

Inteligência artificial e biometria

Paulo Henrique Pisani
paulo.pisani@ufabc.edu.br

10/julho/2024

Tópicos

- Inteligência artificial, Aprendizado de máquina;
- Biometria;
- IA no contexto de Biometria;
- Oportunidades de pesquisa e tópicos futuros.

Inteligência Artificial



You

O que é Inteligência Artificial?



ChatGPT

A Inteligência Artificial (IA) refere-se à capacidade das máquinas de realizar tarefas que normalmente exigiriam inteligência humana. Isso inclui a capacidade de aprender com experiências passadas, reconhecer padrões, tomar decisões, resolver problemas, compreender linguagem natural e até mesmo interagir de forma natural com os seres humanos.

Existem diferentes abordagens e técnicas dentro da IA, incluindo:

1. **Aprendizado de Máquina (Machine Learning):** Um subcampo da IA que se concentra no desenvolvimento de algoritmos que permitem aos computadores aprender com dados e tomar decisões com base nesse aprendizado, sem programação explícita.
2. **Redes Neurais Artificiais:** Modelos computacionais inspirados no funcionamento do cérebro humano, que são usados em muitas aplicações de aprendizado profundo (deep learning).
3. **Processamento de Linguagem Natural (NLP):** Uma área da IA que se concentra na interação entre computadores e humanos por meio da linguagem natural, permitindo a compreensão e a geração de texto e fala.
4. **Visão Computacional:** Capacidade dos computadores de interpretar e entender o conteúdo visual, como imagens e vídeos.
5. **Agentes Inteligentes:** Sistemas de IA que percebem seu ambiente e tomam ações para atingir objetivos, muitas vezes utilizando técnicas como planejamento e raciocínio.

A IA está se tornando cada vez mais presente em nossa vida cotidiana, sendo aplicada em uma ampla gama de setores, desde saúde e finanças até transporte e entretenimento. O objetivo final da IA é criar sistemas capazes de realizar tarefas de forma autônoma, adaptativa e inteligente, em um nível comparável ou até mesmo superior ao humano.



You

O que é Inteligência Artificial?



ChatGPT

A Inteligência Artificial (IA) refere-se à capacidade das máquinas de realizar tarefas que normalmente exigiriam inteligência humana. Isso inclui a capacidade de aprender com experiências passadas, reconhecer padrões, tomar decisões, resolver problemas, compreender linguagem natural e até mesmo interagir de forma natural com os seres humanos.

Existem diferentes abordagens e técnicas dentro da IA, incluindo:

1. Aprendizado de Máquina (Machine Learning): Um subcampo da IA que se concentra no desenvolvimento de algoritmos que permitem aos computadores aprender com dados e tomar decisões com base nesse aprendizado, sem programação explícita.

Exemplo

(<https://teachablemachine.withgoogle.com/>)

The screenshot displays the Teachable Machine web interface. At the top, there is a navigation menu with the 'Teachable Machine' logo. Below this, two class cards are visible: 'Carro' and 'Casa'. Each class card shows '2 Image Samples' and includes 'Webcam' and 'Upload' buttons. A 'Training' panel is positioned to the right of the class cards, featuring a 'Train Model' button and a dropdown menu set to 'Advanced'. To the right of the training panel is a 'Preview' panel with an 'Export Model' button and a message: 'You must train a model on the left before you can preview it here.' At the bottom of the interface, there is a dashed box labeled 'Add a class'. The browser's address bar shows the URL 'https://teachablemachine.withgoogle.com/train/image'. The footer includes a language selector set to 'English' and a version string 'release-2-4-5 - 2.4.5#18d7c1'.

Carros: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/buick-oldtimer-old-car-blue-car-1400243/>
<https://pixabay.com/photos/automobile-cayman-coupe-design-5330343/>

Casas: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/house-building-lonely-house-3386450/>
<https://pixabay.com/photos/house-garage-driveway-architecture-1867187/>

Exemplo

(<https://teachablemachine.withgoogle.com/>)

The screenshot displays the Teachable Machine web interface. The browser address bar shows the URL <https://teachablemachine.withgoogle.com/train/image>. The main interface is divided into several sections:

- Classes:** Two classes are defined: "Carro" and "Casa". Each class has two image samples and options for "Webcam" and "Upload". The "Carro" class samples are a blue Buick Oldtimer and a Cayman coupe. The "Casa" class samples are a house with a driveway and a house with a garage.
- Training:** A "Training" panel shows the status "Model Trained" and an "Advanced" dropdown menu.
- Preview:** A "Preview" panel includes an "Export Model" button, an "Input" toggle set to "ON", and a "File" dropdown menu. Below these are two large blue boxes: "Choose images from your files, or drag & drop here" and "Import images from Google Drive".
- Output:** An "Output" panel shows two colored boxes: an orange box labeled "Carro" and a pink box labeled "Casa".
- Footer:** A dashed box at the bottom left contains the text "Add a class". The bottom right corner shows the language set to "English" and the version "release-2-4-5 - 2.4.5#18d7c1".

