Institute for Systems and Robotics | LISBOA

Interpretability in CAD Systems for Skin Cancer Diagnosis

Catarina Barata

LARSyS Laboratory of Robotics and Engineering Systems

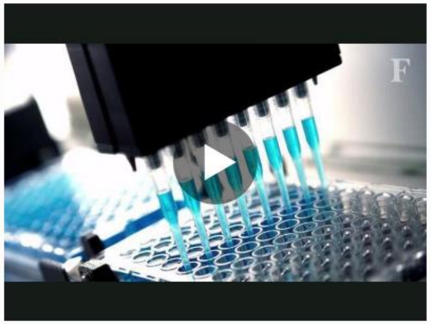
TÉCNICO LISBOA

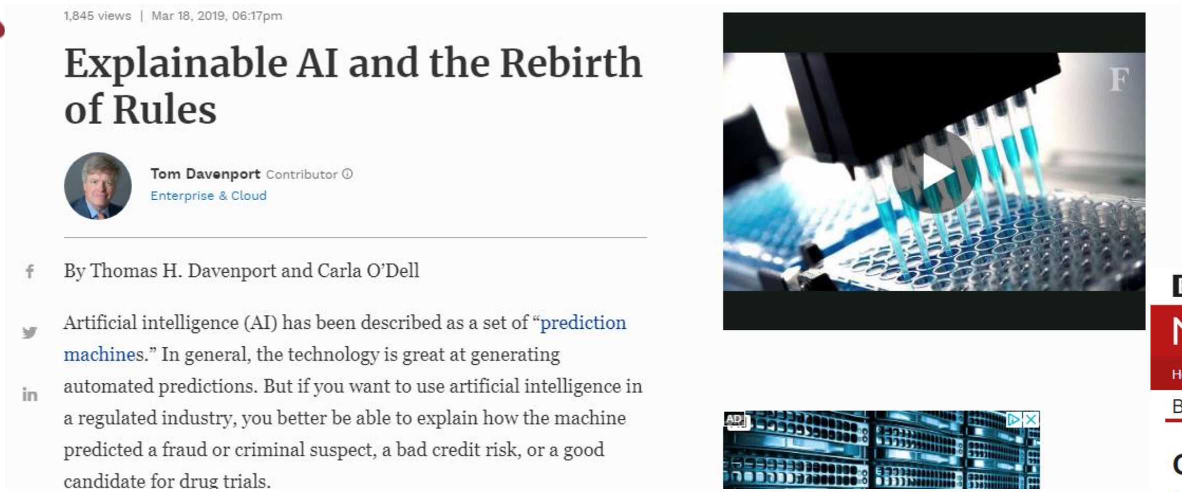
Explainable AI and the Rebirth of Rules

Tom Davenport Contributor
Enterprise & Cloud

By Thomas H. Davenport and Carla O'Dell

Artificial intelligence (AI) has been described as a set of "prediction machines." In general, the technology is great at generating automated predictions. But if you want to use artificial intelligence in a regulated industry, you better be able to explain how the machine predicted a fraud or criminal suspect, a bad credit risk, or a good candidate for drug trials.





1,845 views | Mar 18, 2019, 06:17pm

Explainable AI and the Rebirth of Rules

Tom Davenport Contributor
Enterprise & Cloud

By Thomas H. Davenport and Carla O'Dell

Artificial intelligence (AI) has been described as a set of "prediction machines." In general, the technology is great at generating automated predictions. But if you want to use artificial intelligence in a regulated industry, you better be able to explain how the machine predicted a fraud or criminal suspect, a bad credit risk, or a good candidate for drug trials.

BBC Sign in | News | Sport | Weather | Shop | Reel | Travel | More

NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts

Business | Market Data | Global Trade | Companies | Entrepreneurship | Technology of Busine

Can we trust AI if we don't know how it works?

By Marianne Lehnis
Technology of Business reporter

15 June 2018

Share

GETTY IMAGES

What's it thinking? Will AI become too clever for us?

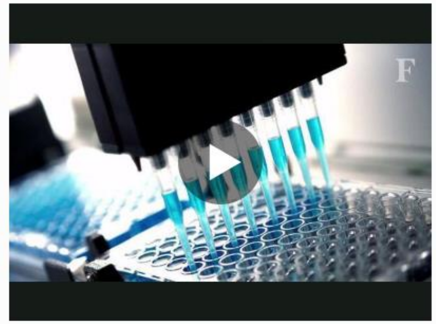
1,845 views | Mar 18, 2019, 06:17pm

Explainable AI and the Rebirth of Rules

Tom Davenport Contributor
Enterprise & Cloud

By Thomas H. Davenport and Carla O'Dell

Artificial intelligence (AI) has been described as a set of "prediction machines." In general, the technology is great at generating automated predictions. But if you want to use artificial intelligence in a regulated industry, you better be able to explain how the machine predicted a fraud or criminal suspect, a bad credit risk, or a good candidate for drug trials.



Tech

More clinical evidence needed to accelerate adoption of AI-enabled decision support: report

by Heather Landi | Jan 28, 2019 12:30pm

Regulatory issues, improved product labeling and patient privacy concerns need to be addressed before AI is safely and widely adopted as part of clinical decision support, according to a team of healthcare and AI experts. (monsitj/iStockPhoto)

About the Author

Heather Landi
Senior Editor

BBC Sign in | News | Sport | Weather | Shop | Reel | Travel | More

NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts | H

Business | Market Data | Global Trade | Companies | Entrepreneurship | Technology of Busine

Can we trust AI if we don't know how it works?

By Marianne Lehnis
Technology of Business reporter

15 June 2018

Share

GETTY IMAGES

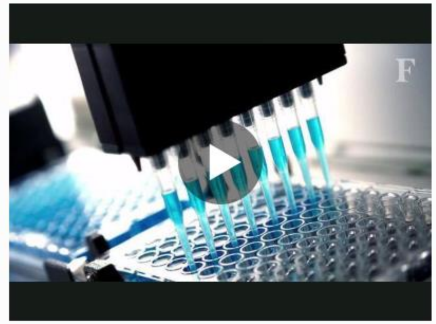
What's it thinking? Will AI become too clever for us?

Explainable AI and the Rebirth of Rules

Tom Davenport Contributor
Enterprise & Cloud

By Thomas H. Davenport and Carla O'Dell

Artificial intelligence (AI) has been described as a set of "prediction machines." In general, the technology is great at generating automated predictions. But if you want to use artificial intelligence in a regulated industry, you better be able to explain how the machine predicted a fraud or criminal suspect, a bad credit risk, or a good candidate for drug trials.



Can we trust AI if we don't know how it works?

By Marianne Lehnis

Tech

More clinical evidence needed to accelerate adoption AI-enabled decision support: report

by Heather Landi | Jan 28, 2019 12:30pm

Regulatory issues, improved product labeling and patient privacy concerns need to be addressed before AI is safely and widely ado[pted as] part of clinical decision support, according to a team of healthcare and AI experts. (monsitj/iStockPhoto)

PROBLEM SOLVE

GDPR regulations put premium on transparent AI

As the EU's GDPR regulations go into effect, enterprises must focus on building transparency in AI applications so that algorithms' decisions can be explained.

George Lawton

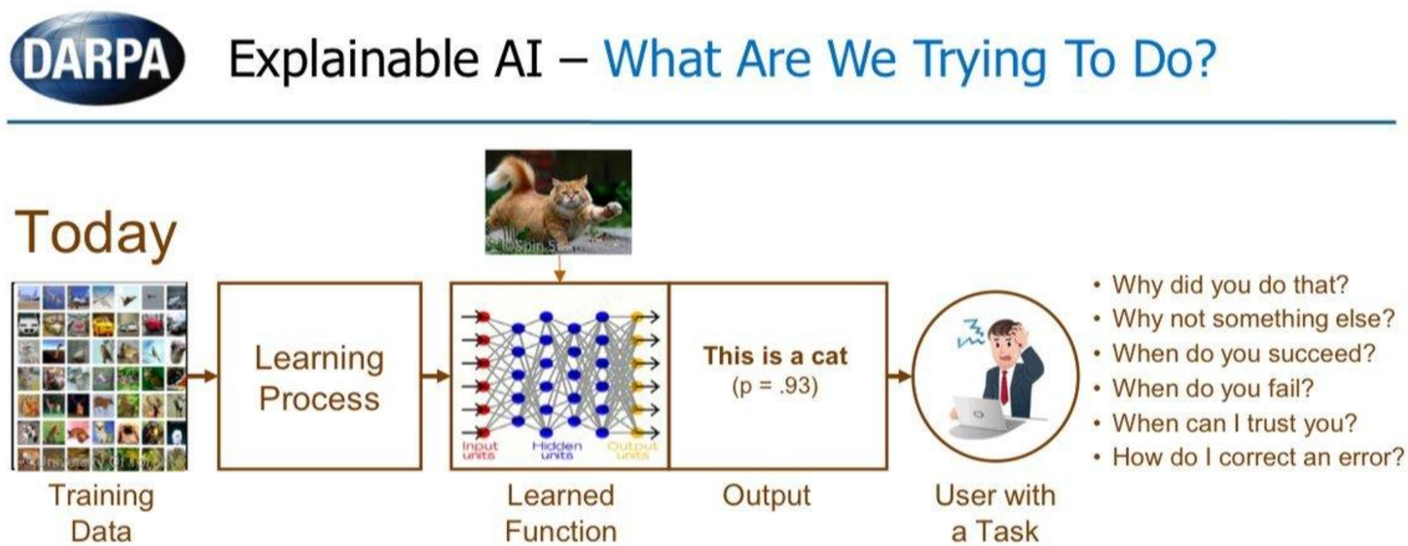
The European Union's new GDPR regulations could shake up the way enterprises craft algorithms to make decisions, particularly when it comes to building transparent AI applications.

"GDPR will impact all industries, and has particularly relevant ramifications for AI developers and AI-enabled businesses," said Dillon Erb, CEO at Paperspace Co., an AI cloud provider.

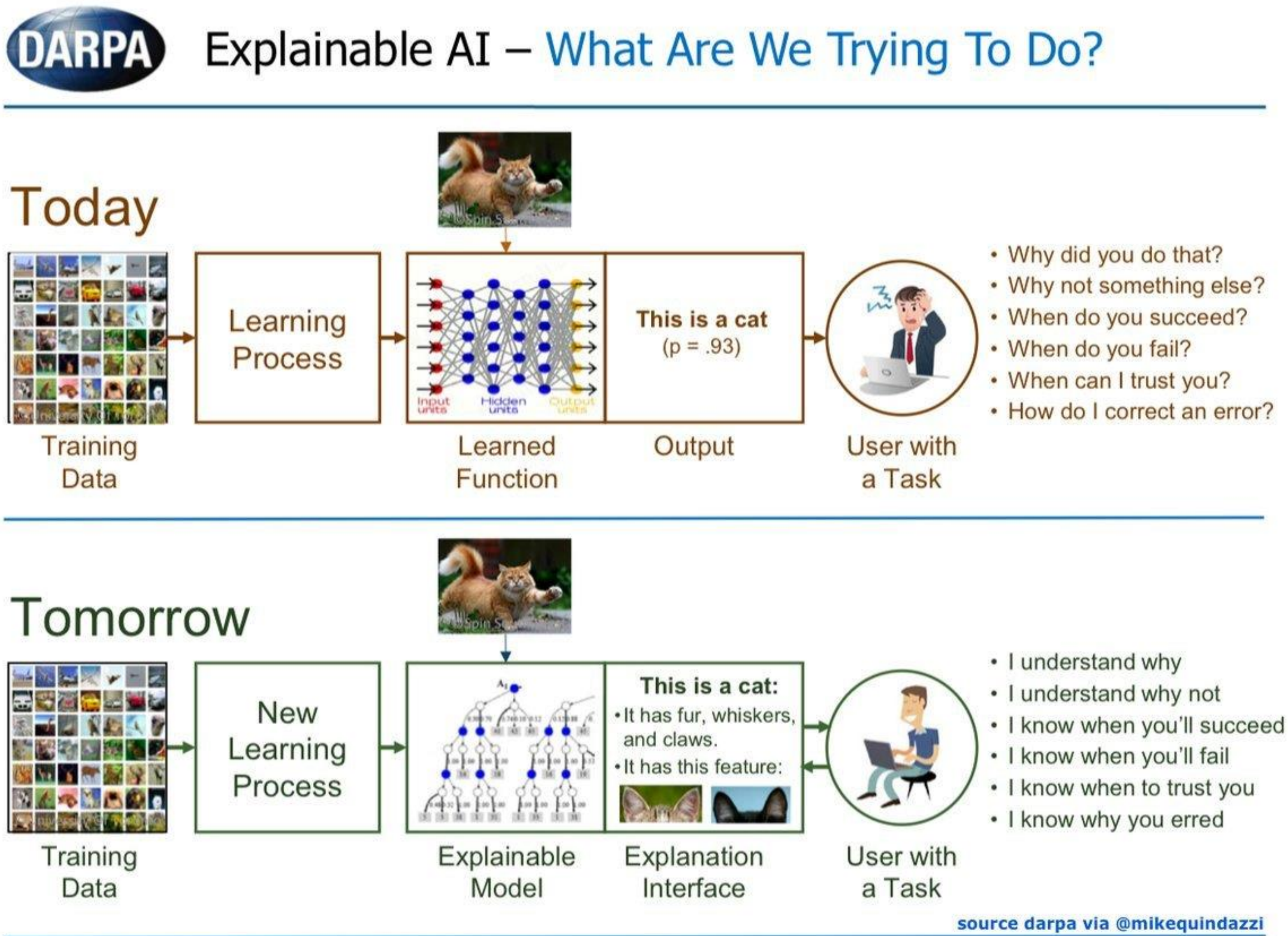
Outline

- What do we mean by explainability and interpretability?
- Interpretability in dermoscopy – A historical perspective
- Where to go next?

