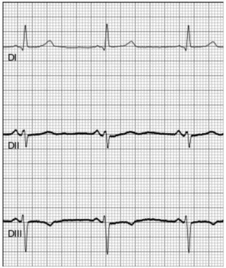
The Three Challenges of Using Deep Neural Networks in Electrocardiography

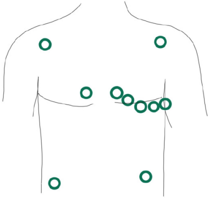
Antônio Horta Ribeiro
Uppsala University

Online series: *Novel insights & applications in AI ECG interpretation*
Germany Chapter of IEEE Engineering in Medicine and Biology Society
24th of May 2023

Outline

The electrocardiogram (ECG) exam





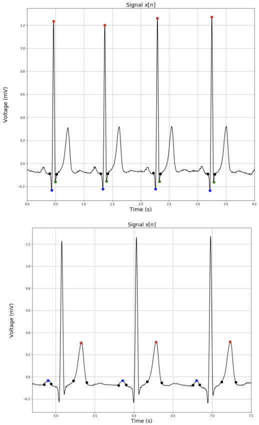
Left: ECG signal **Right:** Electrode placement.

Computational electrocardiography

Figure: Automated ECG interpretation Glasgow Royal Infirmary (1971).

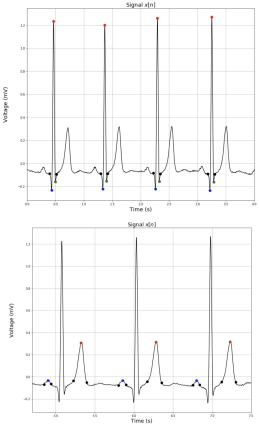
Macfarlane, P.W.; Kennedy, J. "Automated ECG Interpretation—A Brief History from High Expectations to Deepest Networks." Hearts 2021.

Classical ECG automated analysis



ECG segmented using signal processing.
Top: QRS complex. **Bottom:** T and P waves.

Classical ECG automated analysis



ECG segmented using signal processing.
Top: QRS complex. **Bottom:** T and P waves.

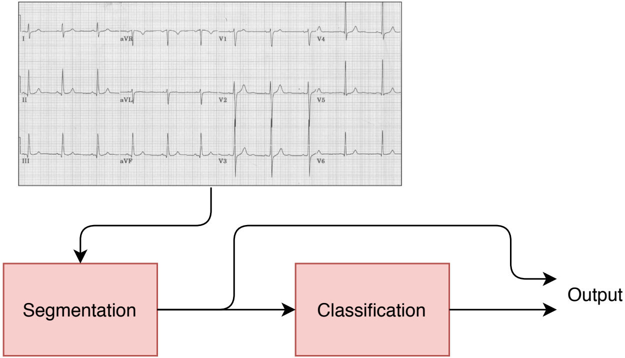
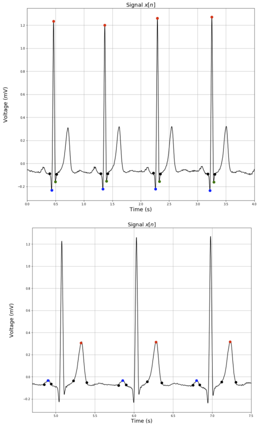


Figure: Two step analysis procedure

Classical ECG automated analysis



ECG segmented using signal processing.
Top: QRS complex. **Bottom:** T and P waves.

