A SkinCare projekt

Somfai Ellák, ELTE IK Mesterséges Intelligencia Tanszék

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE
HUNGARY

PROGRAM FINANCED FROM THE NRDI FUND

Skincare Project 2020-2022

People: - Ellák Somfai (TKP)
- András Lőrincz
- Marcell Pólik
- Benjamin Baffy (TDK)
- Ádám Szijártó (TDK)

Funding: - EIT Digital (2020)
- TKP (as ELTE contribution)

Partners: - Degetel (France)
- Semmelweis University (Dept of Dermatology, Venereology and Dermatooncology)

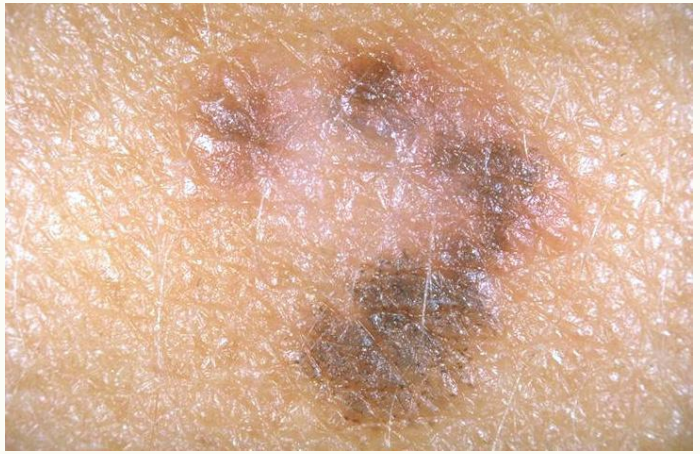
Objectives: - industrial code for skin lesion classification
- related academic projects (next slides)

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

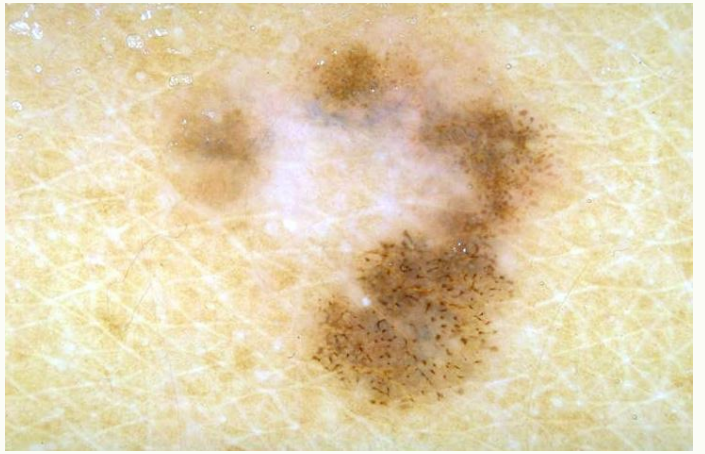
Challenges:

- *deadly melanoma* images resemble *benign naevus* images, even human experts need extensive training

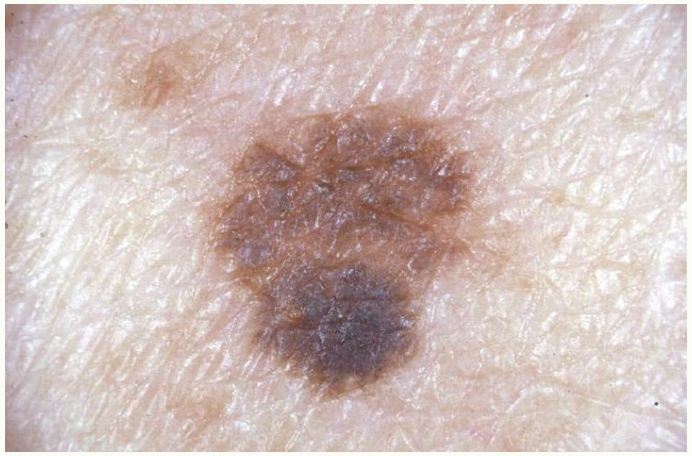
- dermoscopic images show more details, larger public dataset available; but clinical (plain macro) images are more suitable for remote / automated diagnosis

