Real-Time Driver Behaviour Characterization through Rule-based Machine Learning

Fabio Martinelli*, Francesco Mercaldo*,
Vittoria Nardone+, Antonella Santone§, Gigliola Vaglini^

*Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche (CNR), Pisa, Italy
+Department of Engineering, University of Sannio, Benevento, Italy
§Department of Bioscience and Territory, University of Molise, Pesche (IS), Italy
^Department of Information Engineering, University of Pisa, Pisa, Italy
{fabio.martinelli, francesco.mercaldo}@iit.cnr.it, vnardone@unisannio.it, antonella.santone@unimol.it, gigliola.vaglini@unipi.it

Consiglio Nazionale delle Ricerche

STRIVE@SAFECOMP - 18 September, Västerås, Sweden

Outline

1. **Motivation and Background**
2. **The Real-Time Detection Method**
3. **Empirical Evaluation**
4. **Concluding Remarks**

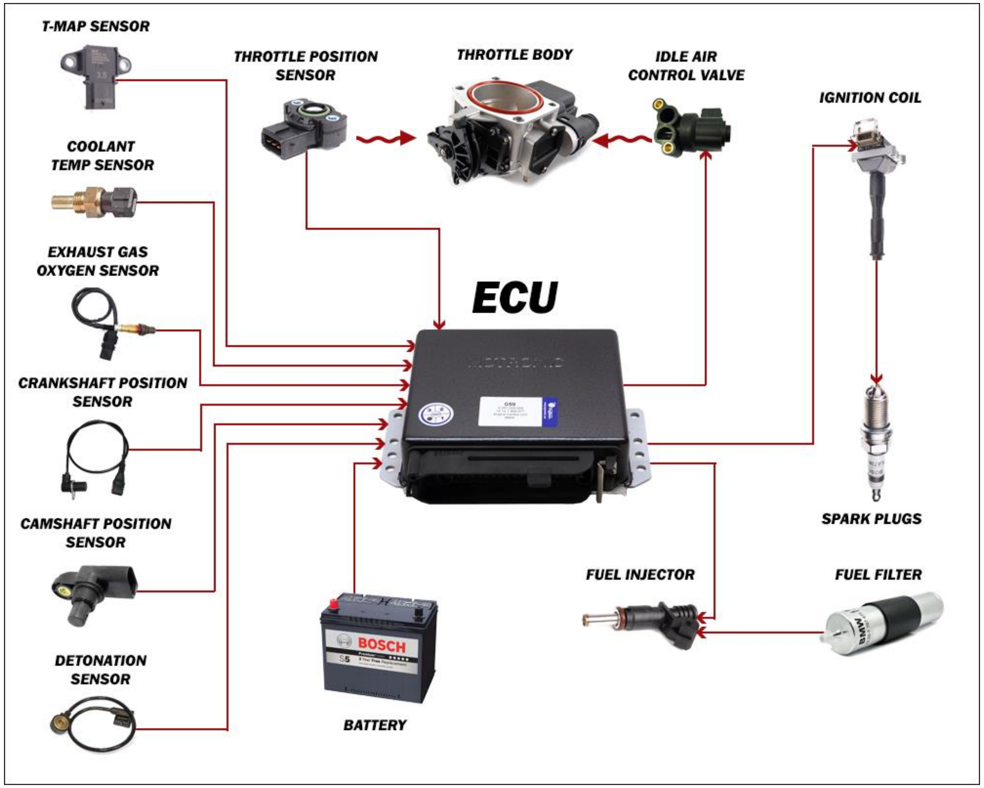
Motivation

- Modern vehicles have lots of connectivity
- Anti-theft
- New Insurance paradigms
  - Pay how you drive
- The Controller Area Network (CAN bus) is a robust vehicle bus standard designed to allow devices to communicate with each other in applications without a host computer.

