Classification History

Esdras Cavalcanti
esdras.almeida@fgv.edu.br

João Felipe
B51067@fgv.edu.br

Marcelo Angelo
B52146@fgv.edu.br

***Abstract*—Machine learning models, especially those for classification, are increasingly present in daily life, yet their conceptual complexity represents a barrier for non-experts. This paper presents the development of an interactive data visualization tool designed to demystify the fundamentals of classification for a lay audience. Using a visual narrative constructed with scrollytelling techniques, the tool guides the user from basic classification concepts, through the historical XOR problem and the rise of Artificial Neural Networks (ANN), to the revolution caused by Convolutional Neural Networks (CNN) in image classification. The approach culminates in an interactive interface that allows the user to draw digits and observe in real-time the classification process by a CNN, fostering experimentation. Additionally, the article discusses the evolutionary trajectory and the growing computational demand of artificial intelligence models, contextualizing the reader on the challenges and the future of the field.**

***Index Terms*—Data Visualization, Machine Learning, Classification, Convolutional Neural Network (CNN), Scrollytelling, AI Literacy.**

I. Introduction

The growing ubiquity of machine learning (ML) models in everyday technologies contrasts with the difficulty of understanding their internal mechanisms by people outside the field of computing. This knowledge gap can generate mistrust and hinder broader discussions about the social and ethical impact of these technologies. The present work aims to mitigate this barrier by focusing on the task of classification, one of the pillars of ML.

The main objective is to present the highest-level concepts in an intuitive and interactive way. The proposed visual narrative follows a logical progression: (1) it introduces what classification is in an abstract way; (2) it presents linear classification and its limitations, such as the XOR problem; (3) it demonstrates how the advent of the Multi-Layer Perceptron (MLP) overcame such limitations; and (4) it explores how Convolutional Neural Networks (CNNs) revolutionized image classification.

To solidify learning and engage the user, an interactive application was developed where it is possible to ”open” a CNN, draw a handwritten digit, and observe the activations of the internal layers up to the final prediction. Finally, the work contextualizes this technological evolution through a visualization of the increasing complexity and computational cost of models, raising questions about the future trends of the field.

This article is structured as follows: Section 2 discusses related works. Section 3 details the methodology and the technologies used. Section 4 presents the results, describing the implemented visual narrative. Section 5 discusses the results and the observed feedback. Finally, Section 6 presents the conclusions and points to directions for future work.

II. Related Work

The area of visualization for ML model explainability has received increasing attention. The work of Wang et al. [1], known as CNN Explainer, offers a detailed and interactive visualization of the internal operations of a CNN, such as convolutions and pooling. Unlike our focus on the general concept of classification, CNN Explainer delves into the mechanisms of a specific architecture, being a powerful tool for students in the field.

The Epoch AI [2] platform proposes a comprehensive data visualization of the evolution of AI models, focusing on metrics such as the computational power (FLOPs) required for training. Although it provides a valuable historical overview, the platform does not delve into discussions about the implications of this exponential growth, one of the points we seek to raise in our approach.

Other initiatives, such as those in 3Blue1Brown YouTube Channel [3], use visual approaches to explain what machine learning is more generally, serving as excellent conceptual starting points. Our work differs by integrating these concepts into a single cohesive and interactive narrative, which goes from the abstract (classification of points) to the concrete (classification of images drawn by the user).

III. Methodology

The construction of the visualization system was guided by the pursuit of interactivity and fluidity. Below, we detail the tools and techniques used in each component of the application.

A. Application Architecture

For the construction of the interface, the Svelte framework was used, known for its reactivity and performance. The data visualizations were mostly built with the D3.js library, due to its flexibility for creating custom and animated graphics. For the interactive drawing section, the HTML5 Canvas element was used to capture user input.

B. Classification Visualization

To illustrate the initial concepts of classification, we employed the scrollytelling technique. Points representing two distinct classes (blue and orange) are animated to assume different spatial configurations as the user advances in the narrative. The smooth transition between states (random distribution, linearly separable, and XOR configuration) was implemented with D3.js. The decision to omit axes and markings was deliberate, aiming to focus on the conceptual representation of data separability, rather than analyzing a real dataset.

C. Convolutional Neural Network (CNN) Visualization

The centerpiece of our tool is the interactive visualization of the CNN. We used a pre-trained model of the VGGNet architecture [4], developed in Python with the TensorFlow library and later converted for execution in the browser with TensorFlow.js.

To visualize the activations of the internal layers, the strategy adopted was to create sequential sub-models. The output of one layer serves as input for the next layer, and the result of each step is rendered on the canvas. To ensure a fluid user experience and avoid freezing the browser during model loading, we implemented a lazy loading strategy, which loads the TensorFlow.js model only when the user approaches this section of the page.