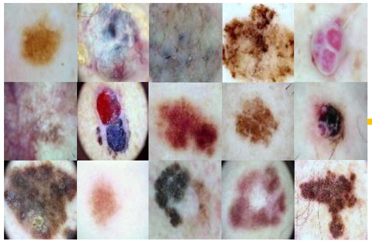
Explainability and Interpretability



Explainability and Interpretability

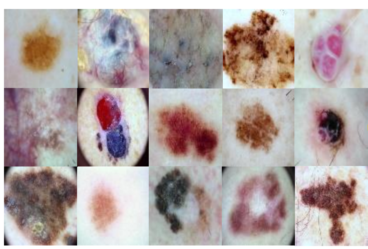


What does it mean for dermoscopy?

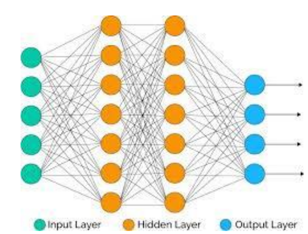


Learning Process

What does it mean for dermoscopy?

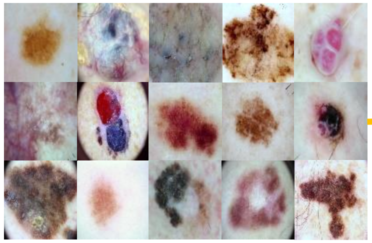


Learning Process

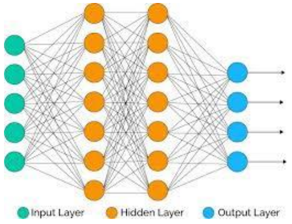


Learned Model

What does it mean for dermoscopy?



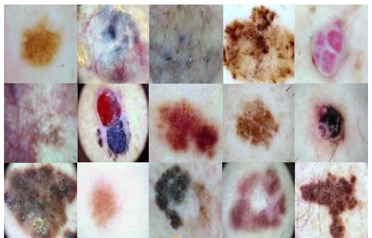
Learning Process



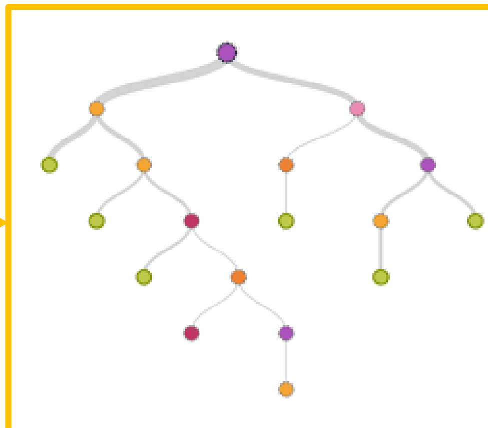
Learned Model

Melanoma (p=0.7)

What does it mean for dermoscopy?



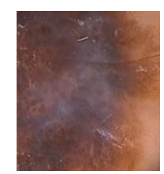
Learning Process



Interpretable
&
Structured

This lesion is a **melanoma** because:

- It is melanocytic;
- It has more than 3 colors.
- This structure was detected:



The Design of an Interpretable Model

- What should we have in mind when designing an interpretable model?

The Design of an Interpretable Model

- What should we have in mind when designing an interpretable model?
  - **The final user! (Dermatologists or Patients)**
  - **This is a collaborative process!**

The Design of an Interpretable Model

- What should we have in mind when designing an interpretable model?
  - **The final user! (Dermatologists or Patients)**
  - **This is a collaborative process!**

- Where should we act to improve interpretability?
  1. Features?
  2. Classifier?
  3. Infer from the black-box model?

The Design of an Interpretable Model

- What should we have in mind when designing an interpretable model?
  - **The final user! (Dermatologists or Patients)**
  - **This is a collaborative process!**

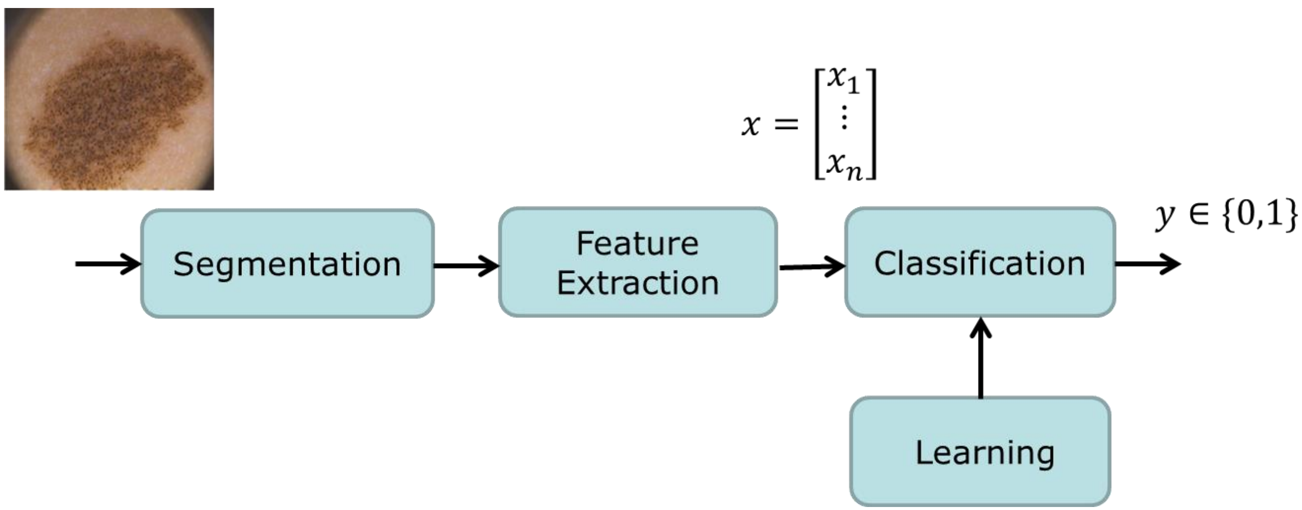
- Where should we act to improve interpretability?
  1. **Clinically Inspired Features**
  2. **Structured & Explainable Classifiers**
  3. **Model Explainability - Visualization**

The Design of an Interpretable Model

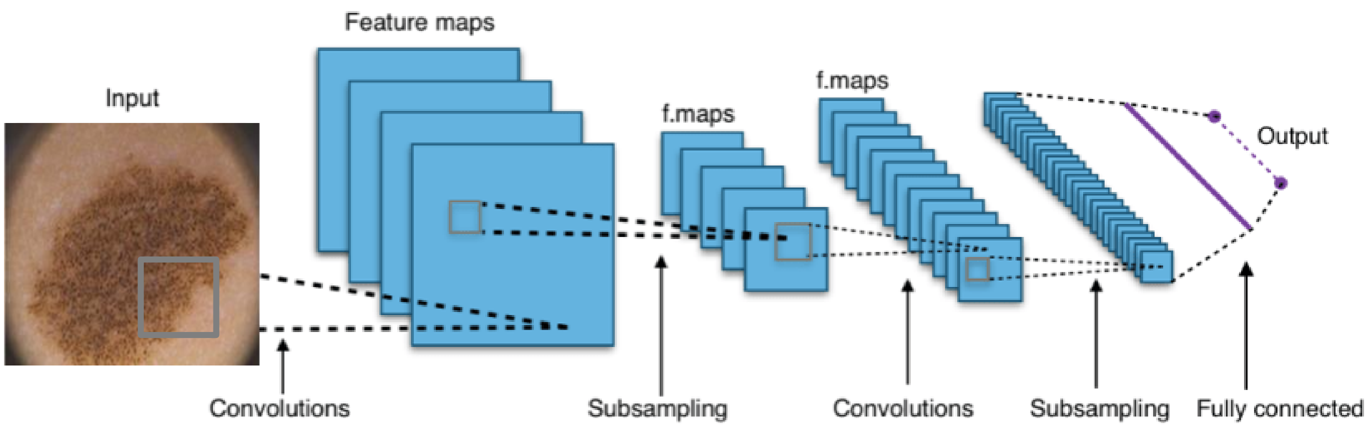
- What should we have in mind when designing an interpretable model?
  - **The final user! (Dermatologists or Patients)**
  - **This is a collaborative process!**

- Where should we act to improve interpretability?
  1. **Clinically Inspired Features**
  2. **Structured & Explainable Classifiers**
  3. **Model Explainability - Visualization**

Dermoscopy Image Diagnosis



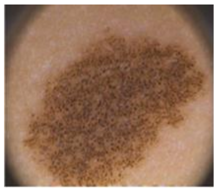
Traditional CAD

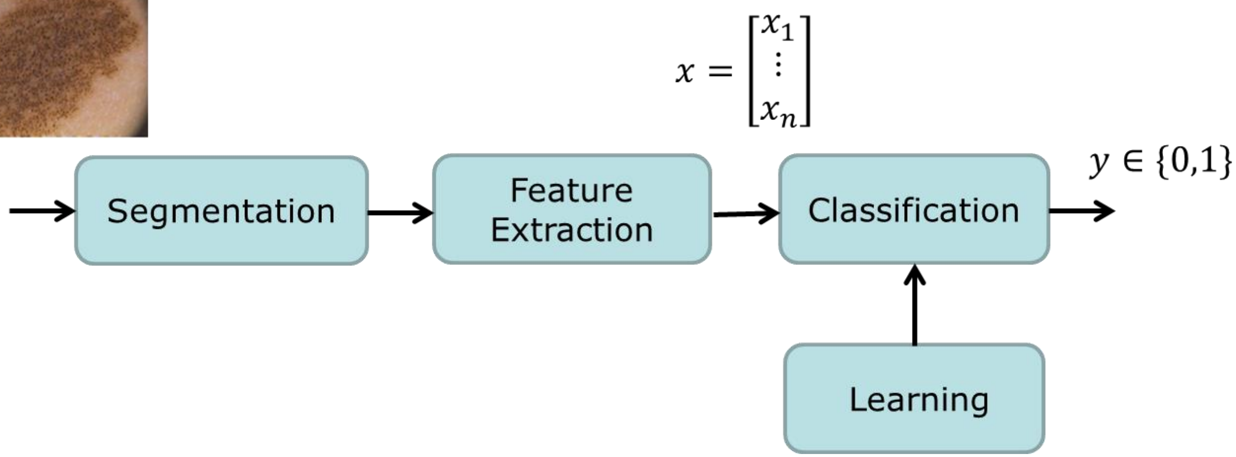


End-to-End CAD

INTERPRETABILITY IN TRADITIONAL CADS

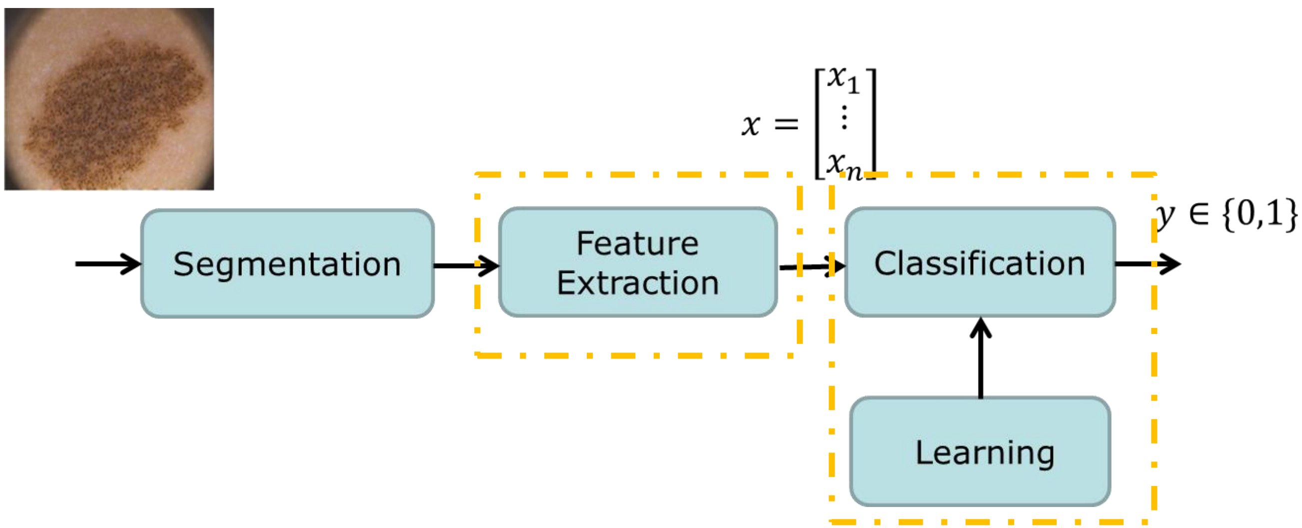
Dermoscopy Image Diagnosis





Traditional CAD

Dermoscopy Image Diagnosis



Traditional CAD

Interpretable Features

- What kind of features is interpretable?
  - Inspired by medical knowledge

Interpretable Features

Traditional Hand-Crafted Features

Asymmetry

- Moments of inertia
- Shape, color, and texture maps
- Centroid location

Border/Shape

- Fractals
- Intensity profiles
- Wavelets

Color

- Color statistics
- Relative colors
- Color quantization
- Different color spaces

Texture

- Gabor filters
- Haralick
- LBP
- Gradient based descritptors

- These features were inspired by medical knowledge.
- But were these features interpretable?