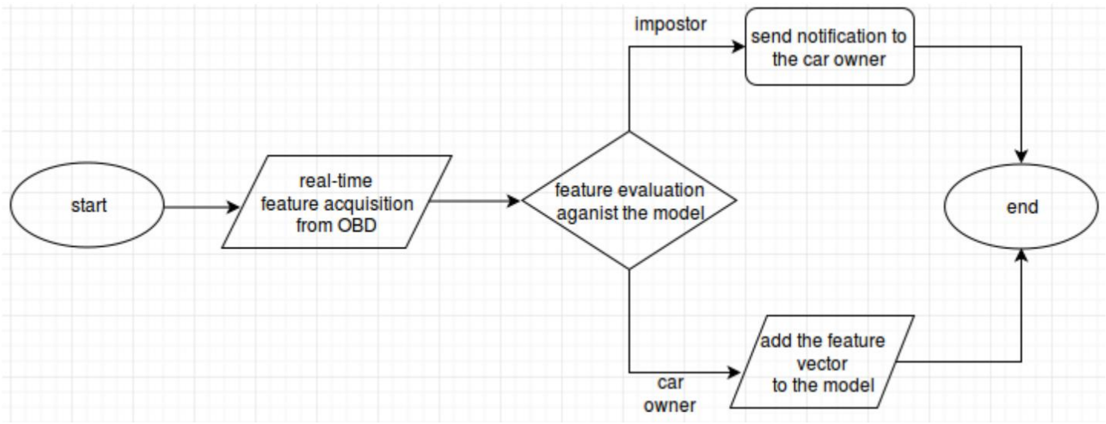
Research Question:

RQ: To what extent OBD II features allow driver detection?

The CAN standard



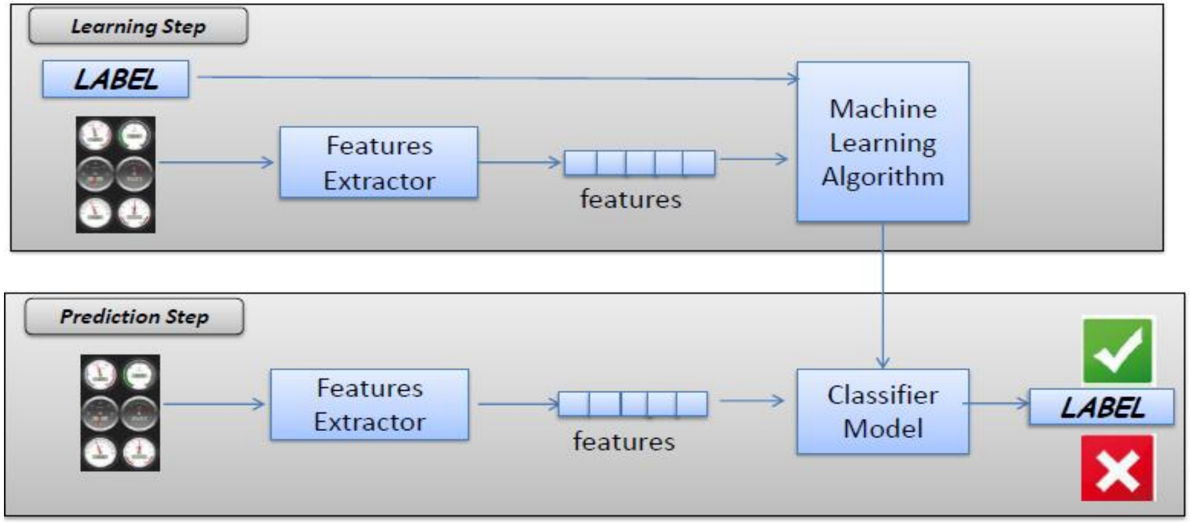
The Real-Time Detection Method



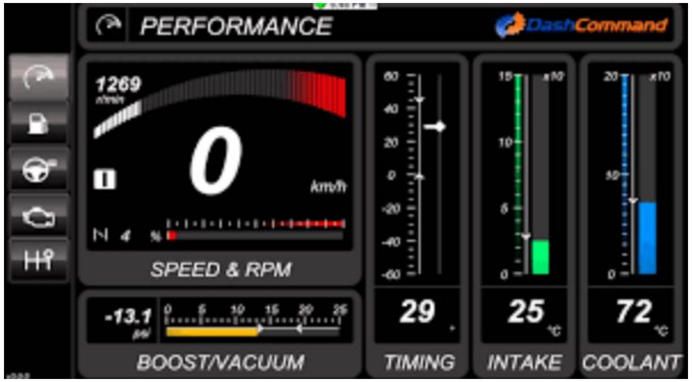
The feature vector

Feature	Description	Info
F1	CO_2 in g/km	(Instantaneous)(g/km)
F2	Engine Load(%)	expressed in g/s
F3	Engine RPM	Revolutions per minute
F4	Fuel flow rate/minute	expressed in cc/min
F5	Fuel Remaining	expressed in liters
F6	Turbo Boost & Vacuum Gauge	expressed in psi

