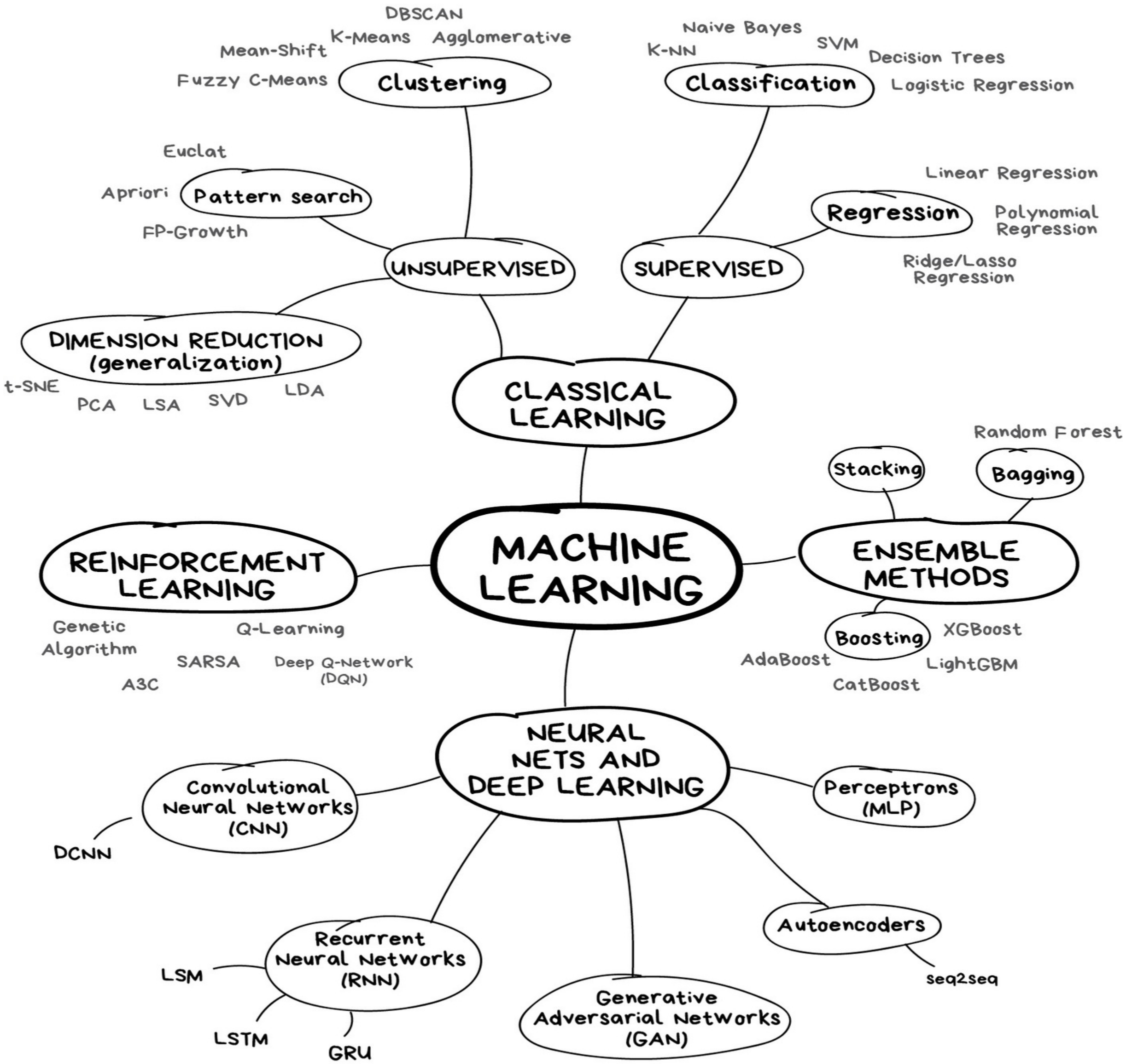
Beyond the headlines:
How to make the best of machine learning models in the wild