Interpretable Features

Traditional Hand-Crafted Features

Asymmetry

- Moments of inertia
- Shape, color, and texture maps
- Centroid location

Border/Shape

- Fractals
- Intensity profiles
- Wavelets

Color

- Color statistics
- Relative colors
- Color quantization
- Different color spaces

Texture

- Gabor filters
- Haralick
- LBP
- Gradient based descritptors

Medical Counterparts (ABCD Rule)

Asymmetry

- Maximum of 2 axes
- Contour, colors, and structures

Border/Shape

- Abrupt ending of pigments
- Analysis of 8 segments

Color

- Identification of up to six colors

Structures

- Identification of up to 5 structures

Interpretable Features

Traditional Hand-Crafted Features

Asymmetry
- Moments of inertia
- Shape, color, and texture maps
- Centroid location

Border/Shape
- Fractals
- Intensity profiles
- Wavelets

Color
- Color statistics
- Relative colors
- Color quantization
- Different color spaces

Texture
- Gabor filters
- Haralick
- LBP
- Gradient based descritptors

- These features were inspired by medical knowledge.
- But they did not have a **true match** with medical findings.

Interpretable Features

- What kind of features is interpretable?
    - **Inspired by medical knowledge**
    - **Have a direct relationship with clinical findings**

Interpretable Features

- What kind of features is interpretable?
  - **Inspired by medical knowledge**
  - **Have a direct relationship with clinical findings**
- How can we extract them?

Interpretable Features

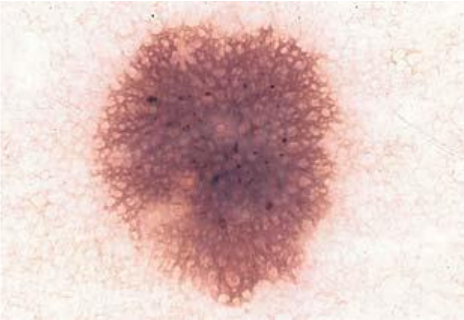
- What kind of features is interpretable?
  - **Inspired by medical knowledge**
  - **Have a direct relationship with clinical findings**

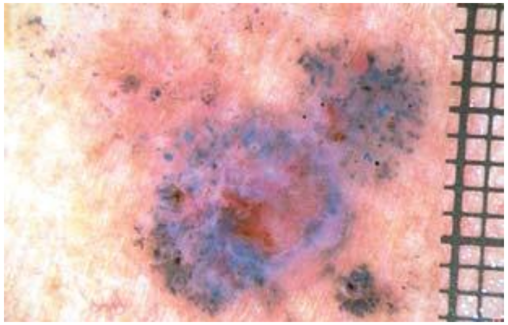
- How can we extract them?
  - Dermatologists use multiple cues to diagnose skin lesions
  - These cues can be seen as **"clinically inspired features"**

Clinically Inspired Features

- Different groups addressed the detection of at least one medical feature.

- There are **three** types of medical features
  - Global patterns (Pehamberger, 1987)

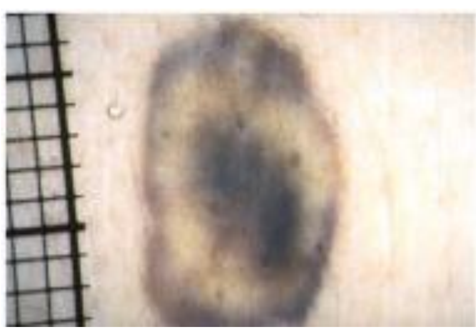


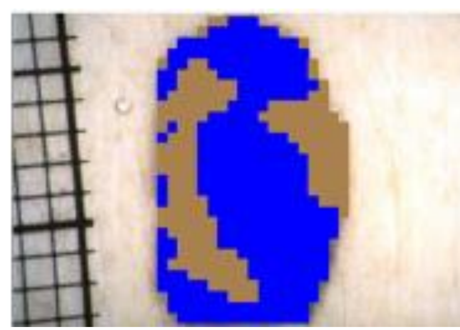


Clinically Inspired Features

- Different groups addressed the detection of at least one medical feature.

- There are **three** types of medical features
  - Global patterns (Pehamberger et al. 1987)
  - Colors (ABCD Rule, Stolz et al. 1994)



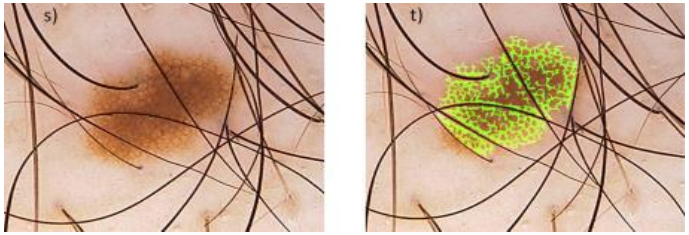


Barata et al., CVIU, 2016

Clinically Inspired Features

- Different groups addressed the detection of at least one medical feature.

- There are **three** types of medical features
  - Global patterns (Pehamberger et al. 1987)
  - Colors (ABCD Rule, Stolz et al. 1994)
  - Dermoscopic structures (ABCD Rule/7-point checklist)



Barata et al. IEEE TBME, 2012

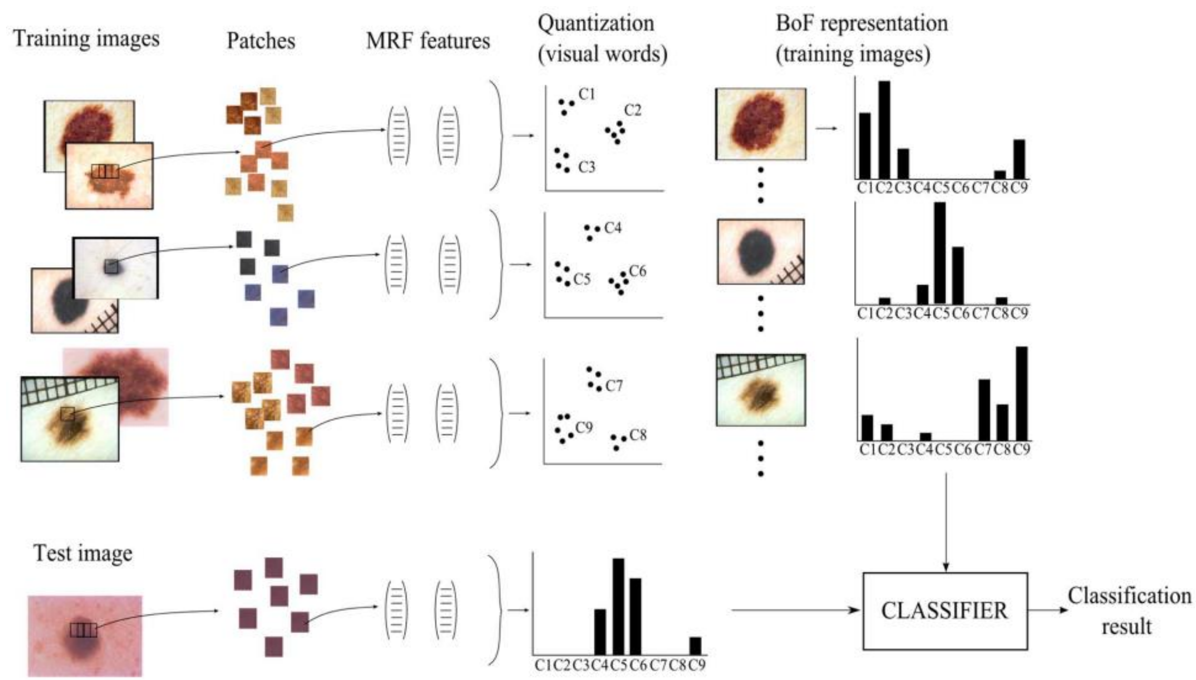
Clinically Inspired Features

- Different groups addressed the detection of at least one medical feature.

- There are **three** types of medical features
  - Global patterns (Pehamberger et al. 1987)
  - Colors (ABCD Rule, Stolz et al. 1994)
  - Dermoscopic structures (ABCD Rule/7-point checklist)

Global Patterns

- Detection of 5 patterns:
  - Globular
  - Homogeneous
  - Reticular
  - Multicomponent



Saéz et al., IEEE TMI, 2014

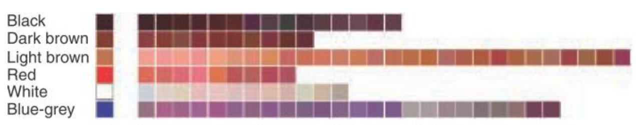
Clinically Inspired Features

- Different groups addressed the detection of at least one medical feature.

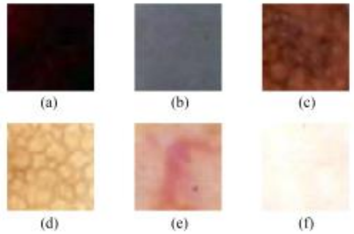
- There are **three** types of medical features
  - Global patterns (Pehamberger et al. 1987)
  - Colors (ABCD Rule, Stolz et al. 1994)
  - Dermoscopic structures (ABCD Rule/7-point checklist)

Detection of Colors

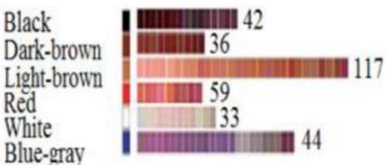
- **Main idea**
  1. Extract representative patches for each color



Seidenari et al., BJD, 2003



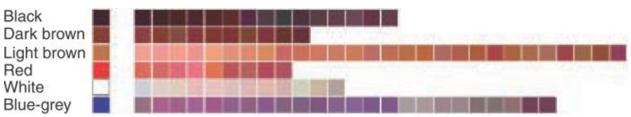
Sáez et al., IEEE JBHI, 2019



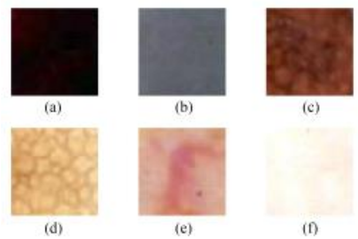
Sabbaghi et al., IEEE JBHI, 2019

Detection of Colors

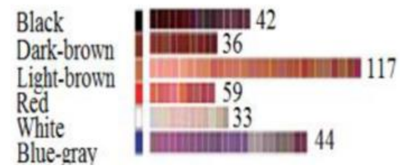
- **Main idea**
  1. Extract representative patches for each color



Seidenari et al., BJD, 2003



Sáez et al., IEEE JBHI, 2019

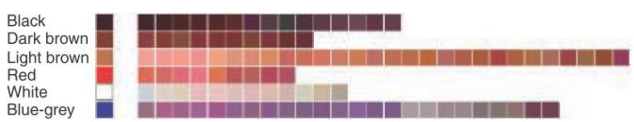


Sabbaghi et al., IEEE JBHI, 2019

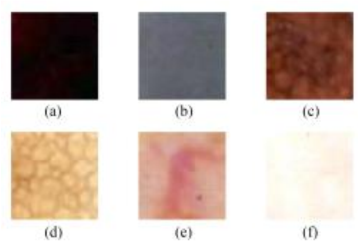
  2. Learn some representation for the pallete

Detection of Colors

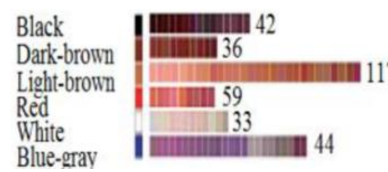
- **Main idea**
  1. Extract representative patches for each color