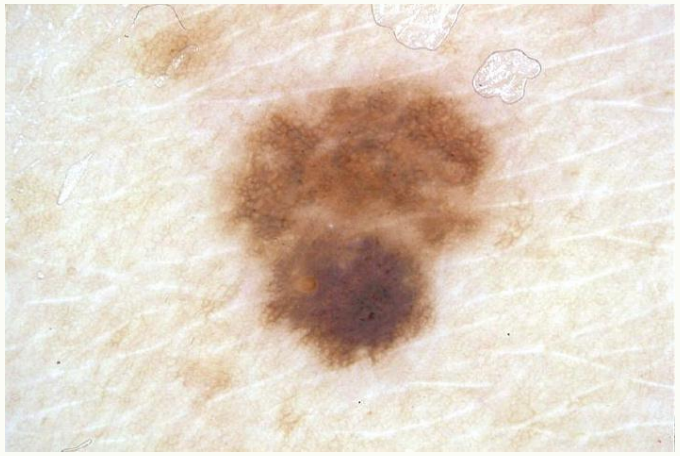


clinical (macro)



dermoscopic



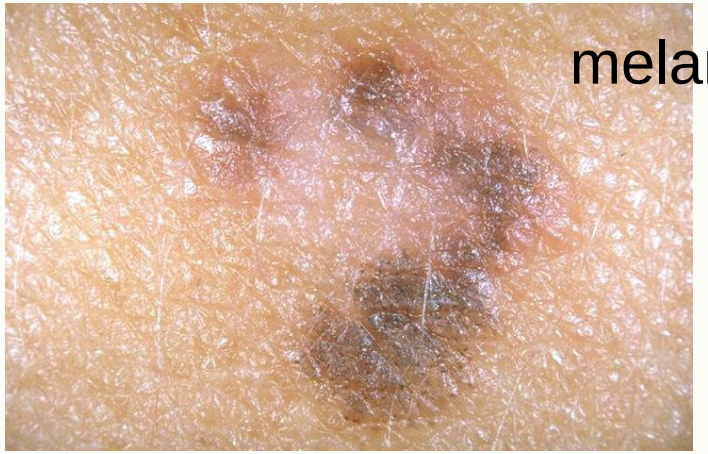


NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

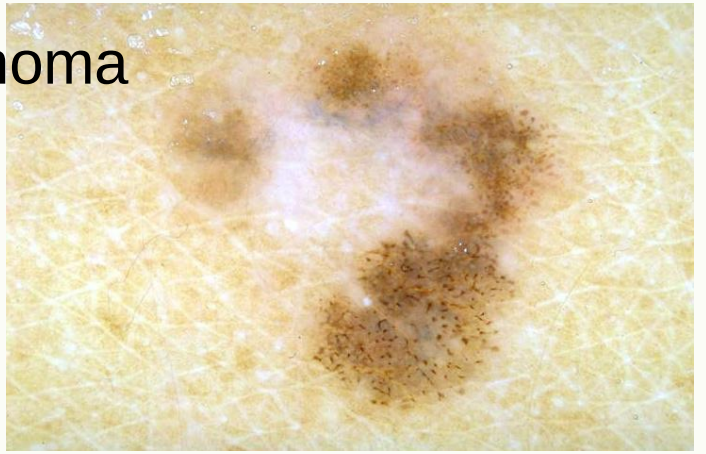
PROGRAM FINANCED FROM THE NRDI FUND

Challenges:

- *deadly melanoma* images resemble *benign naevus* images, even human experts need extensive training
- dermoscopic images show more details, larger public dataset available; but clinical (plain macro) images are more suitable for remote / automated diagnosis