NOURA AL MOUBAYED, DURHAM UNIVERSITY

MACHINE LEARNING FOR BIOMEDICINE

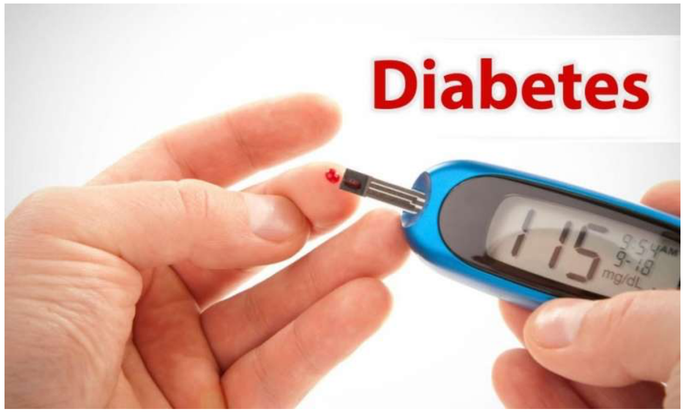
NGSCHOOL 26/10/2019

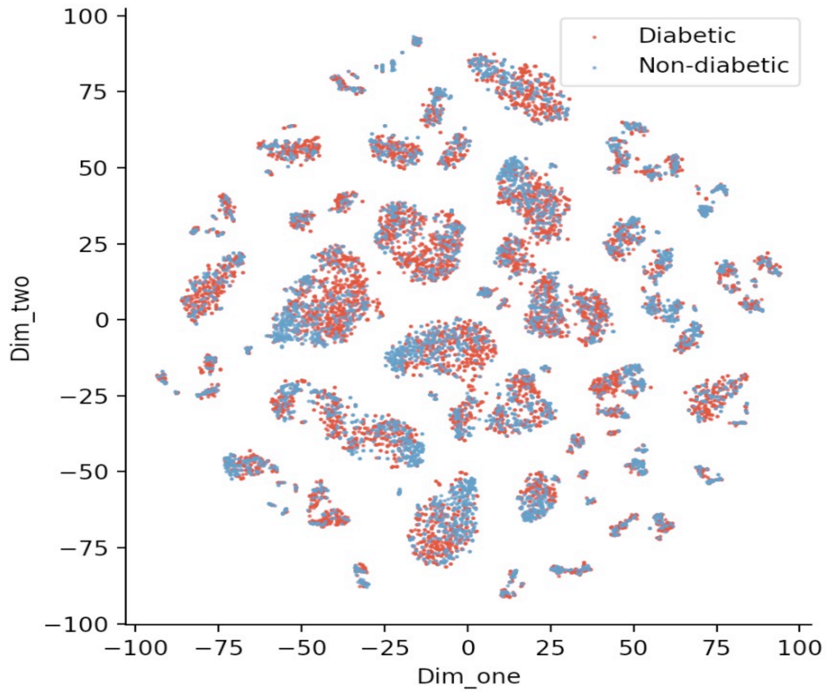
Durham University



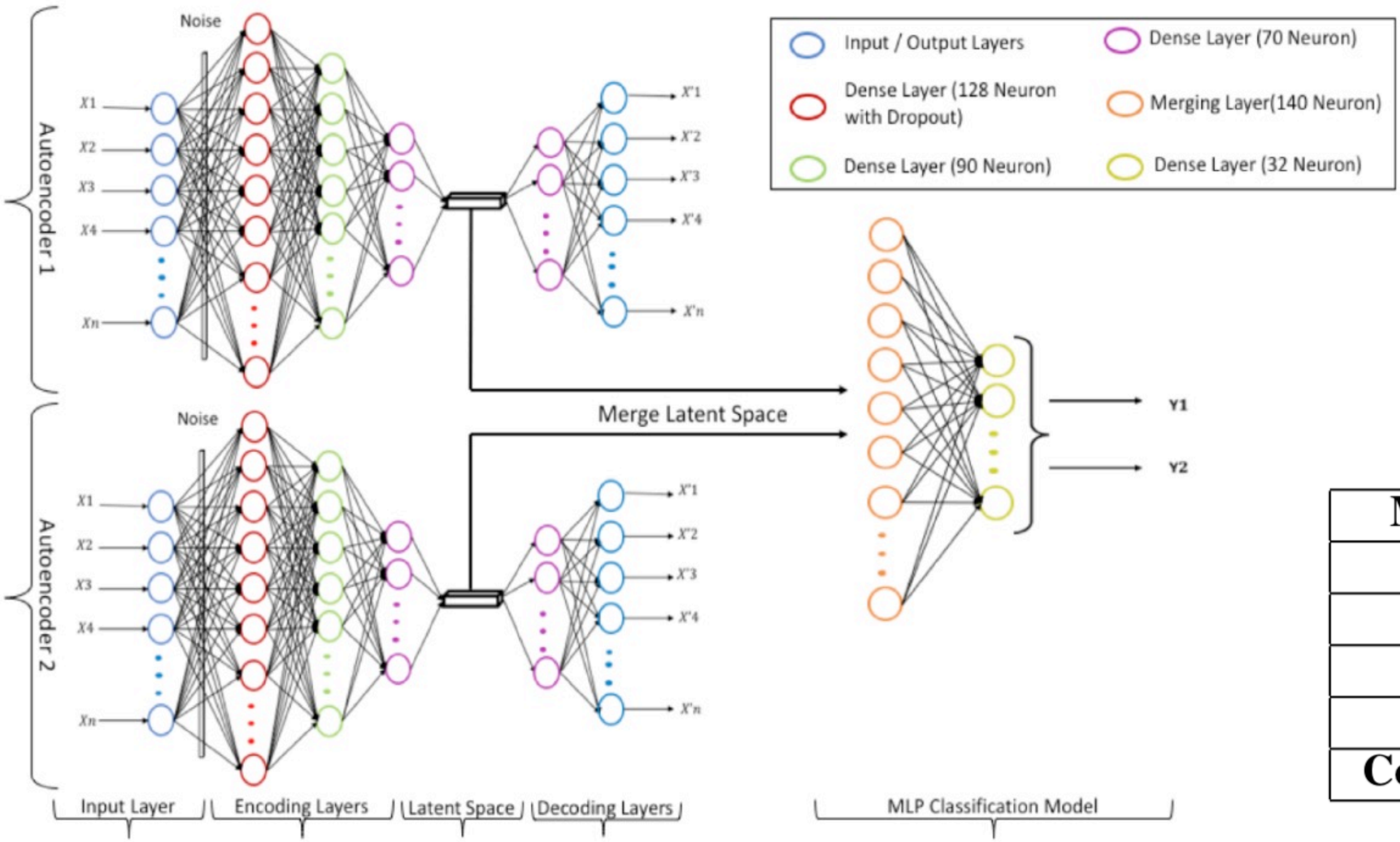
Type-2 Diabetes Mellitus Prediction

- Largest dataset in diabetes research
- Data recorded from 14,609 patients
- 41 million time-stamped lab tests and vital signs



