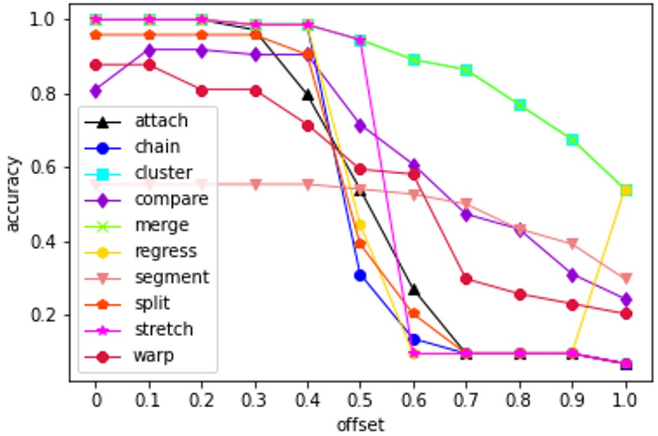
```
mean attach: 0.661319073083779
mean chain: 0.6131907308377896
mean cluster: 0.8128877005347593
mean compare: 0.9607843137254902
mean merge: 0.9572192513368984
mean regress: 0.9625668449197861
mean segment: 0.9607843137254902
mean split: 0.5953654188948306
mean stretch: 0.6541889483065954
mean warp: 0.9607843137254902
```

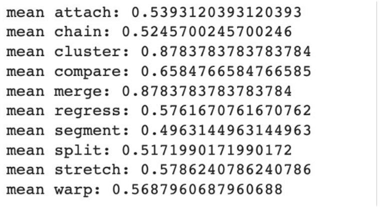
Algorithm Accuracy on Synthetic Data (shift, offset)





```
mean attach: 0.7905844155844156
mean chain: 0.814935064935065
mean cluster: 0.935064935064935
mean compare: 0.6590909090909091
mean merge: 0.9772727272727273
mean regress: 0.8327922077922078
mean segment: 0.8928571428571429
mean split: 0.7727272727272727
mean stretch: 0.8847402597402597
mean warp: 0.8928571428571429
```

