2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 5-8, 2025. Vienna, Austria

Detecting Human Distraction in Manual Control

Y. David Li, Daan M. Pool and Max Mulder

2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 979-8-3315-3358-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/SMC58881.2025.11342582

***Abstract*— InceptionTime neural network models were trained to detect distractions in manual control tasks with pursuit and preview displays. Training and test data were collected in an experiment where ten participants were deliberately distracted from the primary control task using the Surrogate Reference Task. Overall, distractions are easier to detect in pursuit tasks, with test accuracies of around 80% and 60% for pursuit and preview data, respectively. With preview, human controllers see the future target trajectory, which enables them to mitigate distraction effects. Unexpectedly, data with longer distractions from 'hard' secondary tasks are more difficult to classify than 'easy' distractions; an effect attributed to differences in human behavior between the training and test data collection conditions. These results show clear opportunities for neural network models to detect distractions, in real-time, for increasing safety of human-operated vehicles.**

I. Introduction

Distracted driving is one of the main risk factors in road accidents, often related to the use of smartphones and in-vehicle information systems [1]–[3]. Creating a tool that can objectively, and non-intrusively, detect when people are distracted may increase safety on the roads, but also in other domains. For example, detected distractions can be used to timely trigger supporting automation or increased authority of (haptic) shared control systems [4]. Most research on detecting human distractions in manual control focuses on driver distraction. This includes defining what a distraction is, distraction monitoring methods, and mitigation techniques [1], [2]. Distractions can be in the form of a manual, visual or cognitive distraction [5], while detection methods vary from vision to sensor-based approaches, or a combination [3].

This paper aims to verify if detecting visual distractions, using a sensor-based approach as also considered by [1], [2], [6], [7], from basic manual control task data is feasible. A human-in-the-loop experiment was performed where operators performed pursuit and preview tracking tasks [8], [9], where the Surrogate Reference Task (SuRT) [10] was used as a (distracting) secondary task. The collected experiment data are used to train a Neural Network (NN), using the InceptionTime architecture [11], to detect distractions.

In the experiment, participants performed tracking runs with both a pursuit and preview display in three different conditions: 1) no distractions, 2) continuous distractions, and 3) prompted distractions. Data obtained from tracking conditions with no distractions and continuous distractions were used to train the NN as labeled 'normal' and 'distracted' data, respectively. Runs with periodically prompted distractions were used to *test* the accuracy of the trained NNs. Two difficulty levels were used for the distractions by changing the size of the target circle in the SuRT [10].

This paper focuses on describing the experiment method and rationale and is structured as follows. Section II discusses the experiment method, Section III the data processing. Results are presented in Section IV, followed by a Discussion and Conclusions, Section V and Section VI, respectively.

II. Experiment

A. Apparatus

The experiment was conducted in the HMI-Lab, see Fig. 1(a), at TU Delft. Participants were seated (1), held a side-stick in their right-hand (2) for performing the (primary) tracking task presented on a display (Screen 1) in front of them (3), while also performing the (secondary) Surrogate Reference Task (SuRT) presented on a tablet (Screen 2) to their left (4) in some experiment conditions. Participants wore an eye tracker (Pupil Labs Core) to record their head and eye movements, which was used to objectively measure when they were distracted. Fig. 1(b) illustrates that the setup's dimensions were such that participants could never see the primary and secondary task displays at the same time.

B. Primary task: pursuit/preview tracking

Forced-pace pursuit and preview tracking tasks were performed as the primary task, see Fig. 1(c). Participants were instructed to minimize the root mean square (RMS) tracking error, $e(t)$, i.e., the distance between the symbols representing the controlled element (CE) output $x(t)$ and the target $f_t(t)$. Participants controlled the CE, a single integrator (rate control) with a 1.5 gain, with the side-stick.

Pursuit and preview displays differ in the preview time τ_p, i.e, the length of the *future* trajectory of target signal $f_t(t)$ shown on the display, see Fig. 1(c). The preview time is zero for a pursuit display, which negatively affects the human controller (HC) performance [8]. Together, the pursuit and preview displays represent extremes of tracking. Because with a pursuit display the future trajectory is not known (like driving in fog), it was expected that the effects of distraction would be more evident for this display.