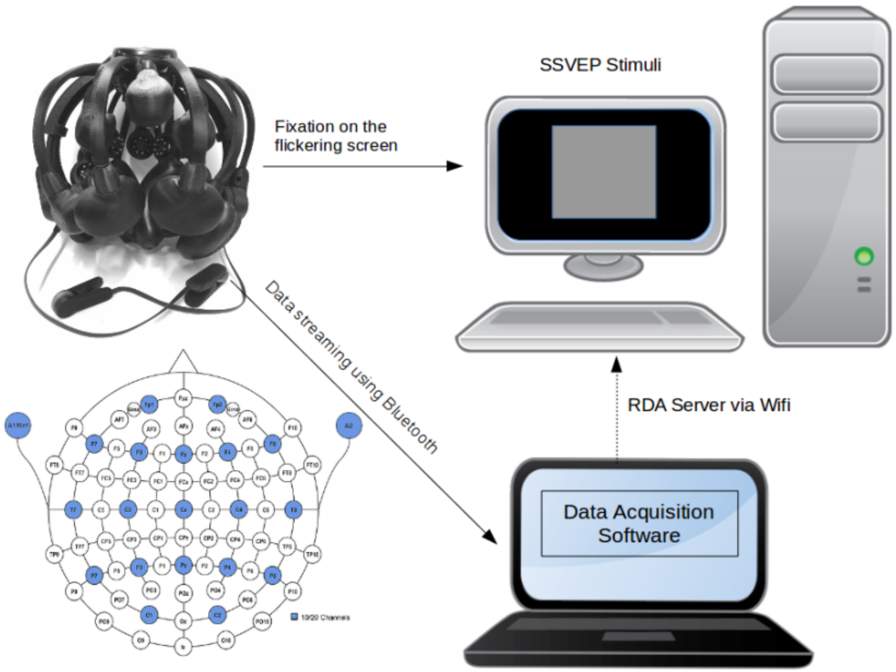


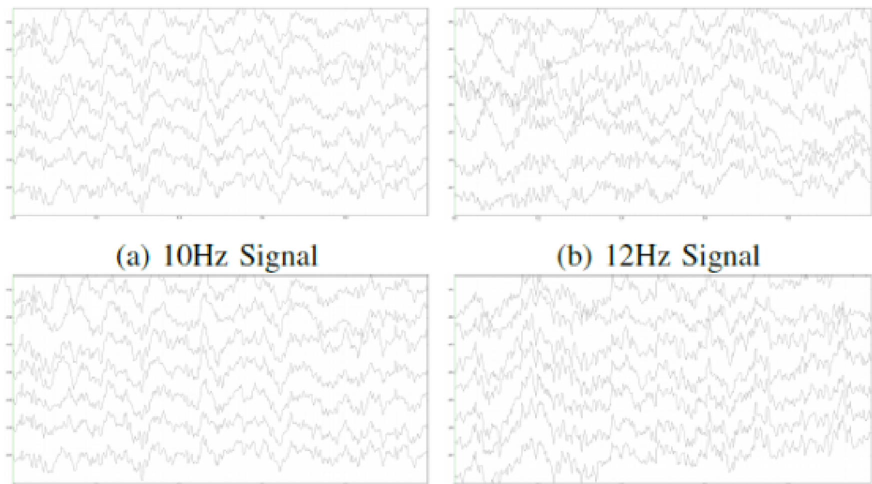
Collaborative Autoencoders



Model	F1-Score
SVM	0.7984
LR	0.7541
MLP	0.8106
DEA	0.7283
Col-DEA	**0.9126**

SSVEP prediction from Dry EEG





(a) 10Hz Signal (b) 12Hz Signal

(c) 15Hz Signal (d) 30Hz Signal

Using Variable Natural Environment Brain-Computer Interface Stimuli for Real-time Humanoid Robot Navigation