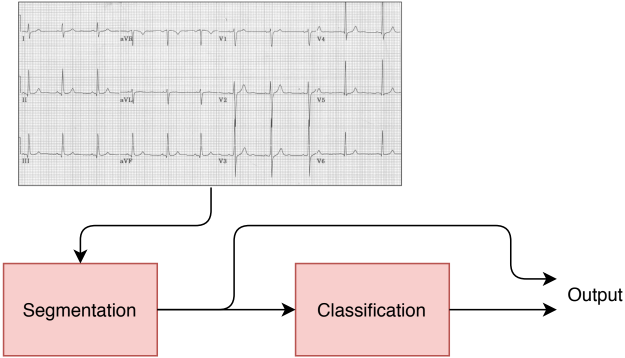
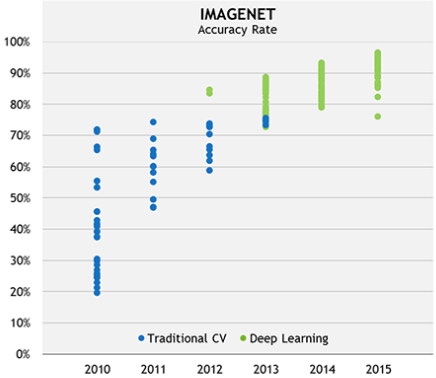


Figure: Two step analysis procedure

P. W. Macfarlane, B. Devine, and E. Clark, "The university of glasgow (Uni-G) ECG analysis program," in Computers in Cardiology, 2005, pp. 451–454. doi: 10.1109/CIC.2005.1588134.

Image classification with deep neural networks

Imagenet



Left: dataset samples. **Right:** Models accuracy on the benchmark.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248–255.

Outline

Deep Neural Networks in Electrocardiography

1. *Can we improve on existing automatic tools?*

   A. H. Ribeiro, M.H. Ribeiro, Paixão, G.M.M., D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr., T. B. Schön, A. L. P. Ribeiro "Automatic diagnosis of the 12-lead ECG using a deep neural network," Nature Communications, 2020

2. *Can we learn to detect new conditions and discover new biomarkers?*

   Stefan Gustafsson, Daniel Gedon, Erik Lampa, **Antônio H. Ribeiro**, Martin J. Holzmann, Thomas B. Schön, Johan Sundström. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. Scientific Reports, 2022.

Telehealth and automatic tools

- The Telehealth Center of Minas Gerais

Telehealth and automatic tools

- The Telehealth Center of Minas Gerais
- More than 4000 ECGs per day

Telehealth and automatic tools

- The Telehealth Center of Minas Gerais
- More than 4000 ECGs per day
- Automatic ECG analysis software used as auxiliary tool

Telehealth and automatic tools

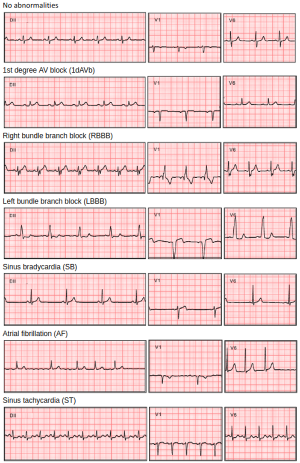
- The Telehealth Center of Minas Gerais
- More than 4000 ECGs per day
- Automatic ECG analysis software used as auxiliary tool
- Better accuracy could enable additional uses

Deep learning for automatic diagnosis

- CODE dataset: historical data 2010 to 2017.
  - 2.3 M ECGs from n =1.6M patients
  - 6 abnormalites

Deep learning for automatic diagnosis

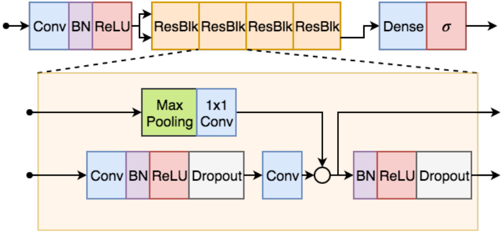
- CODE dataset: historical data 2010 to 2017.
    - 2.3 M ECGs from n =1.6M patients
    - 6 abnormalites
- CODE-test: annotate by 3 cardiologists (n=827).

