Classification of Drivers with HIV-Associated Neurocognitive Disorders using Virtual Driving Test Performance Data

David Grethlein, Venk Kandadai
{david, venk}@diagnosticdriving.com
Diagnostic Driving, Inc.
Philadelphia, PA

Will Dampier
{wnd22}@drexel.edu
Drexel University
Philadelphia, PA

Abstract

In this work we focus on the problem of identifying drivers with *neurocognitive impairment* (NCI), specifically an NCI specific to *people with HIV* (PWH) called *HIV-associated neurocognitive disorders* (HAND) directly from driving simulator data. Since NCI-screening is typically only effective for more progressed forms of HAND, there is a critical need to identify individuals that should be referred to specialists in order to mitigate potentially dangerous driving behaviors and improve their quality of life. Data collected from $(n = 81)$ study participants that used the *virtual driving test* (VDT) platform were analyzed in order to predict which drivers had NCI. Of the $(n = 62)$ PWH participants recruited, $(n = 35)$ had HAND; of the remaining $(n = 19)$ HIV negative participants, $(n = 7)$ had non-HAND NCI (e.g., Parkinson's Disease, Alzheimer's, etc.). In three separate experiments, subsets of VDT data were first selected via Kruskal-Wallis feature ranking and then used as ensemble inputs to classify whether or not drivers had NCI. Within the PWH population, HAND could be classified with 69.4% accuracy and a risk ratio of 2.09 (95% CI 1.52, 2.65); within the HIV negative population, non-HAND NCI could be classified with 84.2% accuracy, risk ratio of 8.25 (6.34, 10.16); and within the combined population, NCI (regardless of causation) could be classified with 63.0% accuracy, risk ratio of 1.67 (1.22, 2.11).

Introduction

In our prior work (Grethlein et al. 2022), a subset of *virtual driving test* (VDT) performance data that was most closely linked to the presence of *HIV-associated neurocognitive disorders* (HAND), a category of NCI that is specific to *people with HIV* (PWH), in a population of $(n = 65)$ PWH study participants was isolated using feature selection. In this work, an additional $(n = 19)$ HIV negative participants were recruited, $(n = 7)$ of whom had non-HAND NCI (e.g., Parkinson's Disease, Alzheimer's, etc.), to act as *control group* for detecting other forms of NCI using VDT performance data. In this work we repeated the same feature selection with the expanded cohort and proceeded to classify the presence of NCI in participants using 5-fold ensembles.

Background

Failure to detect HAND early is associated with a diminished quality of life in PWH and decreased survival (Vance et al. 2014; Kronemer et al. 2017). However, such screens often require trained staff to administer tests, often lack ecological validity to assess impairments to activities of daily living, tend to be sparsely available to in-need communities, and financially prohibitive to low-income patients (Group et al. 2013), and have limited sensitivity when detecting milder forms of HAND (Roebuck-Spencer et al. 2017).

The VDT platform has been shown to be a rich source of explainable data for evaluating driver behavior in several real-world contexts (Grethlein et al. 2020; Walshe et al. 2022). Those using the VDT first undergo a 3 minute practice drive in order to acclimate to the simulated vehicle (e.g. steering wheel sensitivity, turn signal controls, etc.) and then take an assessment drive, lasting roughly 15 minutes, from which performance data is extracted. In this novel research effort, we classify the presence of NCI, particularly HAND in PWH, using VDT performance data.

Methods and Materials

All data for this work was collected under *Small Business Innovation Research* (SBIR) Grant No. R43MH122035 from the *National Institute of Mental Health* (NIMH).

Recruiting Study Participants

The initial group of $(n = 62)$ PWH participants were recruited via telephone interview from the Drexel University *Comprehensive NeuroAIDS Center* (CNAC) cohort (supported by NIMH P30MH092177) from November 2020 to May 2021. The remaining $(n = 19)$ HIV negative participants, $(n = 7)$ of whom had non-HAND NCI (e.g., Parkinson's, Alzheimer's, etc.), were recruited from friends and family of the CNAC cohort from July to September 2021 as the *control group* for this work. PWH and HIV negative participants had similar demographic data (e.g., age, race, years of independent driving). This was done in order to determine if the traits of drivers with HAND could be reliably detected and summarized, distinctly so from other forms of NCI. A more complete report of the recruitment and study visit procedures may be found in (Grethlein et al. 2022).