Carros: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/buick-oldtimer-old-car-blue-car-1400243/>
<https://pixabay.com/photos/automobile-cayman-coupe-design-5330343/>

Casas: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/house-building-lonely-house-3386450/>
<https://pixabay.com/photos/house-garage-driveway-architecture-1867187/>

Exemplo

(<https://teachablemachine.withgoogle.com/>)

The screenshot displays the Teachable Machine web interface. On the left, two classes are defined: 'Carro' and 'Casa', each with two image samples. A 'Training' panel in the center shows 'Model Trained' and 'Advanced' options. On the right, the 'Preview' panel shows the model's output for a house image, with 'Casa' identified at 99% confidence. A blue box highlights the preview area.

Carro

2 Image Samples

Webcam Upload

Casa

2 Image Samples

Webcam Upload

Add a class

Training

Model Trained

Advanced

Preview Export Model

Input ON File

Choose images from your files, or drag & drop here

Import images from Google Drive

Output

Carro

Casa 99%

Carros: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/buick-oldtimer-old-car-blue-car-1400243/>
<https://pixabay.com/photos/automobile-cayman-coupe-design-5330343/>
<https://pixabay.com/photos/car-beetle-volkswagen-blue-car-1283947/>

Casas: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/house-building-lonely-house-3386450/>
<https://pixabay.com/photos/house-garage-driveway-architecture-1867187/>
<https://pixabay.com/photos/iceland-landscape-nature-grass-hut-219182/>

Exemplo

(<https://teachablemachine.withgoogle.com/>)

The screenshot displays the Teachable Machine web interface. On the left, two classes are defined: 'Carro' and 'Casa', each with two image samples. The 'Carro' class includes a blue vintage car and a green car, while the 'Casa' class includes two different house images. A 'Training' panel indicates the model is trained. On the right, the 'Preview' panel shows a selected image of a blue car, which is highlighted with a blue border. Below the image, the 'Output' section shows a bar chart with 'Carro' at 98% and 'Casa' at a lower percentage.

Carros: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/buick-oldtimer-old-car-blue-car-1400243/>
<https://pixabay.com/photos/automobile-cayman-coupe-design-5330343/>
<https://pixabay.com/photos/car-beetle-volkswagen-blue-car-1283947/>

Casas: Imagens usadas de <https://pixabay.com/>:
<https://pixabay.com/photos/house-building-lonely-house-3386450/>
<https://pixabay.com/photos/house-garage-driveway-architecture-1867187/>
<https://pixabay.com/photos/iceland-landscape-nature-grass-hut-219182/>

O que é Aprendizado de Máquina?

- *"Machine learning is a branch of artificial intelligence that enables algorithms to automatically learn from data without being explicitly programmed. Its practitioners train algorithms to identify patterns in data and to make decisions with minimal human intervention."*

<https://www.deeplearning.ai/courses/machine-learning-specialization/>
(Frequently Asked Questions – What is machine learning?)

Treinamento e teste



Carro



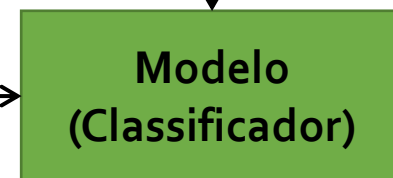
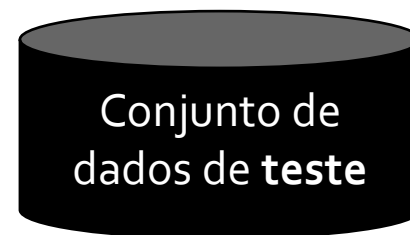
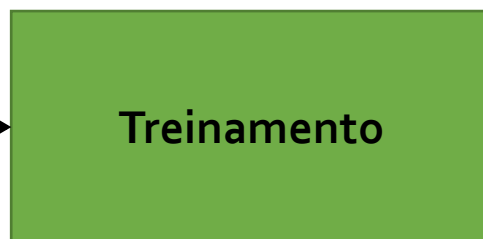
Carro



Casa



Casa



Casa
Carro



Aprendizado de máquina

- Algumas categorias:
 - Supervisionado;
 - Não-supervisionado;
 - Aprendizado por reforço.