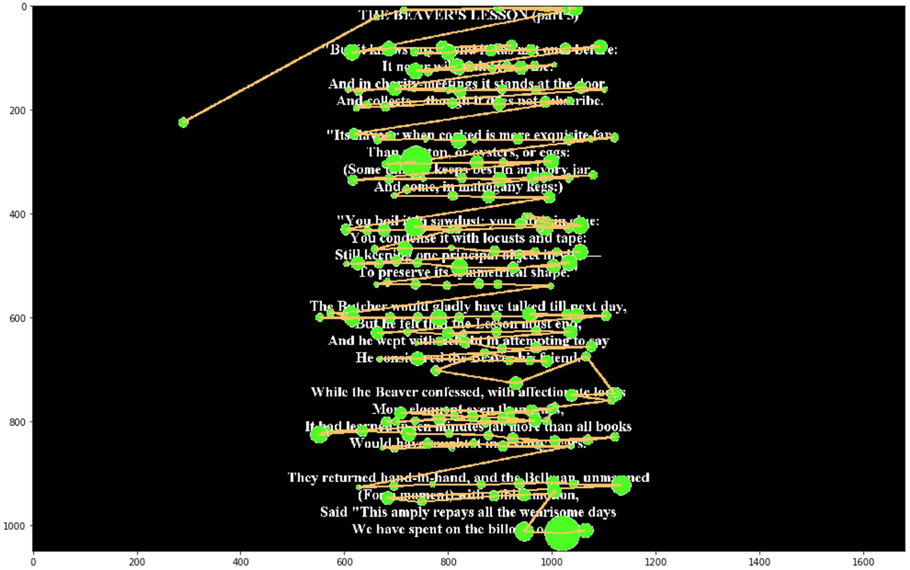




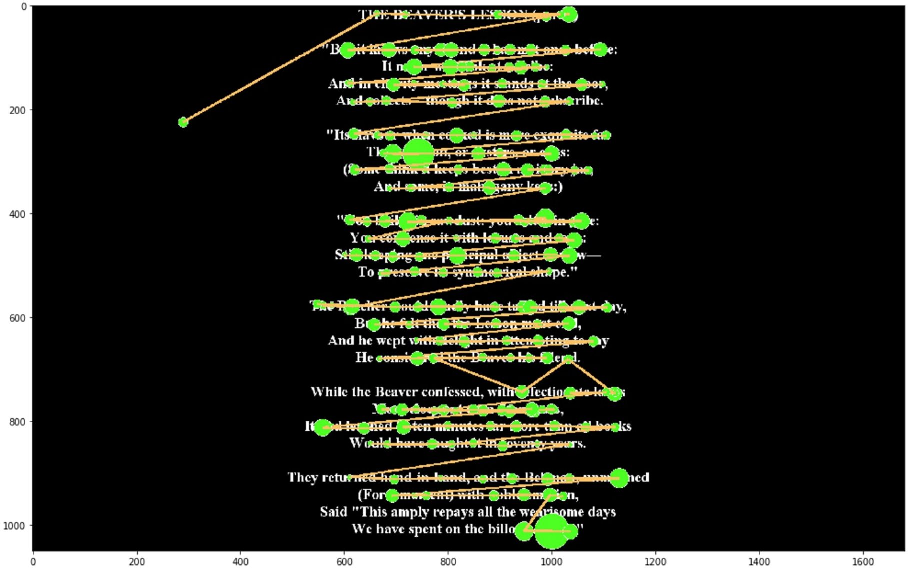
```
mean attach: 0.5393120393120393
mean chain: 0.5245700245700246
mean cluster: 0.8783783783783784
mean compare: 0.6584766584766585
mean merge: 0.8783783783783784
mean regress: 0.5761670761670762
mean segment: 0.4963144963144963
mean split: 0.5171990171990172
mean stretch: 0.5786240786240786
mean warp: 0.5687960687960688
```

Golden Set Generation

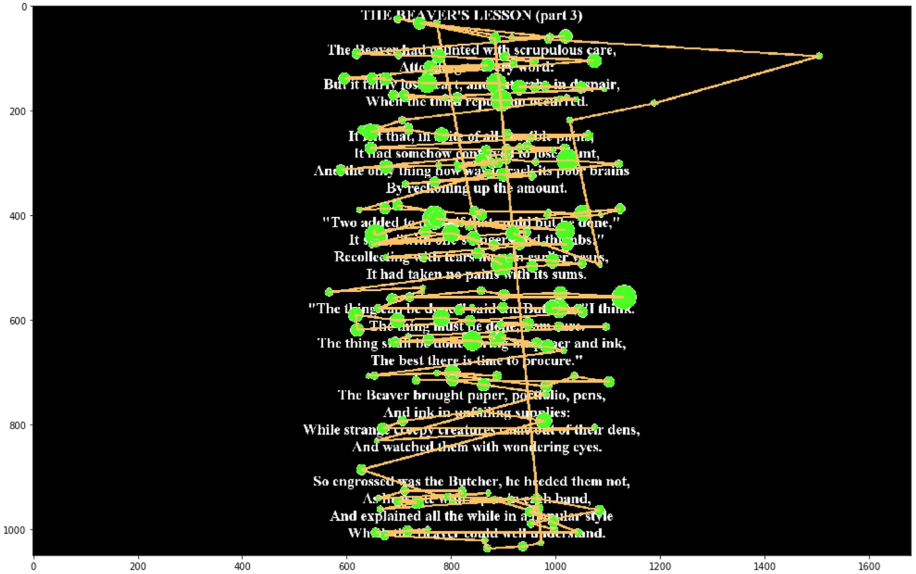
- Fix8 program for manual correction
- Golden Set consists of Json files of manually corrected data
    - Each corresponds to a uncorrected data file
- The Golden Set contains 15 manually corrected trials
- Provided strong insight into the reading behaviors of individuals
    - Influenced how we understood the research we were doing



