Fairness, Explainability & Robustness in Machine Learning

EEIP Lecture Series - 3rd November 2020

Celia Cintas

IBM Research | Africa

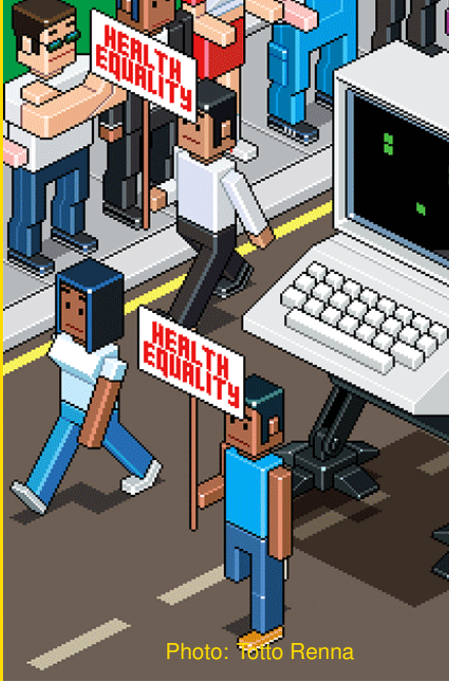


Photo: Tolfo Renna

Table of contents

IBM Research | Africa

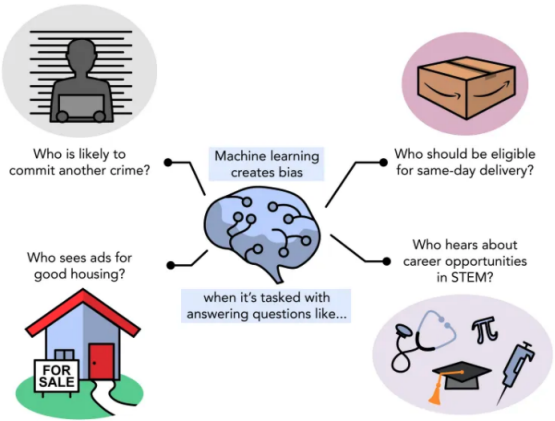
Why is it important?

ML may seem like an **objective** solution

An algorithm is shown a large amount of representative data to learn from. It is important to know that any data we give to the ML model **describes the choices that have already been made in society**.

IBM Research | Africa

Why is it important? (Cont.)



MOTHERBOARD
TECH BY VICE

Google's Sentiment Analyzer Thinks Being Gay Is Bad

Amazon working to address racial disparity in same-day delivery service

By T.C. Sottek | May 8, 2016, 11:38am EDT
Source Bloomberg and Bloomberg

MIT Technology Review

Facebook's ad-serving algorithm discriminates by gender and race

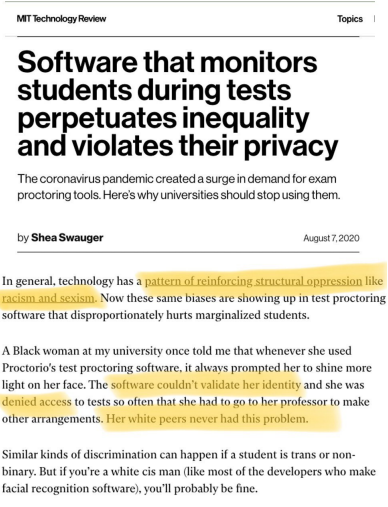
Homes for sale are shown to a higher fraction of white users

The fraction of white users in the ad's audience

IBM Research | Africa

Figure: Nicholas Lue

Inequality in solutions during the COVID-19 crisis: Education



MIT Technology Review Topics

Software that monitors students during tests perpetuates inequality and violates their privacy

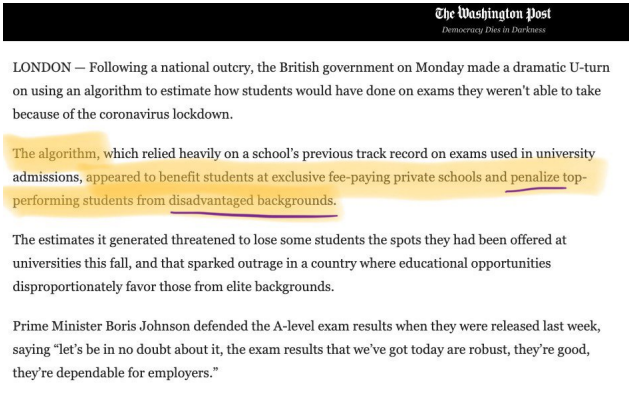
The coronavirus pandemic created a surge in demand for exam proctoring tools. Here's why universities should stop using them.

by **Shea Swauger** August 7, 2020

In general, technology has a pattern of reinforcing structural oppression like racism and sexism. Now these same biases are showing up in test proctoring software that disproportionately hurts marginalized students.

A Black woman at my university once told me that whenever she used Proctorio's test proctoring software, it always prompted her to shine more light on her face. The software couldn't validate her identity and she was denied access to tests so often that she had to go to her professor to make other arrangements. Her white peers never had this problem.

Similar kinds of discrimination can happen if a student is trans or non-binary. But if you're a white cis man (like most of the developers who make facial recognition software), you'll probably be fine.



The Washington Post
Democracy Dies in Darkness

LONDON — Following a national outcry, the British government on Monday made a dramatic U-turn on using an algorithm to estimate how students would have done on exams they weren't able to take because of the coronavirus lockdown.

The algorithm, which relied heavily on a school's previous track record on exams used in university admissions, appeared to benefit students at exclusive fee-paying private schools and penalize top-performing students from disadvantaged backgrounds.

