IMT SCHOOL FOR ADVANCED STUDIES LUCCA

**Data Science & Statistical Learning | II Level Master**

**Health Analytics and Data-Driven Medicine**

UNIVERSITÀ DEGLI STUDI FIRENZE

# Reimagining leprosy elimination with AI analysis of a combination of skin lesion images with demographic and clinical data

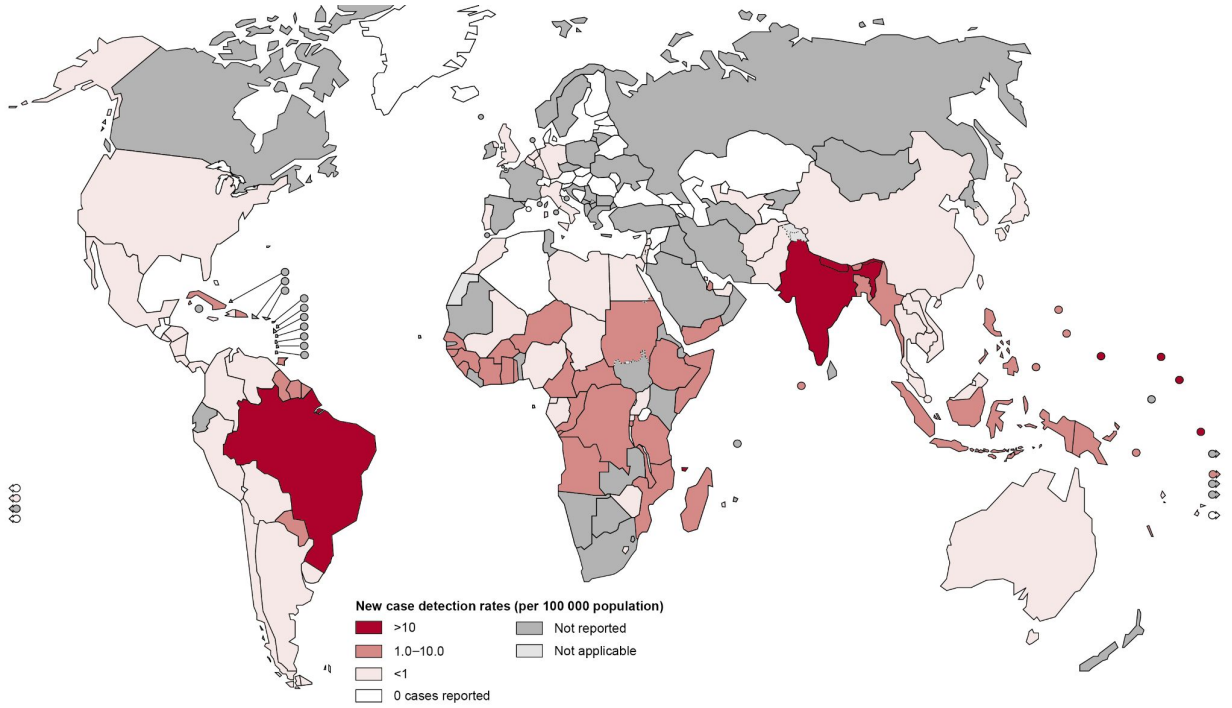

**Prof.** *Giorgio Gnecco*

**Prof.** *Armando Rungi*

*Gianmarco Santoro*

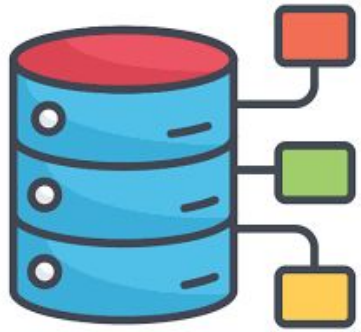

25/01/2024

- **Leprosy**: tropical **infectious disease**
- **Diagnosis often delayed**:
  - Symptoms take from 2 months to 20 years to appear and disease progression is slow
  - Leading to irreversible neurological damage and **continued transmission**
- Mostly affects **underserved populations**
- Largely eliminated, but **~200'000 new diagnosed** annually