Biometria

Biometria

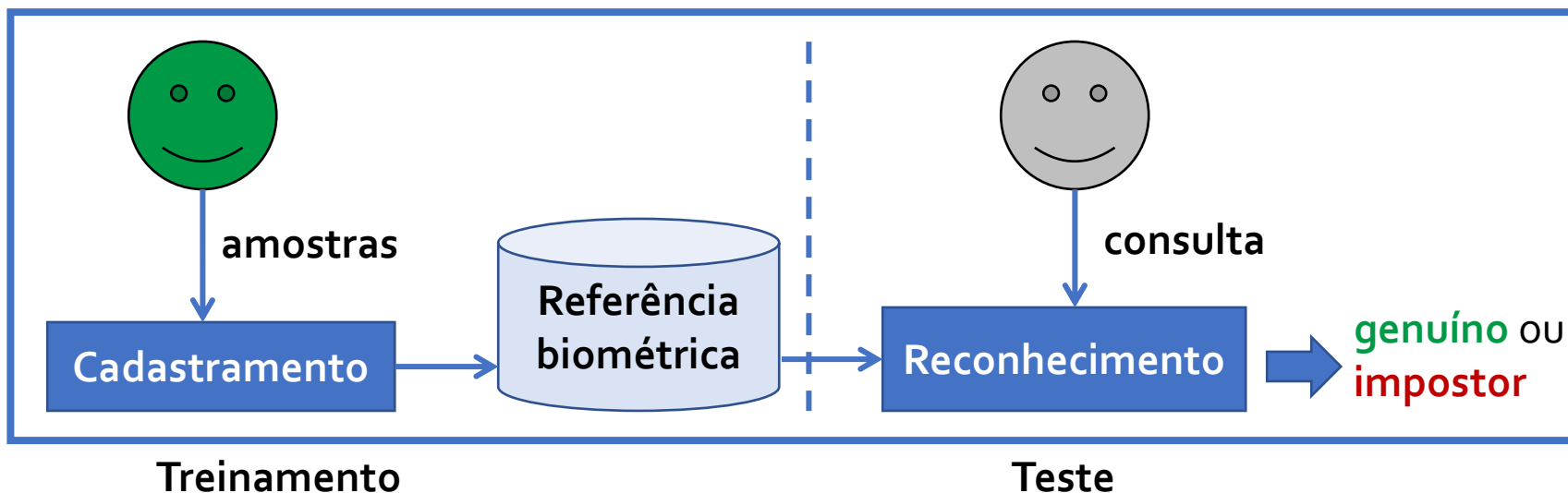
- Reconhecimento automático de indivíduos por meio de suas características biológicas ou comportamentais (JAIN et al., 2016).
- Modalidades:
 - **Físicas:** impressão digital, íris, geometria de face, veias da palma da mão, etc.
 - **Comportamentais:** assinatura, dinâmica da digitação, forma de caminhar, etc.

Aplicações

- Autenticação/reconhecimento de usuários,
investigação forense.
- **Exemplos:**
 - Sistema operacional;
 - Dispositivos móveis/smartphones;
 - ATM/Internet banking;
 - Controle de fronteira.

Sistemas biométricos

- **Definição:** sistemas de **reconhecimento de padrões** que recebem dados biométricos, extraem um **conjunto de características** desses dados, que são então comparados com a **referência biométrica** no banco de dados (JAIN et al., 2004).



Sistemas biométricos

- **Definição:** sistemas de reconhecimento de padrões que recebem dados biométricos, extraem um **conjunto de características** desses dados, que são então comparados com a **referência biométrica** no banco de dados (JAIN et al., 2004).

Um sistema biométrico pode operar no modo **verificação (1:1)** ou no modo **identificação (1:N)**

Treinamento

Teste

Pesquisas em biometria

- Grande parte da pesquisa em biometria foca nos seguintes aspectos (Jain et al., 2016):
 - **Representação de dados biométricos** (extração de características);
 - **Classificação** (dado um conjunto de características biométricas).