Seidenari et al., BJD, 2003



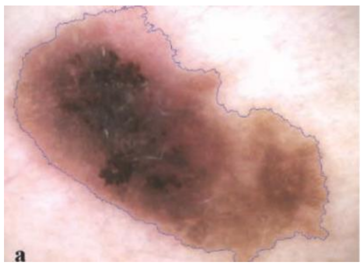
Sáez et al., IEEE JBHI, 2019

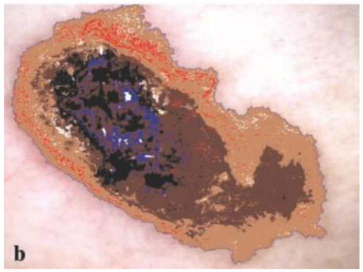


Sabbaghi et al., IEEE JBHI, 2019

  2. Learn some representation for the pallete

  3. Associate new pixels/patches to the pallete





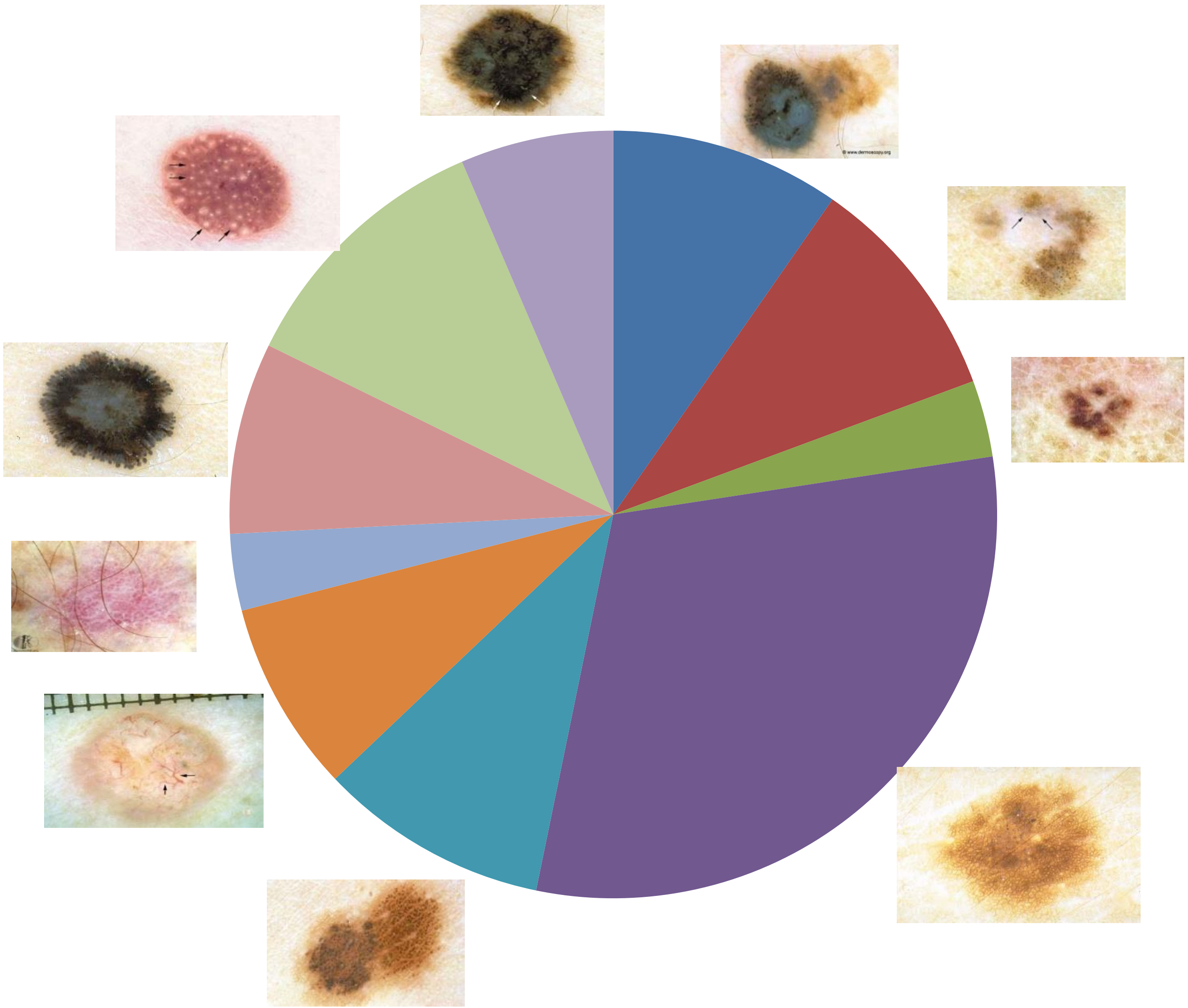
Seidenari et al., BJD, 2003

Clinically Inspired Features

- Different groups addressed the detection of at least one medical feature.

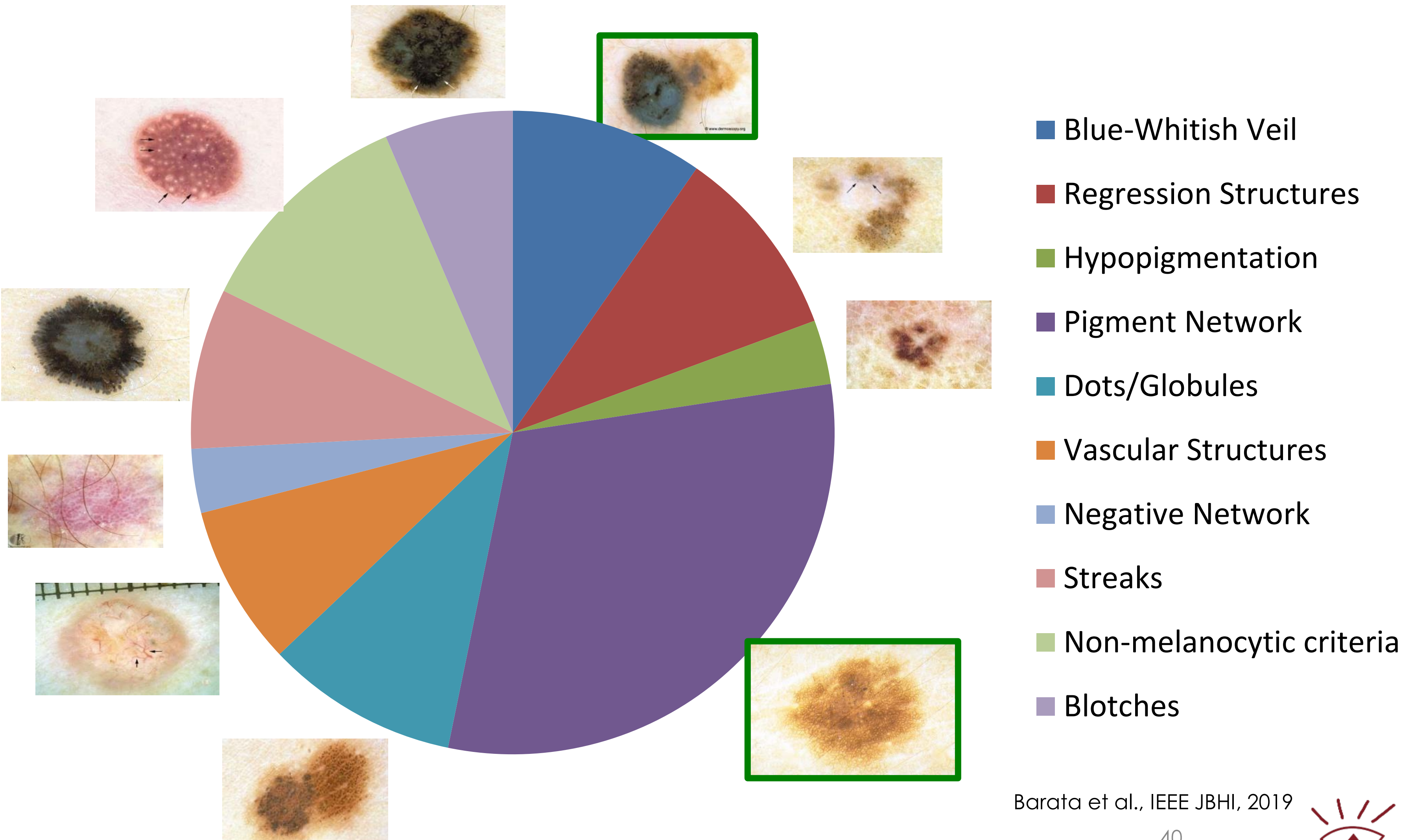
- There are **three** types of medical features
  - Global patterns (Pehamberger et al. 1987)
  - Colors (ABCD Rule, Stolz et al. 1994)
  - Dermoscopic structures (ABCD Rule/7-point checklist)

Detection of Dermoscopic Structures



Barata et al., IEEE JBHI, 2019

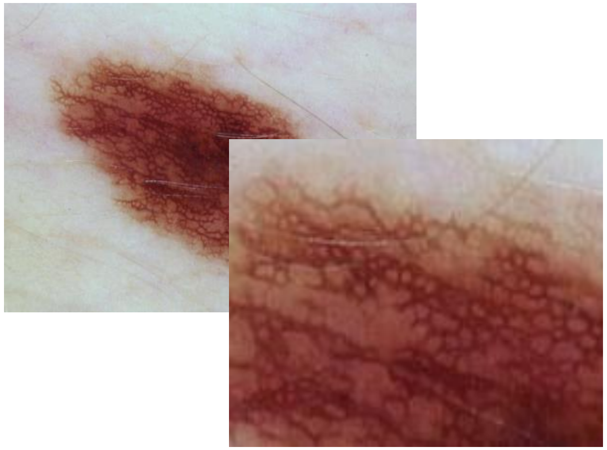
Detection of Dermoscopic Structures



Barata et al., IEEE JBHI, 2019

Pigment Network

- **Main ideas:**
  1. Explore the geometric and color properties of pigment network



PERGAMON

Computerized Medical Imaging and Graphics 22 (1998) 375–389

Computerized Medical Imaging and Graphics

Techniques for a structural analysis of dermatoscopic imagery

Matthew G. Fleming[a], Carsten Steger[b], Jun Zhang[c], Jianbo Gao[c], Armand B. Cognetta[d], Ilya Pollak[e], Charles R. Dyer[f]

[a]Department of Dermatology, Medical College of Wisconsin and Zablocki VA Hospital, Milwaukee, WI, USA
[b]Forschungsgruppe Bildverstehen, Informatik IX, Technische Universität München, Munich, Germany
[c]Department of Computer Science and Electrical Engineering, University of Wisconsin, Milwaukee, WI, USA
[d]Private Practice, Tallahassee, FL, USA
[e]Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA
[f]Department of Computer Science, University of Wisconsin, Madison, WI, USA

Received 15 May 1998; received in revised form 14 September 1998; accepted 14 September 1998

Abstract

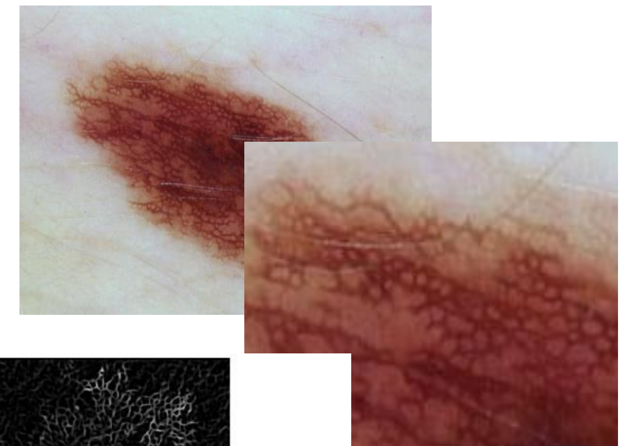
Techniques were developed for automated detection and characterization of dermatoscopic structures, including the pig
brown globules. These techniques incorporate algorithms for grayscale shape extraction based on differential geometry dev
a snake algorithm, and a modification of the region competition strategy of Zhu and Yuille. A novel approach was dev
segmentation of pigmented lesions, based on stabilized inverse diffusion equations. Procedures for detection of air bul
dermatoscopic images are also reported.