- **Accelerating diagnosis** can advance **leprosy elimination**
- **AI** technology has potential to augment health workers in making faster and more accurate diagnosis. It has been successfully applied to other skin conditions
- **AI4Leprosy**: AI diagnosis assistant for leprosy, based on skin **images** and **clinical** data

- 2018-2020: **collected high-resolution** research **protocol images** from leprosy patients and other dermatological diseases along with **clinical and demo data** at a clinic in Brazil
- Patient interviews, clinical consultations, image collections and labelling follow standardized process to **de-identify patients, for privacy**
- All **patients with leprosy-like skin lesions** and above the age of 6 were included, except those who did not provide consent or present for their final diagnosis at visit 3:
    - **228 recruited patients, 222 included** randomly divided in:
        - **182** patients were used **to train** the algorithms in a training dataset
        - **40** patients were separated as an independent **testing** group for validation
    - **1226 collected skin images**
    - **585 sets of metadata** stored in an open-source dataset for other researchers to exploit



Images sources: https://uxwing.com/privacy-icon/ **&** https://www.flaticon.com/free-icon/database_9243391

- **First visit**:
  - **Patients** presenting **skin problems** were examined for leprosy-like lesions
  - Received info on research, completed **consent**, declare **health info**, **demographic** info and potential **symptoms**
- **Second visit**:
  - **Photography** taken
  - Samples for **clinical evaluations** to confirm (or exclude) leprosy diagnosis
- **Third visit**:
  - **Laboratory results with diagnosis** and appropriate treatment for leprosy or other disease
- Images **follow International Skin Imaging Collaboration** (ISIC):
  - Background color, lighting, field of view, focus/depth of field, resolution, scale and color calibration
- Up to **three images** were taken from **each skin lesion**:
  - A **panoramic** to identify the body part, a **close-up photo** and one including **surrounding normal skin**

