

## Twitter Thread by Vladimir Haltakov



**Vladimir Haltakov**

@haltakov

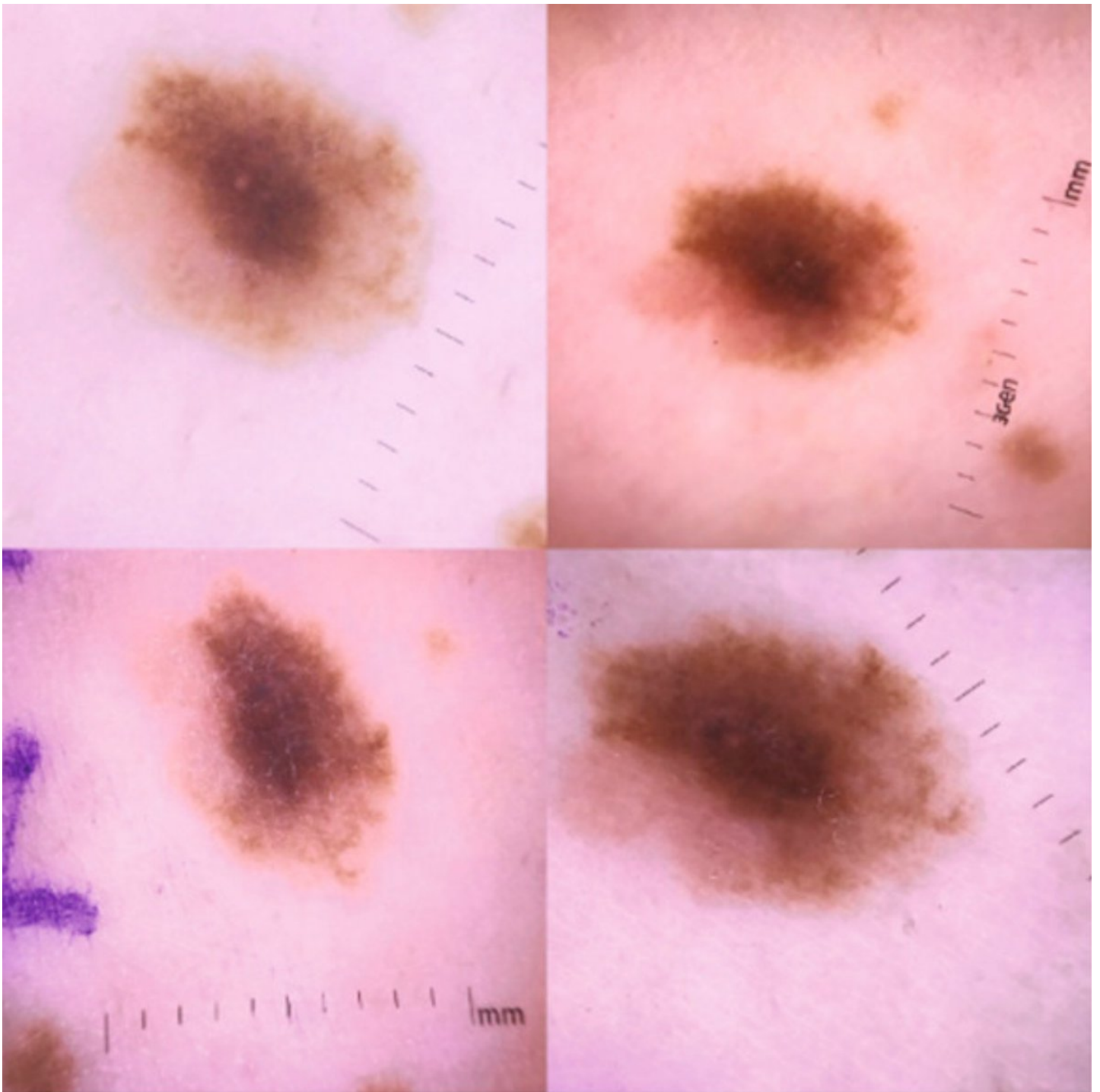


**Dataset bias strikes again! ■**

**This is a great example of dataset bias! This ML method classifies skin lesions as being malignant or not.**

**Can you see the problem from the examples? ■**

**Let me shortly explain ■**



The dataset used to train the model contains images of both malignant and benign lesions. The problem is that the malignant images often contain rulers printed on the skin. They are used by doctors to assess the lesion over time.

Guess what the model learned... ■

The ML model learned to detect the rulers... ■■■■

"In our dataset, images with rulers were more likely to be malignant; thus the algorithm inadvertently "learned" that rulers are malignant."

<https://t.co/H00FFokKHr>