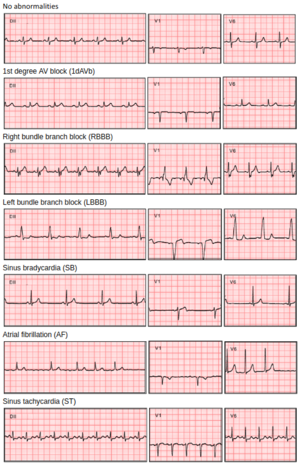


A. H. Ribeiro , M.H. Ribeiro, Paixão, G.M.M., et al. "Automatic diagnosis of the 12-lead ECG using a deep neural network," Nature Communications, 2020

Deep learning for automatic diagnosis

- CODE dataset: historical data 2010 to 2017.
  - 2.3 M ECGs from n =1.6M patients
  - 6 abnormalites
- CODE-test: annotate by 3 cardiologists (n=827).
- Develop and evaluate deep neural network





A. H. Ribeiro , M.H. Ribeiro, Paixão, G.M.M., et al. "Automatic diagnosis of the 12-lead ECG using a deep neural network," Nature Communications, 2020

Results

	F1 Score			
	DNN	cardio.	emerg.	stud.
1dAVb	**0.897**	0.776	0.719	0.732
RBBB	**0.944**	0.917	0.852	0.928
LBBB	**1.000**	0.947	0.912	0.915
SB	**0.882**	**0.882**	0.848	0.750
AF	**0.870**	0.769	0.696	0.706
ST	**0.960**	0.882	0.946	0.873

Performance of the deep neural network (DNN) agains experts.

cardio. $\rightarrow$ 4th year cardiology residents

emerg. $\rightarrow$ 3rd year emergency residents

stud. $\rightarrow$ 5th year Medical students

Myocardial infarction

Can we detect non-STEMI cases from the ECG?

Myocardial infarction

Can we detect non-STEMI cases from the ECG?

Diagnosis:

- ST-elevation MI (STEMI) $\rightarrow$ ECG.
- non-STEMI $\rightarrow$ blood testing

Left: No MI. **Right:** STEMI.

Stefan Gustafsson, Daniel Gedon, Erik Lampa, **Antônio H. Ribeiro**, Martin J. Holzmann, Thomas B. Schön, Johan Sundström. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. Scientific Reports, 2022.

Myocardial infarction

Can we detect non-STEMI cases from the ECG?

Left: No MI. **Right:** STEMI.

Diagnosis:

- ST-elevation MI (STEMI) → ECG.
- non-STEMI → blood testing

Dataset:

- Stockholm emergency: 2007 to 2016
- 400 000 ECGs
- NSTEMI ≈ 1%. STEMI ≈ 0.5%
- MI annotated using blood testing.

Stefan Gustafsson, Daniel Gedon, Erik Lampa, **Antônio H. Ribeiro**, Martin J. Holzmann, Thomas B. Schön, Johan Sundström. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. Scientific Reports, 2022.

Myocardial infarction

Can we detect non-STEMI cases from the ECG?

Diagnosis:

- ST-elevation MI (STEMI) $\rightarrow$ ECG.
- non-STEMI $\rightarrow$ blood testing

Dataset:

- Stockholm emergency: 2007 to 2016
- 400 000 ECGs
- NSTEMI $\approx$ 1%. STEMI $\approx$ 0.5%
- MI annotated using blood testing.

Left: No MI. **Right:** STEMI.

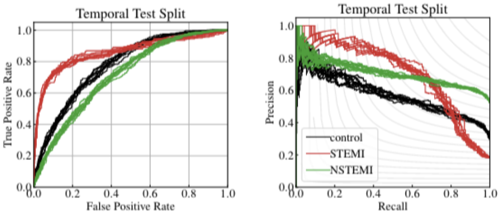


Figure: ROC and precisicion-recall curves AUROC = 0.82 / 0.74 (STEMI/NSTEMI).

Stefan Gustafsson, Daniel Gedon, Erik Lampa, **Antônio H. Ribeiro**, Martin J. Holzmann, Thomas B. Schön, Johan Sundström. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. Scientific Reports, 2022.