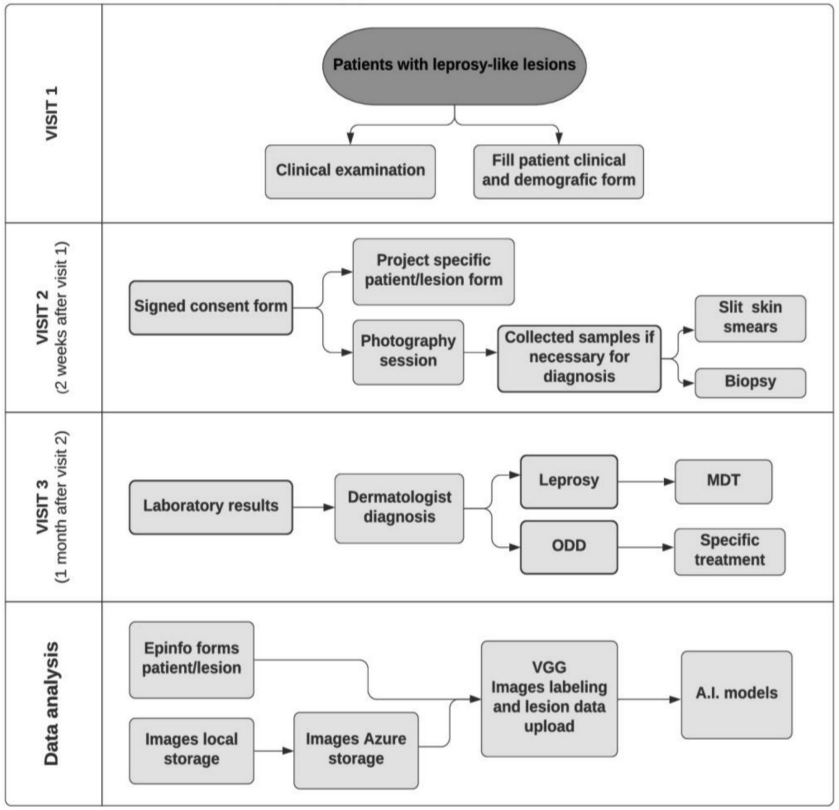
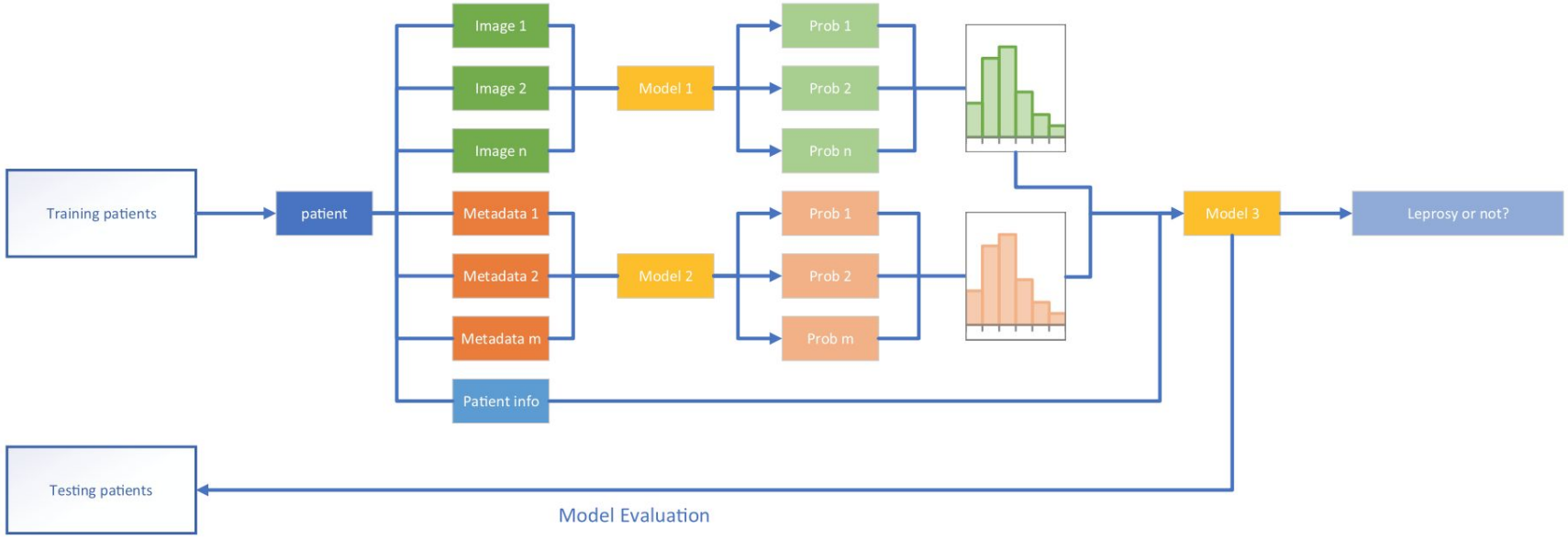



Image source: Paper*




- Developed a **two-step patient-level prediction of leprosy probability**:
  - Model **1**: based on the **skin lesion images**
  - Model **2**: based on **metadata**
- Each model produced a **probability** of leprosy **for each data**. Given that patients could have multiple records, **outputs** are combined from both models **per patient in a histogram**, to represent the predicted probabilities
- Model **3** was trained to combine analysis made in the first step, with the patient info. This last step **established the overall probability** by combining the histograms from Model 1 and Model 2

Image source: Paper*

- For each experiment of Model 1 or 2, used **5-fold cross-validation to evaluate the performance** of each algorithm