Inteligência artificial no contexto de biometria

IA aplicada em biometria

- Técnicas de IA/AM podem ser aplicadas em biometria;
- Exemplos de técnicas de IA/AM usadas em biometria:
 - Deep learning (Sundararajan & Woodard, 2018; Delgado-Santos et al., 2023)
 - k-NN, Random forest (Kambourakis et al., 2021)
 - One-class SVM (Alpar, 2021)
 - etc.

Portanto, podemos entender biometria como uma área em que Inteligência Artificial é aplicada.

Pesquisas envolvendo Inteligência Artificial e Biometria

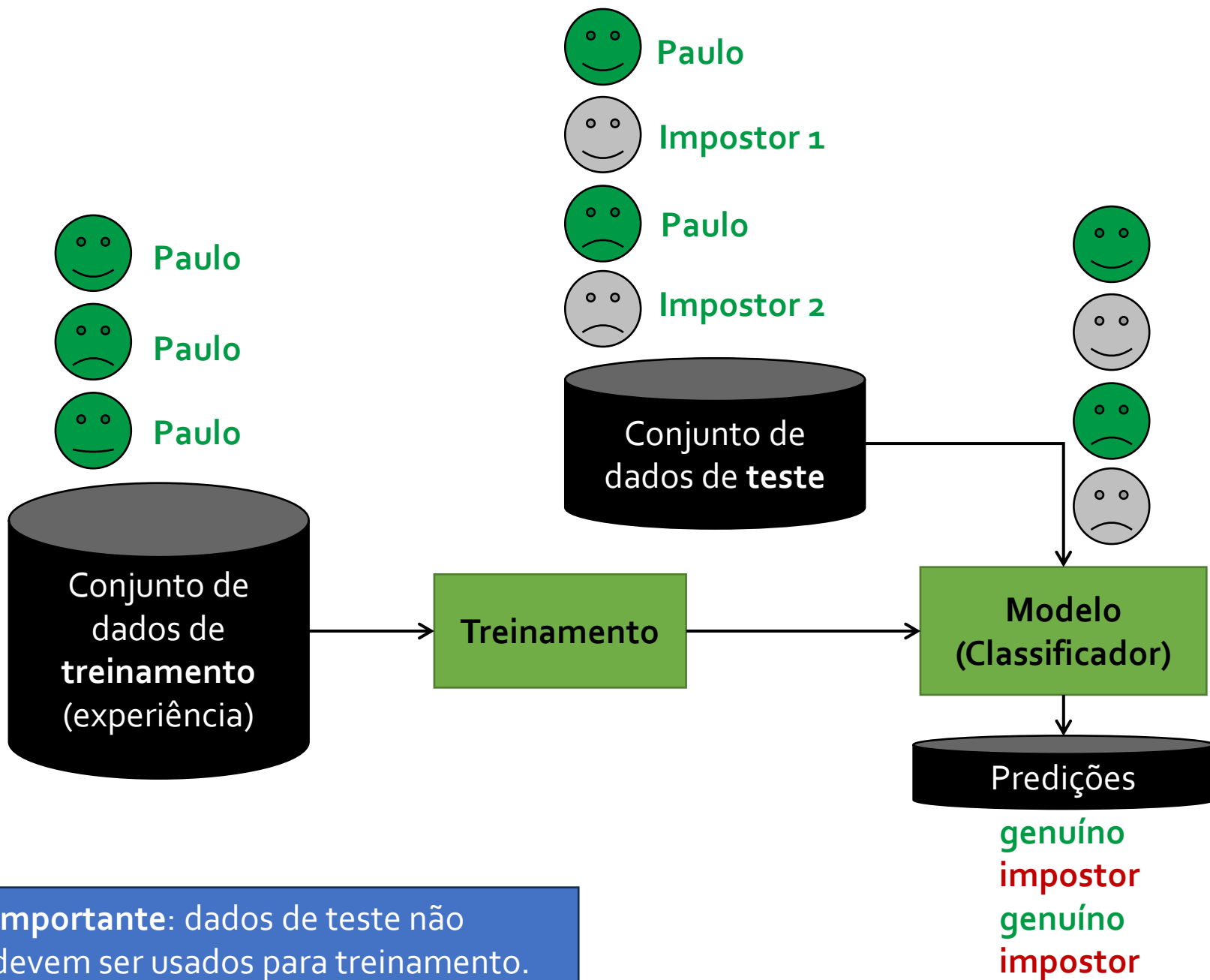
- Vamos discutir diversos aspectos:
 - Coleta de dados;
 - Métricas;
 - Dados para treino e teste;
 - Reprodutibilidade.
- Observação:
 - Diversos aspectos discutidos para biometria também podem ser aplicáveis em outras áreas em que Inteligência Artificial é usada.