Data-driven methods to detect new conditions

- Study of bio-markers from ECG to detect mortality

  E. M. Lima*, **A. H. Ribeiro***, G. M. M. Paixão et al., "Deep neural network estimated electrocardiographic-age as a mortality predictor," Nature Communications, vol. 12, 2021 ***Equal contribution**

Data-driven methods to detect new conditions

- Study of bio-markers from ECG to detect mortality

  E. M. Lima*, **A. H. Ribeiro***, G. M. M. Paixão et al., "Deep neural network estimated electrocardiographic-age as a mortality predictor," Nature Communications, vol. 12, 2021 ***Equal contribution**

- Predicting risk of developing Atrial Fibrilation

  S. Biton, S. Biton, **A. H. Ribeiro**, et al., "Atrial fibrillation risk prediction from the 12-lead ECG using digital biomarkers and deep representation learning," European Heart Journal - Digital Health, 2021

Data-driven methods to detect new conditions

- Study of bio-markers from ECG to detect mortality

  E. M. Lima*, **A. H. Ribeiro***, G. M. M. Paixão et al., "Deep neural network estimated electrocardiographic-age as a mortality predictor," Nature Communications, vol. 12, 2021 ***Equal contribution**

- Predicting risk of developing Atrial Fibrilation

  S. Biton, S. Biton, **A. H. Ribeiro**, et al., "Atrial fibrillation risk prediction from the 12-lead ECG using digital biomarkers and deep representation learning," European Heart Journal - Digital Health, 2021

- Screening Chagas Diasease.

  Carl Jidling, Daniel Gedon, Thomas B. Schön, Claudia Di Lorenzo Oliveira, Clareci Silva Cardos, Ariela Mota Ferreira, Luana Giatti, Sandhi Maria Barreto, Ester C. Sabino, Antônio L. P. Ribeiro, **Antônio H. Ribeiro** "Screening for Chagas disease from the electrocardiogram using a deep neural network" Submitted to Plos Neglected Tropical Diseases, 2023 (preprint: medRxiv).

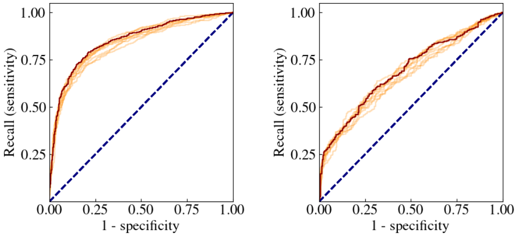


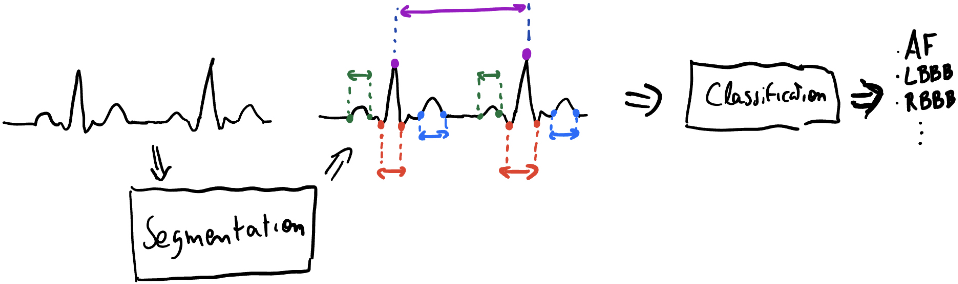
Figure: Performance on predicting Chagas Disease.