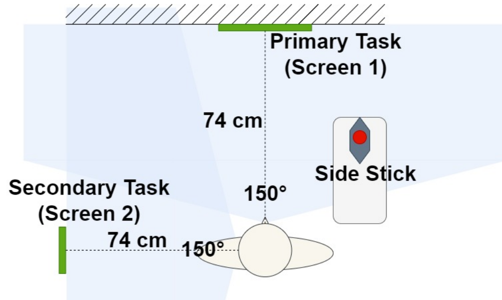
The tracking task matched those tested in earlier experiments [8], [9]. Multisines were used for both the target signal $f_t(t)$ and a disturbance signal $f_d(t)$ acting on the CE; the bandwidth of both signals was set at 2.5 rad/s. Each tracking run lasted 128 seconds, with an 8 seconds run-in time.

C. Secondary task: the Surrogate Reference Task

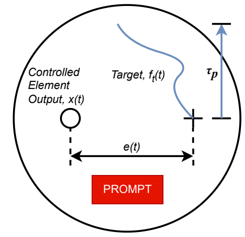
The secondary task acted as a distraction for participants while performing the primary tracking task. Here, the Surrogate Reference Task (SuRT) [10] was used, see Fig. 2.

Control & Simulation, Faculty of Aerospace Engineering, TU Delft, The Netherlands. Email: {d.m.pool, m.mulder}@tudelft.nl

(a) The HMI-Lab simulator



(b) Experiment setup dimensions



(c) Pursuit/preview tracking display

Fig. 1: Experiment setup (a), dimensions (b), and tracking display (c); blue preview line only shown for preview task.

Participants had to select the largest (target) circle (8 mm diameter), in the midst of a set of 49 smaller circles. When the target is selected on the tablet's touchscreen, a new randomized placement of the target and distractor circles is generated. Two SuRT variations were tested: 'easy' and 'hard', which differed in the size of the distractor circles (4 mm for easy, 7 mm for hard) [12]. A predefined scoring scheme [10] was used for secondary task performance.

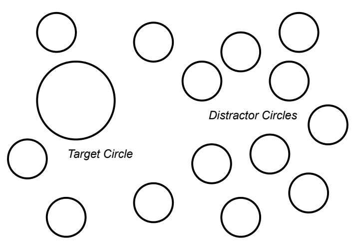
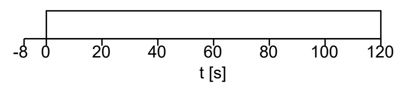
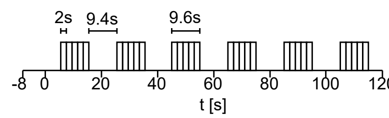


Fig. 2: Example of the SuRT [10] secondary task.



(a) Continuous distraction



(b) Prompted distraction

Fig. 3: Continuous (a) and prompted (b) distractions.

TABLE I: Experiment conditions.

Task	*Conditions*		
Preview (R)	R, n	R, c, e R, p, e	R, c, h R, p, h
Pursuit (S)	S, n	S, c, e S, p, e	S, c, h S, p, h

D. Variation in distractions

Data were collected in three conditions: i) no distraction, ii) continuous distractions, and iii) prompted distractions. Fig. 3(a) shows that in the conditions with continuous distractions, participants were tasked with completing the secondary task continuously, whilst performing the tracking task. They were instructed that both tasks were *equally* important and to achieve a score of at least 2,000 on the SuRT. In the prompted distraction condition, 6 prompted moments of distraction occurred in each trial, see Fig. 3(b). Within each period, a distraction occurred randomly in one of the five 2-second slots. The minimum and maximum time between two consecutive prompts were thus 9.4 s and 24.6 s, respectively. A visual prompt reading '*Please find the large circle!*' appeared on the primary display (Fig. 1(c)) whenever the secondary task had to be performed.

