Introducing Inclusive Construct Label-Centric Approach for Model Performance Enhancement in Autonomous Vehicles

Yashaswini Viswanath
RACE, Reva University
yashaswini.cse@gmail.com

Sudha Jamthe
Stanford University CSP
sujamthe@businessschoolofai.com

Suresh Lokiah
Independent Researcher
Sureshlokiah@gmail.com

***Abstract*—Driver Monitoring Systems (DMS) track movement of people using a camera inside the vehicle using AI to predict driver alertness to decide the safety of the driver and people on the road. Cameras collect huge amounts of data in all light conditions and activities of people inside the car. This data carries a wealth of insights about driver movement. Hence we propose a new label centric-approach by labeling the camera data with inclusive AI constructs for a more expansive annotation of the same dataset instead of the typical model-centric approach or data-centric approach to improve the performance of this AI. We used the DMD multi-model dataset for driver monitoring scenarios which comes with labeling movement to track 8 actions of the human texting with left or right hand, talking, drinking, phone call with left or right hand, reaching the side of the car or combing hair.**

We developed a binary classification CNN model for movement in the car. We tested the model trained against an inclusive AI constructed labeling option on the same dataset where we expanded the labeling to track movement of hair flying, scarf fluttering, hand waving or rubbing eyes. The results showed that the inclusive AI construct improved model performance without any change to the model algorithm tuning. Hence we recommend using a label-centric approach to improve labeling of data from camera streams such as the autonomous vehicle to be inclusive on an expanded construct for labels covering all people of all cultures in all lights, all hair, dress, and actions so AI model performance can be improved by capturing more knowledge by being more inclusive of all humans and their actions inside the vehicle.

Keywords—labeling, annotation, inclusive ai, construct, model performance, label-centric, autonomous vehicles, DMS, Driver Monitoring, Driver Attention, Driver Distraction

I. Introduction

Driver Management Systems (DMS) track movement of people using a camera inside the vehicle and AI to predict driver alertness from this data. Tesla tracks driver's alertness[1][1] in full self-driving mode in the vehicle to decide on safe driver rating for insurance premiums and to allow Full Self Driving capability access to Tesla drivers. Data has more knowledge than is captured by labels available in the training data and if we capture all the knowledge contained in the data, we can train the AI better. AI understands the data from the annotations from the labeling of the data. Labeling is typically outsourced to someone who does annotations without any knowledge of AI. Little thought is given to how bias can be introduced in models at the labeling stage by having a narrow definition of constructs of what the AI is modeled to predict. So our approach in this paper is a label centric-approach. All AI models in the vehicle have to be inclusive of all people to predict the movement and alertness of humans for the driver management system to recognize them as being alert to give control to save human lives inside the car and on the road.

This paper postulates that we can improve the AI model performance by improving labeling of movement with inclusive AI constructs for a more expansive annotation of the same dataset.

OEMs are using AI in the car to track driver alertness as an ADAS driver assistance feature to allow for the car to handover control to the human in case of the failure of the autonomous capability of the car. The success of this AI depends on its accuracy and AI model performance. With AI today, data scientists take one of two approaches to improve the model performance. They take a model centric approach and focus on improving the model with hyper parameter tuning or take a data centric approach and retrain the model with more training data.

A. *Modeling Approaches of Today*

Today ML model performance is improved using two approaches. One is to focus on improving model performance called a model-centric approach. Another is a data-centric approach to focus on the data to train the model with improved data that is reflective of the problem statement. Both of these approaches leave behind valuable knowledge in the data because of gaps in labeling that is not inclusive of all kinds of people and situations that represent an unbiased dataset. Hence we propose the label-centric approach here.

B. *Labeling Approaches and their limitations*

Today Labeling is done by a 3rd party company that does not involve the data scientist and most labels are created with elementary and simple constructs of actions. This leaves behind valuable information in the data that is not labeled and hence the value is lost as introduced by Inclusive AI Movement Construct by Susanna Raj in AI Ethics: Responsible AI And Inclusive AI Masterclass at Business School of AI. [2][2]

---

[1] Simon Alvarez, Tesla introduces Safety Score (Beta) system that incentivizes safe driving, www.teslarati.com, https://www.teslarati.com/tesla-autopilot-safety-scores-explained-fsd-beta/, (accessed on 28 Dec 2021)

[2] Business School of AI, (2021), AI Ethics: Responsible AI And Inclusive AI Masterclass by Susanna Raj [Online], https://businessschoolofai.teachable.com/admin/courses/1441962/curriculum/lectures/34275444

Hence we need a move exhaustive definition of constructs to define classes of labels to maximum value from the data. We call this as “Inclusive AI Construct” and have created a new Inclusive AV dataset for which we share the details below.