Nik Khadijah Nik Aznan, Jason D. Connolly,
Noura Al Moubayed and Toby P. Breckon

Department of {Computer Science, Engineering, Psychology}

Durham University, UK

2019

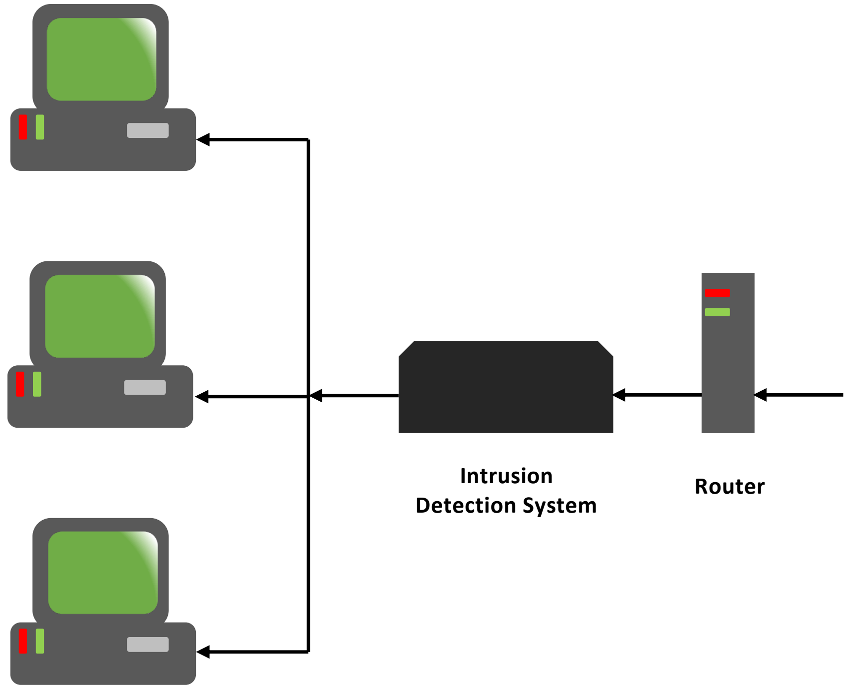
Durham
University

- Network Intrusion Detection

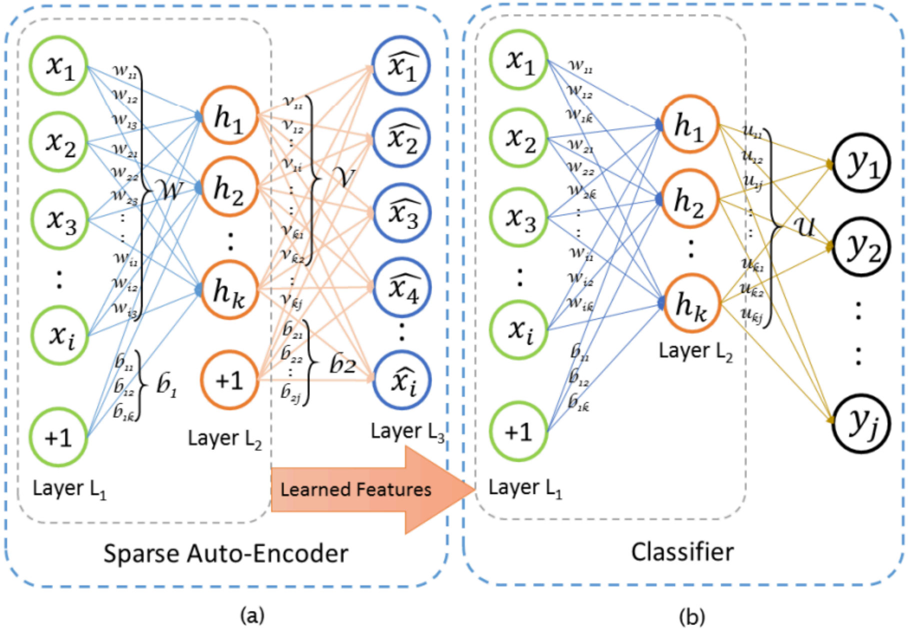
UNSW-NB15 attack categories

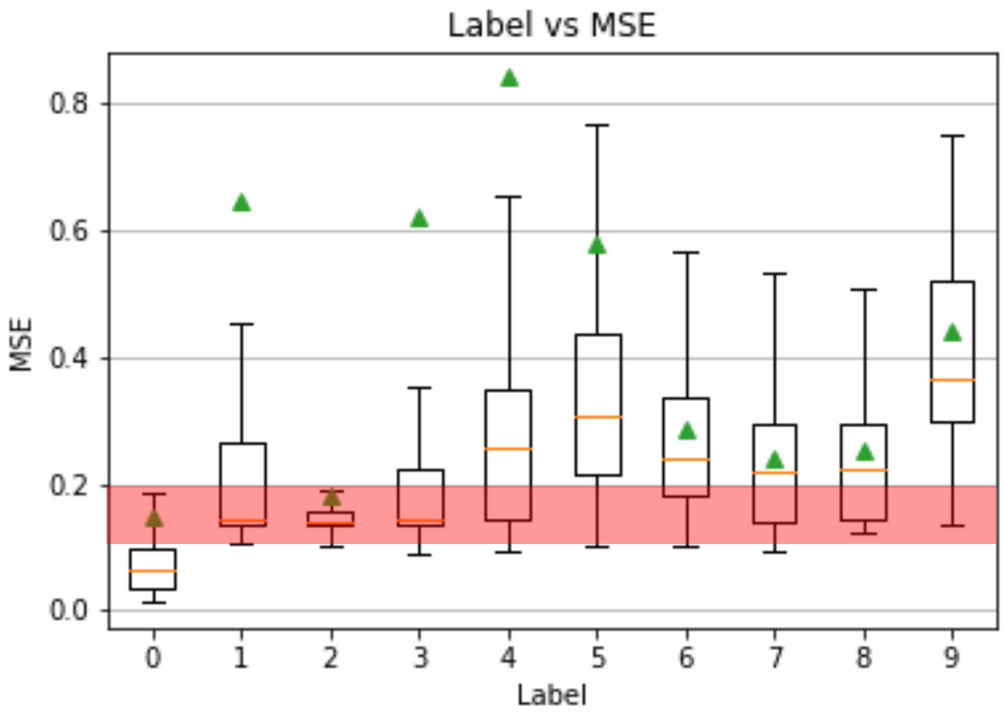
Id	Type	No. Records	Description
0	Normal	2,218,761	Ordinary benign network traffic.
1	Analysis	2,677	Port scanning, spam and html file penetrations methods.
2	Backdoors	2,329	A technique to bypass a security mechanism stealthily.
3	DoS	16,353	An attack which compromises the availability of a service.
4	Exploits	44,525	An attack which exploits a source code vulnerability.
5	Fuzzers	24,246	An automated software testing technique used to find bugs.
6	Generic	215,481	A block-cipher attack without knowledge of structure.
7	Reconnaissance	13,987	A collection of passive information gathering techniques.
8	Shellcode	1,511	A payload used to exploit a software vulnerability.
9	Worms	174	A self-replicating malware that spreads through a network.