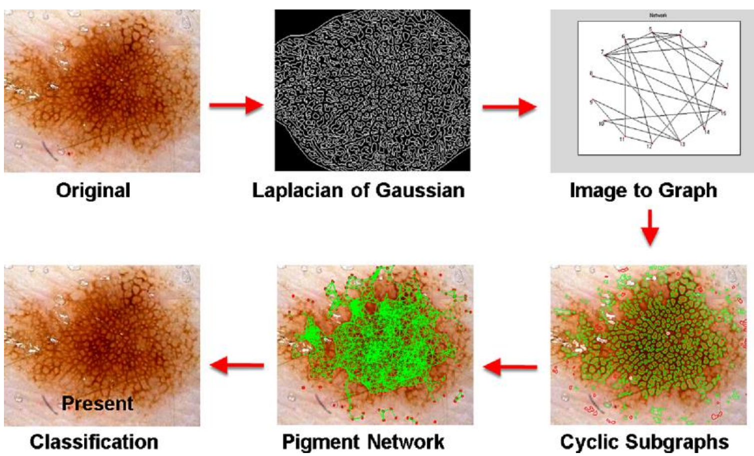
Keywords: Melanoma; Nevus; Dermatoscopy; Dermoscopy; Epiluminescence microscopy; Image analysis

The value of the statiatical approaches should become clearer with time, as more lesions are evaluated and additional groups involved. However, we have been interested in exploring an alternative, structural approach to dermatoscopic image analysis. This approach seeks to model human interpretation more closely, by extracting and assessing the classical dermatoscopic features. If such extraction

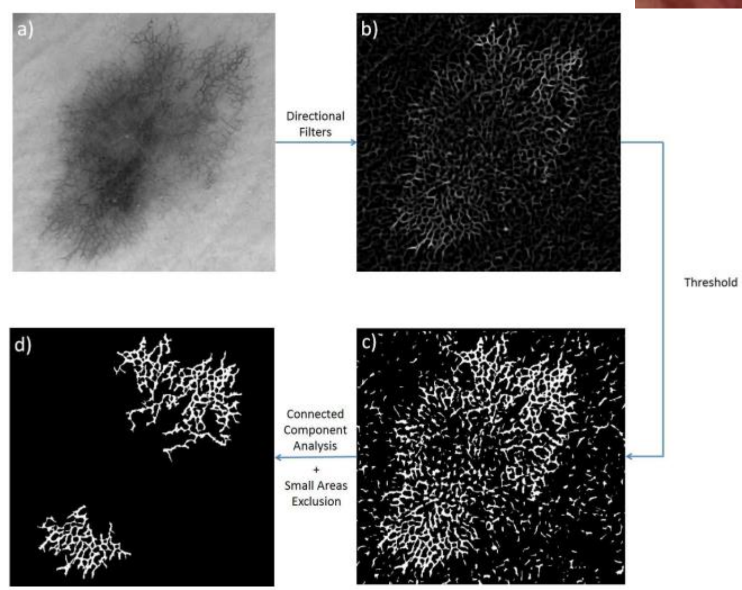
Pigment Network

- **Main ideas:**
  1. Explore the geometric and color properties of pigment network





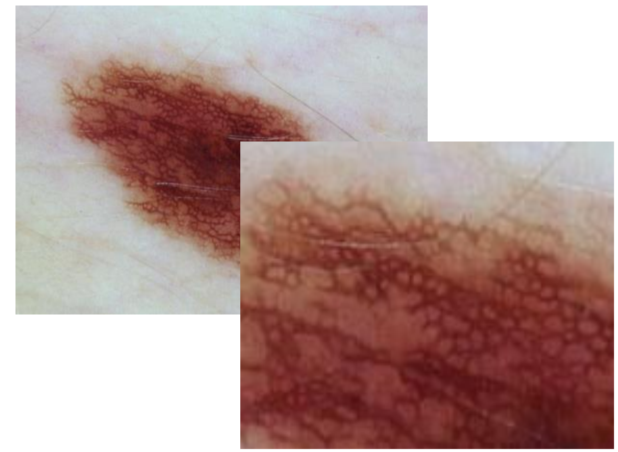
Sadeghi et al., CMIG, 2011

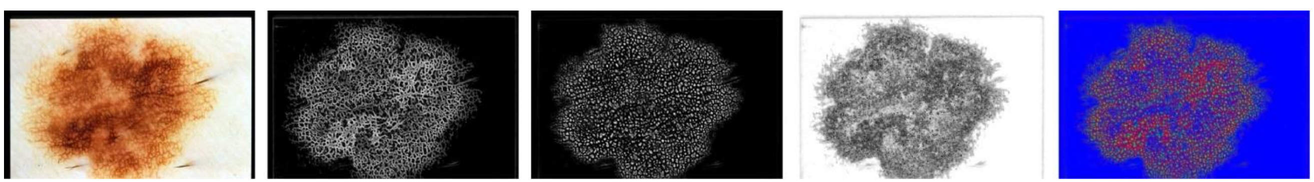


Barata et al. IEEE TBME, 2012

Pigment Network

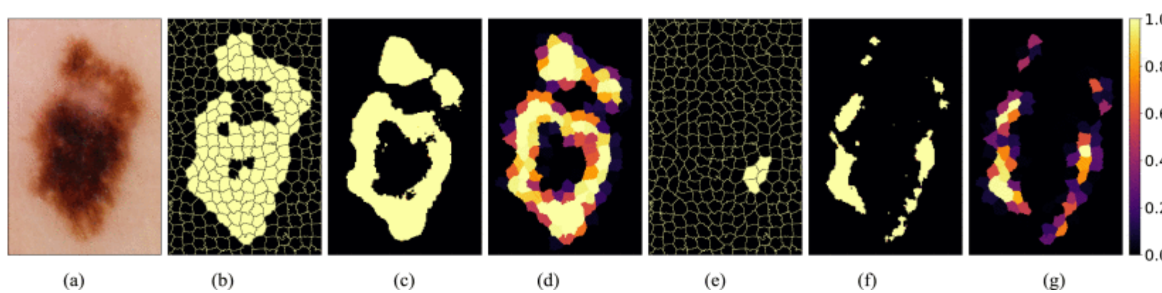
- **Main ideas:**
  1. Explore the geometric and color properties of pigment network
  2. Rely on machine learning algorithms





Garcia-Arroyo et al., CMIG, 2018

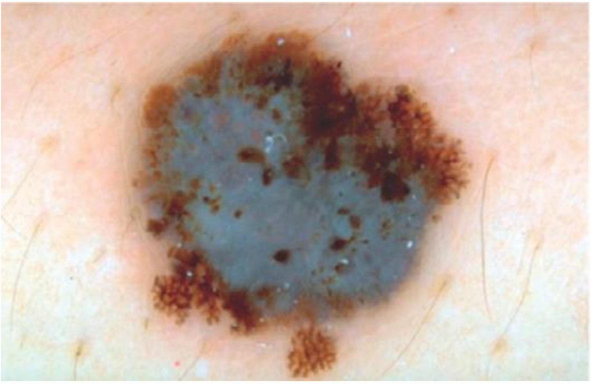
Kawahara et al., IEEE JBHI, 2019



Blue-Whitish Veil

- **Main idea:**
  1. Learn a color palette

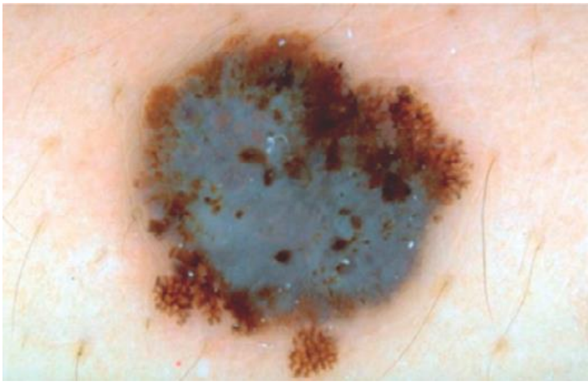
Madooei et al., MICCAI'13



Blue-Whitish Veil

- **Main idea:**
  1. Learn a color palette

Madooei et al., MICCAI'13

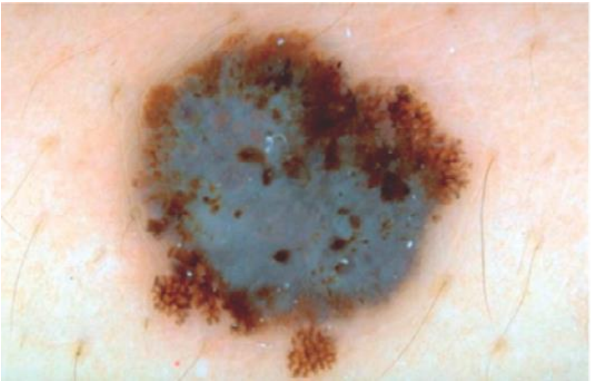


  2. Learn a representation

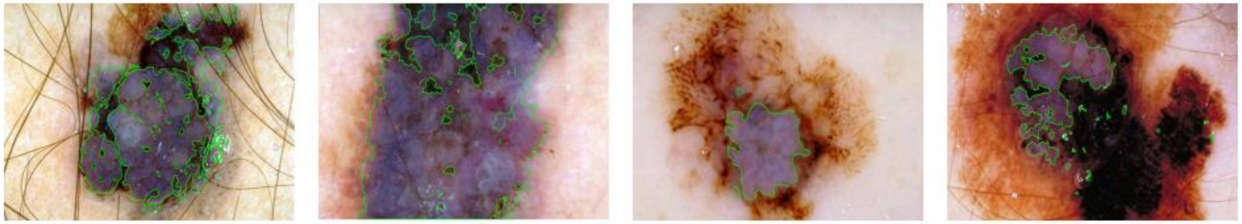
Blue-Whitish Veil

- **Main idea:**
  1. Learn a color palette

Madooei et al., MICCAI'13

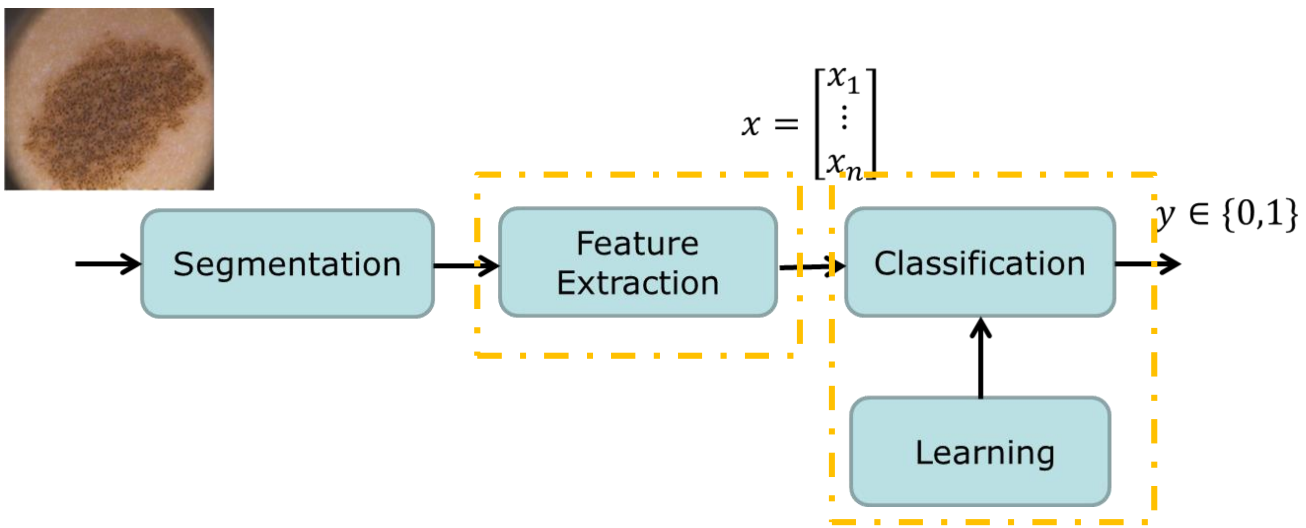


  2. Learn a representation
  3. Match new patches/pixels



Madooei et al., MICCAI'13

Dermoscopy Image Diagnosis



Traditional CAD

Interpretable Classifiers

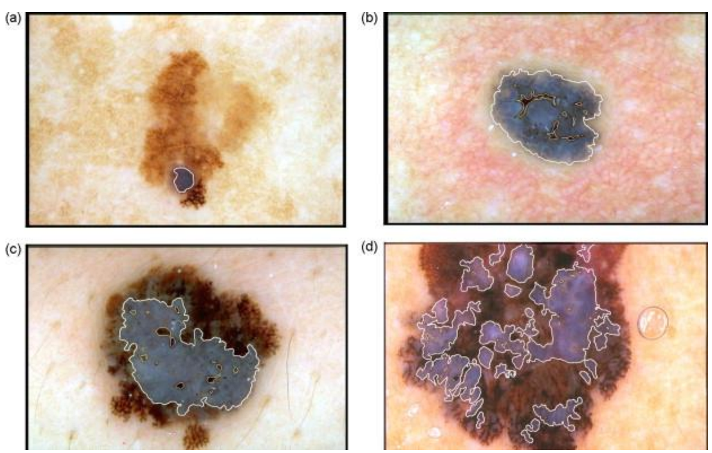
- What is an interpretable classifier?
  - **A classifier that is able to explain its decision**