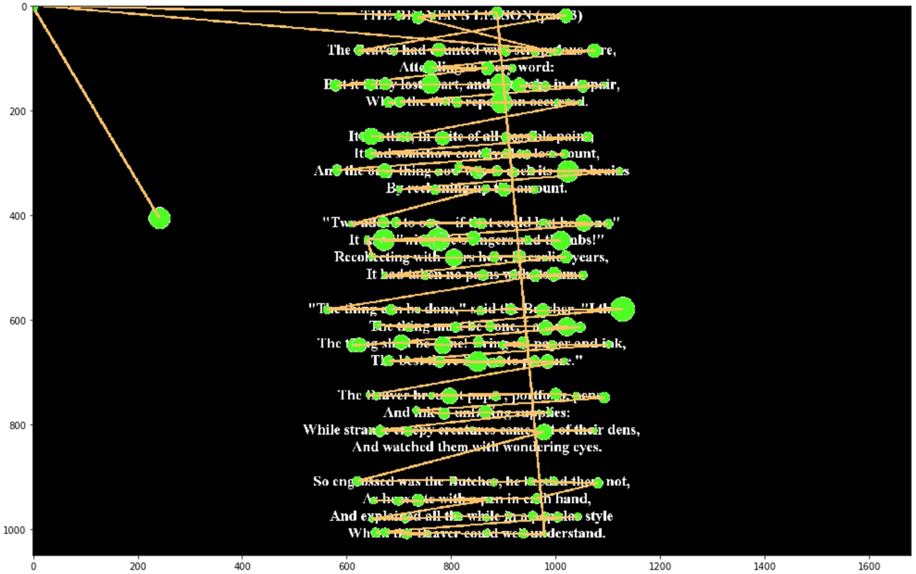
Uncorrected



Corrected



Uncorrected

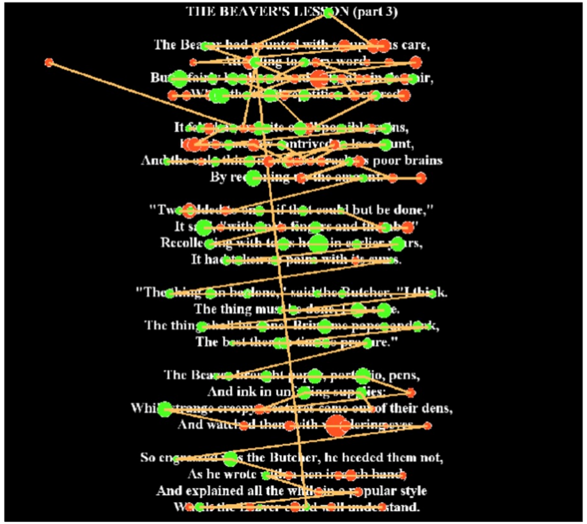


Corrected

Correction Accuracy on GazeBase

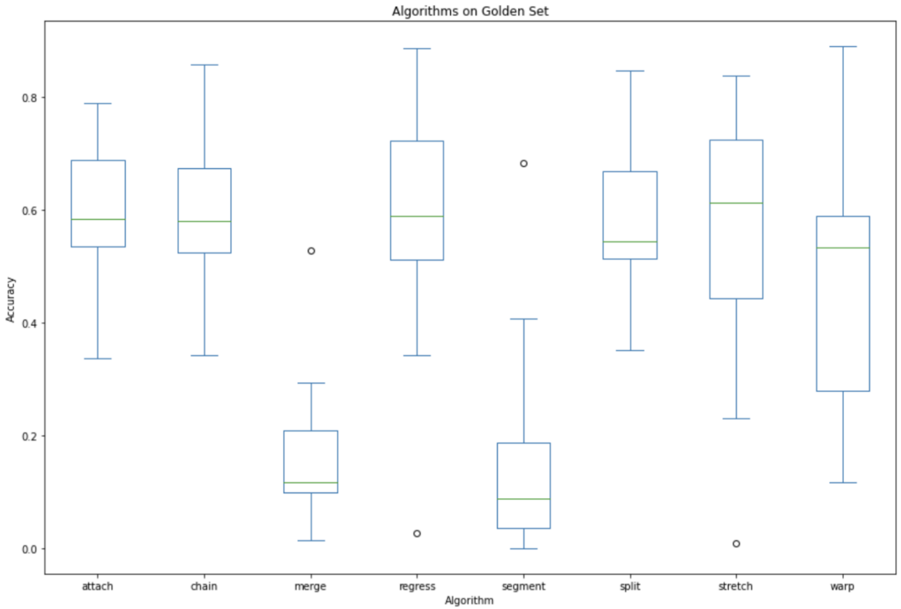
- 8 algorithms developed by other researchers
- Ran each of the 15 Golden Set trials through each algorithm
- Carr’s paper and the issues with real world data and automated algorithms
    - After tuning the algorithms, we received usable accuracy results
