Name of the student: Léa SENGHOR	Name of your Level 1 Selvina GOVINDEN
Source: IEEE	Paper title Machine Learning and End-to-End Deep Learning for Monitoring Driver Distractions From Physiological and Visual Signals

Keywords speci1c to the paper: Machine learning, deep learning, driver distraction, sensors, facial expressions.

I. Introduction

A. Aim of the study

Observation:

➔ Every year, 25,000 people die in car accidents in the EU, mainly as a result of human error.

➔ The more autonomous driving becomes, the more distracted the driver tends to be, as he or she may be doing other things at the same time:

⧫ *Dialing a phone number, sending an SMS, reading, writing and verif1cation of an itinerary on a map ⇒ Considered as* ***DISTRACTION***

➔ Detecting distracted driving= **A very important topic for future vehicles**

To address this issue, we have implemented deep learning and machine learning tools to detect these distractions.

Affective computing: Systems and devices capable of recognizing, processing, analyzing and modeling human affective states.

- Interesting for our subject, as it enriches and facilitates human-machine interaction)
- Example of application: in new vehicles, AI uses acceleration sensors to collect data for navigation and to detect anomalies directly linked to driver behavior, such as fatigue).
- Limit: Not yet able to detect **mental state:** *stress, mental health, cognitive load, distractions,* etc.

B. The importance of monitoring driver distractions

According to Regan et al (2011), distraction is a subcategory of driver inattention:

Driver inattention: (driver distraction): Detour of attention from activities essential to safe driving to a competing activity, which may result in inadequate or non-existent attention to activities essential to safe driving.

- There are therefore **4 types of distraction**: *cognitive distraction, emotional distraction, sensorimotor distraction and mixed distraction.*

Given the growing importance of road safety, this article is designed to meet a number of objectives:

- Identify the best detection characteristics and modalities for the four types of distractions:
cognitive, emotional, sensorimotor and mixed.
- Compare DL architectures and ML algorithms in driver distraction monitoring. To do this, it focuses on physiological signals (e.g. *palm electrodermal activity, heart rate and respiratory rate*) and visual signals (e.g. *eye tracking, pupil diameter, nasal EDA, emotional activation, etc.*). ● Evaluate the effectiveness of different methods of detecting driver distractions.
- Collect data to improve road safety through monitoring and the development of e cient means to combat driver distraction

II. The contributions of this article

Today, many articles focus on how arti1cial intelligence through topics such as: background learning, algorithms, long term short term memory (LSTM) can help in the detection of driver inattention. The most advanced methods for monitoring driving disturbances are based on ML algorithms.
However: First study focusing on driver distraction detection and analyzing end-to-end learning on signals using 1D convolutions and long-term memory neural networks (LSTM). Do an experiment comparing seven classical machine learning (ML) methods and seven end-to-end deep learning (DL) methods.

III. Methodology

A. Description of data collection

1. Types of signals collected

- **Physiological signals:** Measured from living beings, and more particularly from human beings, these signals group together all the signals. The article talks for example about :
- Heart rate: (HR)
- Respiratory rate (BR)
- Palm electrodermal activity: (EDA)

- **Visual signals:**
  - Eye tracking
  - Pupil diameter
  - Facial action units

2. Data collection procedure

- **Previous statistical study:** The data set was collected during a previous driving simulation study.
- More specifically, **multimodal data** the people studied were equipped with driver-worn sensors or car-integrated sensors, such as video cameras.
- Data type:
  - For the seven end-to-end DL architectures: the input images and are based on available DL architectures that have been successfully applied to images (e.g. AlexNet, VGG-1G and ResNet-152),
  - For classic ML models, features were extracted from segmented data.

B. Participants

1. Number and features

Sample taken: Study by Pavlidis et al [7] → analyzed the driving behaviors of G8 volunteers in a driving simulator in the presence of various distractions. Distracted (cognitive, emotional, sensory-motor and mixed) and undistracted sessions.

C. Analysis techniques

1. Model training and validation

Data/ Method	Classical ML High-performance used	DL High-performance used	Better performance to detect distractions
Physiological signals (e.g. nEDA, BR, HR)	XGB, GB	eLSTM, STRNet	XGB achieved the highest score in F1 with 94%.
Visual signals (UA Facial, EMO)	XGB, GB	STRNet	STRN and achieved the highest score with 75%.

Interpretations:

- Among the seven ML models and the seven DL models, the best-performing methods for recognizing and identi1ying distractions from physiological and visual signals are:
  - The extreme gradient boosting classi1cator (XGB) for ML
  - The ResNet spectro-temporal model (STRNet) for DL.

Data from driving simulators where drivers have completed driving sessions are classified into two groups:

1. The "windows" last between 20 and 80 seconds and focus on a specific feature with or without distraction.

2. Sessions including a complete driving session

- The extreme gradient boosting classi1cator (XGB) used in ML shows better performance for driving sessions on the so-called full simulator adapted to physiological signals.
- However, STRNet used in DL shows the best results when applied to specific visual signals *(e.g. AU12 that represents the action of raising the lip corners, as in a smile).*
- Other ML architectures used: Random Forest, Naïve Bayes, Decision tree for ML // CNN, LSTM, FC NN for DL
- Certain characteristics, such as label instability, can bias the interpretation of results and explain why the results of physiological signals (longer latency) are less good than those of visual signals for the detection of distracted driving. **Conclusion:**

- Overall, the XGB model shows the best results of the study, but the STRNet model also performs very well
- Compared to DL models, ML models show better performance in detecting driver distractions at the wheel

DL models provide better classification1 of visual signals. UA (e.g.: raising the lip corners, as in a smile, raising the eyebrow, opening the mouth,) give much more detail on reactions to distractions.