Network Intrusion Detection



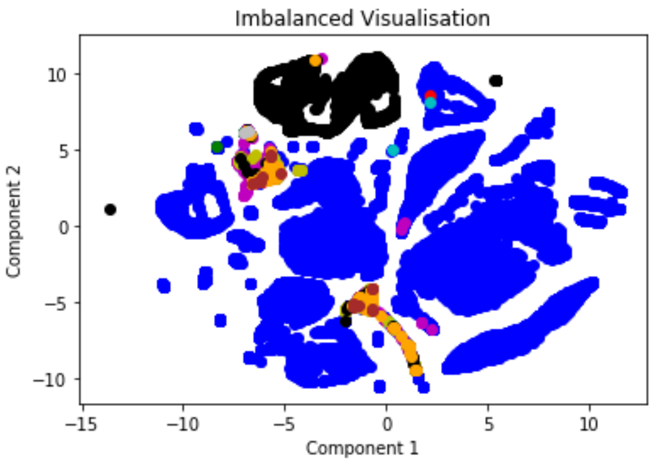
Network Intrusion Detection

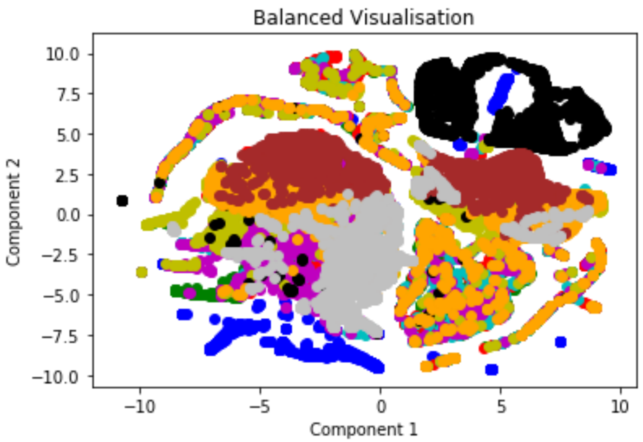




Durham University

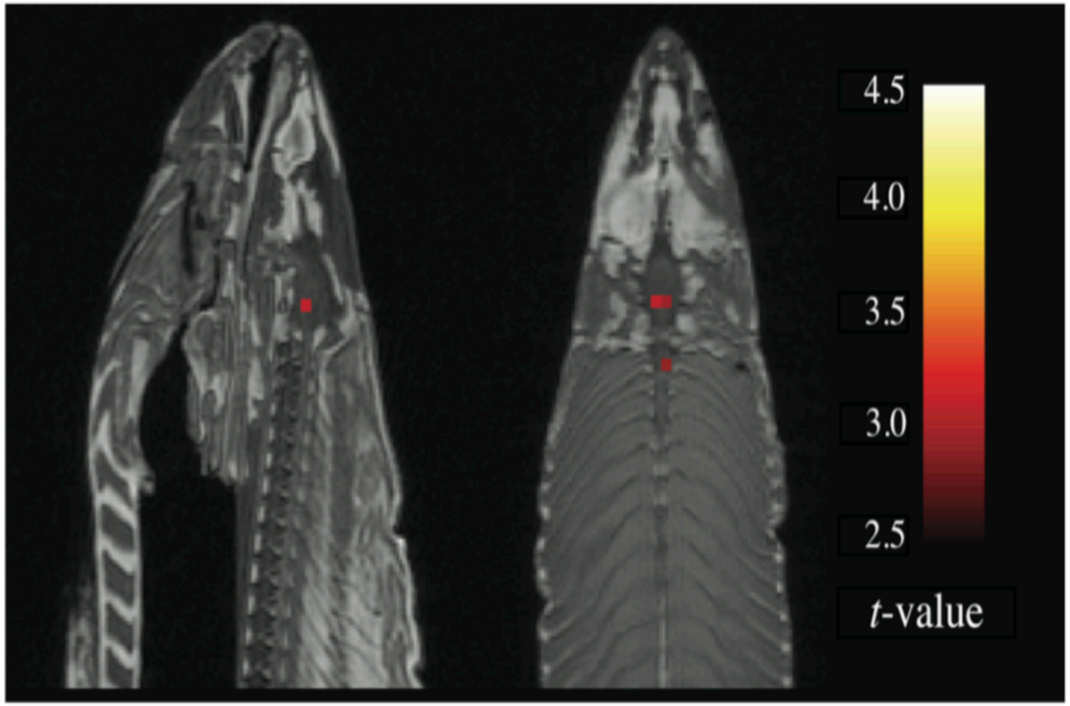
Network Intrusion Detection

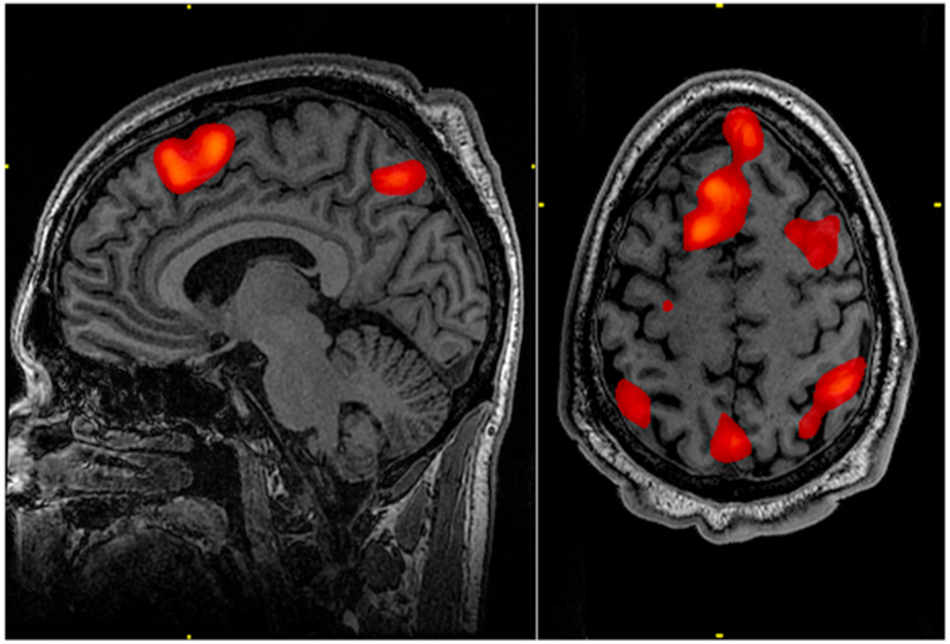




	Artificial Neural Network (ANN)	
	F1	F1
Normal	0.994	0.992
Analysis	0.011	0.534
Backdoors	0.017	0.560
DoS	0.100	0.557
Exploits	0.663	0.648
Fuzzers	0.346	0.916
Generic	0.985	0.986
Reconnaissance	0.665	0.899
Shellcode	0.669	0.993
Worms	0.000	0.993

Dead Fish vs Human Brain

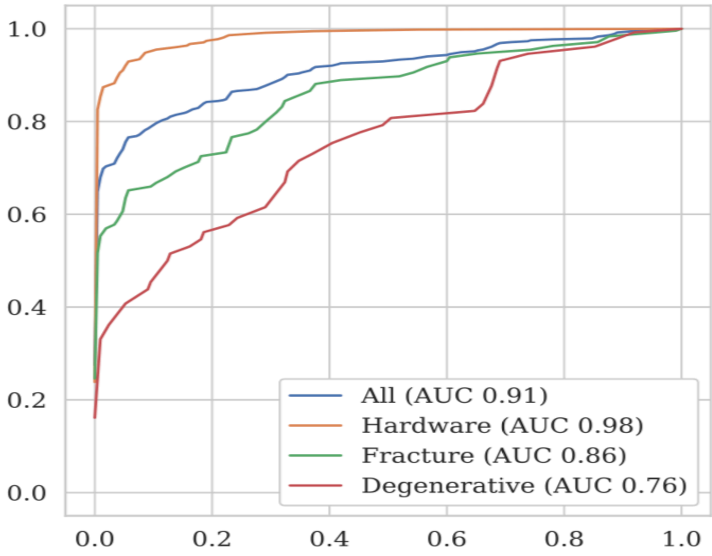


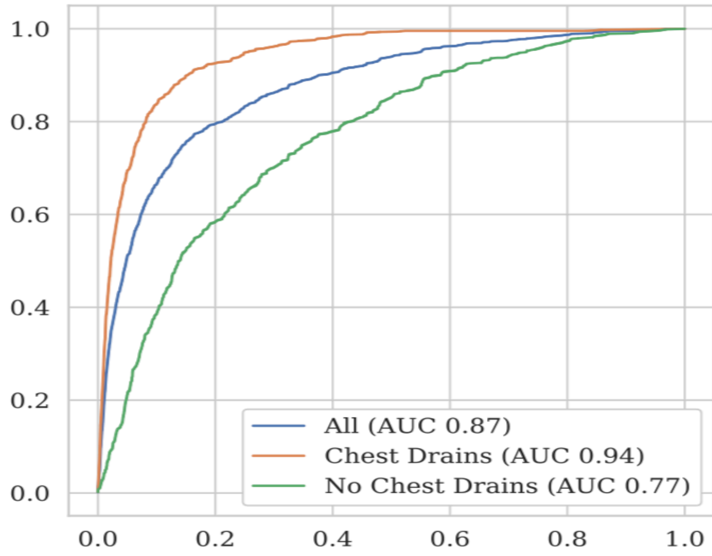


IgNobel Prize in Neuroscience, 2012

Bennett et al. "Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction" Journal of Serendipitous and Unexpected Results, 2010.