D. Visualization of Model Evolution

In this section, we used a dataset from Epoch AI [2], which catalogs notable AI models and their metrics, such as the amount of training FLOPs. We chose to create a simplified visualization of the one presented in the original source, focusing on a timeline that highlights important milestones and the exponential growth of computational demand. The visualization was built with D3.js to allow for interactivity and annotations.

IV. Results: The Visual Narrative

This section describes the user's journey through the created storytelling, detailing each stage of the visualization.

a) Introduction to Classification:: The narrative begins with an explanatory text and a scatter plot with two classes of points (blue and orange) distributed randomly.

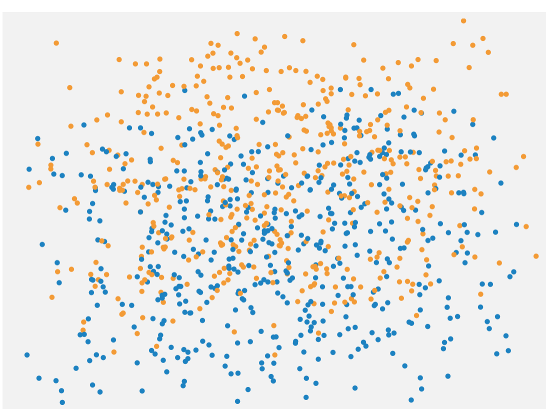

Fig. 1. Introduction to Classification: Random scatterplot

b) Linear Classification:: As you scroll, the text introduces the computational concept of classification. The points on the graph animate to form a linearly separable set, and a line representing a simple classifier, such as the Perceptron [6], divides the space.

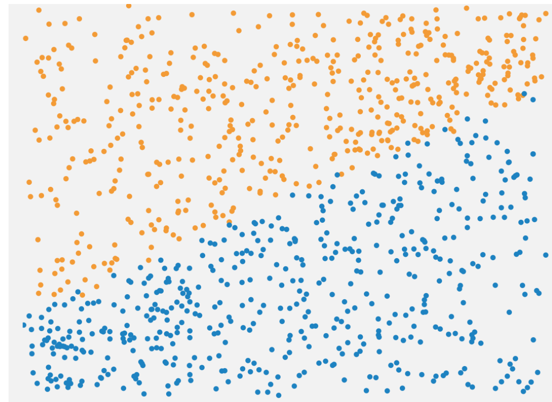

Fig. 2. Linear Classification: Linear separable scatterplot

c) The XOR Problem:: Next, the "AI winter" and the limitation of linear models are mentioned, exemplified by the XOR problem. The points rearrange to form the classic XOR configuration, where a single line is unable to separate the classes.

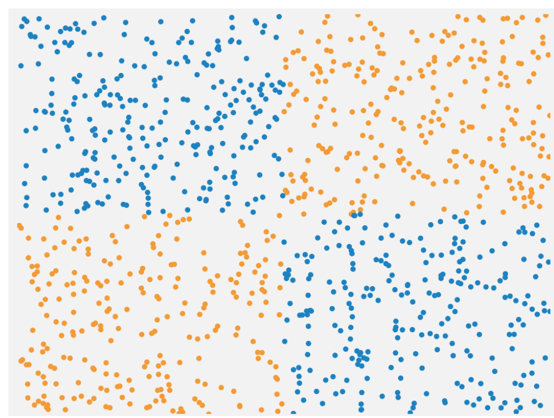

Fig. 3. The XOR Problem: XOR scatterplot

d) MNIST:: A breaf introduction to MNIST to get user on image classification.

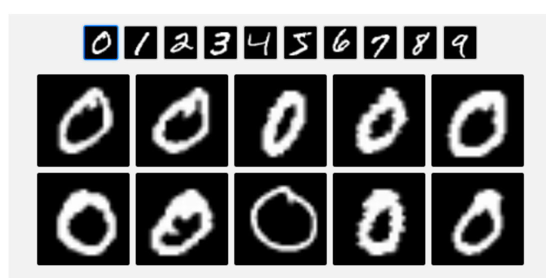

Fig. 4. MNIST: Dataset sample

e) The Solution with Neural Networks and the CNN:: The narrative explains how MLPs solved the XOR problem and how CNNs, an evolution, became the standard for computer vision tasks. At this point, the famous MNIST dataset is introduced. The main interface is then presented: a drawer where the user can draw a number and, next to it, a visual representation of the VGGNet layers. As the drawing is processed, the user can see the activations in each layer, from the input to the output layer that indicates the model's prediction.