C. Introduction to Inclusive AI Construct in Labeling

Data is the new oil is a cliche that is in use today implying that data carries value to create AI to solve problems that were previously possible before AI. Raw data that carries information in it to create value is really not valuable if it cannot be used to train the AI. This is where we see the power of annotations to label all the knowledge in the data. Data that is labeled inclusively to gleam maximum insight from it is really valuable data. This Inclusiveness can be expanded by adding labeling constructs covering all races, genders, cultures, countries and lived experiences of all people who will engage with the AI.

II. EXPERIMENT SETUP

A. Dataset Used For This Research

We used the DMD Driving monitoring multi-modal dataset[3][3] for driver monitoring scenarios. This database comes videos of driver driving from real and simulated environments showcasing a variety of driver distraction actions.

The DMD multi-model dataset for driver monitoring scenarios which comes with labeling movement and we used 8 actions of the human texting with left or right hand, talking, drinking, phone call with left or right hand, reaching the side of the car or combing hair in this research.

We chose to use actions that define movement as a proxy of distraction and hence our paper is focused on tracking the labeling construct of movement and expanding it to be inclusive. We tracked 8 actions of the human texting with left or right hand, talking, drinking, phone call with left or right hand, reaching the side of the car or combing hair.

We got 16000 frames of one video from the DMD Driver Model Dataset. We labeled them as movement and no_movement based on 8 actions and expanded them to 6 more actions that define movement inside a vehicle. See Table 1 shows schematic of the list of movement actions and what was included in the base labeled dataset and what was in the expanded inclusive construct dataset.

Baseline Labeled Dataset: We used one video with 8 simulated driver actions from the DMS Dataset as our baseline labeled data. We generated 16000 frames from this video and used this as our baseline labeled dataset.

See below for sample image frames showing the 8 actions of drivers from the DMS dataset showing driver actions of drinking water, fidgeting phone, etc.

Figure 2: Driver Movement Action act_driver_actions/phonecall_ left from the DMS dataset

Figure 3: Driver movement Action act_driver_actions/drinking from the DMS dataset

Figure 3: Driver no movement action act_gaze_on_road/looking_road from the DMS dataset

Inclusive Construct Labeled Dataset:

We created a new dataset called Inclusive AV Dataset [4][4].

We made this similar to the DMD dataset to augment it with additional driver actions for movement inside a car or autonomous vehicle. This included 6 additional actions to

[3] Ortega, J., Kose, N., Cañas, P., Chao, M.a., Unnervik, A., Nieto, M., Otaegui, O., & Salgado, L. (2020). DMD: A Large-Scale Multi-Modal Driver Monitoring Dataset for Attention and Alertness Analysis. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (Accepted).

show a woman with a scarf flying, long hair waving in the wind, scratching face, clapping, talking in road rage and hand movement that shows up as a blur. This expands the DMS dataset to be more inclusive of more types of people, outfits and situations inside the vehicle. Our goal with this dataset is to showcase more inclusive constructs to define human movement inside the vehicle. This is not an exhaustive dataset and can be expanded to more diverse people and actions in the future.

Figure 4: Driver movement action scratching from the Inclusive AI dataset

Figure 5: Driver movement action hair_flowing from the Inclusive AI dataset to be inclusive of people in the car with hair flowing

Figure 6: Driver movement action scarf_flying distracting driver attention as a movement from the Inclusive AI dataset

The Two Datasets Used in the experiments:

We added the Inclusive AV Dataset to one of 16K frames from gA_1_s4 video file from DMS dataset to create the Inclusive Construct Labeled Dataset.