- We compared the algorithms against each other, across trials
    - This would be the comparison point for if adding classification would increase accuracy

“Regress” Algorithms on GazeBase



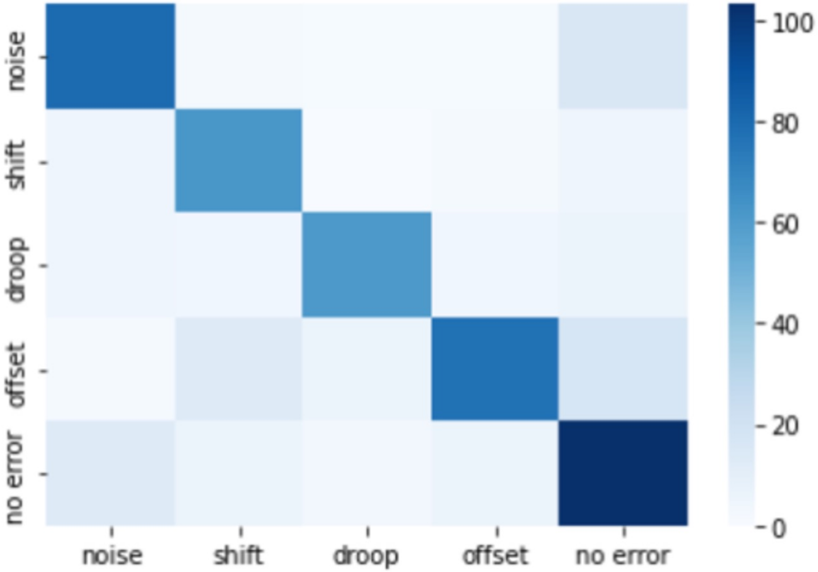
- Correction algorithms are generally able to correct droop, offset and short-distance regressions
- They are not good at correcting multi-line regressions and skipping