The estimates it generated threatened to lose some students the spots they had been offered at universities this fall, and that sparked outrage in a country where educational opportunities disproportionately favor those from elite backgrounds.

Prime Minister Boris Johnson defended the A-level exam results when they were released last week, saying "let's be in no doubt about it, the exam results that we've got today are robust, they're good, they're dependable for employers."

Inequality in solutions during the COVID-19 crisis: Health

Healthcare IT News

AI bias may worsen COVID-19 health disparities for people of color

A new article in the Journal of the American Medical Informatics Association points to the dissemination of "under-developed and potentially biased models" in response to the novel coronavirus.

Stanford SOCIAL INNOVATION Review

Informing and inspiring leaders of social change

SOCIAL ISSUES | SECTORS | SOLUTIONS | MAGAZINE | MORE

Technology

The Problem With COVID-19 Artificial Intelligence Solutions and How to Fix Them

NEWS & FEATURES

Artificial Intelligence, Health Disparities, and Covid-19

How racially biased is AI medicine? Experts are asking if biased algorithms worsen Covid-19's toll on Black Americans.

Visual: FG Trade / Getty Images

| Africa

Why is it hard to fix?

Unknown sources The introduction of bias isn't always obvious during a model's construction.

Missing ethics design Many of the standard practices in deep learning are not designed with bias detection in mind.

The definitions of fairness It's also not clear what the absence of bias should look like.

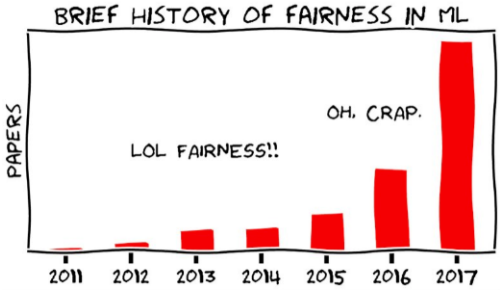


Figure: Moritz Hardt

IBM Research | Africa

Source: MIT Review by Karen Hao

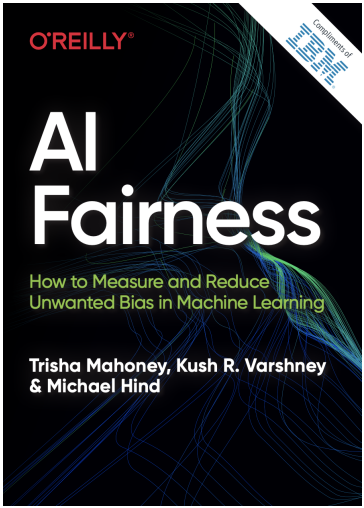
How & when can fairness be improved in our ML pipelines?

1. Detect and Report bias.
2. Mitigation techniques:
   - Pre-processing algorithms can be used if we can modify the training data.
   - We can use In-processing algorithms if we can change the learning procedure for a ML model.
   - If we need to treat the learned model as a black-box and cannot modify the training data or learning algorithm, we will need to use the Post-processing algorithms.
3. Continuous Pipeline Measurement. The probability distribution governing data can change over time, resulting in the training data distribution drifting away from the actual data distribution. Bias measurement and mitigation should be integrated into your continuous pipeline measurements.

IBM Research | Africa

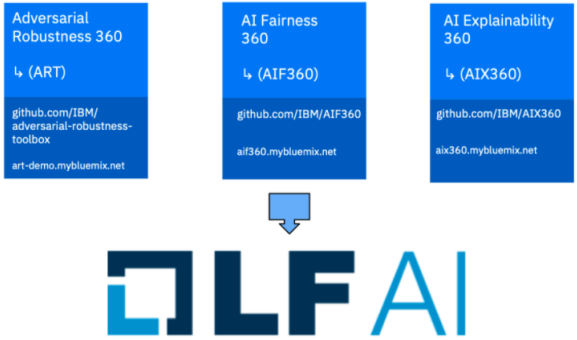
How & when can fairness be improved in our ML pipelines? (Cont.)

Pre-Processing Algorithms Mitigate bias in training data	In-Processing Algorithms Mitigate bias in classifiers	Post-Processing Algorithms Mitigate bias in predictions
Reweighing Modifies the weights of different training examples	**Adversarial Debiasing** Uses adversarial techniques to maximize accuracy and reduce evidence of protected attributes in predictions	**Reject Option Classification** Changes predictions from a classifier to make them more fair
Disparate Impact Remover Edits feature values to improve group fairness	**Prejudice Remover** Adds a discrimination-aware regularization term to the learning objective	**Calibrated Equalized Odds** Optimizes over calibrated classifier score outputs that lead to fair output labels
Optimized Preprocessing Modifies training data features and labels	**Meta Fair Classifier** Takes the fairness metric as part of the input and returns a classifier optimized for the metric	**Equalized Odds** Modifies the predicted label using an optimization scheme to make predictions more fair
Learning Fair Representations Learns fair respresentations by obfuscating information about protected attributes		



IBM Research | Africa

How & when can fairness be improved in our ML pipelines? (Cont.)



1. **AI Fairness 360** https://aif360.mybluemix.net/ examine, report, and mitigate discrimination and bias in machine learning.
2. **AI Explainability 360** https://aix360.mybluemix.net/ algorithms that span the different ways of explaining along with proxy explainability metrics.
3. **Adversarial Robustness Toolbox (ART)** https://art-demo.mybluemix.net/ provides tools to defend and evaluate ML models against adversarial threats.

IBM Research | Africa

Properties that we need to asses in our ML solutions

Fairness ensures that a model's predictions do not unethically discriminate **unprivileged groups**.