E. Ethics approval, participants, and procedures

The experiment (number 3010) has been approved by TU Delft's Human Research Ethics Committee. Ten students and employees from TU Delft's aerospace engineering faculty participated. All were right-handed, had good vision, and were aged between 20 to 55 years.

The experiment had 10 conditions, 5 for each display type, Preview (R) or Pursuit (S), see Table I. Here, the characters 'e' and 'h' represent the 'easy' and 'hard' SuRT tasks, 'c' and 'p' the 'continuous' and 'prompted' distractions, and 'n' the non-distracted condition. The experiment was performed in two separate sessions of 30 tracking runs each, one with each display. The first 5 trials of each session were training runs. Each condition was tested in blocks of 5 consecutive runs; the first run was training, the remaining 4 were used as the measurement data. Latin squares were used to balance order effects, learning, and fatigue.

III. Data Processing

A. Data collection and labeling

The collected primary task data consisted of time series, sampled at 100 Hz. The eye-tracker measured participants' head orientation and provided objective data of when they focused on primary task (Screen 1, 'non-distracted') or looked away to the (SuRT) distractions (Screen 2, 'distracted'). The distractions lasted between 0.3 and 6 seconds. The NN models were trained on the data from the no/continuous distraction runs, and tested on the prompted distraction data.

Before the tracking data could be used for training NN models, they were split into labeled chunks of successive samples. Based on previous work [13], the NN was fed with data chunks of 1.5 seconds, with an overlap of 0.75 s. Naturally, the fixed and time-paced 1.5 s windows did not perfectly correspond with the times participants switched between 'non-distracted' and 'distracted' behavior as measured with the eye tracker. Chunks with *only* non-distracted data were labeled as 'normal', while all chunks with only distracted data *and* or a mix of distracted and normal data were labeled as 'distracted', see [14] for more details.

With 10 participants, 4 runs per condition, and a maximum of 158 1.5-second data chunks (with overlap 0.75 s) extracted from each 120 s run, a maximum of 6,320 data chunks are available per condition. After applying the labeling method to the continuous distraction condition data, a total of 3,694 (easy) and 4,058 (hard) 'distracted' chunks were obtained for the pursuit task, and 3,212 (easy) and 4,314 (hard) 'distracted' chunks for preview. Only these distracted chunks were used in training the NN models, together with the 6,320 'normal' chunks from the 'no distraction' condition.

B. Neural network training and hyperparameters

Following previous research [11], the data chunks used to train the NN models consisted of six tracking signals: the error $e(t)$, the HC input $u(t)$, the system output $x(t)$, and their time derivatives. Data were normalised per run, such that the signal magnitudes could not affect the NN learning process [13]. After data preparation, all chunks used to train the NN were randomly allocated to a training (80%) or validation (20%) set. The validation set was used to quantify how well the model performed on the training data. Eventually, the models were tested using test data from prompted tracking runs.

The hyperparameters for training the NN were not (further) optimized and were set as in [11]: batch size=64, epochs=25, kernel size=64, number of filters=24, maximum learning rate=0.00275, weight decay=0.05, no bottleneck, no batch normalization, no residual connections.

C. Neural network models

Six NN models were trained and validated using a slightly imbalanced dataset consisting of the 'distracted' chunks obtained from the runs with continuous distractions and the 6,320 'normal' data chunks from the 'no distraction' conditions. For both the pursuit and preview tasks, three NN models were trained: 1) a model trained only on the 'easy' distraction data, 2) a model trained only on the 'hard' distraction data, and 3) a 'combined' model trained on both the easy/hard distraction datasets. In the remainder of this paper, these six *models* are referred to with an 'M' in front of the experiment condition symbols listed in Table I. For example, 'MR-e' and 'MR-h' are the preview (R) models trained on the easy and hard data, respectively, while 'MR-C' indicates the corresponding 'combined' model. Similar symbols starting with a 'D' are used to indicate the *datasets*: 'DS-pe' and 'DS-ce' refer to the pursuit (S) data for 'prompted' easy and 'continuous' easy distractions, respectively.