So now we have a baseline labeled dataset with 8 actions defining movement class and rest as no_movement class and an Inclusive Construct Labeled Dataset with additional 6 actions that defines movement class and rest as no_movement class. We used this to compare the two datasets with the same model to test our hypothesis that label_centric approach can improve model performance.

The Actions Defining Movement class in the experiments:

	Movement Dataset Schema		
In baseline label dataset	***In inclusive construct labeled dataset***	***Action defining movement***	***Labeled Dataset***
n		act_gaze_on_road/looking_road	gA_1_s4_phase1
n		act_driver_actions/safe_drive	gA_1_s4_phase1
n		act_gaze_on_road/not_looking_road	gA_1_s4_phase1
n		act_hands_using_wheel/both	gA_1_s4_phase1
n		act_hands_using_wheel/only_left	gA_1_s4_phase1
n		act_driver_actions/change_gear	gA_1_s4_phase1
n		act_hand_on_gear/hand_on_gear	gA_1_s4_phase1
y	y	act_driver_actions/reach_side	gA_1_s4_phase1
y	y	act_driver_actions/hair_and_makeup	gA_1_s4_phase1
y	y	act_driver_actions/phonecall_right	gA_1_s4_phase1
y	y	act_driver_actions/phonecall_ left	gA_1_s4_phase1
y	y	act_talking/talking	gA_1_s4_phase1
y	y	act_driver_actions/texting_right	gA_1_s4_phase1
y	y	act_driver_actions/phone	gA_1_s4_phase1
y	y	act_driver_actions/drinking	gA_1_s4_phase1
n		act_driver_actions/unclassified	gA_1_s4_phase1
n		act_hands_using_wheel/none	gA_1_s4_phase1
y	y	act_driver_actions/texting_left	gA_1_s4_phase1
n		act_hands_using_wheel/only_right	gA_1_s4_phase1
n	y	hand blur	AV_file_phase2
n	y	hair_flowing	AV_file_phase2
n	y	scarf_flying	AV_file_phase2
n	y	Scratching	AV_file_phase2
n	y	Clapping	AV_file_phase2
n	y	self talking road rage	AV_file_phase2

Table 1: Schema of actions defining movement class in the dataset used in the experiments.

B.Model Built for testing hypothesis

We tested the initial round of experiments using a pre-trained image classification model from teachable.withgoogle.com to train a model using two sets of labeled datasets and used 15 epochs, learning rate 0.01, batch size of 16. We used this as the baseline with the 8 actions and the other as an inclusive construct dataset with 14 actions defining movement inside a vehicle. Table 2 shows the initial results of the model developed using teachable.withgoogle.com.

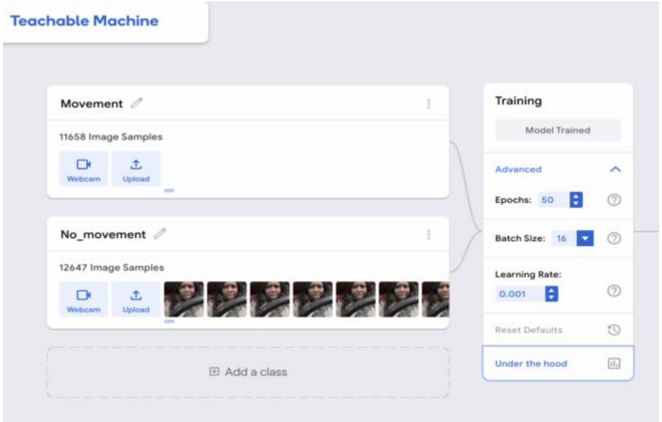


Table 2: Teachablemachine.withgoogle.com image classification classification

We decided to build our own CNN to validate the label-centric approach to model improvement for this research paper.

Convolution Neural Network based Image Classification Model

We developed image classification models using Convolution Neural Networks to use as the model that we will use in our experiments.

Model description:

We created a Convolution Neural Network model with three stages of Convolution Layers each with a max pooling layer. In order to avoid overfitting, the last max pool operation is appended with a dropout layer.This enabled us to create a model that learnt what is movement and what is no movement with trainable params of 553,314. In order to train the model we used DMD dataset and also augmented the dataset with frames captured in a custom setup. The resultant classification model was used for prediction on the test set of images.