Robustness defines expectations for how an ML model will behave upon **deployment in the real world**.

Transparency & Explainability tackles the question of **how** an ML model makes its predictions.

IBM Research | Africa

Current Research Projects

IJCAI-PRICAI YOKOHAMA 2020

Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)

Detecting Adversarial Attacks via Subset Scanning of Autoencoder Activations and Reconstruction Error

Celia Cintas[1*], Skyler Speakman[1*], Victor Akinwande[1], William Ogallo[1], Komminist Weldemariam[1], Srihari Sridharan[1] and Edward McFowland[2]
[1]IBM Research Africa, Nairobi, Kenya.
[2] Carlson School of Management, University of Minnesota, USA.
celia.cintas@ibm.com,skyler@ke.ibm.com, {victor.akinwande1, william.ogallo}@ibm.com, {k.weldemariam, sriharis.sridharan}@ke.ibm.com, mcfowland@umn.edu

MICCAI 2020 Lima PERU

Estimating Skin Tone and Effects on Classification Performance in Dermatology Datasets

Newton M. Kinyanjui,[1,4] Timothy Odonga,[1,4] Celia Cintas,[1] Noel C. F. Codella,[2] Rameswar Panda,[3] Prasanna Sattigeri,[2] and Kush R. Varshney[1,2]
IBM Research, [1]Nairobi, Kenya. [2]Yorktown Heights, NY, USA. [3]Cambridge, MA, USA
[4]Carnegie Mellon University Africa, Kigali, Rwanda

PRESERVATION OF ANOMALOUS SUBGROUPS ON VARIATIONAL AUTOENCODER TRANSFORMED DATA

ICASSP2020 Barcelona

Samuel C. Maina[1] *Reginald E. Bryant*[1] *William O. Ogallo*[1]
Kush R. Varshney[1] *Skyler Speakman*[1] *Celia Cintas*[1]
Aisha Walcott-Bryant [1] *Robert-Florian Samoilescu*[1,2] *Komminist Weldemariam*[1]

[1] IBM Research, Nairobi, Kenya
[2]Politehnica University of Bucharest, Bucharest, Romania

1. Improving existing ML techniques to address global health challenges in developing countries. [CRA+20, DCOWB]
2. Leveraging statistical methods in the space of neural networks' activations to detect anomalies and provide robustness to trained models [CSA+20].
3. Investigating how to ensure that ML models are fair, robust and reliable for everyone [KOC+20].

IBM Research | Africa

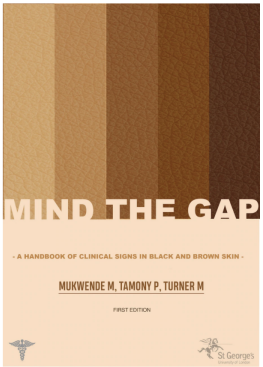
Disparities in ML & Dermatology

1. Are standard **dermatology image datasets** used in ML tasks **biased with respect to skin tone**? Can we quantify this?
2. If so, does the dataset bias lead to **unequal performance** of downstream disease classification?

IBM Research | Africa

Disparities in Dermatology

- In the African American population, melanoma is often diagnosed at an advanced stage with deeper tumors [MSL+17, WEK+11].
- Five year survival rates for Acral Lentiginous Melanoma (ALM) is 82.6% in caucasian population, but only 77.2% in African American patients. [MCH15].
- The paucity of images of skin manifestations in patients with darker skin is problematic, because it may make identification of COVID-19 presenting with cutaneous manifestations more difficult for both dermatologists and the public. [LJZ+20]
- Dermatologists started an international registry to catalog examples of skin manifestations of COVID-19. The registry compiled more than 700 cases, but only 34 and 14 cases respectively for patients with Hispanic and African ancestry. [Rab20]



The cover of the "Mind the Gap" handbook, written by Malone Mukwende, with two of his lecturers, Peter Tamony and Margot Turner.

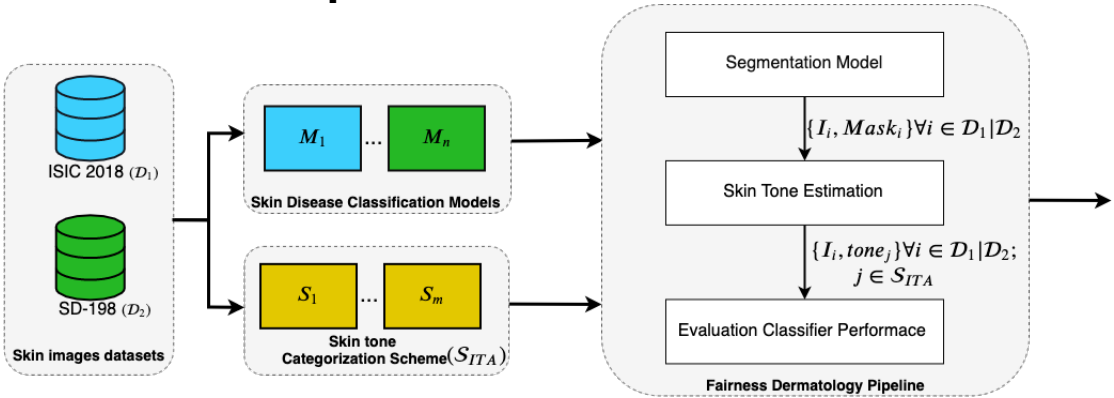
IBM Research | Africa

Machine Learning & Dermatology

- Skin disease diagnosis using machine learning
  1. Benchmark model for melanoma diagnosis outperforms trained dermatologists [CNP+16]
  2. ISIC challenges (https://www.isic-archive.com/)
- Predictive inequity in computer vision with respect to skin type
  1. Automated face image analysis for gender classification [BG18]
  2. Pedestrian detection systems [WHM19]

IBM Research | Africa

Overview : Proposed Framework



Kinyanjui, et al. "Estimating skin tone and effects on classification performance in dermatology datasets."MICCAI 2020 [KOC+20].