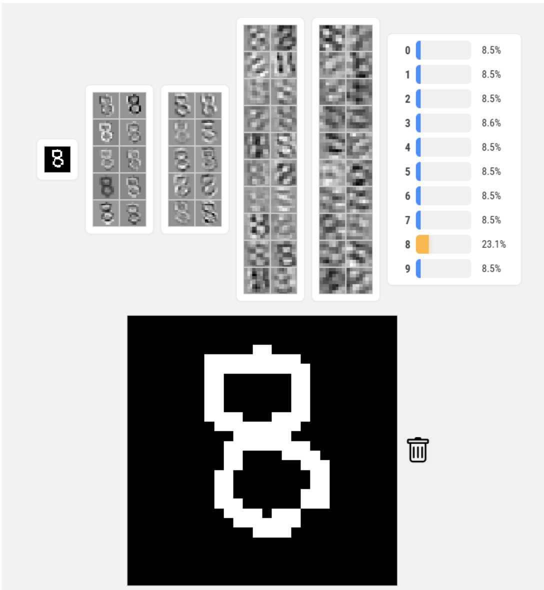

Fig. 5. The Solution with Neural Networks and the CNN: CNN Drawer

f) AI Milestones:: To contextualize the CNN, an interactive timeline, built with D3, lists other notable models that were milestones for classification, segmentation, and detection, such as AlexNet [6] and ResNet [7].

Fig. 6. AI Milestones: Models timeline

g) Model Complexity:: Finally, using data from Epoch AI, a bar or line chart shows the dramatic growth in the need for FLOPs over the years, initiating a discussion about the future and sustainability of developing ever-larger models.

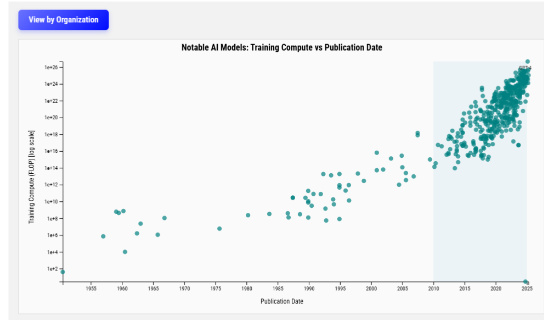

Fig. 7. Model Complexity: FLOPs plot

V. Discussion

The main objective of this project was to create an educational tool for users with no prior knowledge of machine learning. An informal case study, conducted with users close to the authors, indicated that the visual storytelling approach was effective. Participants reported understanding the evolutionary progression of the models and were engaged with the final discussion on the growing complexity of AI.

The interactive CNN drawing section was particularly successful in keeping users engaged. A pattern of behavior was observed where users spent a considerable amount of time trying to ”trick” the model by drawing ambiguous digits or non-numeric shapes. This active search for the model's limits, although not foreseen, proved to be an excellent starting point for discussions about the robustness and explainability of AI models, an active field of research. The fact that the model does not have 100% accuracy (an expected result) made this exploration even richer, as the classification errors served as practical examples of the technology's limitations.

VI. Future Work and Conclusion

This work has demonstrated the potential of visual and interactive narratives for teaching complex machine learning concepts. In conclusion, the tool proved to be successful in its purpose of engaging and educating a non-specialized audience.

Based on the results and feedback received, several directions for future work can be explored:

- **Expansion of Models:** Add visualizations of other architectures, such as an MLP to solve the XOR problem or a Vision Transformer (ViT) [8], allowing for a direct comparison of their operation.
- **Focus on Explainability:** Develop a section dedicated to exploring why the model confuses certain numbers, using explainability techniques such as saliency maps to highlight the regions of the image that most influenced the decision.
- **Enhancement of Visualizations:** Improve the visualization of the Epoch AI data, perhaps connecting it more explicitly to the storytelling through dynamic annotations or scrollytelling.

We believe that by continuing to develop tools that unite design, interactivity, and technical content, it is possible to make education in artificial intelligence more accessible and democratic.

REFERENCES

[1] Link to CNN Explainer: https://poloclub.github.io/cnn-explainer/

[2] Link to Epoch AI: https://epoch.ai/data/notable-ai-models

[3] Link to the 3Blue1Brown video: https://www.youtube.com/watch?v=aircAruvnKk

[4] Simonyan, K., & Zisserman, A. (2014). ”Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv preprint arXiv:1409.1556.

[5] Rosenblatt, F. (1958). ”The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” Psychological Review, 65(6), 386–408.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ”ImageNet Classification with Deep Convolutional Neural Networks.” Advances in Neural Information Processing Systems (NIPS), 25.

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). ”Deep Residual Learning for Image Recognition.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[8] Dosovitskiy, A., et al. (2020). ”An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” arXiv preprint arXiv:2010.11929.