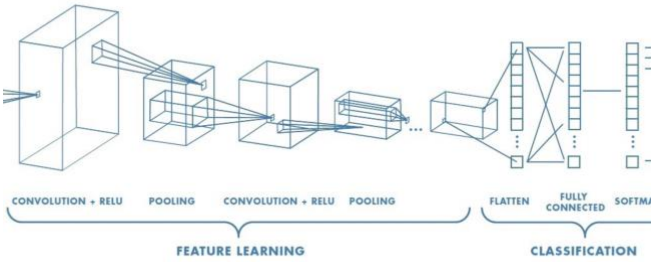


Image Credit: TowardsDataScience.com

C. Data Labeling Toolset

DMS Dataset comes in a VCD format with the annotations explained. It comes in JSON and is provided in VCD structure. It lists the objects and actions annotated by the labeler in each of the DMS Dataset videos. We needed the VCD structure to be flattened to csv formation to be loaded in Pandas dataframe. We needed to do this to summarize the driver actions from the annotation to create the movement and no_movement classes that we needed for the image classification. This helped with feature engineering to train our model. So we. built an open source tool for this research called as VCD Feature Enhancer which we have added to github as an open source contribution.[5][5]

III. EXPERIMENTS

Our hypothesis that we experimented to prove was that a label_centric model will improve model performance. We decided to create a similar CNN model and the same dataset as fixed and ran experiments with the baseline labeled dataset of 8 actions defining movement and compared against the combined dataset of baseline labeled dataset and Inclusive Labeled Dataset of a total of 14 actions defining movement.

To do this, we developed the image classification model using ResNet and tested this using the Baseline Labeled Dataset video file gA_1_s4 which gave us 16K frames from the DMA Dataset. This dataset has 8 actions labeled and we tracked those as movement class and everything else as no_movement class.

A. Experiment Phase 0

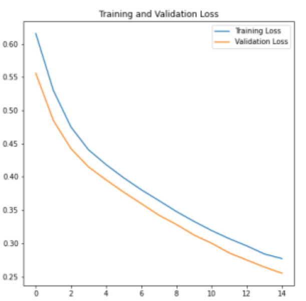
We did the first Phase 0 experiment to create a baseline model and test it out. To do this, we developed the image classification model using ResNet and tested this using the Baseline Labeled Dataset video file gA_1_s4 which gave us 16K frames from the DMA Dataset. This dataset has 8 actions labeled and we tracked those as movement class and everything else as no_movement class.

[5] VCD Flattener. (2022), Business School of AI, Accessed: Dec. 13, 2022, (Online), https://github.com/bsairesearch/DMD-Driver-Monitoring-Dataset/commits/master/vcd_flatten3.ipynb

Class	Model Performance		
	precision	***recall***	***f1-score***
Movement	0.79	0.92	0.85
No Movemnt	0.84	0.61	0.7

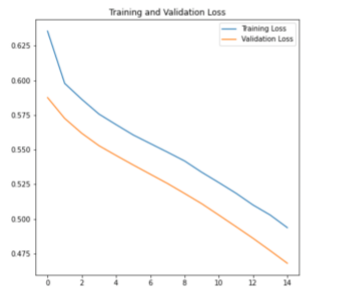
B. *Experiment Phase 1*

We did the first Phase 1 experiment to expand on experiment Phase 0 by repeating the Phase 0 experiment with the baseline labeled dataset and Inclusive AV Dataset but we did a baseline labeling of only the 8 actions defining movement. Table 2 for the schema of actions defining movement class in the dataset used in the experiments. We got the AI confusion matrix to measure model performance to use as baseline. We will use this to compare to the Experiment Phase 3 results to test our label_centric approach hypothesis of model improvement next.



C. *Experiment Phase 2*

We labeled the Inclusive AV Dataset to show the additional 6 actions to label a total of 14 actions defining movement. Then, we did the first Phase 2 experiment to expand on experiment Phase1 by repeating the Phase 1 experiment with the baseline labeled dataset and full labeled Inclusive AV Dataset with labeling of only all 14 actions defining movement. Table 1 shows the schema of actions defining movement class in the dataset used in the experiments.