Clinically Confirming the Presence of NCI

All participants recruited, aged 20-75 (median 54), had previously completed the *comprehensive neuropsychological* *assessment* (CNPA) prior to recruitment and all had prior research study experience. The CNPA consists of a battery of tests administered by a trained psychometrist in a clinical setting, lasting approximately 2 hours long. The component tests of the CNPA evaluated 7 cognitive domains, described in full in (Grethlein et al. 2022), and were selected for their sensitivity to detecting HAND (Carey et al. 2004; de Almeida et al. 2017). The presence or absence of HAND in PWH was confirmed via the Frasceteri criteria (Antinori et al. 2007) using at least 2 psychometric measures per domain. Impairment to at least 3 cognitive domains confirmed the presence of NCI. Given the 7 cognitive domain-specific deficit scores produced from the component tests, the CNPA produced the *NCI status*, a final dichotomized value confirming the presence or absence of NCI, for each participant.

Virtual Driving Test Data

Each participant's self-guided interaction with the VDT workstation was recorded as a 10 Hz multivariate time series, comprising 40 channels, and was reduced to a vector of scalar features offline. In total, 2601 scalar features were extracted from each participant's VDT, as further described in (Grethlein et al. 2022), and used as inputs to the feature selection and classification analyses detailed later in this work.

Detecting NCI in Three Participant Populations

The first set of experiments we report on isolated the VDT performance data that best explained the differences between PWH driving with and without HAND. The second set of experiments were meant to isolate perceivable differences in driving in the control group: between HIV negative drivers with and without non-HAND NCI (e.g., Parkinson's, Alzheimer's). The third set of experiments combined the two populations intending to uncover any differences between driving with and without NCI (regardless of etiology). All VDT features analyses we report on were conducted using *python* (v 3.8.9) and the *scikit-learn* package (v 1.1.1).

Feature Selection

A subset of the 2601 VDT features was selected for each of the 3 populations using the 5-fold consensus-forming method described in (Grethlein et al. 2022).

Classification of VDT Data

The classification of NCI status using VDT data used the exact same stratified 5-fold splitting as the feature selection experiments, with random seed 0 used for generating reproducible results. Only the VDT features selected in the intermediate "top 100 features" ranked list for a split were used to train each individual classifier. The reserved 5^{th} fold of data in each split was later used to validate the split's classifier by predicting the presence of NCI on previously unseen data. We exhausted each population's data examined through cross-validation by validating 5 of the same type of classifier: i.e., *homogeneous ensembles*. Note that these 5 classifiers were each trained on a unique subset of 4 folds of training data, each using potentially different inputs obtained from the 5 intermediate "top 100 features" lists.

We chose to use out-of-the-box classifiers requiring minimal setup and computing resources as this was our first attempt classifying NCI status using the VDT data. We trained and tested the following types of classifiers in order to detect NCI from VDT performance data: *decision tree classifiers* (DTC), *random forest classifiers* (RFC), *multi-layer perceptrons* (MLP), *k-Nearest Neighbors* (KNN), *Logistic Regression* (LogReg), and *Support Vector Machines* (SVM).

We experimented building DTC and RFC ensembles using max-depth parameter values $\{1, 2, 3, 4, 5\}$. DTCs and RFCs (in different ensembles) were trained both with and without *class balancing*, setting the weights of classes to be inversely proportionate to their frequency, to account for lop-sided datasets. Similarly, LogReg and SVM ensembles were tested both with and without class balancing to mitigate class imbalance in our small datasets. SVM ensembles were built using both linear and *radial basis function* (RBF) kernels in order to leverage potentially non-linear relationships in VDT features towards the goal of NCI classification. KNN classifiers were tested using $k = \{1, 3, 5, 7\}$ in order to evaluate whether participants with the most similar VDT data to one another could be leveraged to predict NCI status.