IBM Research | Africa

Sensitive Content Warning

Skin Disease Graphical Content Warning

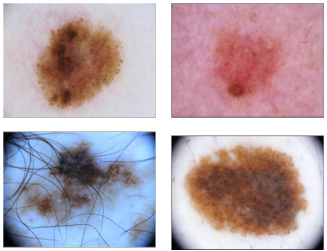
Note that we will show **skin disease examples** that could be **sensitive** or **triggering** to some viewers. We notice this, so viewers can prepare themselves to adequately engage or, if necessary, disengage for their own well-being.

IBM Research | Africa

Datasets

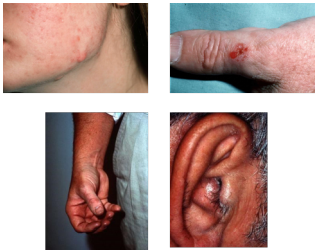
ISIC 2018

- 10015 dermoscopic images
- 7 disease classes
- 2594 images with ground truth segmentation masks for diseased area



SD-198

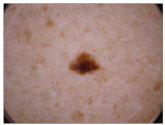
- 6548 clinical images
- 198 disease classes
- No segmentation data

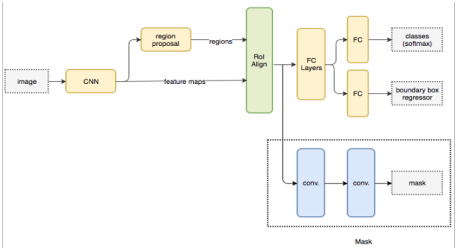


IBM Research | Africa

Segmentation to Obtain Non-Diseased Region

1. Finetune Mask R-CNN model ([HGDG17])
   - Adjust pretrained classifier with a FastRCNNPredictor with 2 classes (background and mask)
   - Adjust mask predictor with new MaskRCNNPredictor with 2 classes and 512 hidden neurons
2. Further apply thresholding techniques on predicted grayscale mask including contour extraction for ISIC2018 and grid search for optimal binary thresholding for SD-136





IBM Research | Africa

Skin Tone Metric of Non-Diseased Region

1. Given non-diseased pixels, characterize them with a skin tone metric
   1. Use individual typology angle (ITA) [WWdPR15], Highly correlated with melanin index
   2. $\text{ITA} = \tan^{-1}\left(\frac{L-50}{b}\right) \times \frac{180}{\pi}$ Where L is luminance and b quantifies amount of yellow.
   3. Use pixels with L and b values within 1 standard deviation to deal with outliers.
2. Bin into categories [CSD+15]

ITA Range	Skin Tone Category	Abbreviation
ITA $> 55^\circ$	Very Light	very_lt
$48^\circ <$ ITA $\leq 55^\circ$	Light 2	lt2
$41^\circ <$ ITA $\leq 48^\circ$	Light 1	lt1
$34.5^\circ <$ ITA $\leq 41^\circ$	Intermediate 2	int2
$28^\circ <$ ITA $\leq 34.5^\circ$	Intermediate 1	int1
$19^\circ <$ ITA $\leq 28^\circ$	Tanned 2	tan2
$10^\circ <$ ITA $\leq 19^\circ$	Tanned 1	tan1
ITA $\leq 10^\circ$	Dark	dark

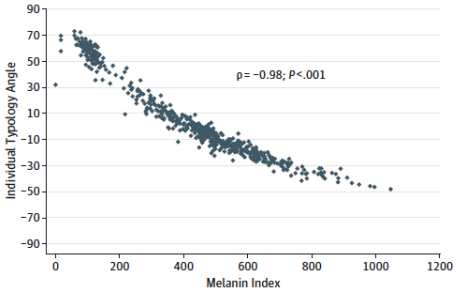


Figure from [WWdPR15].

IBM Research | Africa

Classification of Skin Disease

We replicated state-of-the-art skin disease classification neural networks:

1. Finetune Densenet 201 model pretrained on ImageNet [HLVDMW17].
2. Regularization methods: Dropout and early stopping.

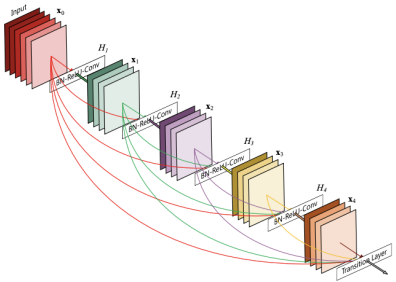


Figure from [HLVDMW17]

Overall accuracy on validation data

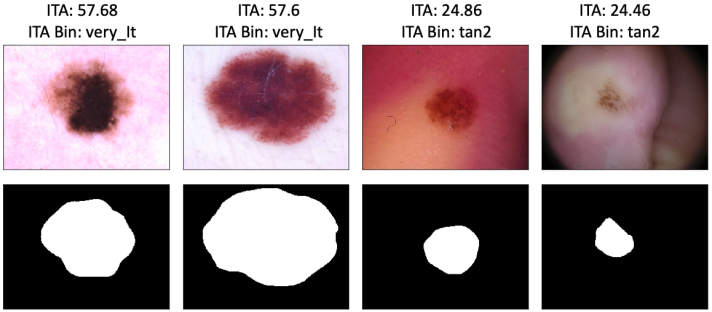
- ISIC2018: balanced accuracy 0.884 (benchmark model 0.885)
- SD-136: accuracy score 0.567 (benchmark model 0.52)