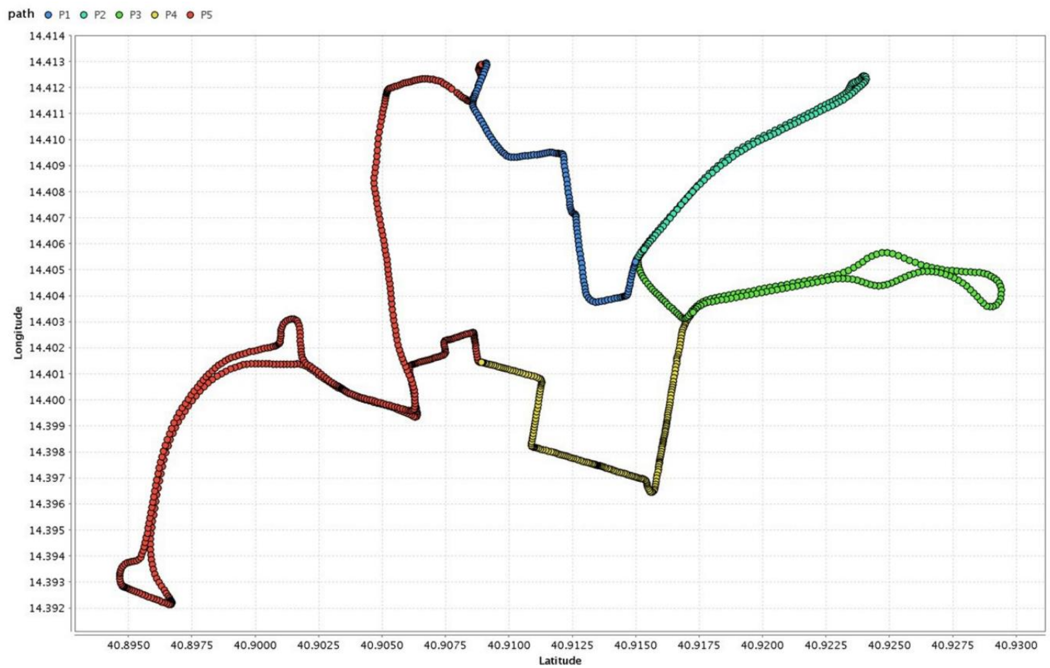
The Rule-based Machine Learning Model



Feature Gathering



Dataset overview



Consiglio Nazionale delle Ricerche

Istituto di Informatica e Telematica del CNR

Empirical Rules

#	Rule
D1	$0.0 <= F1 <= 1160.24169922 \wedge 10.98039246 <= F2 <= 92.94117737 \wedge$ $434.0 <= F3 <= 3799.75 \wedge 0.0 <= F4 <= 171.47810364 \wedge$ $49.42653656 <= F5 <= 49.79877853 \wedge -12.37939644 <= F6 <= -0.19622612$
D2	$0.0 <= F1 <= 1164.5357666 \wedge 10.58823586 <= F2 <= 90.98039246 \wedge$ $468.75 <= F3 <= 3623.0 \wedge 0.0 <= F4 <= 186.95175171 \wedge$ $48.64160156 <= F5 <= 49.00281143 \wedge -12.37939644 <= F6 <= -0.19622612$
D3	$0.0 <= F1 <= 1254.05834961 \wedge 10.98039246 <= F2 <= 98.03921509 \wedge$ $617.25 <= F3 <= 3710.0 \wedge 9.45204163 <= F4 <= 207.32177734 \wedge$ $48.23397446 <= F5 <= 48.59840775 \wedge -12.23435783 <= F6 <= -0.19622612$
D4	$0.0 <= F1 <= 1088.32873535 \wedge 10.58823586 <= F2 <= 95.29412079 \wedge$ $602.75 <= F3 <= 3205.75 \wedge 7.70137358 <= F4 <= 144.01383972 \wedge$ $47.80204773 <= F5 <= 48.14254379 \wedge -12.37939644 <= F6 <= -0.19622612$
D5	$0.0 <= F1 <= 1183.73730469 \wedge 13.72549057 <= F2 <= 81.5686264 \wedge$ $550.75 <= F3 <= 3521.5 \wedge 3.36647606 <= F4 <= 150.65760803 \wedge$ $45.38215256 <= F5 <= 45.7795639 \wedge -11.79924488 <= F6 <= -1.50156593$
D6	$0.0 <= F1 <= 1315.60339355 \wedge 10.58823586 <= F2 <= 92.15686798 \wedge$ $586.75 <= F3 <= 3807.5 \wedge 0.22675589 <= F4 <= 211.76445007 \wedge$ $47.37043381 <= F5 <= 47.75802994 \wedge -12.37939644 <= F6 <= -0.34126377$
D7	$0.0 <= F1 <= 1073.23742676 \wedge 10.98039246 <= F2 <= 100.0 \wedge$ $534.0 <= F3 <= 4060.0 \wedge 1.06496859 <= F4 <= 135.25418091 \wedge$ $46.96563721 <= F5 <= 47.32172775 \wedge -12.52443409 <= F6 <= -0.34126377$