Lastly, we tested *multi-layer perceptron* (MLP) homogeneous ensembles composed of 5 separately trained feed-forward neural networks (Hinton 1990). Since our datasets are all smaller than 100 VDTs, we employed dropout regularization (Srivastava et al. 2014). By imputing 0 for a percentage of neuron inputs during training, we aimed to avoid over-emphasizing any individual weighted connection between layers of neurons which could lead to over-fitting.

The MLP architecture we used in our experiments took the "top 100 features" ranked list for a split as inputs and is as follows. The input layer fed into 2 *fully-connected* (FC) layers of 128 neurons each, both using sigmoid activation. Dropout was imposed after both FC layers. Next, were 3 more hidden FC layers, all using sigmoid activation functions; composed of 256, 128, and 64 neurons, respectively. Finally, the output layer was a single neuron, using sigmoid activation. The entire MLP consisted of 103,681 parameters.

MLPs were trained for 100 epochs using the ADAM optimizer (Kingma and Ba 2014), binary cross-entropy as the loss function, a learning rate of 0.01, and a batch size of 16. We tested MLP homogeneous ensembles using $dropout = \{0\%, 10\%, 20\%, 30\%, 40\%, 50\%\}$. This was done in order to gauge how much regularization was needed for MLPs to distinguish NCI status in participants taking the VDT.

The input data to the DTC and RFC ensembles underwent no normalization nor transformation, as we sought to generate rule sets expressed and interpretable in the same units as the VDT data. The input data to the KNN, LogReg, and SVM ensembles were first unit-normalized (subtract by feature mean, divide by feature standard deviation), and then transformed via *principle component analysis* (PCA); using enough eigenvalues to preserve 90% of the total variance in the training data. Inputs to the MLP ensembles were only normalized, to avoid large fluctuations in the values of weighted connections during training. All validation VDT features data (reserved 5^{th} fold in each split) were similarly transformed (using same normalization and PCA as respective training data, if applicable) before being classified. In total we tested 36 unique homogeneous ensembles, each composed of 5 classifiers, for all 3 populations examined.

Evaluation Metrics

In order to quantify classifier success predicting NCI status, we chose to use *accuracy* (Acc), *risk ratio* (RR) (Sistrom and Garvan 2004) with *95% confidence interval* (95% CI), and *area under the curve* (AUC). Since all 5 classifiers in an ensemble were each validated on a unique 5^{th} fold of data reserved, we report on the evaluation metrics listed above for an entire ensemble.

Results

CNPA Results

All participants had their NCI status confirmed via CNPA during their study visit, prior to taking the VDT, in a clinical setting to ensure their CNPA results from prior visits were still accurate. The prevalence of NCI and cognitive domain impairments in all 3 study populations are listed in Table 1.

Feature Selection Results

None of the 28 VDT features selected in the consensus for detecting non-HAND NCI in the HIV negative population appeared in consensus for HAND in the PWH population (Grethlein et al. 2022). Participants with non-HAND NCI in the HIV negative population were most distinguishable at traffic light intersections where they were instructed to turn left across oncoming traffic, and where a lead vehicle stuttered forward when the light turned green before suddenly stopping and then turning right (see Table 2). Those with non-HAND NCI tended to exhibit more erratic changes in on-screen gaze position, harder braking, and weaker throttling than those without NCI in several on-road scenarios.

Of the 22 VDT features selected in the consensus for detecting NCI (regardless of underlying cause) in the combined population, 4 were shared (ranked 3^{rd}, 11^{th}, 18^{th}, 19^{th}) with the HAND consensus for the PWH population, and none were shared with the consensus for detecting non-HAND NCI in the control group. Participants with NCI in the combined population were best characterized by a higher accumulation of driving errors (e.g., running red lights, stop signs, etc.), and following lead vehicles more closely. The vehicle following behaviors of those driving with NCI was most notable in the crosswalk zone where traffic would stop suddenly for pedestrians crossing the road (see Table 3).