Interpretable Classifiers

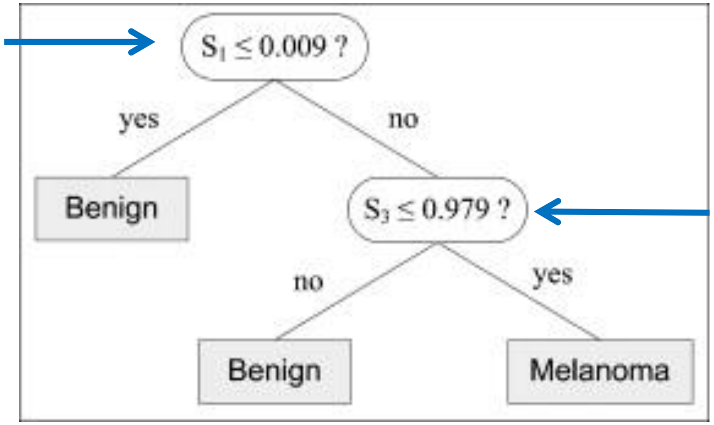
- What is an interpretable classifier?
    - **A classifier that is able to explain its decision based on medical knowledge**

Interpretable Classifiers

- What is an interpretable classifier?
  - **A classifier that is able to explain its decision based on medical knowledge**



Area

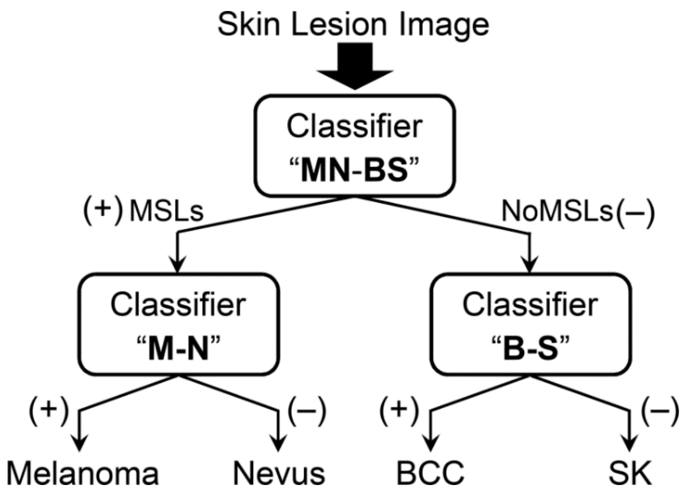


Shape

Celebi et al., CMIG, 2008

Interpretable Classifiers

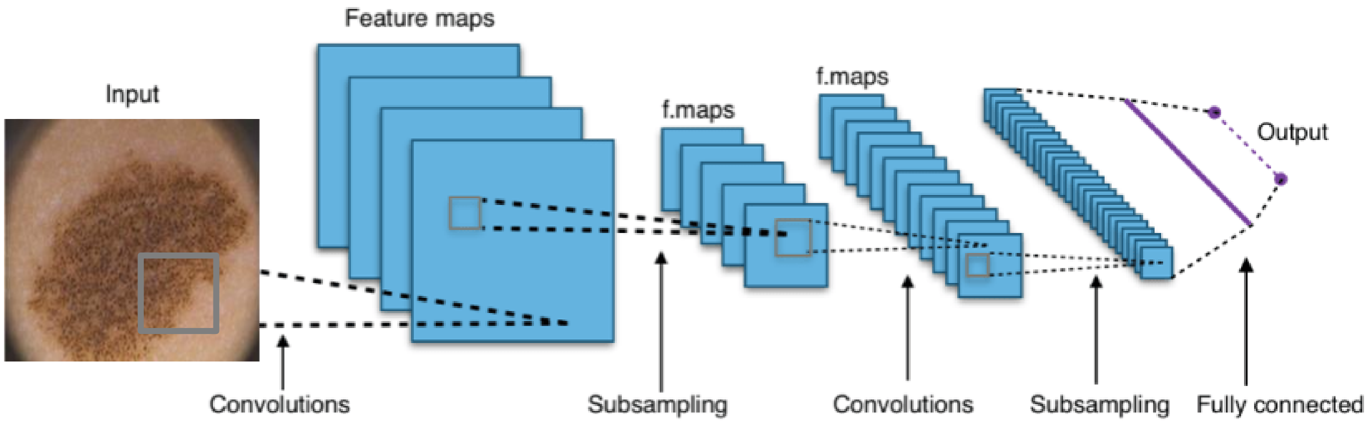
- What is an interpretable classifier?
  - **A classifier that is able to explain its decision based on medical knowledge**
  - **A structured classifier that incorporates medical knowledge**



Shimizu et al., IEEE TBME, 2014

INTERPRETABILITY IN END-TO-END CADS

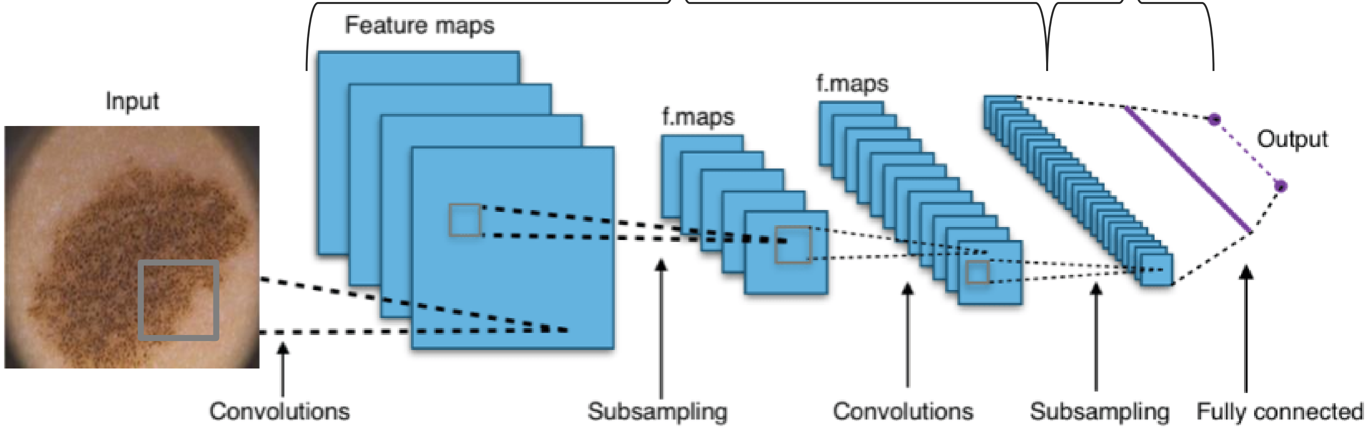
Dermoscopy Image Diagnosis



End-to-End
CAD

Dermoscopy Image Diagnosis

Feature Extraction Decision

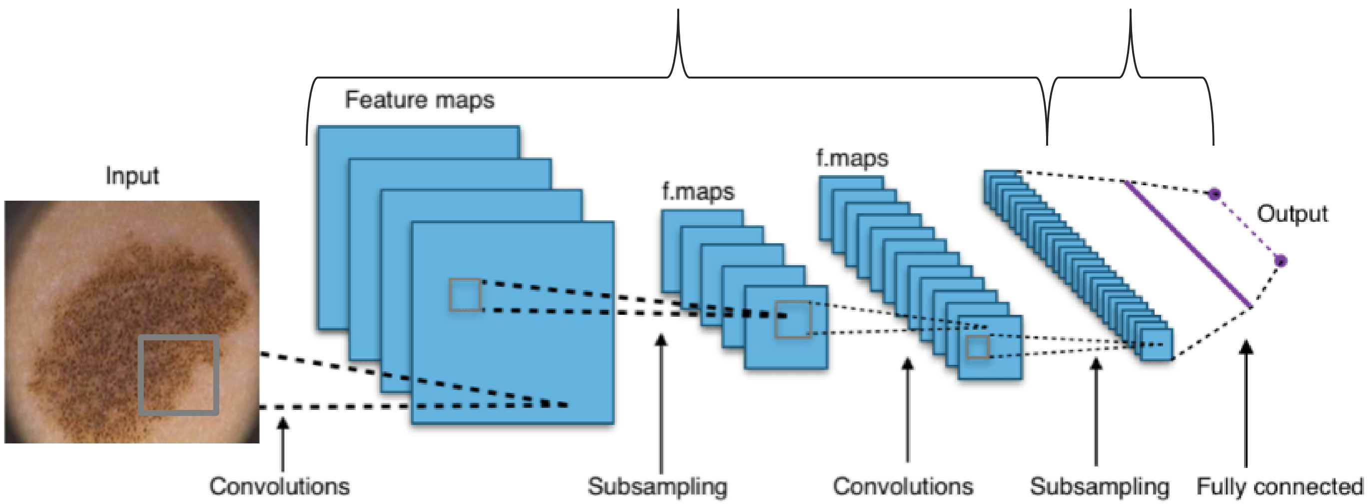


End-to-End
CAD

Dermoscopy Image Diagnosis

- How can we infer interpretability when we do not impose the features nor the classifier?

Feature Extraction Decision

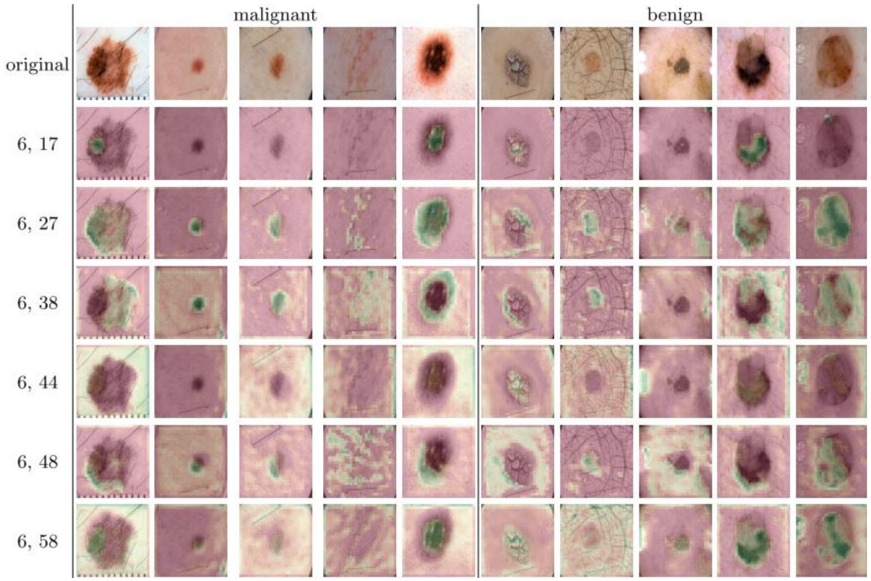


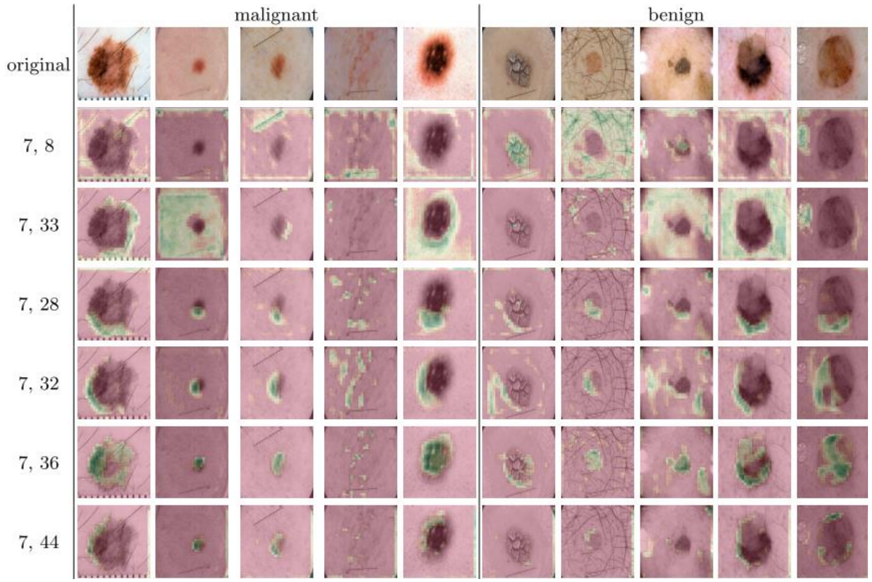
End-to-End CAD

Model Explainability

- Different visualization techniques can be used to
    - **Understand what the network is "seeing"**