IBM Research | Africa

Results

Metrics for segmentation on ISIC 2018

The Mask R-CNN model yields an accuracy of **0.956**, a false negative rate of **0.024**, and a mean absolute error in ITA computation of **0.428** degrees. [KOC+20]

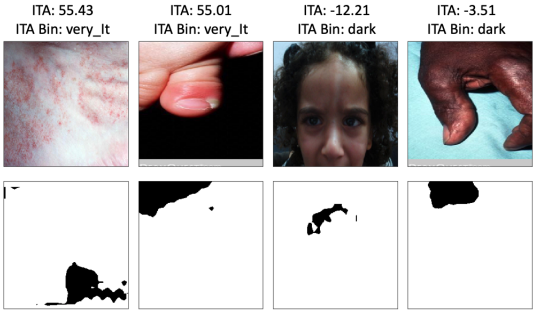


IBM Research | Africa

Results (Cont.)

Metrics for segmentation on SD-136

The segmentation model on the SD-136 dataset yield an accuracy of **0.802**, a false negative rate of **0.076**, and a mean absolute error in ITA computation of **3.572** degrees. [KOC+20]

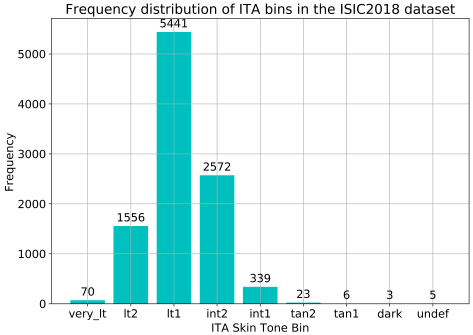


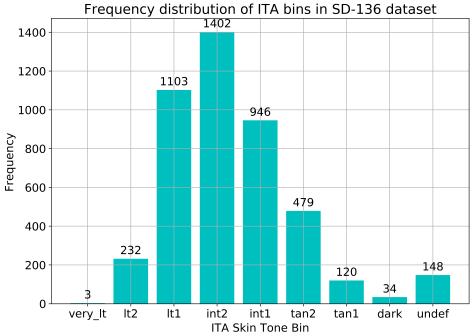
IBM Research | Africa

Results (Cont.)

Skin Tone Distribution

There is underrepresentation of darker skin tones in both datasets



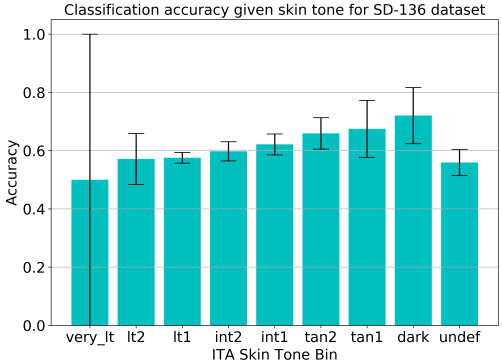


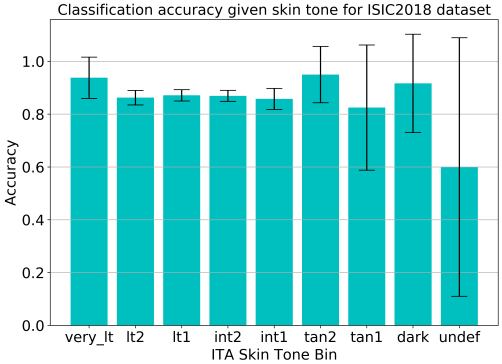
IBM Research | Africa

Results (cont.)

Accuracy by Skin Tone

There is no statistically significant pattern of unequal performance by skin tone

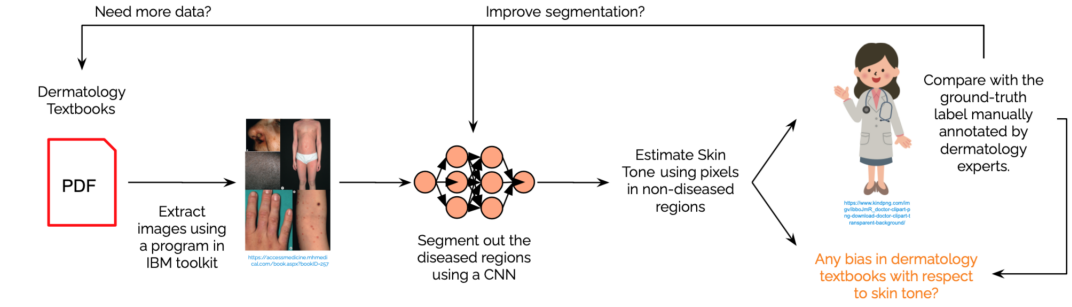




Extensions

Automatic report of skin tone distributions

Kim et al. are currently working on extending the current segmentation and classification models to work on dermatology textbooks and academic paper images.



Further experiments & Open Questions on Fair Dermatology

1. Experiments around stratification of skin tone by disease.
2. How a fair distribution looks like in this case?
3. How can we improve the segmentation process across all skin tones?
4. How can we collaborate with dermatologists across the continent to avoid having only US-centric sources in dermatology?