	Prevalence in Population: n (%)		
	PWH	Control	PWH + Control
# VDTs in Population	62 (100.0%)	19 (100.0%)	81 (100.0%)
Prevalence of NCI / Cognitive Domain Impaired	35 (56.45%)	7 (36.84%)	42 (51.85%)
Processing Speed	9 (14.52%)	5 (26.32%)	14 (17.28%)
Attention & Working Memory	15 (24.19%)	3 (15.79%)	18 (22.22%)
Motor Function	24 (38.71%)	6 (31.58%)	30 (37.04%)
Executive Function	9 (14.52%)	6 (31.58%)	15 (18.52%)
Language	7 (11.29%)	2 (10.53%)	9 (11.11%)
Verbal Memory	34 (54.84%)	3 (15.79%)	37 (45.68%)
Visuospatial Memory	24 (38.71%)	3 (15.79%)	27 (33.33%)

Table 1: Prevalence of NCI and cognitive domain impairments in all 3 study populations.

Classification Results

The evaluation metrics produced by the most accurate ensembles for classifying NCI status in all 3 populations may be found in Table 4. Overall MLP and SVM ensembles tended to yield higher RRs in all 3 populations examined.

The most accurate ensemble for classifying HAND status in the PWH population correctly predicted HAND in 26 of the 35 participants with the condition, only incorrectly flagging the presence of HAND in 10 of the 27 participants without confirmed cases of HAND. Similarly, the most accurate ensemble for the control group correctly predicted NCI in 6 of the 7 participants with confirmed cases; incorrectly flagging NCI in 2 of the 12 participants without NCI. Thirdly, the most accurate ensemble for classifying NCI in the combined population correctly predicted NCI in 26 of the 42 participants with NCI; incorrectly predicting the presence of NCI in 14 of the 39 participants without confirmed cases.

Discussion

The VDT features selected in the consensus for predicting HAND in the PWH population were most closely associated with the attention & working memory, motor function, and executive function cognitive domains (Grethlein et al. 2022).

In contrast, the features selected in the non-HAND consensus for the control group were most closely associated with the processing speed, executive function, and visuospatial memory cognitive domains. This was most clearly expressed in the crosswalk zone, where participants with impairment to these domains may have struggled to react with the sudden start-stop lurching of traffic as pedestrians crossed the road. Similarly, those with NCI displayed more rapid eye movement and jerkier braking when turning across incoming traffic and behind the lead vehicle that suddenly stops after the traffic light at that intersection turns green.

When the two populations were combined, the consensus features selected to predict NCI (regardless of etiology) were most associated with the attention & working memory, executive function, and verbal memory cognitive domains. Those with NCI showed signs of struggling when crossing oncoming traffic, maintaining safe distance with vehicles ahead at traffic-controlled intersections, and at the crosswalk. Also, those with NCI tended to stop at the inactive railroad crossing, possibly mistaking it for a roadway. Difficulties in multi-tasking (e.g., gauging distance to traffic ahead while reading signage) likely contributed to the higher accumulation of driver errors for those with NCI.

All experiments were limited by the binary NCI status (produced by the CNPA) that didn't account for differences in severity of NCI among participants; likely failing to capture the full nuance of how NCI presents in driving behavior. The two-stage feature selection (done to mitigate lop-sided datasets: 2601 features for fewer than 100 VDTs) performed in this work was limited by the lack of a filtering step to remove highly correlated features. As a result, some nearly identical VDT features were highly ranked together, which likely prohibited other informative features from appearing in the final consensus. We omitted such a filtering step to have parity with our previous work. Our analyses were also