The three challenges

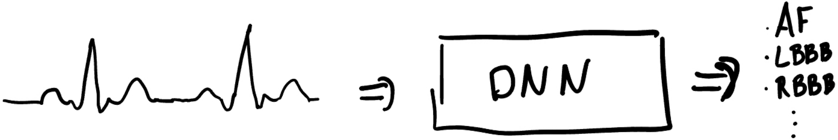
1. Interpretability
2. Robustness
3. Integration into medical practice

Outline

Classical

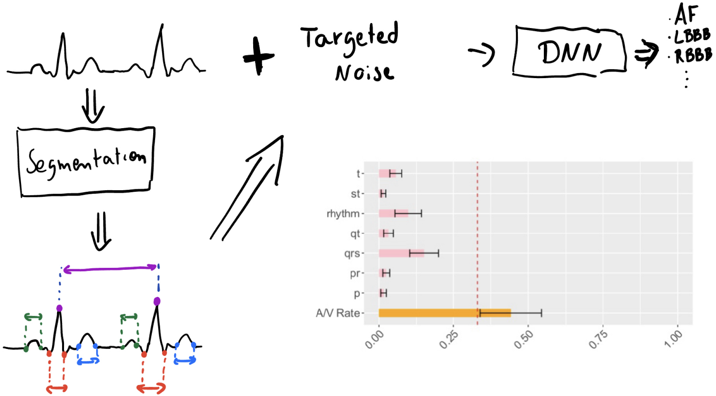


End-to-end



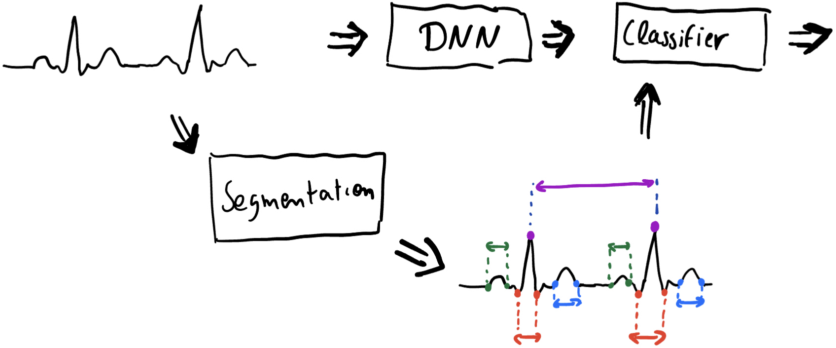
Importance of different segments on DNN prediction

- Add noise to a specific segment of the ECG.
- Use the error to assess the importance of the segments



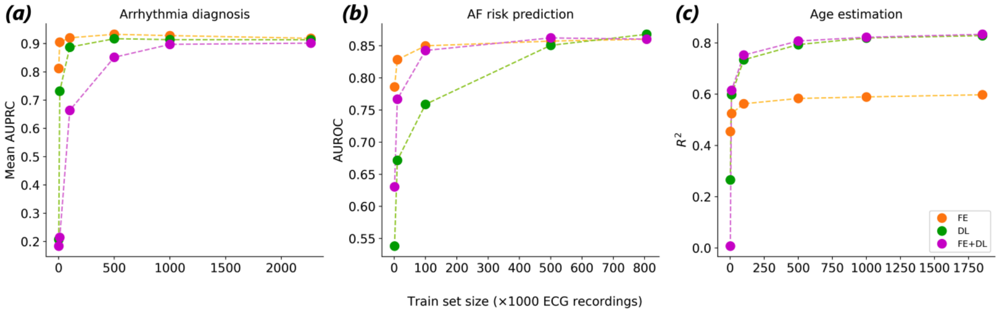
Derick M. Oliveira, **Antônio H. Ribeiro**, João A. O. Pedrosa, Gabriela M.M. Paixao, Antonio Luiz P. Ribeiro, Wagner Meira Jr. "Explaining end-to-end ECG automated diagnosis using contextual features" ECML-PKDD (2020).