Class	Model Performance		
	precision	***recall***	***f1-score***
Movement	0.84	0.94	0.89
No Movemnt	0.82	0.59	0.69

D. *Experiment Results*

We compared the accuracy of the CNN models trained from the baseline labeled dataset and the inclusive AI construct dataset. We found that the accuracy and precision improved through the inclusive labeling independent of model performance from improving the model by a data-centric approach, thus proving that the label_centric approach increased model performance.

As you can see, our goal here was not to improve the overall model performance. To that effect, we did not focus our experiments to improve model performance with a model_centric approach or data_centric approach. We focused only on a label_centric approach and saw improvement in the model performance when trained with a benchmark test dataset. So this leaves room to add incremental model performance by continuing to improve the model performance with model_centric and data_centric approaches additional to the label_centric approach demonstrated by our experiments.

IV. Call for Collaboration

Data is the new oil is a cliche that is in use today implying that data carries value to create AI to solve problems that were previously possible before AI. Raw data that carries information in it to create value is really not valuable if it cannot be used to train the AI. This is where we see the power of annotations to label all the knowledge in the data. Data that is labeled inclusively to gleam maximum insight from it is really valuable data.

In the research for this paper we propose a label-centric approach to model performance improvement. We focused on showing incremental performance improvement by improving labeling with better inclusive constructs. We would like to call for collaborators and future research to build upon our research because much work awaits us to benefit from this research. Future research can show whether model performance will improve if we augment label-centric performance improvements with existing model-centric and date-centric approaches or whether these improvements are independent of each other.

We chose to focus on the camera data watching humans inside the car. Future research can find out if there are certain types of data or environments that support a data-centric approach to create model performance improvements over other approaches.

V. Conclusion

We are introducing a label_centric approach to model performance improvement. We found that the data inside cars and autonomous vehicles captures humans in several actions while only a limited set of actions are labeled and tracked to check for driver attention in Driver Management systems in cars today. This is not inclusive of all races, people, cultures, countries, outfits and movements that is reflective of all people who will be exposed to Driver Attention Systems flagged by the AI in the vehicle. We tested the model performance improvement by comparing the performance of a base model with a set of 8 actions defining movement inside a vehicle against 14 actions defining movement in the same camera data inside the vehicle. We found that we can improve the AI model performance by improving labeling of movement with inclusive AI constructs for a more expansive annotation of the same dataset.

This opens up a new label_centric model improvement approach to AI model performance improvement not limited to autonomous vehicles; the camera data inside the car calls for a more inclusive labeling covering many different people, situations, outfits and activities. We call for collaboration from future researchers to build upon our work to expand the label_centric approach to test for how it augments model_centric and data_centric approaches and in what kind of data situations it lends itself for optimal model performance.

We are very hopeful that with an inclusive construct labeling of data and improving models with a label_centric approach we will be able to uncover knowledge in data across the globe covering diverse set of people of all races, genders, cultures and life situations to make AI solve an even broader set of problems for all mankind.

References

[1] Simon Alvarez, Tesla introduces Safety Score (Beta) system that incentivizes safe driving, www.teslarati.com, https://www.teslarati.com/tesla-autopilot-safety-scores-explained-fsd-beta/, (accessed on 28 Dec 2021)

[2] Business School of AI, (2021),

AI Ethics: Responsible AI And Inclusive AI Masterclass [Online], https://businessschoolofai.teachable.com/admin/courses/1441962/curriculum/lectures/34275444

[3] Ortega, J., Kose, N., Cañas, P., Chao, M.a., Unnervik, A., Nieto, M., Otaegui, O., & Salgado, L. (2020). DMD: A Large-Scale Multi-Modal Driver Monitoring Dataset for Attention and Alertness Analysis. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (Accepted).

[4] Inclusive AV Dataset, Business School of AI, 2022, [Online], Available: ttps://github.com/bsairesearch/inclusive-av-dataset

[5] VCD Flattener. (2022), Business School of AI, Accessed: Dec. 13, 2022, (Online), https://github.com/bsairesearch/DMD-Driver-Monitoring-Dataset/commits/master/vcd_flatten3.ipynb

[6] Teachable Machine, Google, Accessed: Dec. 28, 2022. (Online), https://teachablemachine.withgoogle.com/