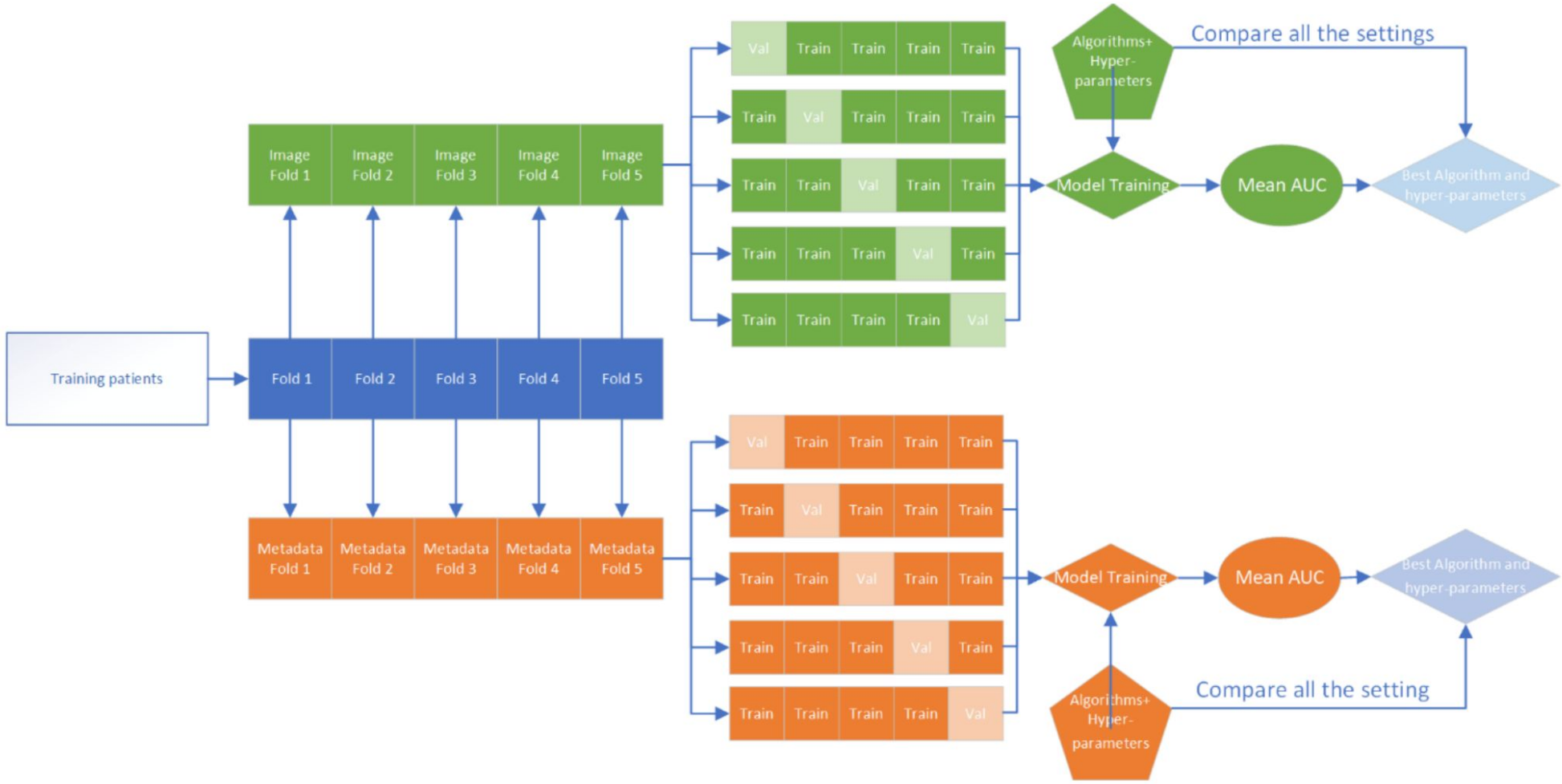



Image source: Paper*

IMT SCHOOL FOR ADVANCED STUDIES LUCCA
UNIVERSITÀ DEGLI STUDI FIRENZE

- **Model 1** combined the following settings:
  1. **Neural Network** architecture: **Inception-v4** or **ResNet-50**
  2. **Tuning** strategy: **tune all** (fine-tune the complete neural net model) or **freeze** (train only the output layer)
  3. **Input** image type: **close-up only** or **all images**
  4. **Optimizer**: **stochastic gradient descent**
- Final algorithm to train Model 1: **fine-tuned ResNet-50 using close-up images**, 74.6% **AUC**
- **Performance** of the ResNet-50 is **lower** when **both close-up and other images** were included

