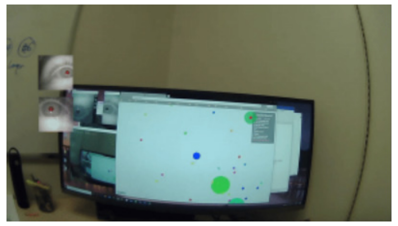
GAZE-BASED ACTIVITY RECOGNITION WITH A LSTM

1. BACKGROUND

Detecting gaze-based activities such as reading, browsing, writing and gaming can have many applications. One of these use cases is to detect if a user is distracted behind their laptop and remind them to get back to work.

Productive



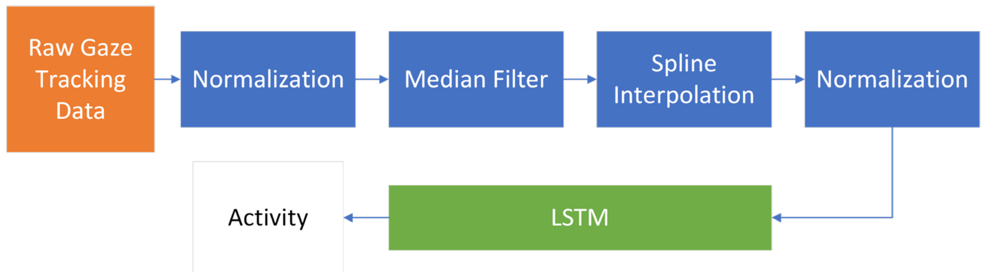
Distracted

Images gathered from [1]

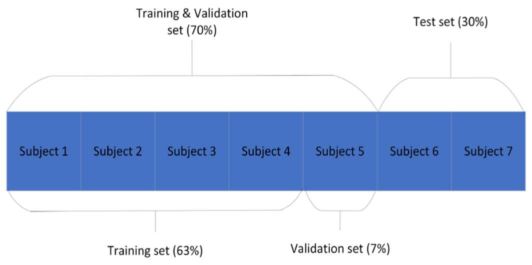
Detecting gaze-based activities has currently mostly been done using conventional machine learning methods. Deep learning methods such as the long-short term memory neural network (LSTM), show high potential in this area, due to their automatic feature extraction. Besides this, the LSTM shows potential, due to its ability to learn from long sequences of input data. Thats is why the research question is: Can a LSTM be used for gaze-based activity recognition?

2. METHODOLOGY

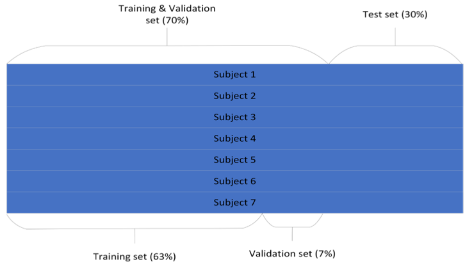
To answer the question stated previously, some steps have to be taken. These steps are data pre-processing, designing and implementing the LSTM, determining and tuning hyper parameters of the LSTM and evaluating the LSTM. In the figure below an overview of the classification pipeline can be found.



3. DATA SPLIT

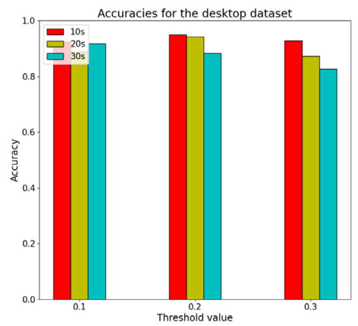


Subject-independent



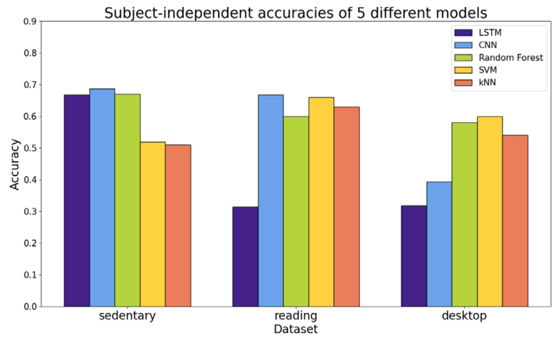
Subject-dependent

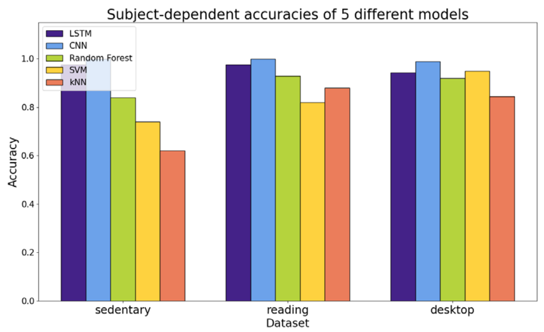
4. RESULTS



In the figure on the left, the user dependent accuracies for different values of hyper parameters on the desktop data set can be found. It can be seen that the window size of 10s and threshold value of 0.2 performs best. The same results were found when the model was trained and evaluated on the other two data sets.

In the figure on the left, the subject independent accuracies of the LSTM compared to other models can be found. The model unperforms when compared to the other models, especially on the smaller data sets.





In the figure on the left, the subject dependent accuracies of the LSTM compared to other models can be found. This figure shows that the LSTM performs better than all conventional machine learning methods.

5. CONCLUSION

The best performing model turned out to be a neural network with a single LSTM layer and a smaller window size of ten seconds for splitting the data. It was found that the LSTM is highly suited when predicting with subject-dependent data, with an average accuracy of 96.61%. The model is less suited when using subject-independent data, with an average accuracy of 43.34% it being the worst performing model on subject-independent data.

6. FUTURE WORK

Bigger data set

The fact that the performance decreases on smaller data sets is probably because the model is overfitting to the small amount of training data. This could also partially explain why the model is performing worse on the bigger window size, since a bigger window size results in less data points. Further research with a bigger data set could be conducted to test this hypothesis.

Calibrating the model

To achieve higher accuracy's when classifying activities of a new subject, the model could be retrained on a small data sample of that subject. This could be done by letting the subject perform all the activities for a short amount of time and using that data to continue training the model before actually using it. Research could be conducted to test the feasibility of this method.

Testing more hyper-parameters

Currently only the window size and filter threshold were used as hyper-parameters. However, there are more parameters which could be adjusted to increase the performance of the model. Examples of these hyper-parameters are overlap in the sliding window, batch size and amount of units in the LSTM layer.

REFERENCES

[1] Guohao Lan, Bailey Heit, Tim Scargill, and Maria Gorlatova. GazeGraph: Graph-Based Few-Shot Cognitive Context Sensing from Human Visual Behavior, page 422â435. Association for Computing Machinery, New York, NY, USA, 2020.

Author: *Kasper Vaessen* *k.e.vaessen@student.tudelft.nl*
Responsible Professor: *Guohao Lan* *G.Lan@tudelft.nl*
Supervisor: *Lingyu Du* *Lingyu.Du@tudelft.nl*