IBM Research | Africa

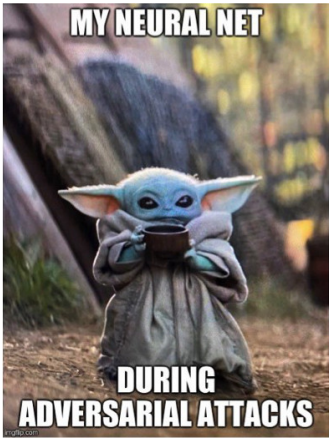
Subset Scanning for Out Of Distribution Detection

1. Can we detect when a sample looks **odd or new** to our ML model?
2. Can we detect if a sample is **generated** or **modified**?
3. Can we know **what** makes a sample odd or fake?

IBM Research | Africa

Why is important to detect adversarial attacks?

Reliably detecting attacks in a given set of inputs is of high practical relevance due to the vulnerability of neural networks to adversarial examples. These altered inputs create a security risk in applications with real-world consequences, such as self-driving cars, robotics and financial services.

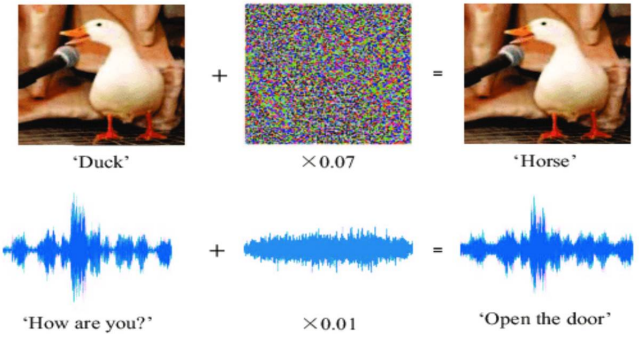


IBM Research | Africa

What is an Adversarial Attack?

white-box an attacker has complete access to the model, including its structure and trained weights. E.g. Basic Iterative Method (BIM) [KGB16], Fast Gradient Signal Method (FGSM) [GSS15], DeepFool (DF) [MDFF16].

black-box an attacker can only access the outputs of the target model. (e.g HopSkipJumpAttack [CJW19]).

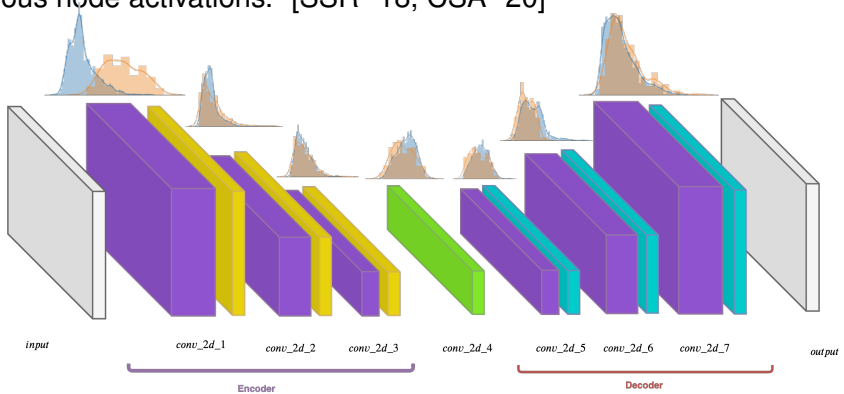


IBM Research | Africa

Picture: https://whataftercollege.com/machine-learning/adversarial-attack-machine-learning/

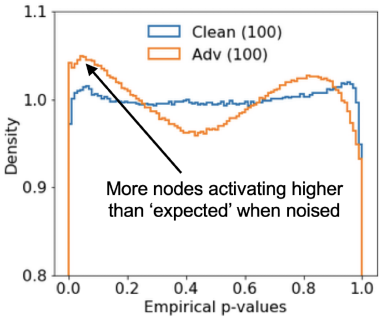
Subset Scan for Anomalous Pattern Detection

We propose an unsupervised method for detecting adversarial attacks in inner layers of autoencoder (AE) networks by maximizing a non-parametric measure of anomalous node activations. [SSR+18, CSA+20]



IBM Research | Africa

Subset Scanning for Anomalous Pattern Detection (Cont.)



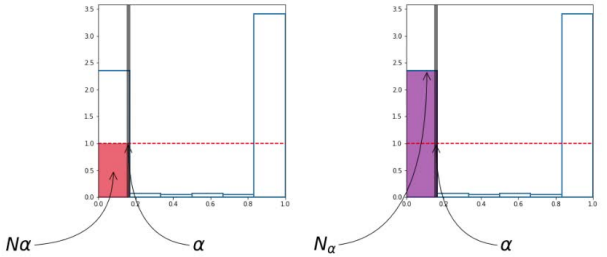
Assumption

Activations from adversarial images have a different distribution of p-values than benign/clean samples.

p-value is the proportion of background activations (H_0), drawn from the same node for several clean samples, greater than the activation from a test sample.

IBM Research | Africa

Subset Scanning for Anomalous Pattern Detection (Cont.)



$$\max_{\alpha} \varphi(\alpha, N_\alpha, N) = \frac{N_\alpha - N\alpha}{\sqrt{N}} \quad (1)$$

Where N_α is the number of p-values less than α
N is the number of p-values
α is the level of significance

How we score new images?

Scoring functions operate on an evaluation image in order to measure how much the p-values deviate from uniform.

IBM Research | Africa

Subset Scanning for Anomalous Pattern Detection (Cont.)

Subset scanning (SS) treats the pattern detection problem as a search for the *most anomalous* subset of activations.