The performance of the six NN models was assessed based on their accuracy, using the validation dataset *after* training:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

where T and F denote true and false classification of samples for positives (P, distracted) and negatives (N, non-distracted), respectively. Each model was trained 10 times to see how well the neural network is able to classify the training data over multiple training runs. Furthermore, Receiver Operating Characteristic (ROC) curves are obtained for all models based on their (prompted) test data performance only.

Predictions made by the model were based on a *fixed* decision threshold of 0.5 for the probability of a chunk belonging to a certain class. A probability close to 1 means that the chunk was considered to match the 'distracted' class, a probability close to 0 corresponds with class 'normal'.

IV. Results

A. Primary task performance

The RMS of error $e(t)$ was computed for complete tracking runs in the three conditions used to train the NN: the non-distracted runs and the easy and hard continuous distraction runs. The RMS($e(t)$) results are shown in Figs. 4 and 5 for the pursuit and preview data, respectively, for all participants together (left, $N = 40$) and all individual participants (right, $N = 4$). Clearly, performance is best in the preview tasks, as expected [8]. For both displays, performance deteriorates when distracted, especially for the hard distractions; also this reduction is less for preview.

The RMS of the HC input $u(t)$ data was also calculated, but showed no strong trends (see [14] for details). Overall, RMS($u(t)$) was lower for the preview display, as expected [8], and the spread in the data increases when distractions are introduced, especially for the hard distractions.

B. Secondary task performance

Participants were instructed to reach a score of 2,000 or higher in the SuRT task in the conditions with continuous distractions. Three of the ten subjects did not reach this level for the hard distractions. The average scores were 2,950 for the easy distractions (both displays) and 2,200 and 2,190 for the hard distractions, for pursuit and preview, respectively.

Also the time it took for participants to respond to the distraction (response time) and find the target circle (response duration) were measured for the prompted distraction runs. The response time was on average 1.1 seconds for all participants and did not differ much between displays and easy/hard distraction types, see [14] for details.

Fig. 6 shows the response duration, for both displays and the easy/hard distractions. No differences between the pursuit and preview data were found. For the easy distractions, it took participants on average 0.82 s to find the target, and in 96.3% (pursuit) or 90.0% (preview) of cases the target was found within 2 seconds. For the hard distractions, these

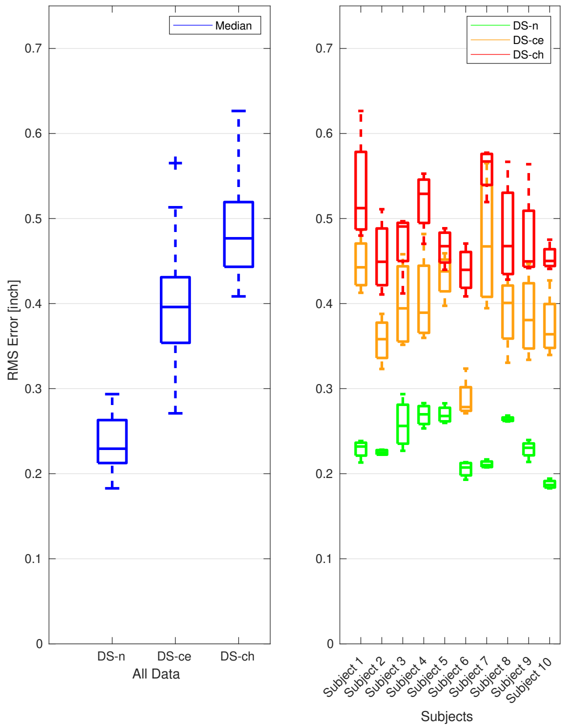


Fig. 4: RMS error for the DS-n, DS-ce, and DS-ch data.