D8	$0.0 <= F1 <= 941.16192627 \wedge 0.0 <= F2 <= 94.11764526 \wedge$ $0.0 <= F3 <= 3415.25 \wedge 7.36353827 <= F4 <= 113.75313568 \wedge$ $42.30183029 <= F5 <= 42.72509003 \wedge -11.74805641 <= F6 <= 0.0$
D9	$0.0 <= F1 <= 1119.75854492 \wedge 12.94117641 <= F2 <= 93.33333588 \wedge$ $715.25 <= F3 <= 3551.5 \wedge 12.2160759 <= F4 <= 157.69128418 \wedge$ $45.93281174 <= F5 <= 46.31466293 \wedge -11.94428253 <= F6 <= -0.34126377$
D10	$0.0 <= F1 <= 1091.08703613 \wedge 13.33333397 <= F2 <= 90.98039246 \wedge$ $472.0 <= F3 <= 3181.25 \wedge 7.24365759 <= F4 <= 163.60809326 \wedge$ $46.58006668 <= F5 <= 46.93938065 \wedge -12.08932018 <= F6 <= -0.77637672$

Experimental Results

Driver	FP Rate	Precision	Recall	F-Measure	RocArea
D1	0	1	1	1	1
D2	0	0.999	1	0.999	1
D3	0	1	1	1	1
D4	0	1	1	1	1
D5	0	1	1	1	1
D6	0	1	1	1	1
D7	0	1	1	1	1
D8	0	1	1	1	1
D9	0	1	0.999	1	1
D10	0	1	1	1	1

$$PR = \frac{TP}{TP+FP}; \; RC = \frac{TP}{TP+FN};$$
$$Fm = \frac{2PR\,RC}{PR+RC}; \; Acc = \frac{TP+TN}{TP+FN+FP+TN}$$

a	b	c	d	e	f	g	h	i	j	classified as
2398	0	0	0	0	0	0	0	0	0	a = D1
0	2922	0	1	0	0	0	0	0	0	b = D2
0	0	3073	0	0	0	0	0	0	0	c = D3
0	0	0	3106	0	0	0	0	0	0	d = D4
0	0	0	0	3846	0	0	0	0	0	e = D5
0	0	0	0	0	2794	0	0	0	0	f = D6
0	0	1	0	0	0	3009	0	0	0	g = D7
0	0	0	0	0	0	0	4151	0	0	h = D8
0	1	0	0	0	0	0	2	3267	0	i = D9
0	1	0	0	0	0	0	0	0	2833	j = D10

Conclusion and Future Work

- We propose a method aimed to discriminate between the car owner and impostors using a rule-based machine learning algorithm.
- We evaluated the proposed method on a dataset of 10 drivers.
- High performances in terms of precision and recall are achieved.
- We plan to investigate whether formal verification techniques are useful to improve the results.

Thanks for your attention

- We are grateful for receiving comments, observations, suggestions, and collaborations with other research groups which could improve our research.

francesco.mercaldo@iit.cnr.it