| Mean ACC & AUC | | ResNet-50 | | Inception-v4 | |
|---|---|---|---|---|---|
| | | Tune All | Freeze | Tune All | Freeze |
| All | ACC (SD)* | 0.6138 (0.040) | 0.5723 (0.051) | 0.5828 (0.070) | 0.5212 (0.050) |
| | AUC (SD) | 0.6760 (0.057) | 0.6003 (0.094) | 0.6144 (0.106) | 0.5487 (0.045) |
| Close-up | ACC (SD) | 0.6660 (0.099) | 0.5790 (0.064) | 0.5834 (0.092) | 0.5661 (0.059) |
| | AUC (SD) | 0.7456 (0.113) | 0.6542 (0.104) | 0.6590 (0.099) | 0.5919 (0.089) |

***Table 1*: The performance comparison for Model 1 (images only) using ResNEt-50 and Inception-v4 neural network architectures.**

* ACC – accuracy; AUC- area under curve; SD – standard deviation.

Image source: Paper*

IMT SCHOOL FOR ADVANCED STUDIES LUCCA
UNIVERSITÀ DEGLI STUDI FIRENZE

- **Model 2**, tested 3 machine learning methods:
  - **Elastic-net logistic regression (LR)**; **XGBoost (XGB)**; **Random forests (RF)**
  - **Temperatures** and **diameters** are **time-consuming** to measure in practice: evaluated the impact of **removing these two features by creating a new subset**
  - Final algorithm: **elastic-net logistic regression using the subset features achieved the highest AUC score 88.0%**
  - Removing **temperature** and **diameter of skin lesions** only **minimally influenced performance** of the algorithms

| Mean score | Elastic-net Regression (LR) | | XGBoost (XGB) | | Random forest (RF) | |
|---|---|---|---|---|---|---|
| | Full | Subset | Full | Subset | Full | Subset |
| ACC (SD) | 0.813 (0.058) | 0.817 (0.06) | 0.808 (0.086) | 0.818 (0.075) | 0.779 (0.088) | 0.818 (0.073) |
| AUC (SD) | 0.881 (0.082) | 0.880 (0.080) | 0.849 (0.086) | 0.846 (0.092) | 0.836 (0.071) | 0.863 (0.090) |
| Sensitivity (SD) | 0.841 (0.118) | 0.845 (0.115) | 0.818 (0.067) | 0.85 (0.092) | 0.795 (0.129) | 0.845 (0.090) |
| Specificity (SD) | 0.791 (0.173) | 0.794 (0.174) | 0.79 (0.177) | 0.784 (0.155) | 0.763 (0.177) | 0.789 (0.158) |

***Table 2*: The performance comparison for Model 2 using features extracted from the form of the lesion for Elastic-net logistic regression (LR), XGBoost (XGB) and Random forests (RF) machine learning methods. For the full data analysis, data included 15 predictors collected at the clinical evaluation as described in the methods section. For subset analysis, temperature and diameter features were excluded.**

Image source: Paper*

- **Model 3**: elastic-net logistic regression, XGBoost and Random Forest
- **Random Forest** and **XGBoost** using **patient information alone** delivered the highest **AUC:**
  - Better at capturing nonlinear relationships than elastic-net logistic regression
- Possible reason why the **inclusion of Model 2 outputs don't increase performance** as expected:
  - **Overlapping info between metadata and patient info** such as sensory loss or pruritus. They characterized similar aspects: for example, 'nodule' would be recorded both on **lesion document** and **patient info**
- Using Model 1 outputs alone yields ~70% AUC, while Model 2 ones or patient info gives a model with an AUC > 90%