Feature Maps

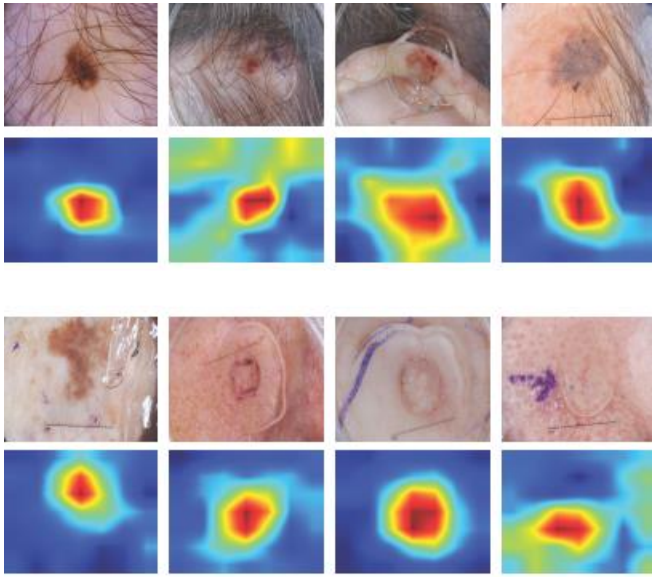


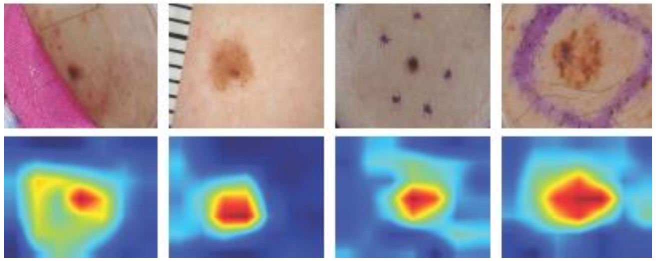


Van Molle et al., MICCAI-W, 2018

Model Explainability

- Different visualization techniques can be used to
  - **Understand what the network is "seeing"**
  - **Understand what guides the decision**





Zhang et al., IEEE TMI, 2019

Class Activation Maps

Model Explainability

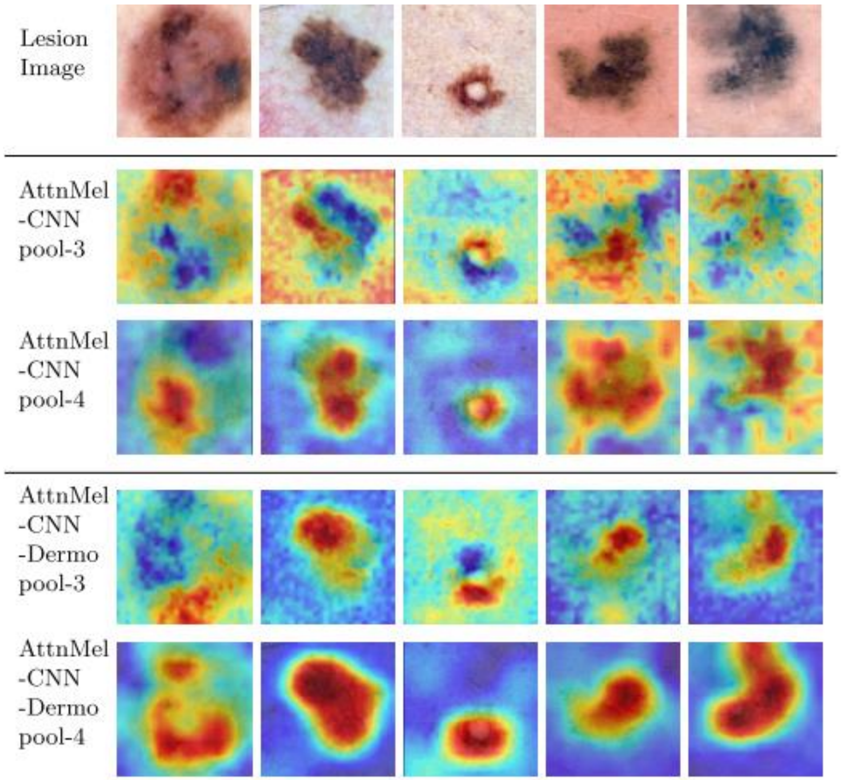
- Different visualization techniques can be used to
  - Understand what the network is “seeing”
  - Understand what guides the decision
- These techniques **improve explainability** but **may not lead to interpretability**!

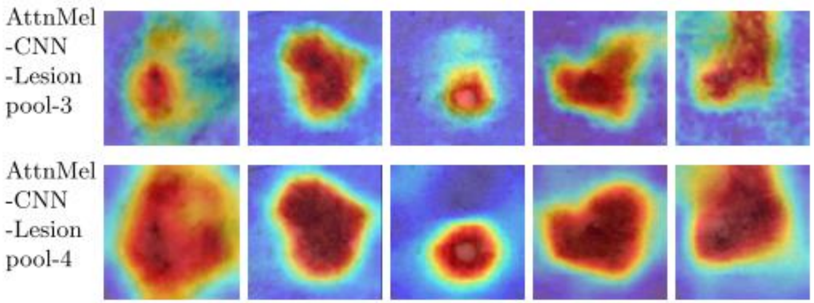
Incorporating Medical Features

- Can we incorporate medical features in DNNs?

Incorporating Medical Features

- Can we incorporate medical features in DNNs?





Yan et al., IPMI, 2019

Attention Regularized with Segmentation Masks

Incorporating Medical Features

- Can we incorporate medical features in DNNs?

IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 23, NO. 2, MARCH 2019 547

DermaKNet: Incorporating the Knowledge of Dermatologists to Convolutional Neural Networks for Skin Lesion Diagnosis

Iván González-Díaz, *Member, IEEE*

Abstract—**Traditional approaches to automatic diagnosis of skin lesions consisted of classifiers working on sets of hand-crafted features, some of which modeled lesion aspects of special importance for dermatologists. Recently, the broad adoption of convolutional neural networks (CNNs) in most computer vision tasks has brought about a great leap forward in terms of performance. Nevertheless, with this performance leap, the CNN-based computer-aided diagnosis (CAD) systems have also brought a notable reduction of the useful insights provided by hand-crafted features. This paper presents DermaKNet, a CAD system based on CNNs that incorporates specific subsystems modeling properties of skin lesions that are of special interest to dermatologists aiming to improve the interpretability of its diagnosis. Our results prove that the incorporation of these subsystems not only improves the performance, but also enhances the diagnosis by providing more interpretable outputs.**

Index Terms—**Skin lesion analysis, melanoma, convolutional neural networks, dermoscopy, CAD.**

improve it by providing valuable information about the clinical case, and serving as filtering tools that automatically detect those cases with a high confidence of benignity, which can have a great impact in the final amount of moles that must be analyzed by the clinicians.

However, despite the research efforts devoted to the topic, these systems have yet to become part of everyday clinical practice. From our point of view, there are two factors currently hampering the adoption of CAD systems by dermatologists. Firstly, the lack of large, open, annotated datasets, containing images of lesions gathered by different medical institutions and a great variety of dermatoscopes, has undermined the generalization capability of developed CAD systems, leading to poor results when applied to different datasets. Additionally, it has prevented standard and fair comparisons between proposed methods, thus hindering the scientific advances in the field. Secondly, most of CAD systems simply provide a tentative diagnosis to the clinicians, which does not actually help them much in practice.

538 IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 23, NO. 2, MARCH 2019

Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets

Jeremy Kawahara, Sara Daneshvar, Giuseppe Argenziano, and Ghassan Hamarneh, *Senior Member, IEEE*

Abstract—**We propose a multitask deep convolutional neural network, trained on multimodal data (clinical and dermoscopic images, and patient metadata), to classify the 7-point melanoma checklist criteria and perform skin lesion diagnosis. Our neural network is trained using several multitask loss functions, where each loss considers different combinations of the input modalities, which allows our model to be robust to missing data at inference time. Our final model classifies the 7-point checklist and skin condition diagnosis, produces multimodal feature vectors suitable for image retrieval, and localizes clinically discriminant regions. We benchmark our approach using 1011 lesion cases, and report comprehensive results over all 7-point criteria and diagnosis. We also make our dataset (images and metadata) publicly available online at http://derm.cs.sfu.ca.**

Index Terms—**Classification, convolutional neural networks, deep learning, dermatology, melanoma, skin, 7-point checklist.**

dermoscopy compared to the unaided eye. However, accurate diagnosis is challenging for non-experts.

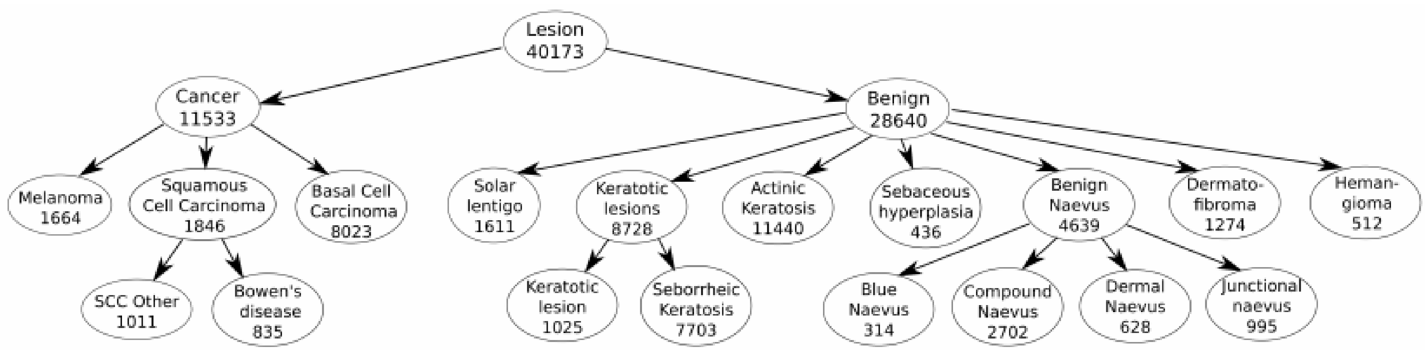
Pattern analysis, which subjectivity assesses multiple subtle lesion features, is commonly used by experienced dermatologists to distinguish between benign and malignant skin tumours. To simplify diagnoses, rule-based diagnostic algorithms such as the ABCD rule [5] and the 7-point checklist [6] have been proposed and are commonly accepted [7]. In this work we focus on the 7-point checklist, which requires identifying seven dermoscopic criteria (Table I) associated with melanoma, where each criteria is assigned a score. The lesion is diagnosed as melanoma when the sum of the scores exceeds a given threshold [6], [8]. Although some literature recommends pattern analysis over the 7-point checklist [9], some works report a trade-off between melanoma sensitivity and specificity. For example, among dermatology residents, the 7-point checklist was

Structured & Explainable Decision

- How can we improve the interpretability of the classifier?

Structured & Explainable Decision

- How can we improve the interpretability of the classifier?
    - **Some authors explored taxonomies**

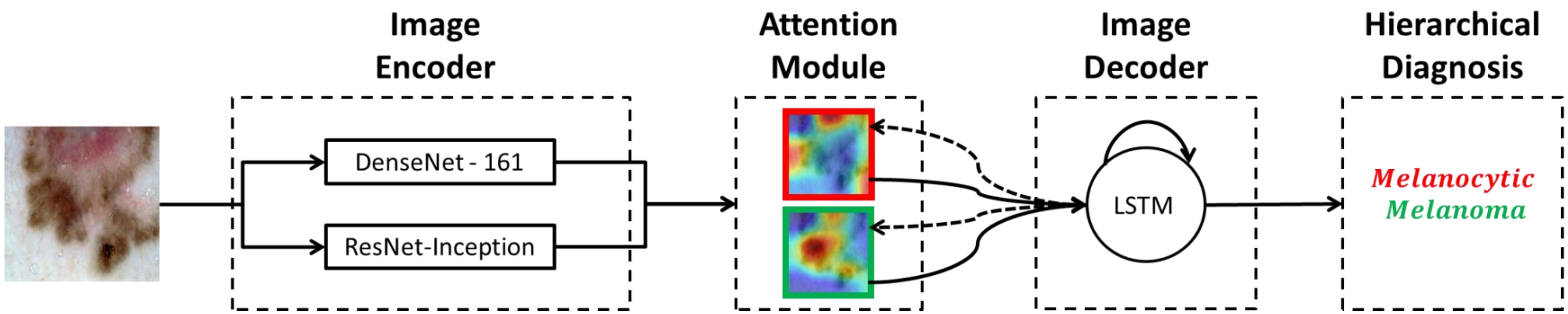


Demyanov et al., ISBI, 2017

Proposal of a **Tree-loss function** to train the DNN

Structured & Explainable Decision

- How can we improve the interpretability of the classifier?
  - **Some authors explored taxonomies**