Correction Accuracy

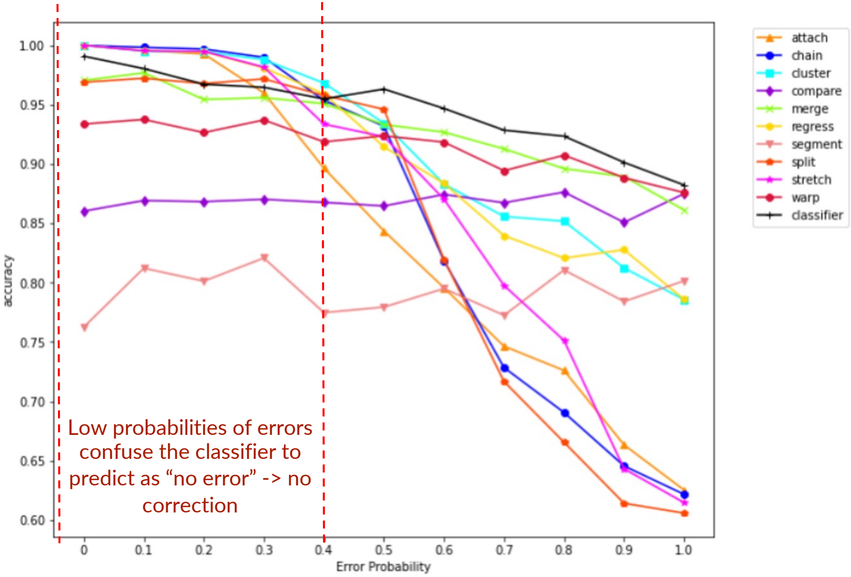


NN Classifier Accuracy

- Test set was generated by random selection of error type, regression/skipping probability, and scale of error
- Epochs = 30; batch_size = 30
- ~97% training, ~81% testing accuracy



Did classification help correction on Synthetic data?

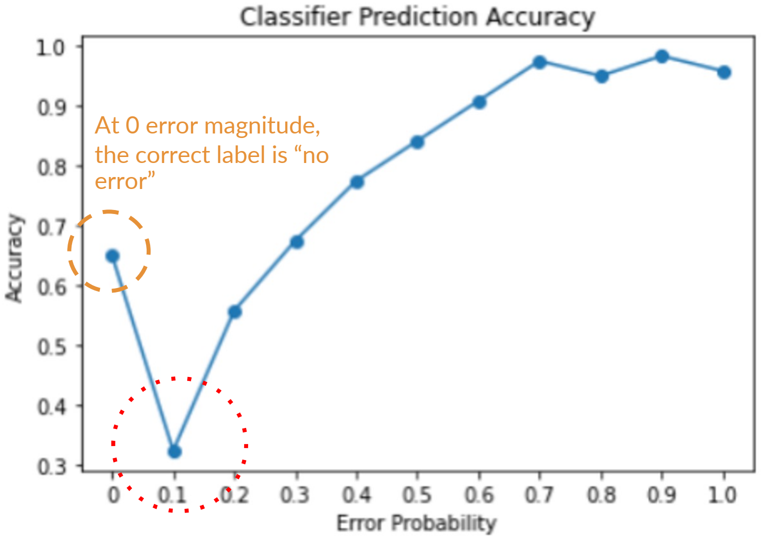


Low probabilities of errors confuse the classifier to predict as “no error” -> no correction

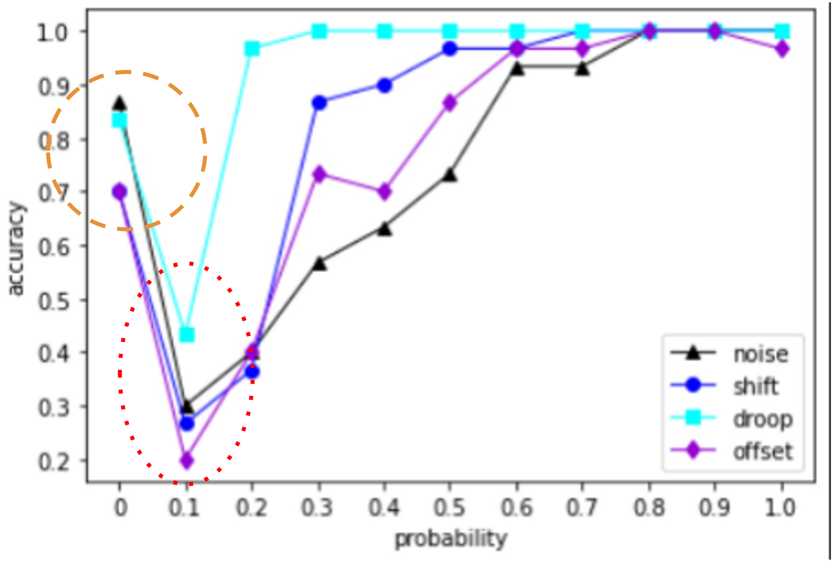
- The average of all four types of errors repeated for 30 times (120 samples) at each error magnitude
- Comparatively low accuracy of the classifier model at low error magnitude due to low prediction accuracy when error magnitudes are small