Rank	Consensus VDT Performance Features Ranked via 5-Fold Median KW H-Value	Feature Units	All VDTs ($n = 19$)		NCI: Present ($n = 7$)		NCI: Absent ($n = 12$)		|%| Median Difference	KW H-Value	KW p-value
			Median	IQR	Median	IQR	Median	IQR			
1	Minimum forward jerk at lead-car sudden stop intersection	mph/sec^2	-71.08	8.65	-79.51	5.72	-68.88	10.16	13.38%	9.39	2.18E-03
2	Maximum forward acceleration at lead-car sudden stop intersection	mph/sec	9.62	1.39	10.73	0.53	9.31	1.96	13.21%	8.68	3.22E-03
3	Minimum vehicle speed at lead-car sudden stop intersection	mph	-0.08	5.25	-0.12	0.01	-0.03	20.12	77.36%	7.37	6.62E-03
4	Minimum difference of vehicle speed with posted speed limit at lead-car sudden stop intersection	mph	-20.08	1.87	-20.12	0.01	-20.03	14.98	0.46%	7.37	6.62E-03
5	Minimum ratio of vehicle speed with posted speed limit at lead-car sudden stop intersection	$ratio$	0.00	0.26	-0.01	0.00	0.00	0.79	77.36%	7.37	6.62E-03
6	Maximum brake pedal depression at second left turn across oncoming traffic intersection	$\%\ depressed$	0.37	0.92	0.95	0.11	0.03	0.36	96.42%	6.91	8.57E-03
7	Maximum forward jerk at lead-car sudden stop intersection	mph/sec^2	117.85	81.25	164.97	18.79	91.74	55.43	44.39%	6.72	9.52E-03
8	Minimum change in on-screen gaze position within the crosswalk	$pixels/sec$	29.86	32.72	4.68	19.45	39.89	10.73	88.27%	6.63	1.00E-02
9	Maximum distance to the right of road median at final right turn stop sign intersection	$meters$	6.30	2.06	4.83	1.73	6.63	0.86	27.18%	6.62	1.01E-02
10	Mean horizontal change in on-screen gaze position during box-truck following task	$pixels/sec$	364.98	122.55	523.36	208.65	304.68	136.95	41.78%	6.22	1.26E-02
11	Minimum change in on-screen gaze position in crosswalk zone	$pixels/sec$	3.87	16.67	0.00	0.00	15.95	18.64	100.00%	6.21	1.27E-02
12	Mean horizontal change in on-screen gaze position at lead-car sudden stop intersection	$pixels/sec$	620.88	160.20	728.33	139.31	550.17	87.81	24.46%	6.13	1.33E-02
13	Mean standard deviation in change in on-screen gaze position over a 1 second period at lead-car sudden stop intersection	$pixels/sec$	1066.51	288.34	1249.27	237.61	983.79	203.34	21.25%	6.13	1.33E-02
14	Standard deviation of brake pedal depression at at second left turn across oncoming traffic intersection	$\%\ depressed$	0.14	0.25	0.26	0.06	0.01	0.12	97.99%	6.04	1.40E-02
15	Mean throttle pedal depression at left turn merge onto main road intersection	$\%\ depressed$	0.17	0.13	0.10	0.08	0.24	0.13	59.60%	6.00	1.43E-02
16	Minimum forward acceleration at lead-car sudden stop intersection	mph/sec	-16.62	9.83	-22.30	3.02	-13.14	7.61	41.09%	6.00	1.43E-02
17	Amount of time spent actively braking at second left turn across incoming traffic intersection	$seconds$	1.42	4.37	4.94	8.20	0.06	1.64	98.74%	5.76	1.64E-02
18	Mean throttle pedal depression at first left turn across incoming traffic intersection	$\%\ depressed$	0.14	0.07	0.11	0.02	0.16	0.07	36.04%	5.69	1.70E-02
19	Mean throttle pedal depression at second left turn across incoming traffic intersection	$\%\ depressed$	0.22	0.17	0.17	0.11	0.29	0.11	39.09%	5.56	1.84E-02
20	Minimum change in on-screen gaze position at first right turn stop sign intersection	$pixels/sec$	0.00	9.81	0.00	0.00	9.23	11.01	100.00%	5.40	2.01E-02
21	Minimum horizontal change in on-screen gaze position at first right turn stop sign intersection	$pixels/sec$	0.00	1.52	0.00	0.00	1.38	1.76	100.00%	5.40	2.01E-02
22	Minimum vertical change in on-screen gaze position at first right turn stop sign intersection	$pixels/sec$	0.00	1.13	0.00	0.00	0.39	2.15	100.00%	5.40	2.01E-02
23	Number of stops made by vehicle at second left turn across incoming traffic intersection	$count$	0.00	1.00	1.00	0.50	0.00	0.25	100.00%	5.40	2.01E-02
24	Median difference of vehicle speed to posted speed limit at second left turn across incoming traffic intersection	mph	-21.83	12.32	-33.15	10.78	-19.72	6.58	40.50%	5.01	2.51E-02
25	Median forward jerk of vehicle in the crosswalk zone	mph/sec^2	-1.57	3.80	-3.38	12.56	0.00	2.44	100.00%	4.87	2.73E-02
26	Median forward acceleration of vehicle in the crosswalk zone	mph/sec	0.22	0.47	0.45	1.67	0.00	0.29	100.00%	4.87	2.73E-02
27	Mean vehicle speed in second left turn across incoming traffic intersection	mph	15.10	8.31	8.59	7.75	17.55	7.00	51.03%	4.86	2.75E-02
28	Mean ratio of vehicle speed to posted speed limit in second left turn across incoming traffic intersection	$ratio$	0.40	0.21	0.24	0.21	0.47	0.18	48.86%	4.86	2.75E-02