Hidden Stratification Problem





	Sensitivity/Recall
Cancer (human)	95%
Cancer (AI)	95%
High-risk subtype (human)	100%
High-risk subtype (AI)	0%

Rayner,et al, Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging, 2019

The Institute of
Advanced Research
Computing

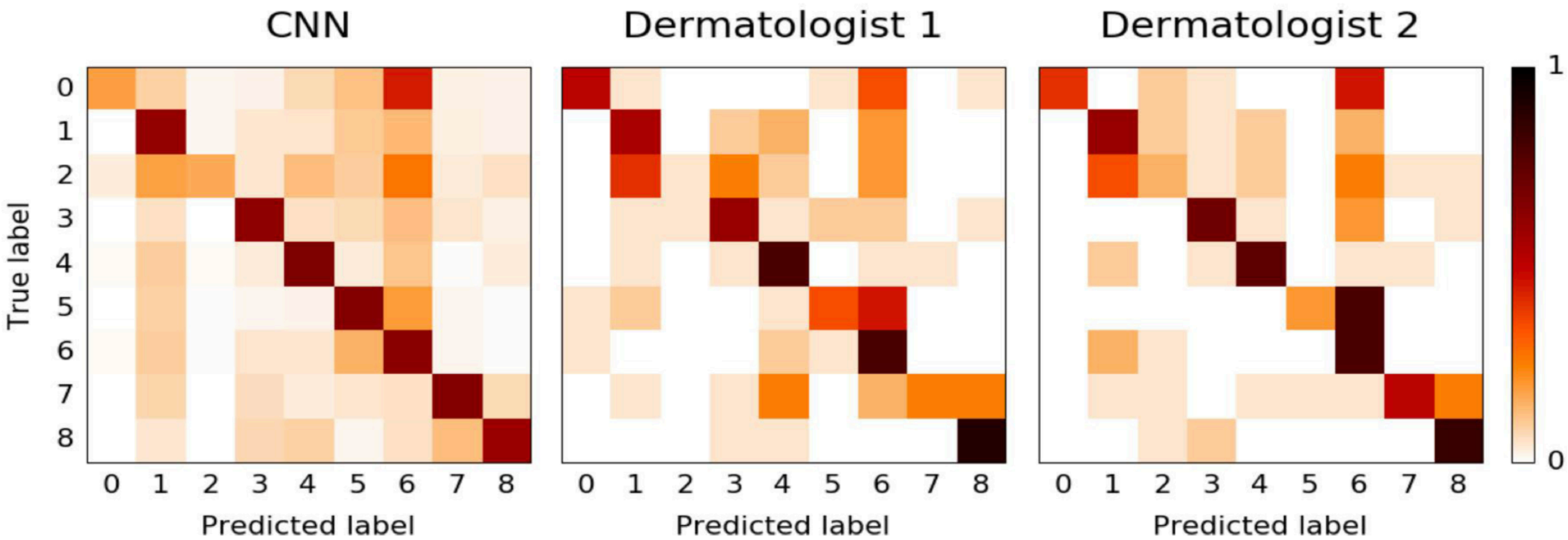
Rayner,et al, Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging, 2019

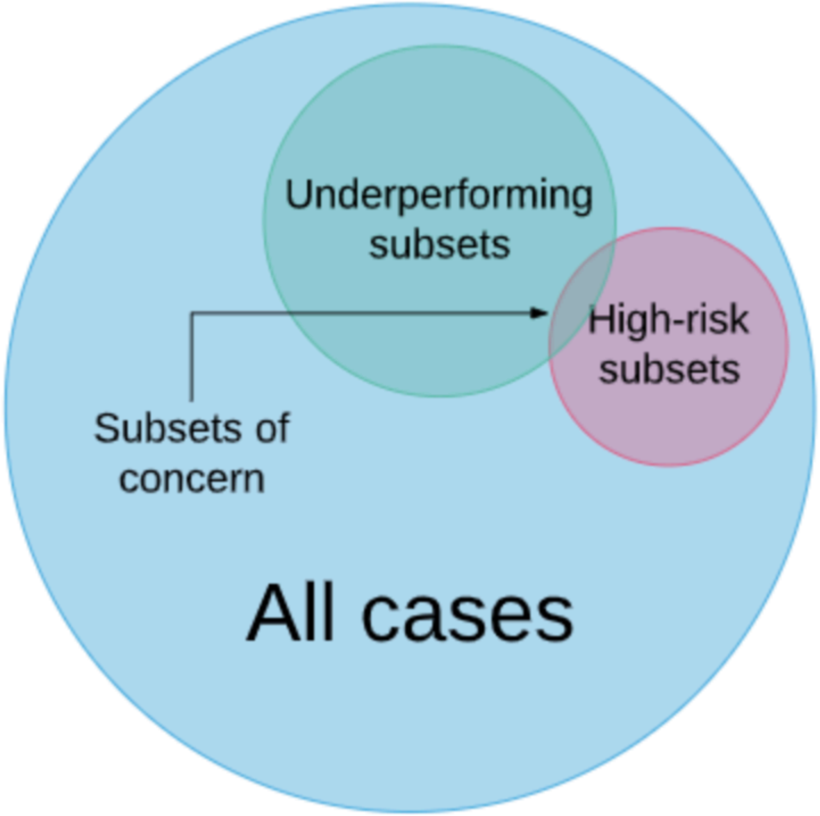
LETTER

doi:10.1038/nature21056

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]

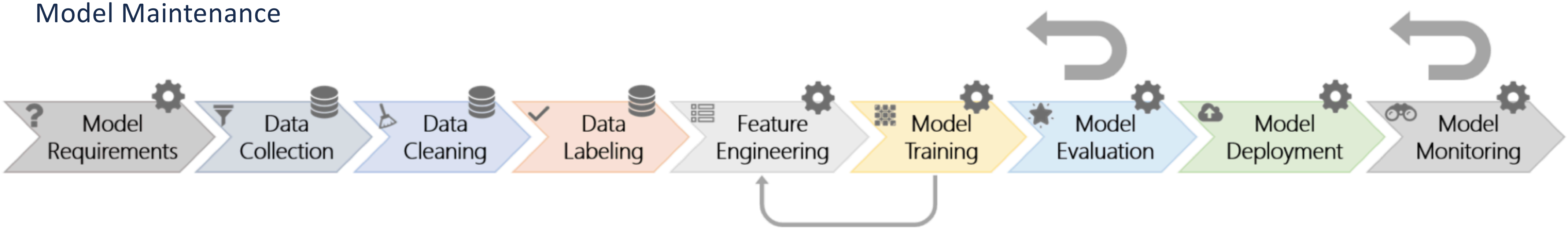




“For an AI system to be judged safe for task X, we need to know the performance in the subsets of concern Y and Z.”