| | Elastic-net Regression-LR (%) | | | | XGBoost - XGB (%) | | | | Random Forest - RF (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ACC (SE) | AUC (SE) | SEN (SE) | SP (SE) | ACC (SE) | AUC (SE) | SEN (SE) | SP (SE) | ACC (SE) | AUC (SE) | SEN (SE) | SP (SE) |
| Model 1 outputs | 72 (0.070) | 73 (0.083) | 78 (0.098) | 68 (0.099) | 65 (0.075) | 71 (0.086) | 67 (0.111) | 64 (0.102) | 65 (0.075) | 71 (0.084) | 67 (0.111) | 64 (0.102) |
| Model 2 outputs | 92 (0.041) | 95 (0.025) | 94 (0.054) | 91 (0.061) | 90 (0.047) | 92 (0.03) | 89 (0.074) | 91 (0.061) | 90 (0.047) | 92 (0.039) | 89 (0.074) | 91 (0.061) |
| Patient info | 88 (0.052) | 96 (0.020) | 72 (0.105) | 100 (0) | 95 (0.034) | 99 (0.006) | 94 (0.054) | 95 (0.044) | 95 (0.034) | 98 (0.01) | 89 (0.074) | 1 (0) |
| Model 1 & 2 outputs | 80 (0.063) | 89 (0.054) | 83 (0.087) | 77 (0.089) | 80 (0.063) | 88 (0.053) | 78 (0.098) | 82 (0.082) | 82 (0.060) | 89 (0.054) | 78 (0.09) | 86 (0.073) |
| Model 1 outputs + patient info | 78 (0.066) | 86 (0.061) | 78 (0.09) | 77 (0.089) | 65 (0.075) | 79 (0.071) | 72 (0.106) | 59 (0.104) | 75 (0.068) | 85 (0.065) | 72 (0.105) | 77 (0.089) |
| Model 2 outputs + patient info | 90 (0.047) | 96 (0.019) | 89 (0.074) | 91 (0.061) | 88 (0.052) | 93 (0.034) | 83 (0.088) | 91 (0.061) | 88 (0.052) | 96 (0.021) | 83 (0.087) | 91 (0.061) |
| All | 88 (0.052) | 92 (0.039) | 83 (0.087) | 91 (0.061) | 78 (0.066) | 87 (0.058) | 72 (0.106) | 82 (0.082) | 80 (0.063) | 90 (0.049) | 72 (0.105) | 86 (0.073) |

***Table 3*: The performance comparison for Model 3 using the all the training patients and logistic regression (LR), XGBoost (XGB) and random forest (RF). All metrics were obtained by validating the models on a separate testing dataset from 40 patients. Results are shown for Models 1 and 2 separately or combined. Patient info means the information collected on the patient information document.**

*ACC-accuracy; AUC area under curve; SEN-sensitivity; SP-specificity; SE-standard error.

Image source: Paper*

IMT SCHOOL FOR ADVANCED STUDIES LUCCA
UNIVERSITÀ DEGLI STUDI FIRENZE

- To **interpret** the **model**: applied elastic-net logistic regression with repeated 10-fold cross validation, on the complete dataset
- The feature importance was **measured by the decrease in model accuracy** when changing the features' values
- **Model findings** are **aligned with clinical observations**: e.g. sensory loss is typical of leprosy, lesions rarely cause pruritus
- Given small sample size, **elastic-net logistic regression with Model 2 outputs and patient information is the one selected**:
  - **AUC 96.4%**
  - Simpler and more interpretable than XGBoost or Random Forest, even if they gave better AUC performance
- **Final model**: elastic-net logistic regression with repeated 10-fold cross validation **on the complete dataset including testing patients**