Table 2: Consensus of 28 VDT performance features selected and then ranked via 5-Fold median KW H-value that were most associated with non-HAND NCI (e.g., Parkinson's, Alzheimer's, etc.) status in the HIV negative population.

Rank	Consensus VDT Performance Features Ranked via 5-Fold Median KW H-Value	Feature Units	All VDTs ($n = 81$)		NCI: Present ($n = 42$)		NCI: Absent ($n = 39$)		|%| Median Difference	KW H-Value	KW p-value
			Median	IQR	Median	IQR	Median	IQR			
1	Median difference in vehicle heading with road following direction in construction zone	$degrees$	1.39	1.39	1.77	2.16	1.18	0.85	33.52%	7.47	6.26E-03
2	Amount of time spent coasting in inactive railroad crossing zone	$seconds$	1.67	3.41	0.88	2.1	2.78	3.6	68.23%	7.29	6.94E-03
3	Number of collisions with other vehicles to the side of or behind participant vehicle over whole assessment drive	$count$	0	0	0	1	0	0	UNDEF	7.27	7.02E-03
4	Minimum time to collision with other vehicles in 10 meters of road leading up to crosswalk	$seconds$	10.00	0.00	10.00	0.00	10.00	0.00	0.00%	7.01	8.09E-03
5	Mean time to collision with other vehicles in 10 meters of road leading up to crosswalk	$seconds$	10.00	0.00	10.00	0.00	10.00	0.00	0.00%	7.01	8.09E-03
6	Standard deviation of time to collision with other vehicles in 10 meters of road leading up to crosswalk	$seconds$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	7.01	8.09E-03
7	Minimum difference in vehicle heading with road following direction over the whole assessment drive	$degrees$	3.70E-04	5.77E-04	2.33E-04	4.22E-04	5.56E-04	9.24E-04	58.17%	6.48	1.09E-02
8	Minimum time to collision with lead vehicle in 10 meters of road leading up to crosswalk	$seconds$	10.00	0.00	10.00	0.00	10.00	0.00	0.00%	6.10	1.35E-02
9	Mean time to collision with lead vehicle in 10 meters of road leading up to crosswalk	$seconds$	10.00	0.00	10.00	0.00	10.00	0.00	0.00%	6.10	1.35E-02
10	Standard deviation of time to collision with lead vehicle in 10 meters of road leading up to crosswalk	$seconds$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	6.10	1.35E-02
11	Mean forward jerk of vehicle in inactive railroad crossing zone	mph/sec^2	-0.73	15.69	-4.36	11.76	3.81	11.95	187.44%	5.97	1.46E-02
12	Number of instances driving where time to collision with vehicles less than 5 seconds in crosswalk zone	$count$	0	0	0	0	0	0	UNDEF	5.93	1.48E-02
13	Number of instances driving where time to collision with vehicles less than 5 seconds in 10 meters of road leading up to crosswalk	$count$	0	0	0	0	0	0	UNDEF	5.93	1.48E-02
14	Amount of time spent driving where time to collision with vehicles less than 5 seconds in crosswalk zone	$seconds$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	5.92	1.50E-02
15	Distance driven where time to collision with vehicles less than 5 seconds in crosswalk zone	$miles$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	5.92	1.50E-02
16	Amount of time spent driving where time to collision with vehicles less than 5 seconds in 10 meters of road leading up to crosswalk	$seconds$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	5.92	1.50E-02
17	Distance driven where time to collision with vehicles less than 5 seconds in 10 meters of road leading up to crosswalk	$miles$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	5.92	1.50E-02
18	Minimum rotation of the steering wheel from resting position during the ambulance interaction	$\%\ rotation$	-0.03	0.02	-0.03	0.03	-0.02	0.02	23.79%	5.75	1.65E-02
19	VDT error score: linear combination of frequencies of high-level driving errors over whole assessment drive	$error\ score$	32.00	46.00	45.5	65.5	27	36.5	40.66%	5.35	2.08E-02
20	Number of instances driving where time to collision with lead vehicle less than 3 seconds at first traffic light intersection	$count$	0	0	0	0	0	0	UNDEF	5.17	2.30E-02
21	Amount of time spent driving where time to collision with lead vehicle less than 3 seconds at first traffic light intersection	$seconds$	0.00	0.00	0.00	0.00	0.00	0.00	UNDEF	5.16	2.31E-02
22	Median forward acceleration in 10 meters of road leading up to crosswalk	mph/sec	0.00	0.63	0.03	0.63	0.00	0.68	100.00%	4.97	2.58E-02