Noise -> split
Droop -> regress
Shift -> merge
Offset -> merge/cluster

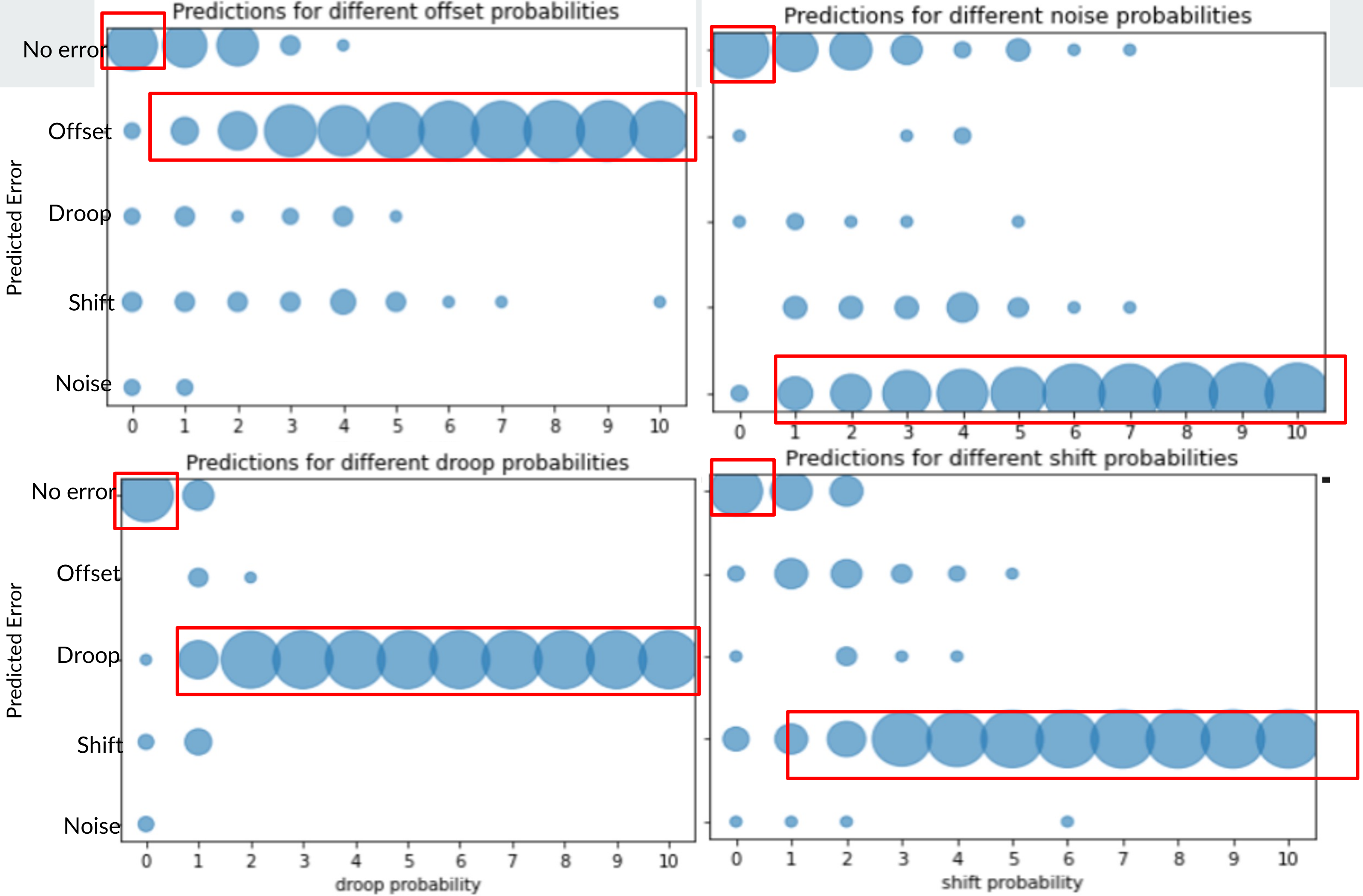
Classifier's confusion at low error probability