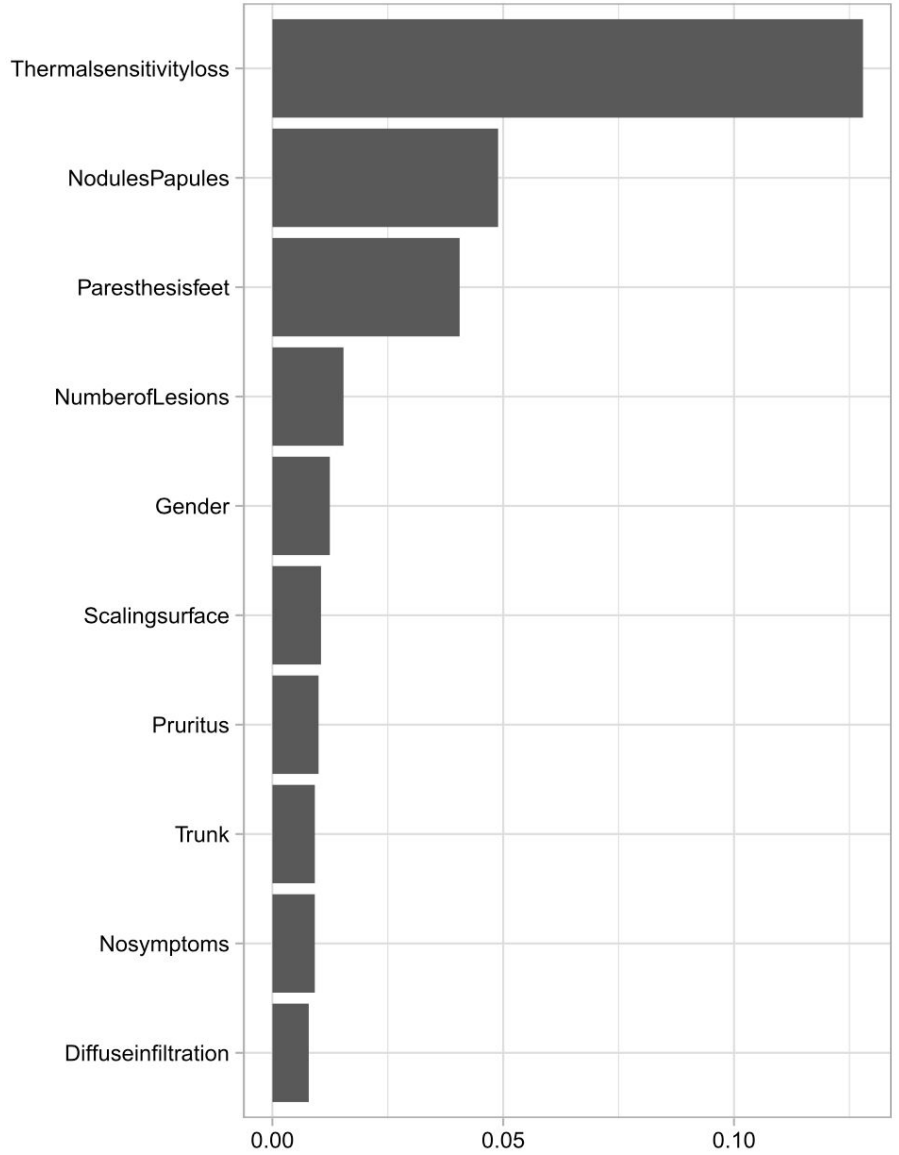
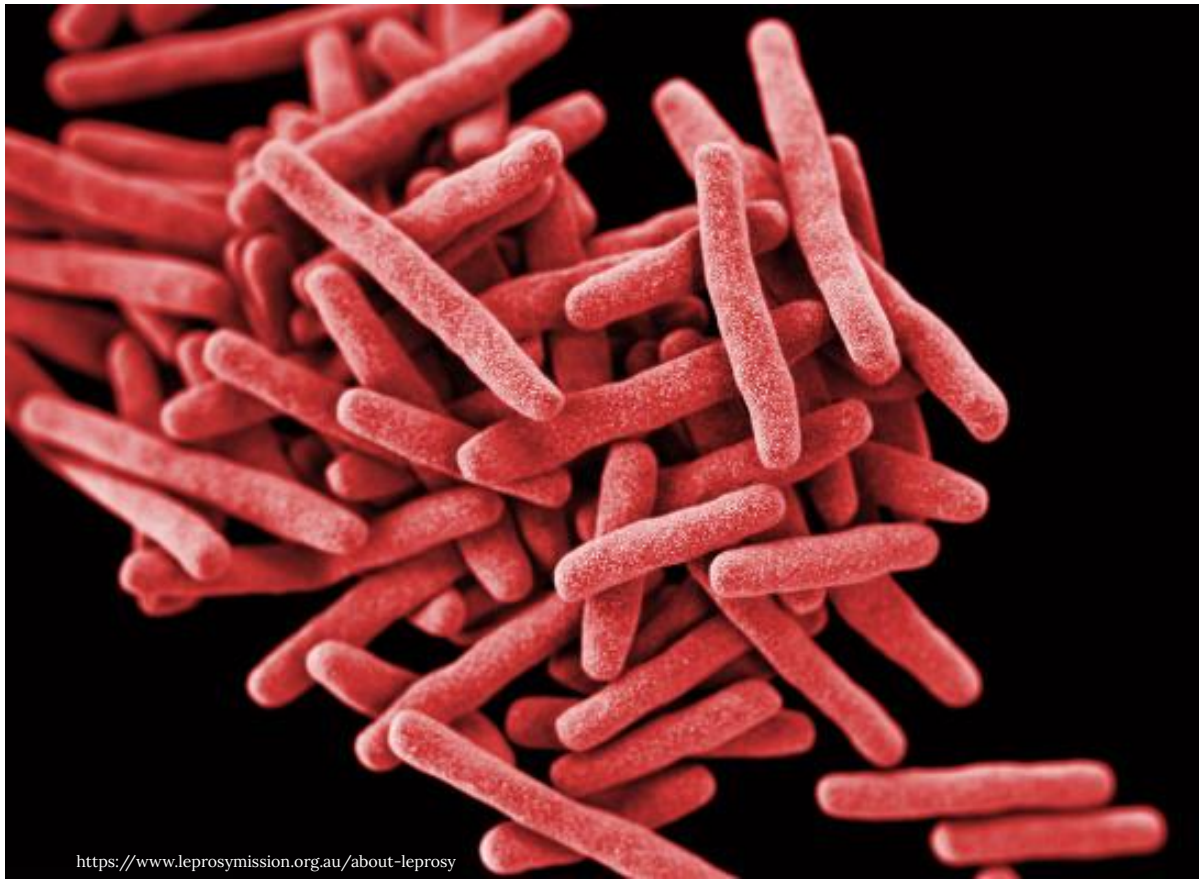