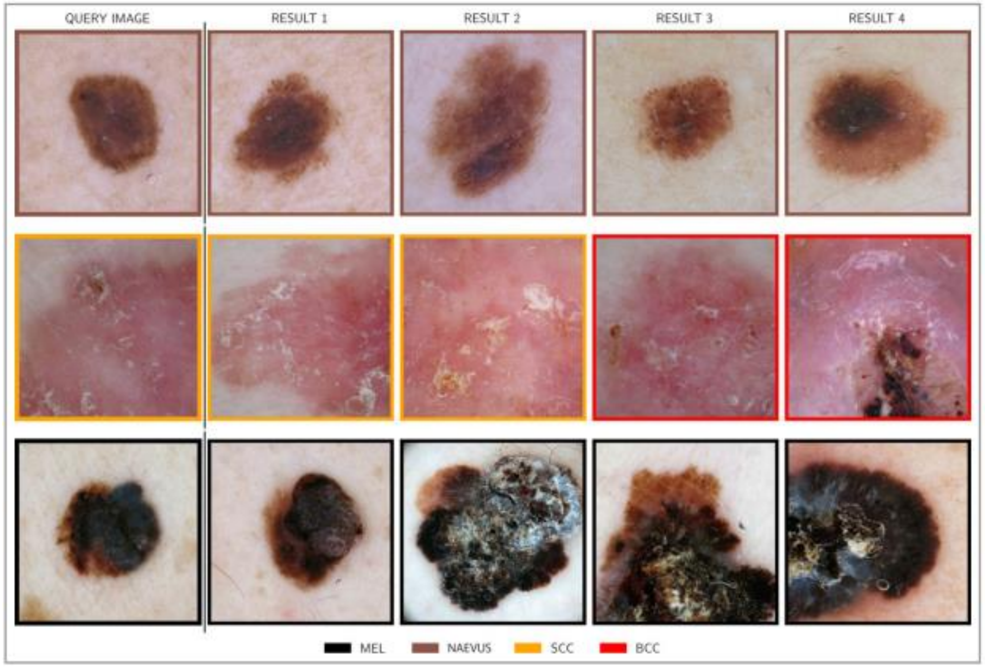


Barata et al., ISIC@CVPR, 2019

Fusion of **structured classifier** with **visualization**

Structured & Explainable Decision

- How can we improve the interpretability of the classifier?
  - **Some authors explored taxonomies**
  - **Other explored content based image retrieval (CBIR)**



Decision Based on CBIR

Tschandl et al., BJD, 2018

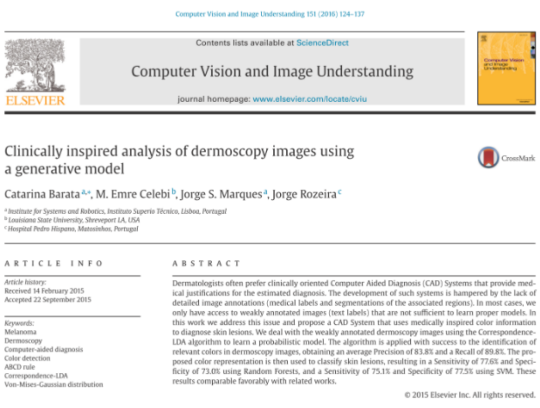
CHALLENGES & FUTURE

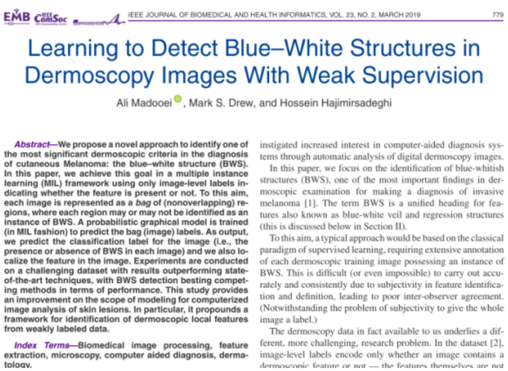
Some Usual Misconceptions

- Interpretable methods require a **great amount of detailed annotations**

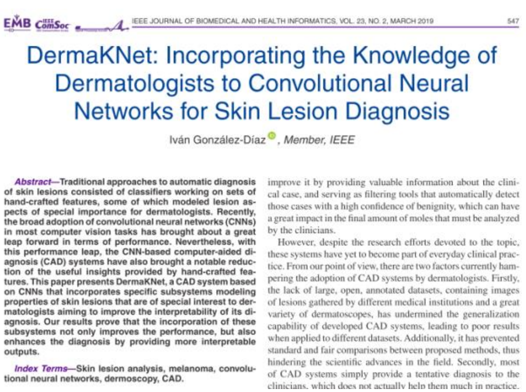
Some Usual Misconceptions

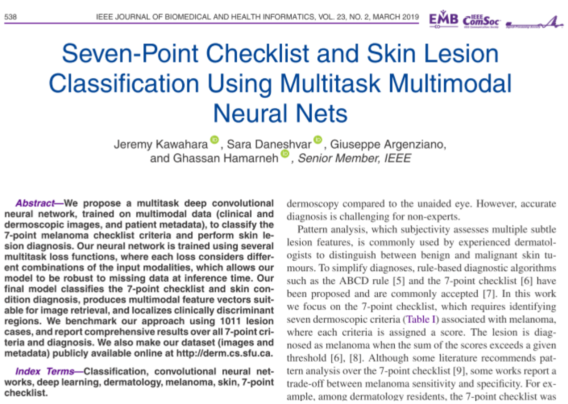
- Interpretable methods require a **great amount of detailed annotations**





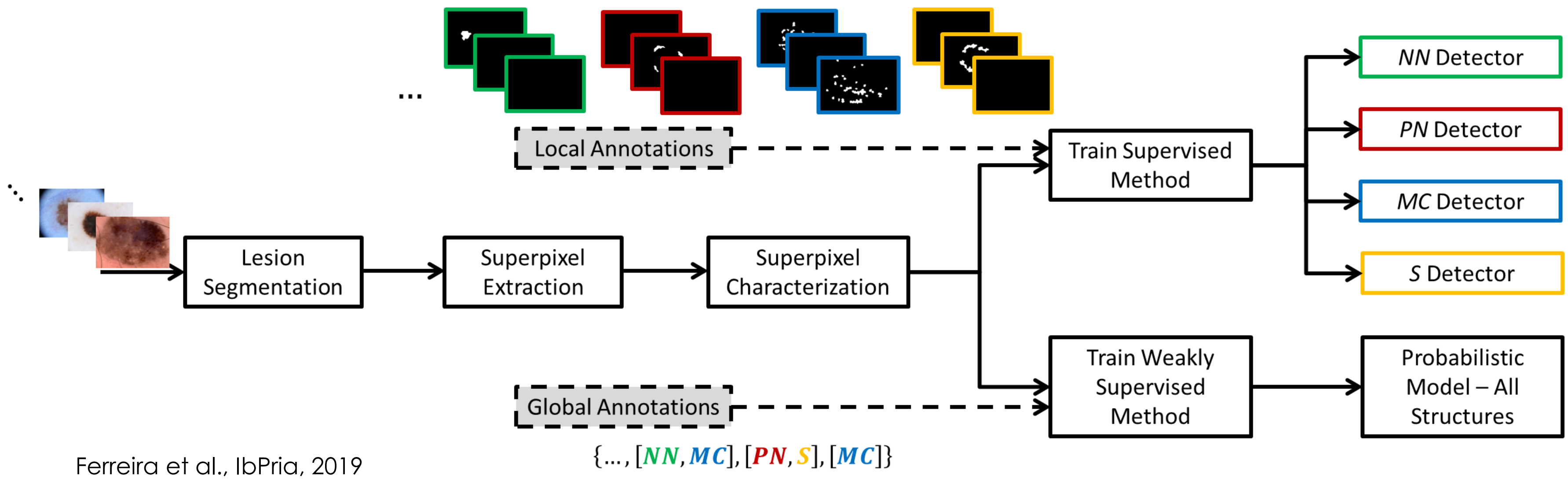
All of these works use weakly annotated sets!!





Some Usual Misconceptions

- Interpretable methods require a **great amount of detailed annotations**



Ferreira et al., IbPria, 2019

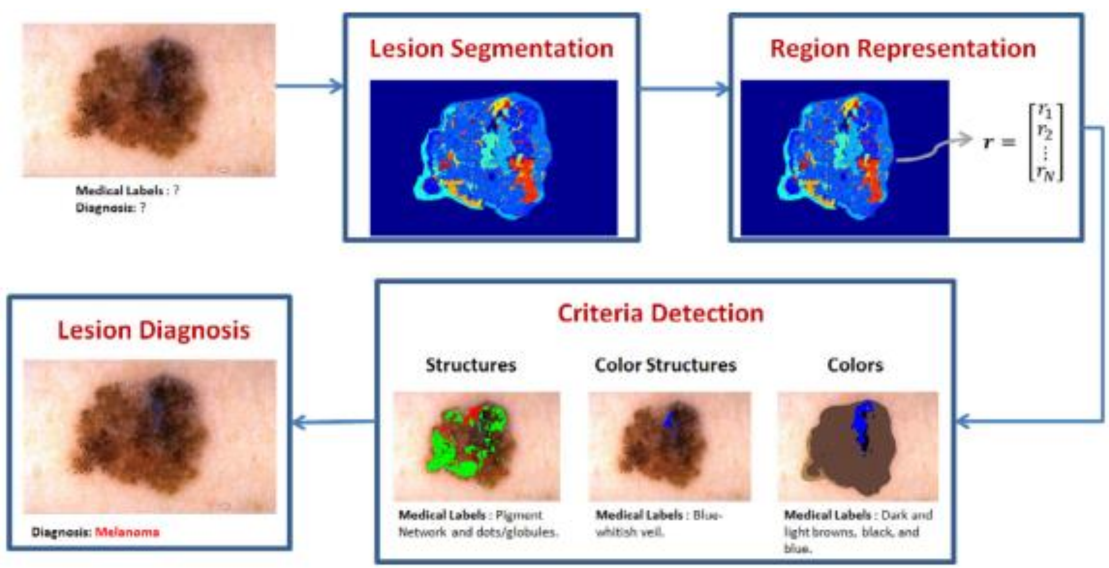
Method	Sensitivity	Specificity	BACC	#Annoations
Supervised	84,6%	69,2%	76,9%	$\approx 460k$
Weakly-Supervised	73,3%	76,0%	74,7%	2000

Some Usual Misconceptions

- ~~Interpretable methods require a **great amount of detailed annotations**~~
- It is **not possible** to apply **clinically inspired features** to **automatic diagnosis**

Some Usual Misconceptions

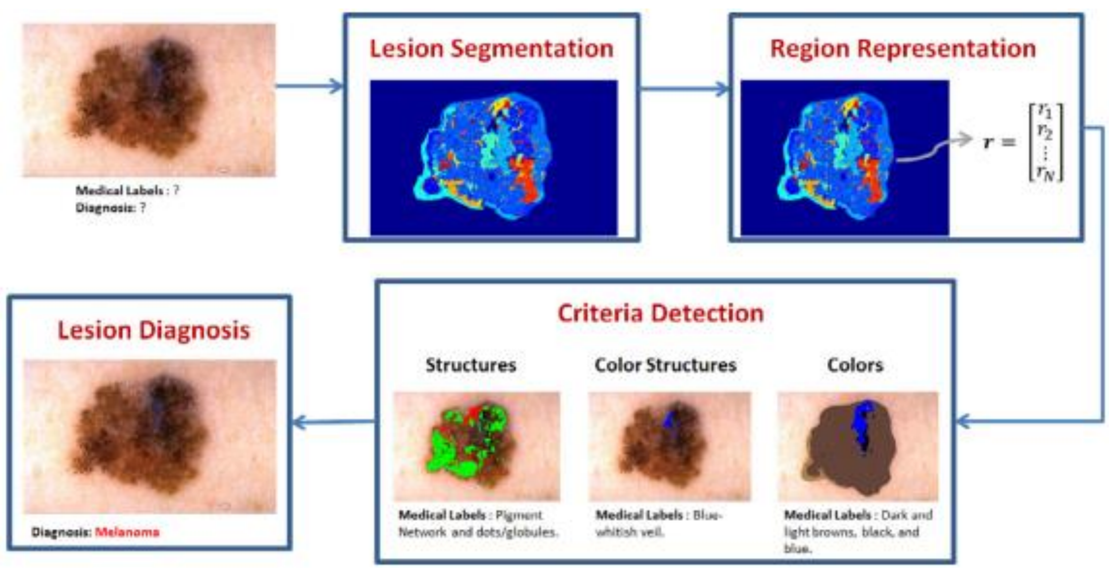
- ~~Interpretable methods require a **great amount of detailed annotations**~~
- It is **not possible** to apply **clinically inspired features** to **automatic diagnosis**



Barata et al., PR, 2017

Some Usual Misconceptions

- ~~Interpretable methods require a **great amount of detailed annotations**~~
- ~~It is **not possible** to apply **clinically inspired features** to **automatic diagnosis**~~



Barata et al., PR, 2017

What lies ahead?

- How can we combine model explainability and interpretation?
    - **Fine grained attention/activation maps**
    - **Learn to translate the maps into medical terms**

What lies ahead?

- How can we combine model explainability and interpretation?
    - **Fine grained attention/activation maps**
    - **Learn to translate the maps into medical terms**
- How relevant is our data?
    - **Identify the most difficult/misleading examples**
    - **Leverage the available data**

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?