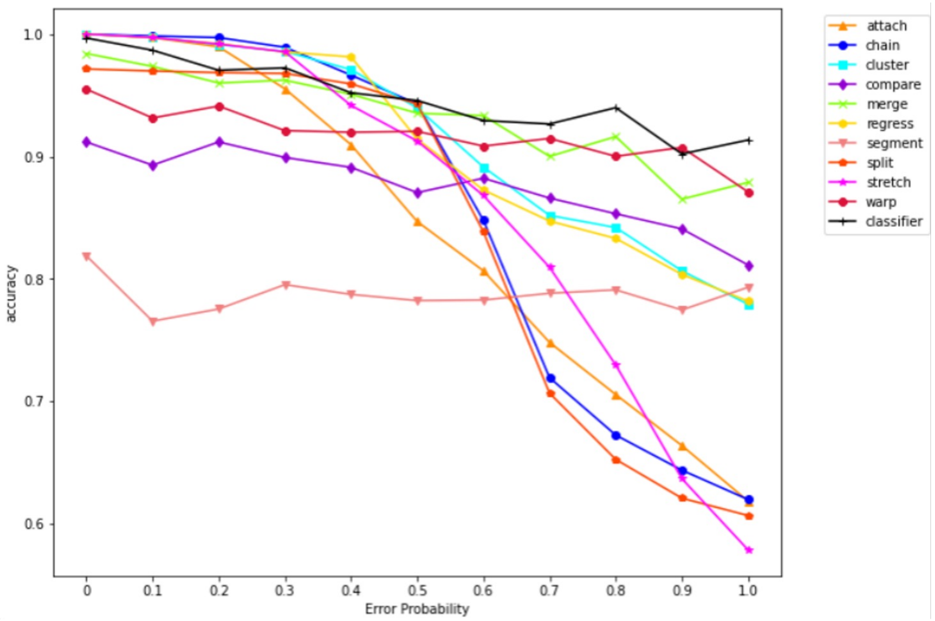
At 0 error magnitude, the correct label is "no error"



The classifier tends to predict trials with error magnitude < 0.4 as "no-error"


Predictions for different offset probabilities
Predictions for different noise probabilities
Predictions for different droop probabilities
Predictions for different shift probabilities
Predicted Error
Predicted Error
No error
Offset
Droop
Shift
Noise
No error
Offset
Droop
Shift
Noise
0
1
2
3
4
5
6
7
8
9
10
droop probability
shift probability