Coleta de dados

- Verificar se existem conjuntos de dados já disponíveis;
- Comitê de Ética em Pesquisa (CEP)
 - Na UFABC: <https://cep.ufabc.edu.br/>
- Exemplo de artigo (autores da Unicamp): A. Ferreira, G. Santos, A. Rocha and S. Goldenstein, "User-Centric Coordinates for Applications Leveraging 3-Axis Accelerometer Data," in IEEE Sensors Journal, vol. 17, no. 16, pp. 5231-5243, 15 Aug.15, 2017, doi: 10.1109/JSEN.2017.2723840.

accelerometer and rotation vector data are obtained from an LG Nexus 5 smartphone. Having the proper institutional review board authorization in our university¹ to collect biometric data from volunteers, each one had to walk with the smartphone in a front pocket jeans for about five minutes. For data collection, we developed an Android App, which samples the accelerometer and rotation vector from the

¹Authorization number 1.459.131 and CAAE 53035216.6.0000.5404 from March, 2016.



Métricas

- **FNMR** (False Non-match Rate) / **FRR** (False Rejection Rate)

$$FNMR = \frac{\textit{number of rejected genuine attempts}}{\textit{number of genuine attempts}}$$

$$FRR = FTA + FNMR \times (1 - FTA)$$

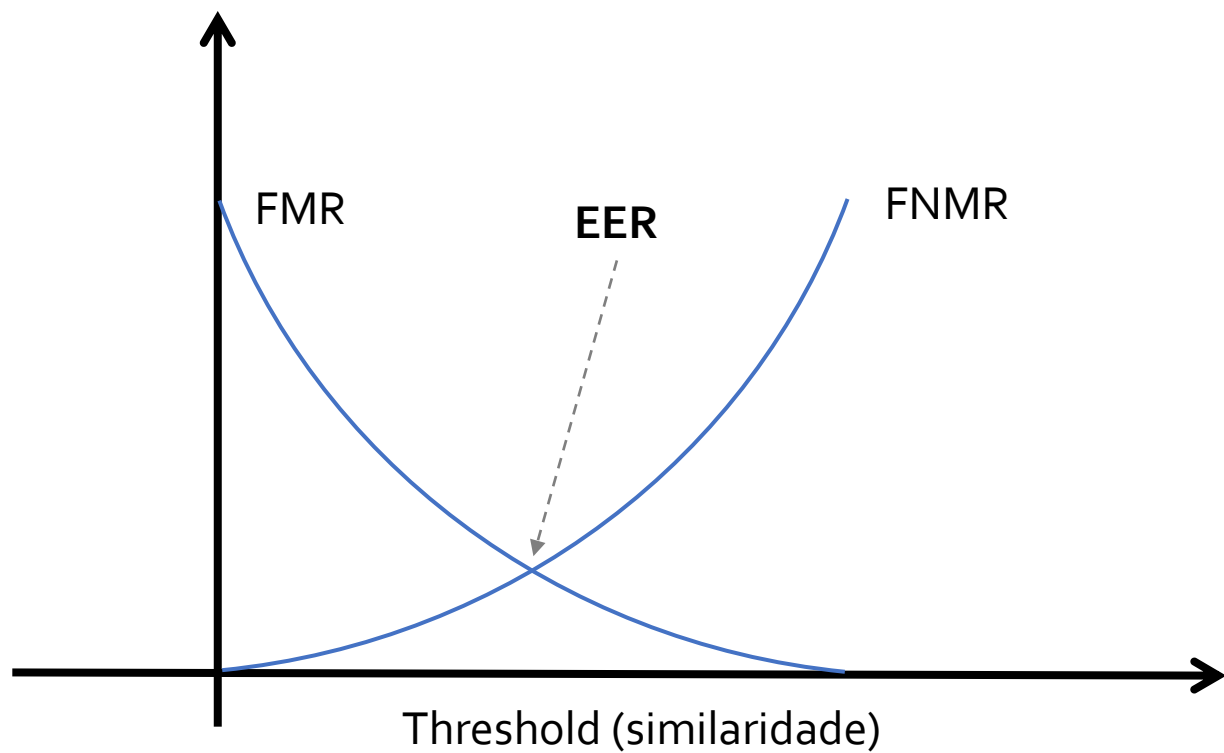
- **FMR** (False Match Rate) / **FAR** (False Acceptance Rate)

$$FMR = \frac{\textit{number of accepted impostor attempts}}{\textit{number of impostor attempts}}$$

$$FAR = FMR \times (1 - FTA)$$

Métricas

- **EER** (Equal Error Rate):
 - Ajuste em que $FMR = FNMR$.