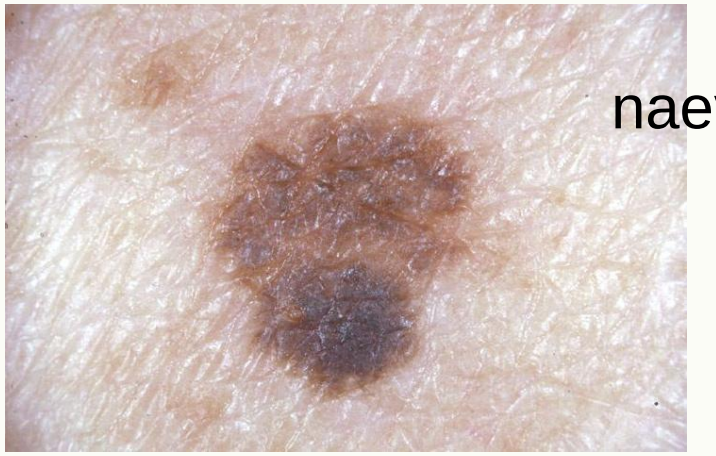


melanoma

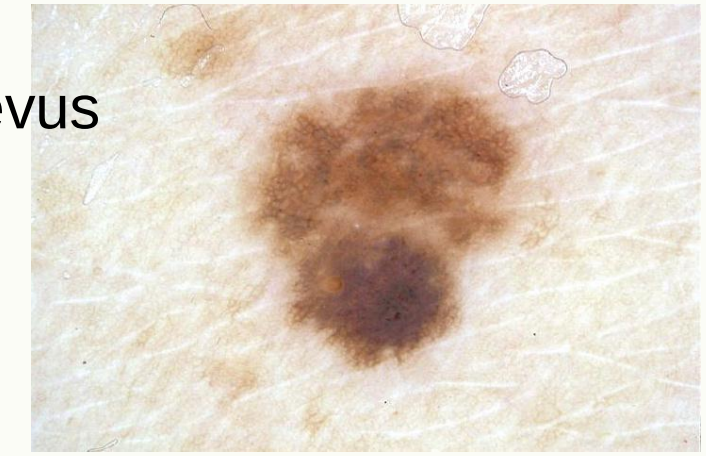


clinical (macro)

dermoscopic



naevus



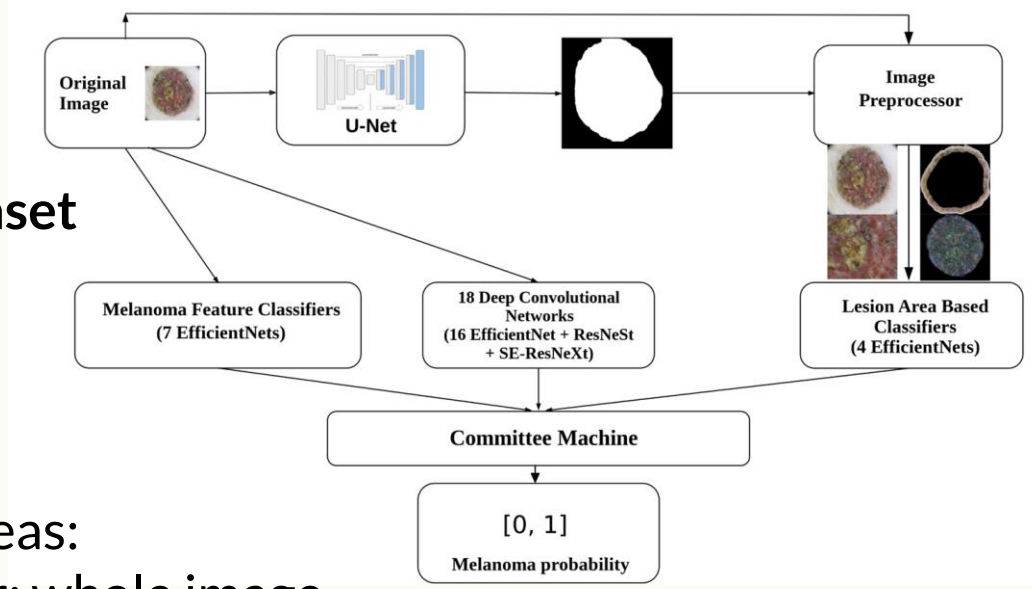
NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROGRAM FINANCED FROM THE NRDI FUND

Handling dataset dependence with model ensembles - skin lesion classification

- dataset dependence: a model trained on samples from clinic A performs better on *unseen* images from the same clinic than on images from clinic B - even though both should sample the same distribution

- solution: (1) **ensemble of heterogeneous models**
  (2) **trained on heterogeneous dataset**



- our model ensemble: 29 deep CNNs
  - 18 models of 2020 champion (brute force)
  - 3+1 feature classifiers - focused on key image areas: border, highlighted asymmetry, magnified center; whole image
  - 7 classifiers trained on dermatologically relevant features, such as blue-white veil, pigment network type etc.
  - model fusion: trained shallow NN