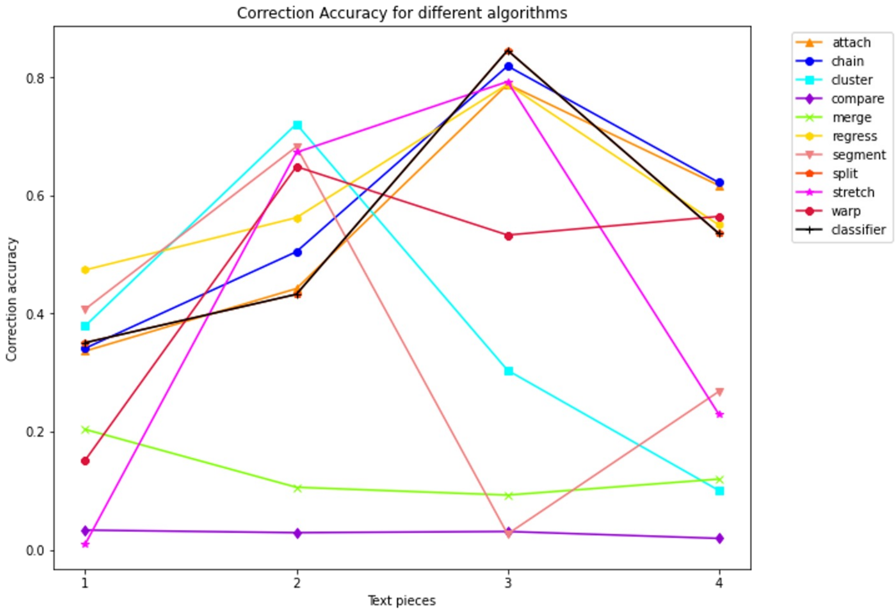
“Hacking” for a better accuracy



Use “chain” when predicted as “no error”

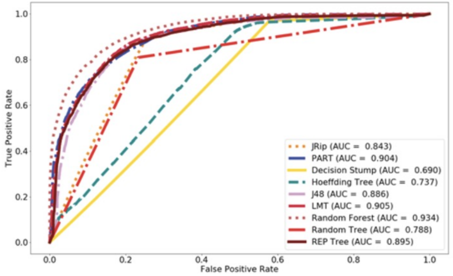
Classifier with real data

- “Noise is the most prominent error in every trial”
- The accuracy of using classifier strictly follows “split”
- There is not a single algorithm that performs the best among every trial due to the complexity of real-data



Rule-Based Approach

- Existing literature is very limited.
- Peripheral research using decision trees and rule-based components cite up to 84% accuracy.
- Serves as a comparison to NN-based approach.



Fuhl, W., Castner, N., & Kasneci, E. (2018, October). Rule-based learning for eye movement type detection. In Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data (pp. 1-6).

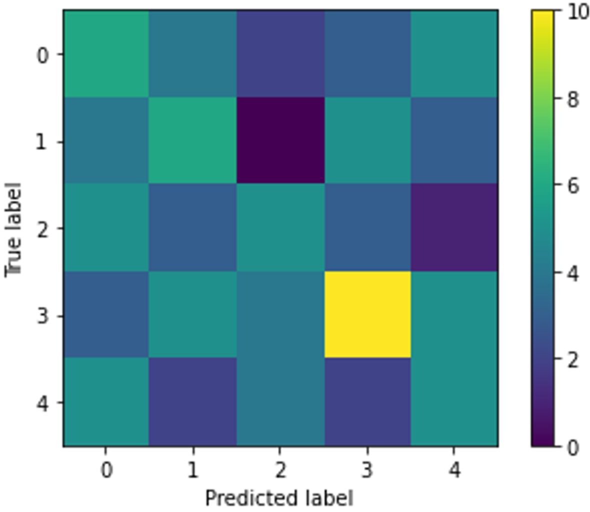
De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., Jayarathna, S., Michalek, A. M., & Jayawardena, G. (2019, July). A rule-based system for ADHD identification using eye movement data. In 2019 Moratuwa Engineering Research Conference (MERCon) (pp. 538-543). IEEE.

Method

- Basic implementation. No rule extraction was used.
- Chooses class based on the most prominent error type, i.e the most number of times a rule for a particular error type has been triggered.

Results

RID	word_length < 4	x_offset	y_offset	angle < 180	angle > 180
0 no error	no	no	no	no	no
1 droop	no	no	yes	no	yes
2 offset	no	yes	yes	no	yes
3 noise	no	yes	yes	yes	yes
4 shift	no	yes	yes	yes	yes



Future Work

- Rule-based **time series** algorithms
- Rule extraction using decision tree, PRISM, etc
- Real-time correction algorithms
- Machine learning model to look at how people correct these trials
- Other Neural Network models

Thank you! Questions?