"The most important question by far when someone wants to deploy a machine learning model is, how often do you want to deploy this? If the answer is once, that is a bad answer. You should never deploy a machine learning model once. You should deploy it never or prepare to deploy it over and over and over and over and over again, repeatedly forever, ad infinitum."

Model Maintenance



S Amershi et al, Software engineering for machine learning: a case study, Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice, *2019*

What do all these faces have in common?

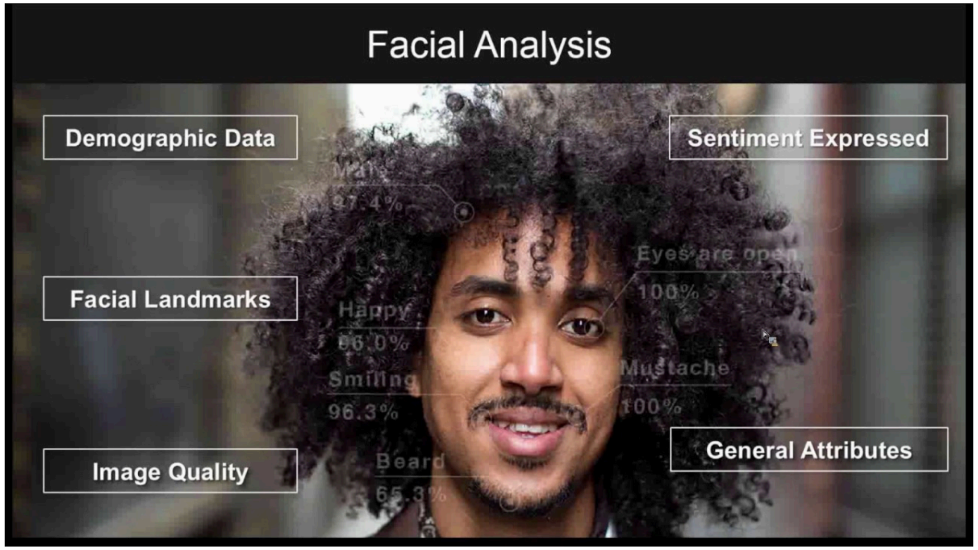
Deep Fake

https://thispersondoesnotexist.com

Durham University

MIT researchers: Amazon's Rekognition shows gender and ethnic bias

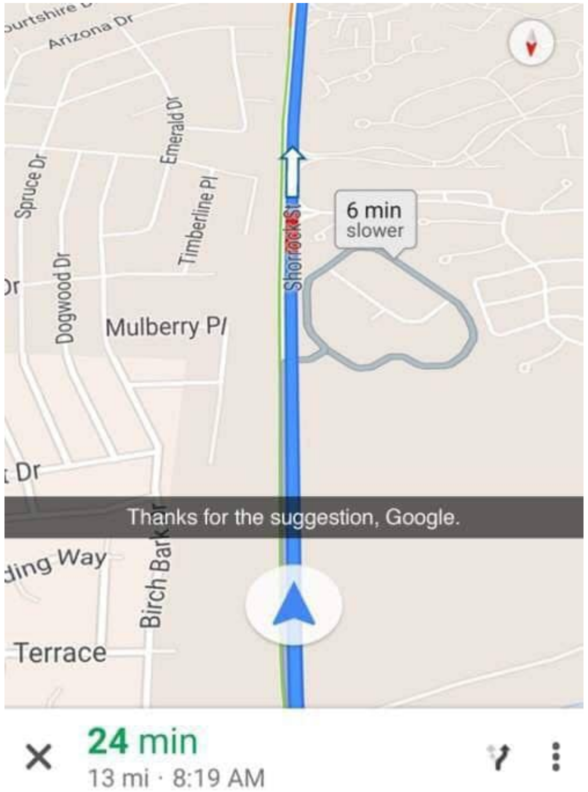
KYLE WIGGERS @KYLE_L_WIGGERS JANUARY 24, 2019 7:50 PM



Above: Amazon's facial recognition service, Amazon Rekognition.

28 Congressional representatives were misidentified as criminals. A majority of the false matches — 38 percent — were people of colour.

Foolish ML



Me: *Spent days and hours writing a code*

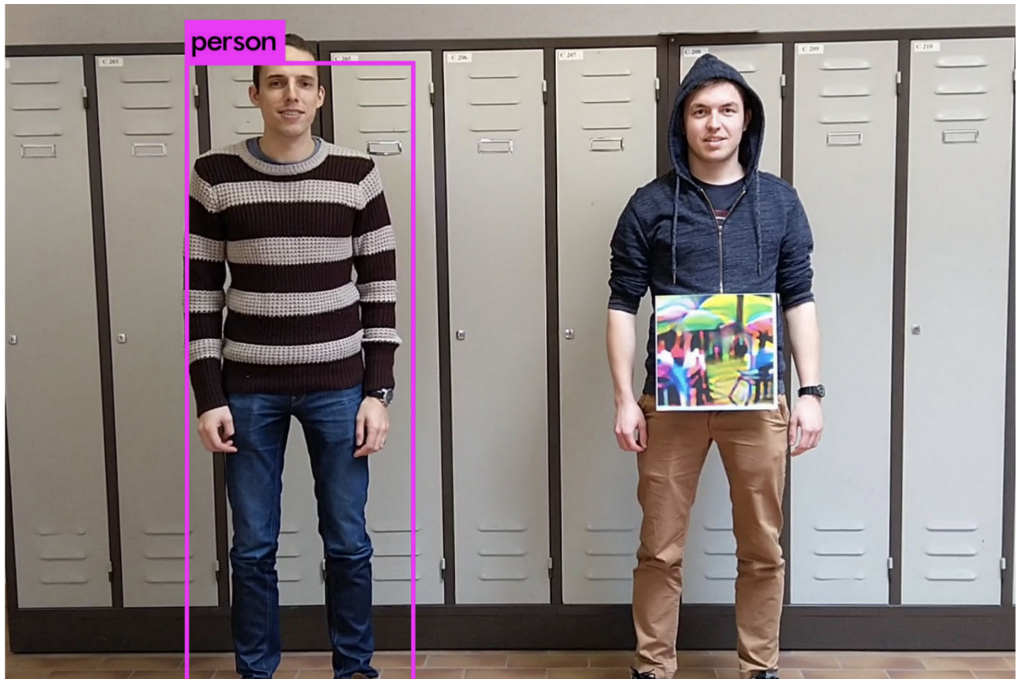
My CNN:

Parents: **If all your friends jumped into the well, Will you?**

Kid: **NO!**

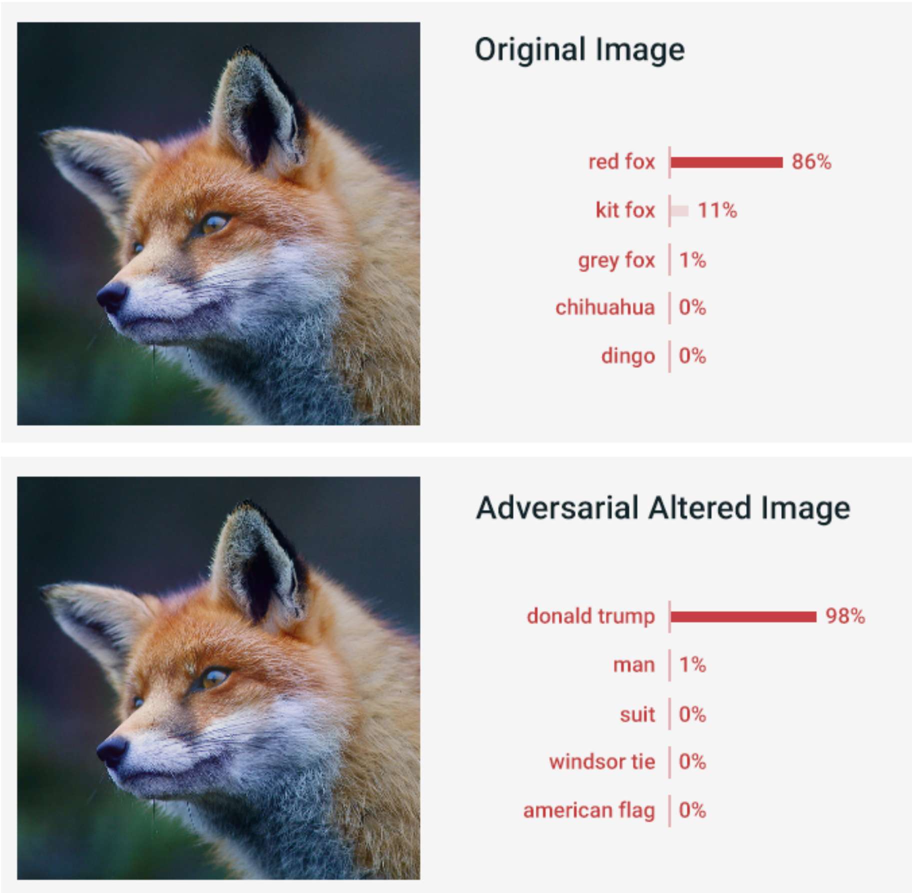
Machine Learning: **????**

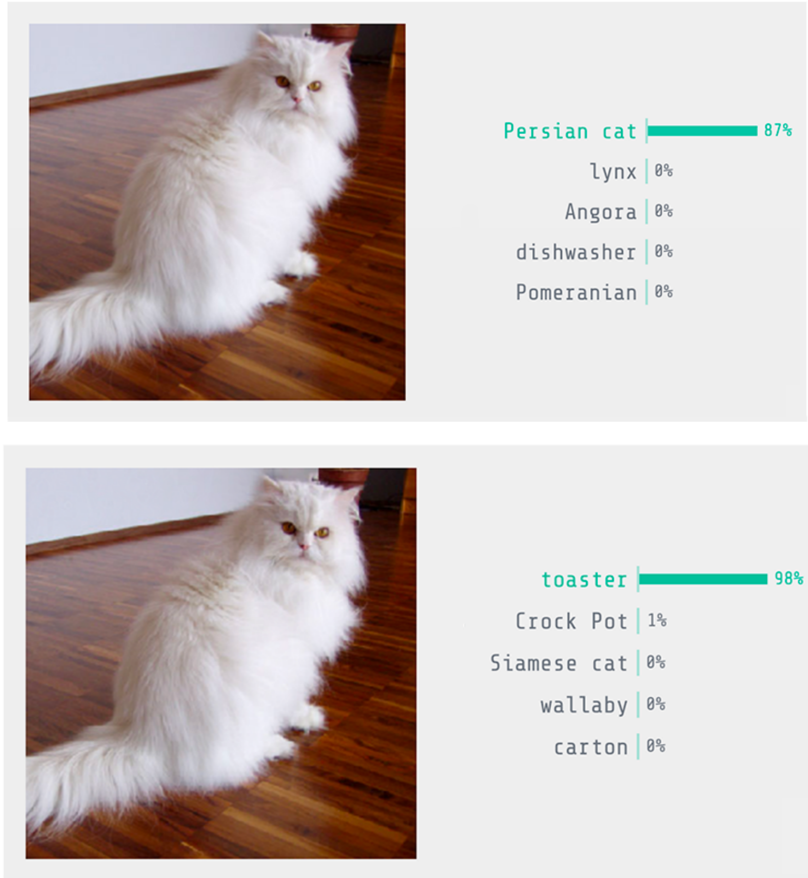
Foolish ML



Simen Thys, Fooling automated surveillance cameras: adversarial patches to attack person detection, 2019

Foolish ML


Original Image
red fox 86%
kit fox 11%
grey fox 1%
chihuahua 0%
dingo 0%
Adversarial Altered Image
donald trump 98%
man 1%
suit 0%
windsor tie 0%
american flag 0%


Persian cat 87%
lynx 0%
Angora 0%
dishwasher 0%
Pomeranian 0%
toaster 98%
Crock Pot 1%
Siamese cat 0%
wallaby 0%
carton 0%

When Predictive performance is not enough

Why?
-Fairness is critical
-Consequences can be far-reaching
-Cost of a mistake is high
-New hypothesis is observed
-GDPR
-Right to Explanation

Transparent solutions
-Rule based
-GAM(Generalized Additive Models)
-LIME (Locally Interpretable Model Agnostic Explanations)
-Naïve Bayes
-Regression Models
-Shapley Values

White-box
-Shallow Ensembles

Black-box
-Deep Learning
-Gradient Boosting
-SVM