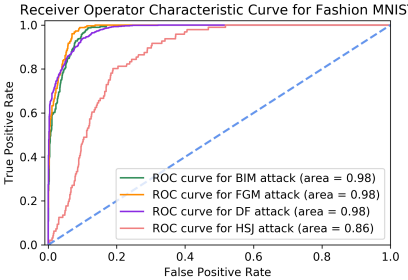
NPSS maximization

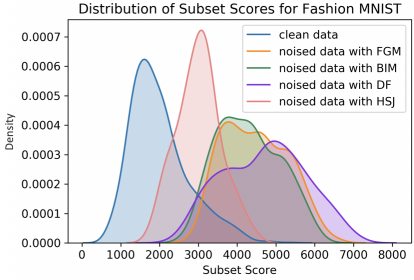
Scoring functions may be viewed as set functions that operate on subsets of nodes. We search for the highest scoring subset of nodes that maximize the deviance from uniform.

$$F(S) = \max_{\alpha} F_{\alpha}(S) = \max_{\alpha} \varphi(\alpha, N_{\alpha}(S), N(S)) \tag{2}$$

IBM Research | Africa

Results in Inner Layers





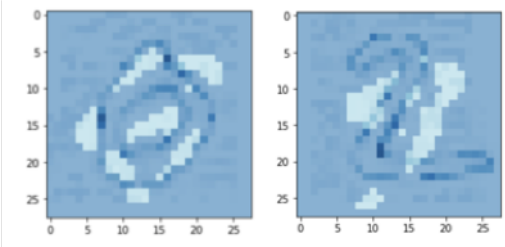
ROC curves & Distribution of subset scores

For each of the noised cases as compared to the scores from test sets containing all natural images for layer *Conv2d_1*. Distribution of subset scores for test sets of images over *Conv2d_1*. Clean images had lower scores than noised images.

IBM Research | Africa

Results over the Reconstruction Error

The results over the RE depend on the AE performance. If an autoencoder's loss is high, it is more difficult to separate between clean and noised samples in the reconstruction space because the most anomalous subset of reconstructed pixels of a clean image may be higher due to chance.



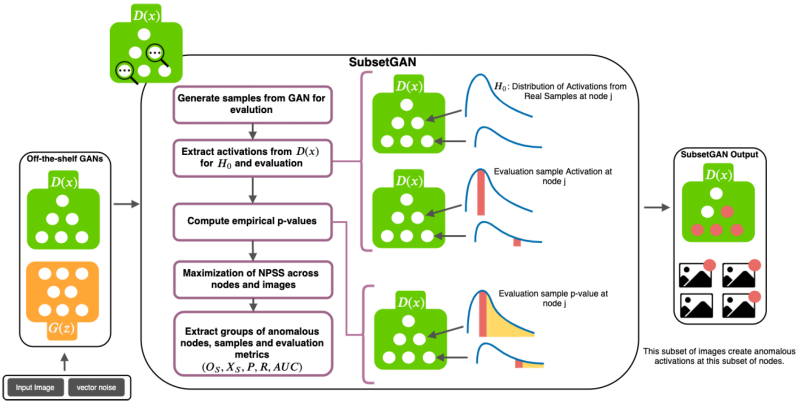
Datasets	Attacks	Detection Power (AUROC)		
		Ours RE	Mean RE	One-SVM
F-MNIST	BIM	**0.698**	0.641	0.478
	FGSM	**0.672**	0.630	0.497
	DF	**0.599**	0.477	0.534
	HSJ	**0.956**	0.935	0.546
MNIST	BIM	**0.998**	0.751	0.624
	FGSM	**0.983**	0.725	0.624
	DF	**0.992**	0.574	0.637
	HSJ	**0.999**	0.619	0.537

Explainability

Subset Scanning over the reconstruction error space is an interesting technique to inspect **which pixels** of the reconstructed image belong to the **most anomalous subset**.

IBM Research | Africa

Fake Content Detection as an Anomalous Pattern Problem



Sample generation from PGGAN & StarGAN [KALL17, CCK+18].

IBM Research | Africa

Further experiments & Open Questions on Robust detection of OOD

- Can we utilize the information from a subset of nodes to make models more **robust**?
- Can we **detect** when a **new class** appears at Inference time?
- Can we **inform how** a sample **is generated** with nodes visualization?

IBM Research | Africa

My two cents :)

1. It is crucial that the groups that develop & research technological solutions for sectors such as education, health, etc., are interdisciplinary.
2. Researchers and developers have to be as diverse than their end users. -or more so
3. The operational constraints and limitations of production models must be clearly and explicitly define.
4. The models to be used in production must make explicit in which context they work and their limitations, be transparent, clarify what biases were evaluated and what are the mitigation techniques used.

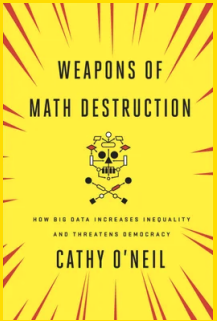
IBM Research | Africa

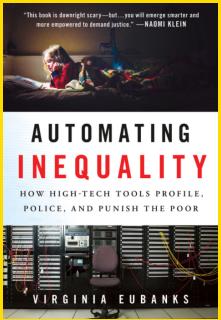
Interesting Resources and Materials

- Tutorials & Books
  1. Fairness in machine learning (NIPS 2017)
  2. 21 fairness definitions and their politics (FAT* 2018)
  3. Fairness and machine learning: Limitations and Opportunities (https://fairmlbook.org/) [BHN19]
- Lectures
  1. Berkeley CS 294: Fairness in machine learning (https://fairmlclass.github.io/)
  2. Cornell INFO 4270: Ethics and policy in data science
  3. Princeton COS 597E: Fairness in machine learning
- Conferences
  1. ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) (https://facctconference.org/)
- Initiatives and Communities
  1. MD4SG (http://md4sg.com/)
  2. Trustworthy ML (https://www.trustworthyml.org/)

IBM Research | Africa

Gracias! Thanks! Asante!