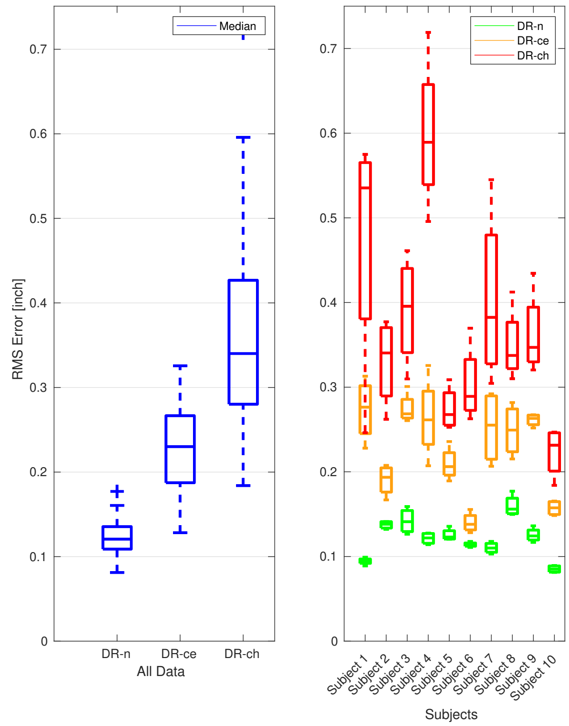


Fig. 5: RMS error for the DR-n, DR-ce, and DR-ch data.

numbers were 1.75 s, 55.4% and 52.9%, respectively. Hence, as was expected, hard distractions took longer to process.

C. NN models' validation performance

The six neural network models introduced in Section III were trained on the data of all 10 participants. For each model, ten repeated training runs were performed; the final validation accuracies are listed in Table II. The validation accuracy is high for all models, especially for the pursuit models. For both displays, models trained with only the 'hard' distractions yield the best validation accuracy; this is explained by the more severe distractions resulting in greater differences between the 'normal' and 'distracted' data.

TABLE II: Validation accuracy for trained NN models.

Model	**MS-e**	**MS-h**	**MS-C**
Max. accuracy	99.24%	100.00%	99.49%
Min. accuracy	98.37%	99.81%	99.06%
Average accuracy	98.74%	99.93%	99.27%
Model	**MR-e**	**MR-h**	**MR-C**
Max. accuracy	96.13%	99.82%	97.80%
Min. accuracy	94.50%	99.66%	96.50%
Average accuracy	95.71%	99.77%	97.27%

D. NN models' test performance

To test the accuracy of the six NN models trained to detect distractions, data chunks of the runs with *prompted* distractions were used. In these runs, participants were distracted at six randomized moments during each tracking run, see Fig. 3. Fig. 7 shows an example prompted easy distraction run, with in the top graph the target signal $f_t(t)$, CE output $x(t)$ and the six prompted distractions. Following a prompted

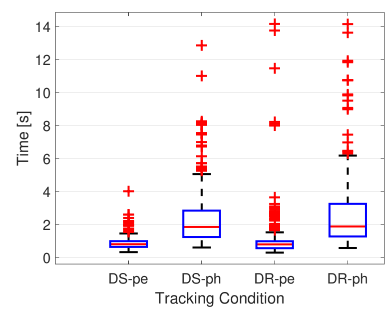


Fig. 6: Duration to complete the secondary task.

distraction, the tracking error $e(t) = f_t(t) - x(t)$ increases, as expected; the subject is distracted from the primary tracking display. The middle graph shows the corresponding HC input signal $u(t)$, which shows 'gaps' in HC control activity when a distraction occurs. In the bottom graph the eye tracker data are shown (blue bars) together with the detection probability of the NN (blue line) and the detection threshold (black horizontal dashed line) fixed at 0.5. Clearly, the NN showed very good test performance for this run; all distractions were detected without FPs.

The NN test performance for detecting the hard distractions is, however, found to be much worse (Fig. 8). Not all distractions are detected and the NN shows severe problems classifying non-distracted data as 'normal' (FPs). This result opposes the superior validation accuracy found for the hard distractions (Table II) and points to a difference in HC control behavior between the continuous (training) and prompted (test) data.