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

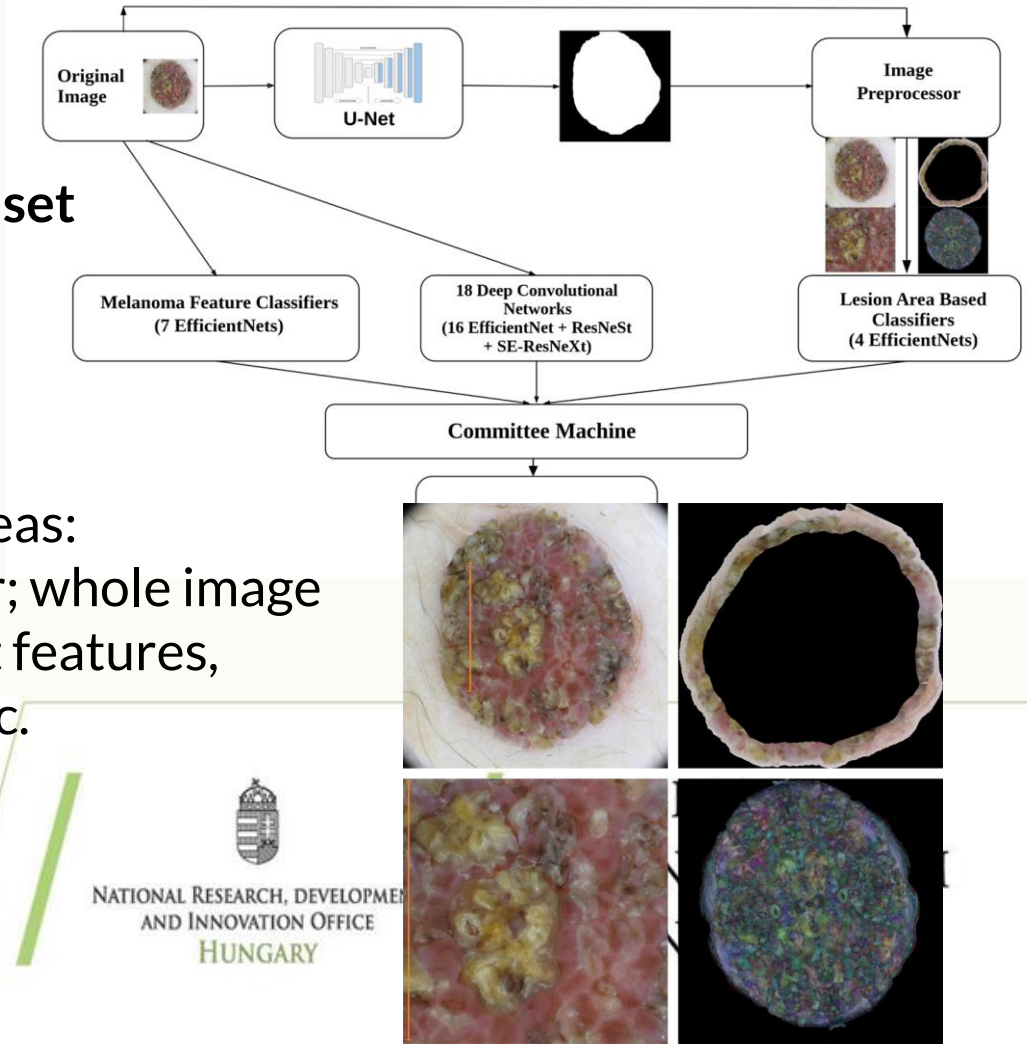
PROGRAM FINANCED FROM THE NRDI FUND

Handling dataset dependence with model ensembles - skin lesion classification

- dataset dependence: a model trained on samples from clinic A performs better on *unseen* images from the same clinic than on images from clinic B - even though both should sample the same distribution

- solution: (1) **ensemble of heterogeneous models**
  (2) **trained on heterogeneous dataset**

- our model ensemble: 29 deep CNNs
  - 18 models of 2020 champion (brute force)
  - 3+1 feature classifiers - focused on key image areas: border, highlighted asymmetry, magnified center; whole image
  - 7 classifiers trained on dermatologically relevant features, such as blue-white veil, pigment network type etc.
  - model fusion: trained shallow NN



NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE
HUNGARY

Handling dataset dependence with model ensembles - skin lesion classification

- performance: 93% balanced accuracy on training set (out-of-fold), 84-91% on other datasets

Balanced accuracy (%)	predict on out-of-fold		predict on other dataset		
	C-ISIC	DERM7D	DERM7D	PH2	Semmelweis
AVG-18	89.12		79.2	84.1	77.0
Calib.Prob.AVG-18	93.45		82.5	91.9	81.8
Committee-18	93.44		83.6	90.9	83.2
Committee-22	93.56		84.8	90.9	83.1
Committee-29	93.51	84.8		91.2	83.6
estimated error	±0.04	±0.2	±0.2	±0.3	±0.4