Métricas

- **HTER (Half Total Error) / Acurácia balanceada**

$$HTER = \frac{FNMR + FMR}{2}$$

$$BAcc = 1 - HTER$$

Alguns aspectos importantes

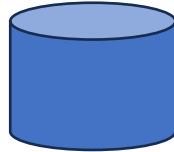
- Considerar questão temporal;
- Há diferentes grupos de usuários (e.g. alguns indivíduos são mais fáceis de serem reconhecidos do que outros):
 - Doddington et al., 1998
 - Yager et al., 2010
 - Mhenni et al., 2020
- Repetições;
- Usar mesmos dados para todos os algoritmos avaliados/comparados.

Reprodutibilidade

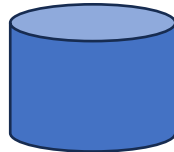
- Repeatability, Reproducibility, Replicability
- Discussão sobre definição dos termos:
 - Barba, 2018
 - Heroux et al., 2018
 - <https://www.ncbi.nlm.nih.gov/books/NBK547546/>
 - <https://www.acm.org/publications/policies/artifact-review-and-badging-current>

Reprodutibilidade

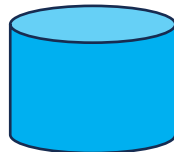
```
from sklearn import datasets, neighbors
X_digits, y_digits = datasets.load_digits(return_X_y = True)
n_exemplos = len(X_digits)
n_treino = int(0.7 * n_exemplos)
X_treino = X_digits[:n_treino]
y_treino = y_digits[:n_treino]
X_teste = X_digits[n_treino:]
y_teste = y_digits[n_treino:]
knn = neighbors.KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_treino, y_treino)
print(knn.score(X_teste, y_teste))
```



```
from sklearn import datasets, neighbors
X_digits, y_digits = datasets.load_digits(return_X_y = True)
n_exemplos = len(X_digits)
n_treino = int(0.7 * n_exemplos)
X_treino = X_digits[:n_treino]
y_treino = y_digits[:n_treino]
X_teste = X_digits[n_treino:]
y_teste = y_digits[n_treino:]
knn = neighbors.KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_treino, y_treino)
print(knn.score(X_teste, y_teste))
```



```
from sklearn import datasets, neighbors
X_digits, y_digits = datasets.load_digits(return_X_y = True)
n_exemplos = len(X_digits)
n_treino = int(0.7 * n_exemplos)
X_treino = X_digits[:n_treino]
y_treino = y_digits[:n_treino]
X_teste = X_digits[n_treino:]
y_teste = y_digits[n_treino:]
knn = neighbors.KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_treino, y_treino)
print(knn.score(X_teste, y_teste))
```



- Níveis (Gundersen et al., 2018):
 - **R1**: mesma implementação do método de IA + mesmos dados = mesmos resultados
 - **R2**: outra implementação do método de IA + mesmos dados = mesmos resultados
 - **R3**: outra implementação do método de IA + outros dados = resultados consistentes

Reprodutibilidade

- Alguns artigos discutem esse problema em IA:
 - Algumas questões que podem dificultar a verificação das conclusões obtidas: **código não publicado** e **sensibilidade às condições de treinamento** (Hudson, 2018);
 - **Números aleatórios (pseudo-aleatórios)** podem impactar algoritmos de aprendizado de máquina (Ahmed & Lofstead, 2022);
 - Sugestões para **documentação de pesquisa em IA** (Gundersen et al., 2018);
 - **Crise de reprodutibilidade** (Ball, 2023).

Oportunidades de pesquisa e
tópicos futuros

Algumas oportunidades de pesquisa e tópicos futuros

- Sistemas biométricos adaptativos (Pisani et al. , 2019; Ryu et al., 2023);
- Bias and Fairness (Jain et al., 2022);
- Regulação;
- Mais tópicos para pesquisas futuras em Jain et al., 2022; Berghoff et al., 2021.