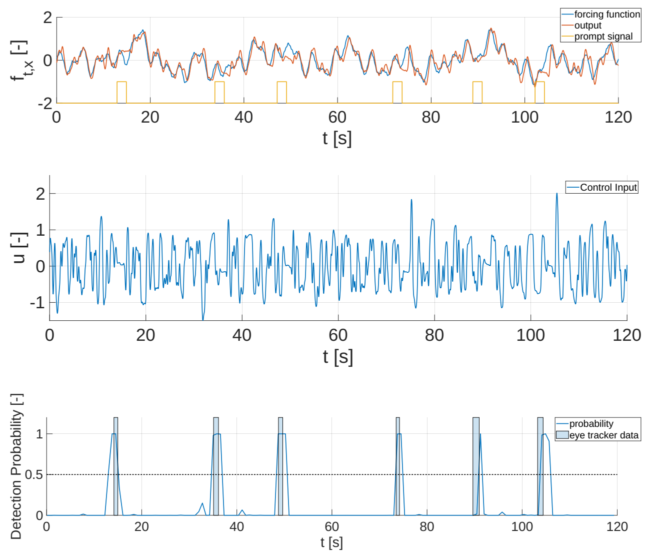


Fig. 7: Example tracking signals and detection probability for a prompted 'easy' distraction run (DS-pe).

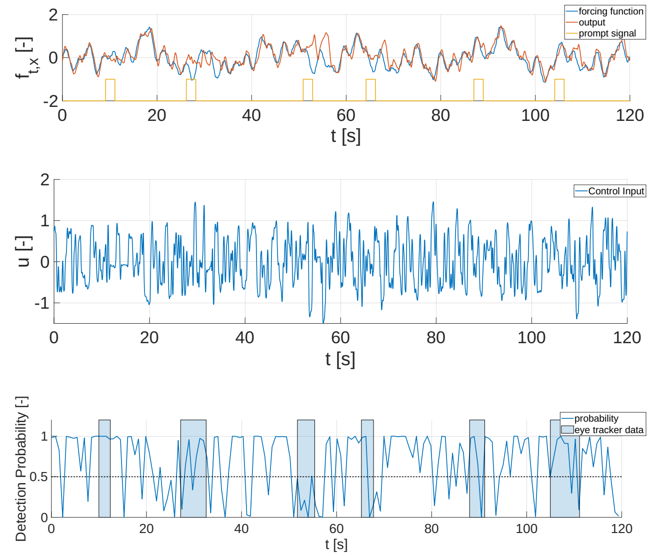


Fig. 8: Example tracking signals and detection probability for a prompted 'hard' distraction run (DS-ph).

The test accuracy of all NN models is summarized in Fig. 9, both for all ten subjects separately and on average. All models show variable performance across the different participants. Overall, the pursuit models (MS, green and red) show superior performance compared to the preview models (MR, blue and orange). For both displays, the detection of hard distractions (red and orange) is much worse, also for the NNs trained on the combined easy/hard data (MS-C and MR-C). The average test accuracies are summarized in Table III.

TABLE III: Average test accuracy for trained NN models.

Model	Data	Accuracy	Model	Data	Accuracy
MS-e	DS-pe	80.78%	MR-e	DR-pe	57.36%
MS-h	DS-ph	48.09%	MR-h	DR-ph	33.78%
MS-C	DS-pe	80.78%	MR-C	DR-pe	61.66%
	DS-ph	41.82%		DR-ph	29.15%

Recall that the classification threshold was fixed at 0.5 in all previous analyses. To investigate the effects of varying this threshold, ROC curves were generated, see Fig. 10 for all trained NNs. The corresponding Area Under the Curve (AUC) values are summarized in Table IV. The MS-e model has the best performance (AUC = 0.8688), followed by the MS-C model when tested on easy distraction data only (AUC = 0.8509). The lowest AUC values are found for NNs trained to detect the hard distractions. The ROC curves in Fig. 10 show superior detection performance for pursuit than for preview; the preview models also show little to no difference when compared to each other.

V. Discussion

Six 'one-size-fits-all' neural network models have been trained to detect distractions for HCs performing pursuit or preview tracking tasks. The high validation accuracy of all models (>94.5%) indicates that there is indeed a distinct difference between 'normal' and 'distracted' tracking data. As

TABLE IV: AUC for all NNs on different datasets.