- the ensemble can be re-trained for clinical (macro) images - less image details, smaller public datasets transfer learning - for deepnets, for committee machine, or both?
  - naive approach (blue ←) yields poor results: fine-tuning dataset too small
  - best solution (red): use adaptive sharpness aware minimization with fine-tuning

transf. learning in deepnets	committee machine	DERM7M (train+test)	PAD-UFES (test)	MED-NODE (test)	Average over 3 datasets
no	plain	70.2±1.1	71.6±0.7	**89.4**±0.9	77.1±0.5
no	SAM	69.7±1.2	71.2±1.1	**89.4**±1.1	76.8±0.7
plain	plain	72.0±0.6	69.2±1.0	81.4±1.1	74.2±0.5
plain	ASAM	73.2±0.6	73.1±1.1	87.5±1.3	77.9±0.6
ASAM	plain	74.4±1.5	72.4±0.4	86.6±1.2	77.8±0.7
ASAM	ASAM	**74.9**±1.1	**74.1**±0.5	88.7±1.0	**79.2**±0.5

- manuscript submitted, under peer review

HUNGARY
PROGRAM
FINANCED FROM
THE NRDI FUND

Thickness prediction of melanoma lesions

- lesion thickness: proxy for tumor progression
  application: prioritizing the most urgent cases for treatment

- used single CNN to compare with literature results
  important: (1) reduce the effects of sample imbalance, (2) proper augmentation

- achieved 71% balanced accuracy for 3-way classification (3 discreet thickness classes)
  better than literature results
  Except one (Jaworek-Korjakowska et al. 2019): claimed 87% accuracy
  BUT: we discovered they made a mistake at data preparation, causing data leak to test set (reproduced their work). After fixing data leak: only 68% accuracy

- outlook for improvement:
  - use larger training set
  - use ensemble of CNNs

- manuscript submitted, under peer review

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

Thickness prediction of melanoma lesions

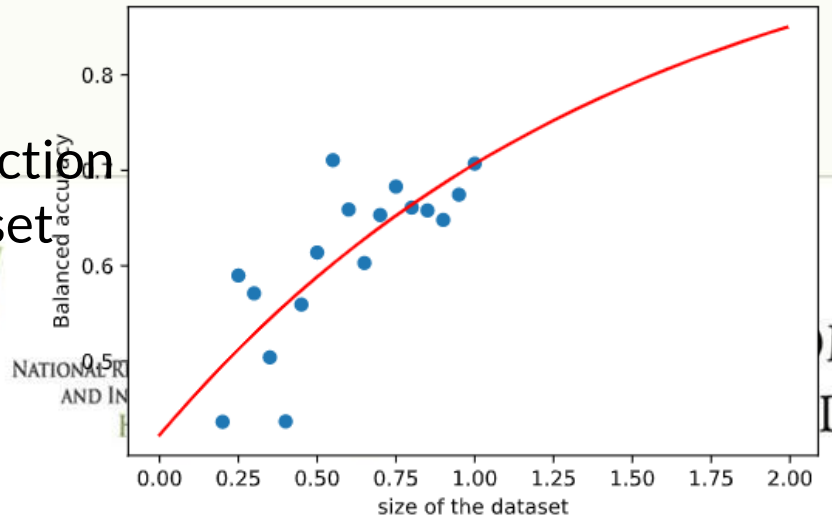
- lesion thickness: proxy for tumor progression
  application: prioritizing the most urgent cases for treatment

- used single CNN to compare with literature results
  important: (1) reduce the effects of sample imbalance, (2) proper augmentation

- achieved 71% balanced accuracy for 3-way classification (3 discreet thickness classes)
  better than literature results
  Except one (Jaworek-Korjakowska et al. 2019): claimed 87% accuracy
  BUT: we discovered they made a mistake at data preparation, causing data leak to test set (reproduced their work). After fixing data leak: only 68% accuracy

- outlook for improvement:
  - use larger training set
  - use ensemble of CNNs

performance projection for larger training set



- manuscript submitted, under peer review

Köszönöm a figyelmet!

www.elte.hu

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROGRAM FINANCED FROM THE NRDI FUND