- ➔ ML is learning from observations/data, there are different types of ML, some are: Supervised machine learning, unsupervised machine learning and reinforcement learning.
- ➔ ML will help systems cope with human limitations/factors, for instance, one can make a system that perceives if a driver is falling asleep and make it, so the driver is awaken and the car is stopped.

There is a learning loop in AI – HF, as machines propose decisions to humans. Machines learn to make better decisions based on data made by humans.

To make a system that detects fatigue in a human:

(1) one must collect, combine, clean data
(2) train a model using ML tools
(3) evaluate the model

To detect if a human is tired, we teach a model to look for **indicators**, these indicators are the **conceptualization** of the **concept**. The **concept** in this example is the fact of being tired.

To make these models, we must collect data for the training and testing steps. This is called **operationalization**, the way we are collecting data on the indicators.

To have accurate data, the tests must ask the right questions, and all participants must be in the same state when answering a question.

Variables in acquisition protocol, there are **manipulated variables** (independent), this usually is the target you want to predict. There are **measured variables** (dependent) this is the data you measure (example: eye-movements and blinking per minute).

When designing models for the perception of human characteristics, one has within and between participants designs. Within designs are models that are modelled to fit a specific human. Where between participant models can be used for every human with the same accuracy.

Confounding variables are factors that might influence the measured variables, for instance one might blink more times per minute due to factors that are not fatigue. These variables will add noise to your measurements. These can also be present in training data.

When collecting data there are many challenges, multimodal learning, noisy targets, imbalanced classes etc..

Unbalanced classes are when one of the classes is overrepresented in the collected data, for instance you have 10 times more data of the heart rate at rest, versus when a person is under stress → this is a problem because when training your model, your system will **overcorrect** to fit the data, this is due to the way the cost function is computed. This overcorrection results in a shift of a classifier threshold. (image 1) This can be a real problem when there is overlapping data points (image 2)

To fix this problem, one can:

(1) under-sample the majority class, this results in fewer data points
(2) over sample the minority class by creating, fake data, this fake data can be a copy of real data, or averages of multiple real values
(3) apply a mix of solution (1) and solution (2)

- ➔ We also need to measure the performance of a ML classifier, a random classifier should have an accuracy of 50%.
- ➔ We cannot simply use accuracy to determine the effectiveness of our model, we must look deeper
- ➔ The first task is to compute the confusion matrix

The confusion matrix looks like this:

	Estimated 0	Estimated 1	Sum
Ground Truth 0	4	1	5
Ground Truth 1	2	3	5
Sum	6	4	10

To compute the accuracy, one simply adds the diagonal values 4+3, and divides by bottom right 10 → 7/10 = 70%

PRECISION:

However, we can compute the precision, which is done the same way using the columns:

- ➔ the precision for class 0 is 4/6 = 66%, and for class 1 is 3/4 = 75%

What the precision tells us, is the accuracy of our estimations for each class, this will remove any bias that would be given by the global accuracy, for instance, one might have a good global accuracy, however if one of the classes has a lower accuracy than the other, it means our classifier isn't really working as intended.

- ➔ If we had a unbalanced classes, our precision computation for the unbalanced class would be 100%, which means we can verify if our classes are unbalanced by computing the precision.

RECALL:

One can also compute the recall, by using the rows:

- ➔ Recall for class 0 is 4/5 = 80%, and for class 1 is 3/5 = 60 %

The recall tells us the number of individuals from a given class, that were actually estimated correctly by the model.

- ➔ Unbalanced classes would be identified by the recall computation as well.

By using both the recall and precision computations one can see where the model is not working, and by knowing where the problem is, one can more easily fix the problem

COHEN'S KAPPA:

This computation is R = Pe(i) * Pgt(i), for i=0,1 → Pe(i) = Sum(estimated(i))/ total || Pgt(i) = sum(GT(i))/total

K = (A-R)/(1-R) → if K=0: same as random classifier, K<0: worse than random classifier, K=1: perfect classifier

If we have many classes, these computations can be adapted for such problems!

Inter-participant variability is when the baseline of a measured concept is different depending on the human, for instance if we are measuring heart rate among humans of different health, one has to keep in mind that the baseline heartrate will be different for each human, this means that a state of stress isn't always at the same hartrate.

IMAGE 1

Logistic regression: effect of class imbalance

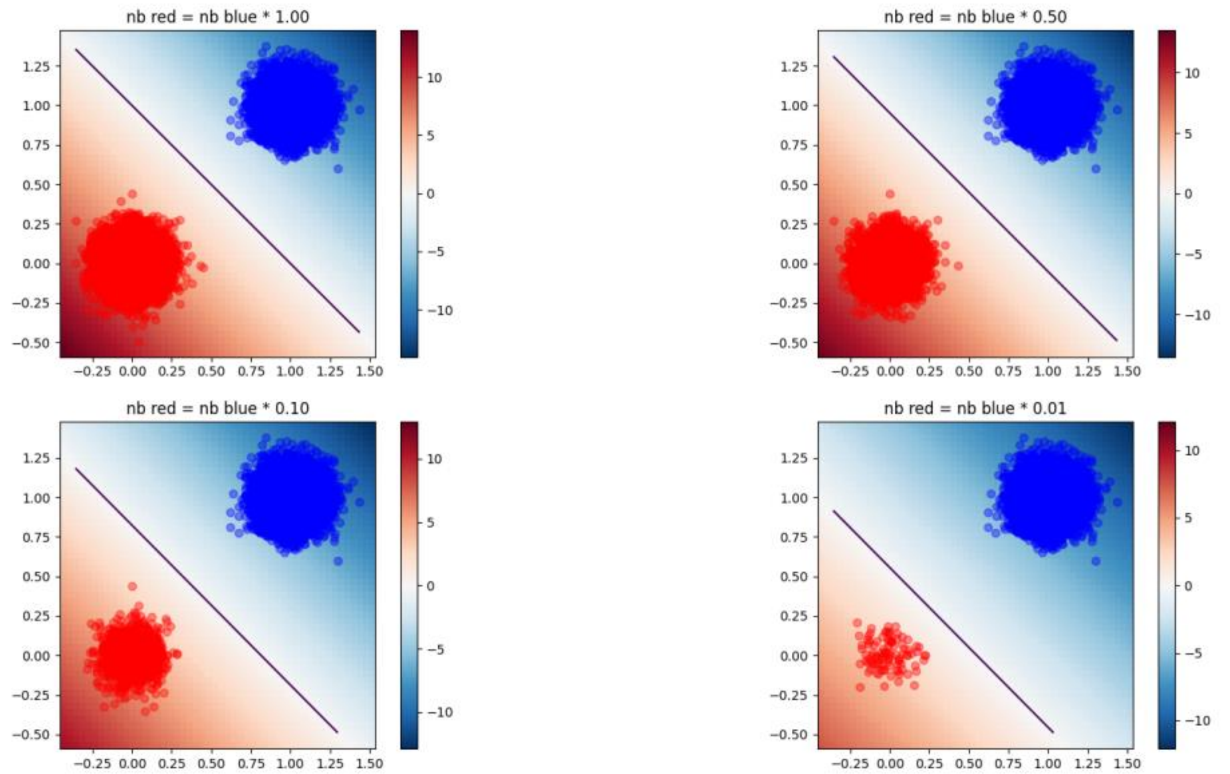

IMAGE 2

Logistic regression: effect of class imbalance