Merging engineered features with deep neural networks



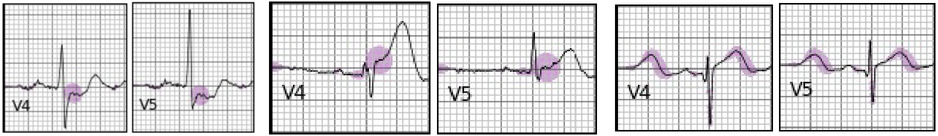
Eran Zvuloni, Jesse Read, **Antônio H. Ribeiro**, Antonio Luiz P. Ribeiro, Joachim A. Behar. "On Merging Feature Engineering and Deep Learning for Diagnosis, Risk-Prediction and Age Estimation Based on the 12-Lead ECG." IEEE Transactions on Biomedical Engineering, 2023.

Merging engineered features with deep neural networks



Eran Zvuloni, Jesse Read, **Antônio H. Ribeiro**, Antonio Luiz P. Ribeiro, Joachim A. Behar. "On Merging Feature Engineering and Deep Learning for Diagnosis, Risk-Prediction and Age Estimation Based on the 12-Lead ECG." IEEE Transactions on Biomedical Engineering, 2023.

Saliency maps



Grad-CAM plots. **(Left)** STEMI with typical ST-segment elevation highlighted. **(Middle)** STEMI with another feature highlighted that is not typical to doctors (establishing the relevance of it would need additional study). **(Right)** NSTEMI with typical but unspecific ST-segment depression highlighted.

Stefan Gustafsson, Daniel Gedon, Erik Lampa, **Antônio H. Ribeiro**, Martin J. Holzmann, Thomas B. Schön, Johan Sundström. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. Scientific Reports, 2022.

Limitations

1. Contextual features

Limitations

1. Contextual features
   - Difficult to generate targeted noise in a meaningful way.

Limitations

1. Contextual features
   - Difficult to generate targeted noise in a meaningful way.
   - Segmentation is a bottleneck.

Limitations

1. Contextual features
    - Difficult to generate targeted noise in a meaningful way.
    - Segmentation is a bottleneck.
2. Saliency maps

Limitations

1. Contextual features
   - Difficult to generate targeted noise in a meaningful way.
   - Segmentation is a bottleneck.
2. Saliency maps
   - The visualization method is also a black box.

Limitations

1. Contextual features
   - Difficult to generate targeted noise in a meaningful way.
   - Segmentation is a bottleneck.
2. Saliency maps
   - The visualization method is also a black box.
   - It is hard to test the method itself is working or not.

Limitations

1. Contextual features
   - Difficult to generate targeted noise in a meaningful way.
   - Segmentation is a bottleneck.
2. Saliency maps
   - The visualization method is also a black box.
   - It is hard to test the method itself is working or not.
   - Manual process.

Interpretability

- The telehealth effort to structure ECG reading of their doctors.

Interpretability

- The telehealth effort to structure ECG reading of their doctors.
- Some tasks cannot be fully explainable

Interpretability

- The telehealth effort to structure ECG reading of their doctors.
- Some tasks cannot be fully explainable
- Important to translate AI-ECG effort into real scientific knowledge.

Interpretability

- The telehealth effort to structure ECG reading of their doctors.
- Some tasks cannot be fully explainable
- Important to translate AI-ECG effort into real scientific knowledge.
- Methods do not need to be interpretable to be useful. But they need to be robust!

Outline

Robustness: *refers to the ability of a model or method to perform well in the presence of various types of disturbances, errors or violations of assumptions.*

Robustness: *refers to the ability of a model or method to perform well in the presence of various types of disturbances, errors or violations of assumptions.*

Goal: *Obtain models that perform well across different types of equipment and populations with different socio-economic conditions, genetic material and different comorbidities.*

Adversarial examples

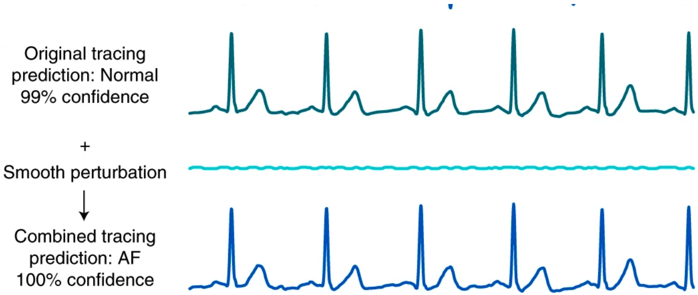


Figure: Effect of adversarial examples on ECG Classification.