Table 3: Consensus of 22 VDT features most associated with NCI (HAND or non-HAND) status in the combined populations.

Population Examined	Homogeneous Ensemble	Acc	AUC	RR (%95 CI)
HAND in PWH	SVM w/RBF + class balancing	69.35%	0.32	2.09 (1.52, 2.65)
non-HAND NCI in Control	MLP w/20% dropout	84.21%	0.90	8.25 (6.34, 10.16)
NCI in PWH + Control	MLP w/0% dropout	62.96%	0.62	1.67 (1.22, 2.11)

Table 4: Evaluation metrics of most accurate ensembles.

limited by the small number of study participants (fewer than 100); likely causing over-fit in the MLP ensembles and casting doubt on how generalizable our results are.

Conclusions

Our current findings suggest that HAND can be detected in PWH with moderate reliability, risk ratio of 2.09 (1.52, 2.65), using VDT data by observing the accumulation of driving errors (e.g. running red lights and stop signs), MVCs, and vehicle control (e.g. lane position and driving off-road). More acutely, non-HAND NCI can be detected in HIV negative drivers with greater reliability, risk ratio of 8.25 (6.34, 10.16), using VDT data by tracking erratic on-screen gaze, poorer vehicle acceleration (e.g., weak throttling and overzealous braking) and below-average speed management (e.g., driving well below the posted speed limit). Then NCI in the combined population could be detected with less reliability; producing a risk ratio of 1.67 (1.22, 2.11) by keeping tabs on the accumulation of driving errors and the distance maintained with vehicles ahead.

In future work, we intend to recruit a larger study population to see if our results hold with more participants taking VDTs. We also plan on filtering out VDT features found to be highly correlated in the training data before the selection process. Next, we intend to build a consensus using *Shapley values* (Lundberg et al. 2020) to quantify the relative contribution of each feature towards accurate classifications. Additional research is also required to determine if the VDT can alleviate strain on the NCI-screening process as a whole (e.g., higher user acceptance, or more screens performed).