Model	Data	AUC	Model	Data	AUC
MS-e	DS-pe	0.8688	MR-e	DR-pe	0.6405
MS-h	DS-ph	0.6758	MR-h	DR-ph	0.6225
MS-C	DS-pe	0.8509	MR-C	DR-pe	0.6436
	DS-ph	0.7452		DR-ph	0.6105
	DS-pC	0.7897		DR-pC	0.6452

these differences – reduced control inputs $u(t)$ and increased tracking errors $e(t)$ (directly following) when distracted – can be 'seen' when looking at the data oneself, verification of the benefit from NNs compared to more 'naive' classification methods is important future work. Continuously distracting subjects seems a viable way of collecting a large number of chunks of 'distracted' tracking data. For tracking runs with easy or hard distractions, 54.62% and 66.23% of the total number of samples were labeled as distracted, respectively. While tracking runs with continuous distractions allow participants to decide for themselves when and how many times they will complete the secondary task, we found that a minimum score should be set to encourage subjects to do the secondary task. Hence, the experiment successfully collected a controlled dataset that can be of further use for real-time HC distraction detection methods.

Despite very high validation accuracies, NN test performance reduced to values between 29.15% and 61.66% (preview runs) and 41.82% and 80.78% (pursuit runs) for the prompted distraction data, which the NNs had not 'seen' during training. These numbers are rather low, especially for the preview display. With preview, HCs can 'look into the future' for their primary task and can use this to time their switch to the secondary task *without* much loss in tracking performance. As a result, tracking errors increase more and more consistently with a pursuit display and NNs can also better detect HC distractions for this task.

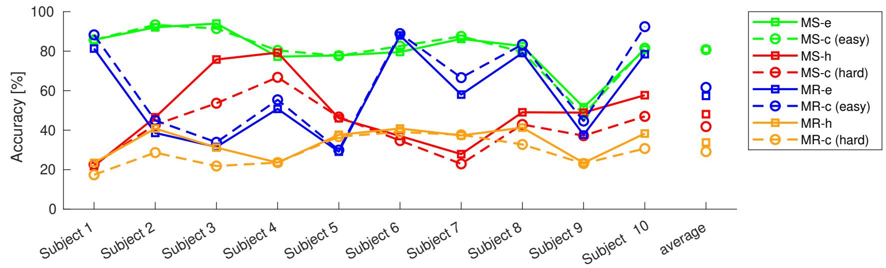


Fig. 9: Test accuracy of all six models, both for individual subjects and averaged across all subjects.

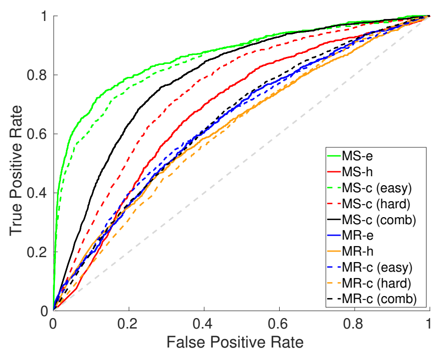


Fig. 10: ROC curves for all trained NNs.

Contrary to our expectation, the NNs performed *better* when detecting easy distractions than for hard distractions. A likely explanation for this result is a 'domain shift': due to the increased secondary task difficulty, the 'normal' and 'distracted' HC behavior in the (prompted) test data may no longer sufficiently match the corresponding training data, which was collected in separate tracking runs without distractions and with continuous distractions, respectively. Another factor contributing to this result may be the chosen labeling method. Chunks of data were labeled as 'distracted' *also* if the chunk contained a *mix* of non-distracted and distracted data. More strict labeling methods and elimination of mixed samples may lead to better results.

NN models performance varied much between participants. Hence, subject-specific models may result in a meaningful improvement over the 'one-size-for-all' approach followed here. Indeed, a preliminary analysis in [14] shows that personalized NN models show increased accuracy for 6 out of 10 and 7 our of 10 participants in the pursuit and preview conditions, respectively. However, this approach also requires much more data *per* participant to train and validate the NNs. Alternatively, it is recommended to explore subject-specific NN hyperparameters, here taken from previous work [11], or use an individualized decision threshold, here fixed at 0.5.