Image source: Paper*

- **Achievements**:
  - Provided an **open-source dataset** of high-quality skin images and clinical data to evaluate the feasibility and accuracy of AI driven models in predict skin diseases
  - Data security and privacy: defined a **protocol avoiding individual recognition**
  - Probability models: recognize leprosy with high AUC 96.4% **combining Elastic-net Regression model 2 outputs and patient information**
- **Considerations:**
  - Several other **skin diseases smartphone apps** have been developed, increasing speed and accuracy in diagnosis
  - Based on this evidence, **next AI4Leprosy** will be trained by collecting images and data **through a smartphone app**:
    - To further **improve the model on lower quality images** and ultimately better **mimic real-world** settings
  - Training the model **in frontline settings** to remediate any **selection bias** considering only one leprosy center
  - A **cardinal sign** for leprosy such as **sensitivity loss** contributed significantly to the AI algorithm
  - **Model experiences more difficulties in diagnosis from skin images** than from the metadata or patient info
  - Training on **other skin types** will be essential, even though Brazil has a quite diverse population that allowed image collection from a variety of skin types; still, growing the overall **amount of data** will improve the efficiency and the contribute of the model 1 in detecting the disease

IMT SCHOOL FOR ADVANCED STUDIES LUCCA

UNIVERSITÀ DEGLI STUDI FIRENZE




# Thank you

**Images and text have been gathered from the paper***: "*Reimagining leprosy elimination with AI analysis of a combination of skin lesion images with demographic and clinical data*" | *Raquel R Barbieri, Yixi Xu, Lucy Setian, et al.* | *Laboratorio de Hanseníase Instituto Oswaldo Cruz, Brazil; Microsoft, United States; Novartis Foundation, Switzerland; University of Basel, Switzerland; University Medical Center Mainz, Germany; University of Aberdeen, Scotland; Retired, United States* | *The Lancet Regional Health - Americas 2022;9: 100192 Published online 3 February 2022*"