@RTFMCelia @ celia.cintas@ibm.com

IBM Research | Africa

References I

- Joy Buolamwini and Timnit Gebru, *Gender shades: Intersectional accuracy disparities in commercial gender classification*, Proc. Conf. Fair. Account. Transp., February 2018, pp. 77–91.
- Solon Barocas, Moritz Hardt, and Arvind Narayanan, *Fairness and machine learning*, fairmlbook.org, 2019, http://www.fairmlbook.org.
- Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo, *Stargan: Unified generative adversarial networks for multi-domain image-to-image translation*, IEEE CVPR, June 2018.
- Jianbo Chen, Michael I Jordan, and Martin J Wainwright, *Hopskipjumpattack: A query-efficient decision-based attack*, arXiv preprint arXiv:1904.02144 3 (2019).

IBM Research | Africa

References II

- Noel C. F. Codella, Quoc-Bao Nguyen, Sharath Pankanti, David A. Gutman, Brian Helba, Allan C. Halpern, and John R. Smith, *Deep learning ensembles for melanoma recognition in dermoscopy images*, IBM J. Res. Dev. 61 (2016), no. 4/5, 5.
- Celia Cintas, Ramya Raghavendra, Victor Akinwande, Aisha Walcott-Bryant, Charity Wayua, and Komminist Weldemariam, *Decision platform for pattern discovery and causal effect estimation in contraceptive discontinuation*, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence Demos (2020), 5288–5290.

IBM Research | Africa

References III

- Celia Cintas, Skyler Speakman, Victor Akinwande, William Ogallo, Komminist Weldemariam, Srihari Sridharan, and Edward McFowland, *Detecting adversarial attacks via subset scanning of autoencoder activations and reconstruction error*, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence Main Track (2020), 876–882.
- Giuseppe R. Casale, Anna Maria Siani, Henri Diémoz, Giovanni Agnesod, Alfio V. Parisi, and Alfredo Colosimo, *Extreme UV index and solar exposures at Plateau Rosà (3500 m a.s.l.) in Valle d'Aosta Region, Italy*, Sci. Total Environ. 512–513 (2015), 622–630.
- Diana Diaz, Celia Cintas, William Ogallo, and Aisha Walcott-Bryant, *Towards automatic generation of context-based abstractive discharge summaries for supporting transition of care*.

IBM Research | Africa

References IV

- Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, *Explaining and harnessing adversarial examples*, CoRR abs/1412.6572 (2015).
- Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B Girshick, *Mask r-cnn. corr abs/1703.06870 (2017)*, arXiv preprint arXiv:1703.06870 (2017).
- Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, *Densely connected convolutional networks*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
- Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, *Progressive growing of gans for improved quality, stability, and variation*, arXiv preprint arXiv:1710.10196 (2017).

IBM Research | Africa

References V

- Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio, *Adversarial examples in the physical world*, CoRR abs/1607.02533 (2016).
- Newton M Kinyanjui, Timothy Odonga, Celia Cintas, Noel CF Codella, Rameswar Panda, Prasanna Sattigeri, and Kush R Varshney, *Fairness of classifiers across skin tones in dermatology*, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 320–329.
- JC Lester, JL Jia, L Zhang, GA Okoye, and E Linos, *Absence of skin of colour images in publications of covid-19 skin manifestations*, British Journal of Dermatology (2020).
- Michael A. Marchetti, Esther Chung, and Allan C. Halpern, *Screening for acral lentiginous melanoma in dark-skinned individuals*, JAMA Dermatol. 151 (2015), no. 10, 1055–1056.

IBM Research | Africa

References VI

- Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard, *Deepfool: a simple and accurate method to fool deep neural networks*, Proceedings of the IEEE CVPR'16, 2016, pp. 2574–2582.
- Krishnaraj Mahendraraj, Komal Sidhu, Christine S. M. Lau, Georgia J. McRoy, Ronald S. Chamberlain, and Franz O. Smith, *Malignant melanoma in African–Americans: A population-based clinical outcomes study involving 1106 African–American patients from the surveillance, epidemiology, and end result (SEER) database (1988—2011)*, Medicine 96 (2017), no. 15, e6258.
- Roni Caryn Rabin, *Dermatology has a problem with skin color*, Aug 2020.
- Skyler Speakman, Srihari Sridharan, Sekou Remy, Komminist Weldemariam, and Edward McFowland, *Subset scanning over neural network activations*, arXiv preprint arXiv:1810.08676 (2018).

IBM Research | Africa

References VII

- Xiao-Cheng Wu, Melody J. Eide, Jessica King, Mona Saraiya, Youjie Huang, Charles Wiggins, Jill S. Barnholtz-Sloan, Nicolle Martin, Vilma Cokkinides, Jacqueline Miller, Pragna Patel, Donatus U. Ekwueme, and Julian Kim, *Racial and ethnic variations in incidence and survival of cutaneous melanoma in the United States, 1999-2006*, J. Am. Acad. Dermatol. 65 (2011), no. 5, S26.e1–S26.e13.
- Benjamin Wilson, Judy Hoffman, and Jamie Morgenstern, *Predictive inequity in object detection*, arXiv:1902.11097, February 2019.
- Marcus Wilkes, Caradee Y. Wright, Johan L. du Plessis, and Anthony Reeder, *Fitzpatrick skin type, individual typology angle, and melanin index in an African population*, JAMA Dermatol. 151 (2015), no. 8, 902–903.

IBM Research | Africa