References

Antinori, A.; Arendt, G.; Becker, J.; Brew, B.; Byrd, D.; Cherner, M.; Clifford, D.; Cinque, P.; Epstein, L.; Goodkin, K.; et al. 2007. Updated research nosology for hiv-associated neurocognitive disorders. *Neurology* 69(18):1789–1799.

Carey, C. L.; Woods, S. P.; Gonzalez, R.; Conover, E.; Marcotte, T. D.; Grant, I.; and Heaton, R. K. 2004. Predictive validity of global deficit scores in detecting neuropsychological impairment in hiv infection. *Journal of clinical and experimental neuropsychology* 26(3):307–319.

de Almeida, S. M.; Kamat, R.; Cherner, M.; Umlauf, A.; Ribeiro, C. E.; de Pereira, A. P.; Franklin, D.; Heaton, R. K.; and Ellis, R. J. 2017. Improving detection of hiv-associated cognitive impairment: comparison of the international hiv dementia scale and a brief screening battery. *Journal of acquired immune deficiency syndromes (1999)* 74(3):332.

Grethlein, D.; Winston, F. K.; Walshe, E. A.; Tanner, S.; Kandadai, V.; and Ontañón, S. 2020. Simulator pre-screening of underprepared drivers prior to licensing on-road examination: Clustering of virtual driving test time series data. *Journal of Medical Internet Research* 22.

Grethlein, D.; Pirrone, V.; Devlin, K. N.; Dampier, W.; Szep, Z.; Winston, F. K.; Ontañón, S.; Walshe, E. A.; Malone, K.; Tillman, S.; Ances, B. M.; Kandadai, V.; Kolson, D. L.; and Wigdahl, B. 2022. Examining virtual driving test performance and its relationship to individuals with hiv-associated neurocognitive disorders. *Frontiers in Neuroscience, Translational Neuroscience* 16.

Group, M. E. W.; Antinori, A.; Arendt, G.; Grant, I.; Letendre, S.; Muñoz-Moreno, J. A.; Eggers, C.; Brew, B.; Brouillette, M.-J.; Bernal-Cano, F.; et al. 2013. Assessment, diagnosis, and treatment of hiv-associated neurocognitive disorder: a consensus report of the mind exchange program. *Clinical Infectious Diseases* 56(7):1004–1017.

Hinton, G. E. 1990. Connectionist learning procedures. In *Machine learning*. Elsevier. 555–610.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kronemer, S. I.; Mandel, J. A.; Sacktor, N. C.; and Marvel, C. L. 2017. Impairments of motor function while multitasking in hiv. *Frontiers in human neuroscience* 11:212.

Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; and Lee, S.-I. 2020. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence* 2(1):56–67.

Roebuck-Spencer, T. M.; Glen, T.; Puente, A. E.; Denney, R. L.; Ruff, R. M.; Hostetter, G.; and Bianchini, K. J. 2017. Cognitive screening tests versus comprehensive neuropsychological test batteries: a national academy of neuropsychology education paper. *Archives of Clinical Neuropsychology* 32(4):491–498.

Sistrom, C. L., and Garvan, C. W. 2004. Proportions, odds, and risk. *Radiology* 230(1):12–19.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15(1):1929–1958.

Vance, D. E.; Fazeli, P. L.; Ball, D. A.; Slater, L. Z.; and Ross, L. A. 2014. Cognitive functioning and driving simulator performance in middle-aged and older adults with hiv. *Journal of the Association of Nurses in AIDS Care* 25(2):e11–e26.

Walshe, E. A.; Elliott, M. R.; Romer, D.; Cheng, S.; Curry, A. E.; Seacrist, T.; Oppenheimer, N.; Wyner, A. J.; Grethlein, D.; Gonzalez, A. K.; and Winston, F. K. 2022. Novel use of a virtual driving assessment to classify driver skill at the time of licensure. *Transportation Research Part F: Traffic Psychology and Behaviour* 87.