VI. Conclusion

Using collected experiment data from pursuit and preview tracking, InceptionTime neural networks were trained to detect human controller distractions in manual control tasks. When distracted, human controllers temporarily neglect the primary task: their control input reduces and tracking errors increase. Distractions are easier to detect during pursuit tracking compared to preview tracking, with test accuracies of around 80% and 60%, respectively. Surprisingly, distractions caused by easy secondary tasks are *easier* to detect than those caused by hard tasks. This is explained by larger variations between the training and test data with harder distractions, leading to frequent difficulties for the NNs to classify data as 'normal' non-distracted behavior.

Data Availability

The experiment data presented and analyzed in this paper is publicly available from the 4TU.ResearchData repository: https://doi.org/10.4121/60f781c4-bcfa-4959-8404-47e53101730c

References

[1] F. Tango and M. Botta, "Real-time detection system of driver distraction using machine learning," *IEEE Trans. on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 894–905, 2013.

[2] K. Torkkola, N. Massey, and C. Wood, "Driver Inattention Detection through Intelligent Analysis of Readily Available Sensors," in *7th IEEE Conf. on Intelligent Transportation Sys.*, 2004, pp. 326 – 331.

[3] A. Kashevnik, R. Shchedrin, C. Kaiser, and A. Stocker, "Driver Distraction Detection Methods: A Literature Review and Framework," *IEEE Access*, vol. 9, pp. 60 063–60 076, 2021.

[4] D. A. Abbink, M. Mulder, and E. R. Boer, "Haptic shared control: smoothly shifting control authority?" *Cognition, Technology & Work*, vol. 14, no. 1, pp. 19–28, 2012.

[5] NHTSA, "Overview of the National Highway Traffic Safety Administration's Driver Distraction Program," U.S. Department of Transportation, Tech. Rep. DOT-HS-811299, 2010.

[6] T. Ersal, H. J. A. Fuller, O. Tsimhoni, J. L. Stein, and H. K. Fathy, "Model-based analysis and classification of driver distraction under secondary tasks," *IEEE Trans. on Intell. Transp. Syst.*, vol. 11(3), 2010.

[7] K. Nakano and B. Chakraborty, "Real-time distraction detection from driving data based personal driving model using deep learning," *Int. J. Intell. Transp. Syst. Res.*, vol. 20, no. 1, pp. 238–251, 2022.

[8] K. van der El, D. M. Pool, M. M. van Paassen, and M. Mulder, "Effects of target trajectory bandwidth on manual control behavior in pursuit and preview tracking," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 1, pp. 68–78, 2020.

[9] M. Mulder, D. M. Pool, D. A. Abbink, E. R. Boer, P. M. T. Zaal, F. M. Drop, K. van der El, and M. M. van Paassen, "Manual Control Cybernetics: State-of-the-Art and Current Trends," *IEEE Trans. on Human-Machine Systems*, vol. 48, no. 5, pp. 468–485, 2018.

[10] DLR Institute of Transportation Systems, "SuRT - Mobile." [Online]. Available: https://apkcombo.com/surt-mobile/de.lapoehn.surt/

[11] G. H. A. Verkerk, "Classifying Human Manual Control Behaviour in Tracking Tasks with Various Display Types Using the Inception Time CNN," Master's thesis, TU Delft, 2021.

[12] T. Petzoldt, H. Otto, and J. Krems, "The Critical Tracking Task: A Potentially Useful Method to Assess Driver Distraction?" *Human Factors*, vol. 56, 06 2014.

[13] R. Versteeg, D. M. Pool, and M. Mulder, "Classifying Human Manual Control Behaviour using LSTM Recurrent Neural Networks," *IEEE Trans. on Human-Machine Systems*, vol. 54, no. 1, pp. 89–99, 2024.

[14] Y. D. Li, "Detection of Distractions in Human Manual Control Tasks Using Machine Learning," Master's thesis, TU Delft, 2023.