Source: Han, X., Hu, Y., Foschini, L. et al. Deep learning models for electrocardiograms are susceptible to adversarial attack. Nature Medicine 26, 360–363 (2020).

Spurious features

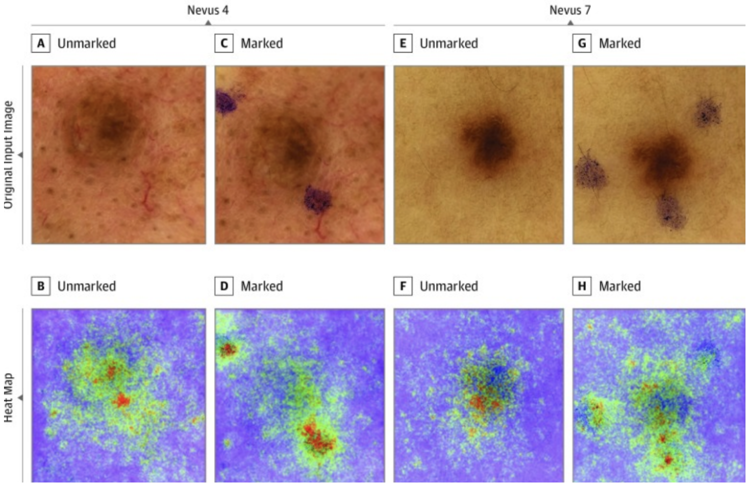


Figure: Learning unwanted features.

Source: Julia K. Winkler, Christine Fink, et al. "Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition". JAMA Dermatology (2019).

How to define robustness?

- Adversarial robustness (i.e., worst-case scenario analysis)

$$\text{Adv. error} = \max_{\|\Delta x\| \leq \delta} \ell(y, f(x + \Delta x))$$

How to define robustness?

- Adversarial robustness (i.e., worst-case scenario analysis)

$$\text{Adv. error} = \max_{\|\Delta x\| \leq \delta} \ell(y, f(x + \Delta x))$$

- Robustness to Natural disturbances.

How to define robustness?

- Adversarial robustness (i.e., worst-case scenario analysis)

$$\text{Adv. error} = \max_{\|\Delta x\| \leq \delta} \ell(y, f(x + \Delta x))$$

- Robustness to Natural disturbances.
  - Evaluation on robustness benchmarks (i.e., Imagenet C/P)

How to define robustness?

- Adversarial robustness (i.e., worst-case scenario analysis)

$$\text{Adv. error} = \max_{\|\Delta x\| \leq \delta} \ell(y, f(x + \Delta x))$$

- Robustness to Natural disturbances.
  - Evaluation on robustness benchmarks (i.e., Imagenet C/P)
  - Often lack of formal definitions!

    N. Drenkow, N. Sani, I. Shpitser, and M. Unberath, "A Systematic Review of Robustness in Deep Learning for Computer Vision: Mind the gap?" arXiv: 10.48550/arXiv.2112.00639.

Out-of-distribution generalization

Goal, obtain a small generalization gap:

$$\mathbb{E}_{(x,y)\sim\widetilde{\mathbb{P}}}\left[\ell(y, f(x))\right] - \mathbb{E}_{(x,y)\sim\mathbb{P}}\left[\ell(y, f(x))\right].$$

- Limitation: *No-free-lunch theorems*
  S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.
- Pitfalls of simple notions of distance between distributions (KL divergence).