Referências

- Paula Delgado-Santos, Ruben Tolosana, Richard Guest, Farzin Deravi, Ruben Vera-Rodriguez, Exploring transformers for behavioural biometrics: A case study in gait recognition, *Pattern Recognition*, Volume 143, 2023, 109798, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2023.109798>.
- Orcan Alpar, Biometric keystroke barcoding: A next-gen authentication framework, *Expert Systems with Applications*, Volume 177, 2021, 114980, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2021.114980>.
- Kambourakis, G., Damopoulos, D., Papamartzivanos, D., and Pavlidakis, E. (2016) Introducing touchstroke: keystroke-based authentication system for smartphones. *Security Comm. Networks*, 9: 542–554.
doi: 10.1002/sec.1061.
- Kalaivani Sundararajan and Damon L. Woodard. 2018. Deep Learning for Biometrics: A Survey. *ACM Comput. Surv.* 51, 3, Article 65 (May 2019), 34 pages. <https://doi.org/10.1145/3190618>

Referências

- G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation," Proc. Int'l Conf. Spoken Language Processing, 1998.
- N. Yager and T. Dunstone, "The Biometric Menagerie," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, pp. 220-230, Feb. 2010, doi: 10.1109/TPAMI.2008.291.
- A. Mhenni, C. Rosenberger and N. E. B. Amara, "Splitting Wolves Category in Doddington Zoo: Impacts on Keystroke Dynamics," 2020 International Conference on Cyberworlds (CW), Caen, France, 2020, pp. 243-248, doi: 10.1109/CW49994.2020.00046.

Referências

- Heroux, Michael A., Barba, Lorena, Parashar, Manish, Stodden, Victoria, and Taufer, Michela. Toward a Compatible Reproducibility Taxonomy for Computational and Computing Sciences.. United States: N. p., 2018. Web. doi:10.2172/1481626.
- Barba LA. arXiv, 1802.03311. 2018. [December 2018]. Terminologies for Reproducible Research. <https://arxiv.org/abs/1802.03311>
- Gundersen, O. E., Gil, Y., & Aha, D. W. (2018). On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications. *AI Magazine*, 39(3), 56-68. <https://doi.org/10.1609/aimag.v39i3.2816> (discute sugestões para documentação de pesquisa em IA)
- Ball P. Is AI leading to a reproducibility crisis in science? *Nature*. 2023 Dec;624(7990):22-25. doi: 10.1038/d41586-023-03817-6. PMID: 38052897.

Referências

- Matthew Hutson, Artificial intelligence faces reproducibility crisis. *Science* 359, 725-726(2018). DOI: 10.1126/science.359.6377.725
- Hana Ahmed and Jay Lofstead. 2022. Managing Randomness to Enable Reproducible Machine Learning. In Proceedings of the 5th International Workshop on Practical Reproducible Evaluation of Computer Systems (PRECS '22). Association for Computing Machinery, New York, NY, USA, 15–20. <https://doi.org/10.1145/3526062.3536353>
- Anil K. Jain, Karthik Nandakumar, Arun Ross, 50 years of biometric research: Accomplishments, challenges, and opportunities, *Pattern Recognition Letters*, Volume 79, 2016, Pages 80-105, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2015.12.013>.
- A. K. Jain, D. Deb and J. J. Engelsma, "Biometrics: Trust, But Verify," in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 3, pp. 303-323, July 2022, doi: 10.1109/TBIOM.2021.3115465.

Referências

- Paulo Henrique Pisani, Abir Mhenni, Romain Giot, Estelle Cherrier, Norman Poh, André Carlos Ponce de Leon Ferreira de Carvalho, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2019. Adaptive Biometric Systems: Review and Perspectives. ACM Computing Surveys 52, 5, Article 102 (October 2019), 38 pages. DOI: <https://doi.org/10.1145/3344255>
- Riseul Ryu, Soonja Yeom, David Herbert, Julian Dermoudy, The design and evaluation of adaptive biometric authentication systems: Current status, challenges and future direction, ICT Express, 2023, ISSN 2405-9595, <https://doi.org/10.1016/j.icte.2023.04.003>.
- C. Berghoff, M. Neu and A. von Twickel, "The Interplay of AI and Biometrics: Challenges and Opportunities," in Computer, vol. 54, no. 9, pp. 80-85, Sept. 2021, doi: 10.1109/MC.2021.3084656.