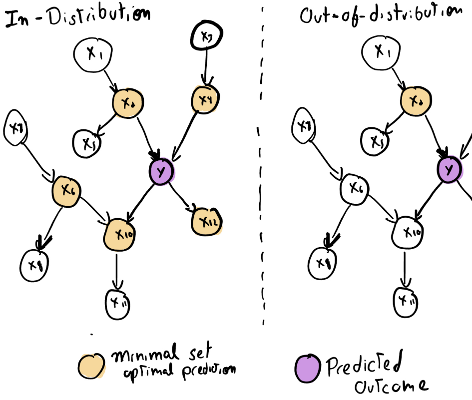
Conterfactual robustness: The ability of model keeping performance on data generated from a counterfactually-altered version of the data generating process.

Robustness from a causal perspective

Causal structure optimize the the problem

$$\min_{f} \max_{\widetilde{\mathbb{P}} \text{ invariant}} \mathbb{E}_{(x,y)\sim\widetilde{\mathbb{P}}}\left[y - f(x)\right] \rightarrow f = \mathbb{E}\left[y | x_{Pa(y)}\right]$$

M. Rojas-Carulla, B. Scholkopf, R. Turner, and J. Peters, "Invariant Models for Causal Transfer Learning," Journal of Machine Learning Research, no. 19, pp. 1–34, 2018.



Some methods for robust design

1. Adversarial training

   A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards Deep Learning Models Resistant to Adversarial Attacks," International Conference for Learning Representations (ICLR), 2018.

2. Disentangle representation learning (i.e., β-VAE)

   Higgins I , Matthey L, Pal A, Burgess CP, Glorot X, Botvinick M, Mohamed S, Lerchner A. Beta-VAE: learning basic visual concepts with a constrained variational framework. In: 5th International Conference on Learning Representations. Toulon, France: Conference Track Proceedings; 2017

   Rutger R van de Leur, et al., Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders, European Heart Journal - Digital Health, Volume 3, Issue 3, September 2022, Pages 390–404

3. Causal learning

   J. Peters, D. Janzing, and B. Schölkopf, Elements of causal inference: foundations and learning algorithms. 2017.

   Castro, D.C., Walker, I. and Glocker, B. Causality matters in medical imaging. Nat Commun 11, 3673, 2020.

Outline

Machine learning and telehealth

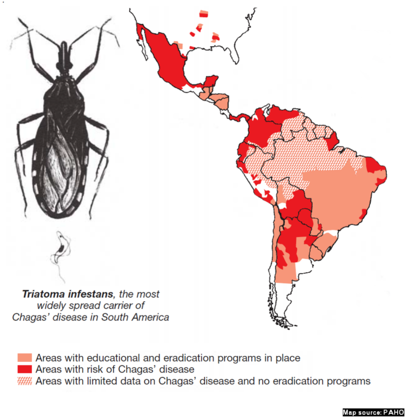
Better automatic tools could:

- Queue the exams more effectively.
- Be used only on exams where it is very certain.
- Detect misdiagnosis.

Screening for diseases

Example: screening for Chagas disease

- 6 million people infected.
- Diagnosed with blood test.
- early diagnosis and treatment halt progression.
- Low detection rates



Carl Jidling, Daniel Gedon, Thomas B. Schön, Claudia Di Lorenzo Oliveira, Clareci Silva Cardos, Ariela Mota Ferreira, Luana Giatti, Sandhi Maria Barreto, Ester C. Sabino, Antônio L. P. Ribeiro, **Antônio H. Ribeiro** "Screening for Chagas disease from the electrocardiogram using a deep neural network" Submitted to Plos Neglected Tropical Diseases, 2023 (preprint: medRxiv).

Summary

- Deep neural networks for automatic ECG analysis
  - Improve existing analysis tools
  - Expand the potential of the exam (screen for diseases and find new biomarkers)
- Challenges
  - Interpretability.
  - Robustness.
  - Integration into medical practice.

Thank you!

✉ antonio.horta.ribeiro@it.uu.se
🌐 antonior92.github.io