Artificial Intelligence in Radiology

HARVARD MEDICAL SCHOOL | BRIGHAM AND WOMEN'S HOSPITAL | DANA-FARBER CANCER INSTITUTE

Ahmed Hosny

Physics Seminar - Department of Radiation Oncology, Massachusetts General Hospital
Tuesday, October 9th 2018

I declare no conflict of interest.

Deep Learning

Applications in Medical Imaging

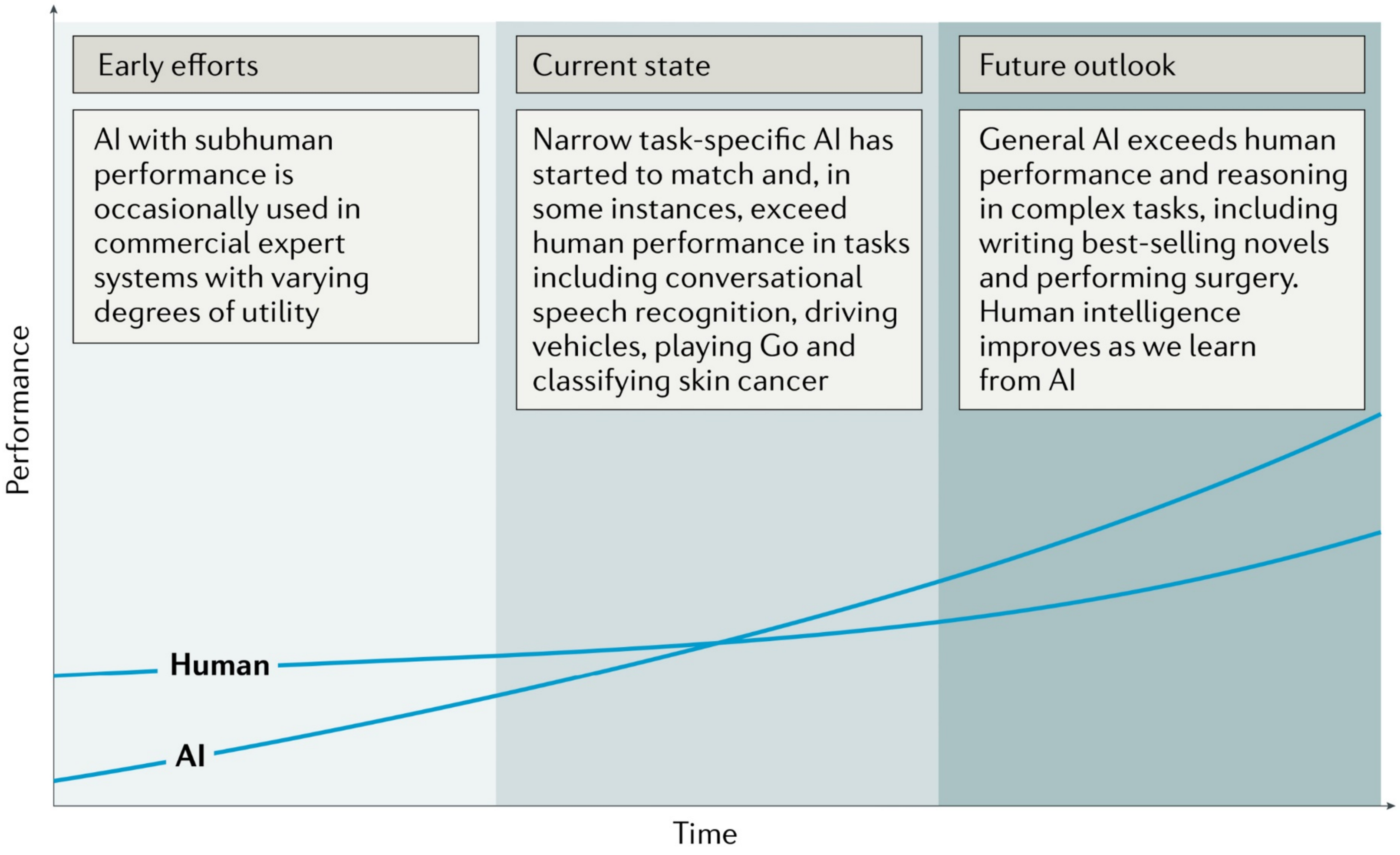
Challenges

Deep Learning

Applications in Medical Imaging

Challenges

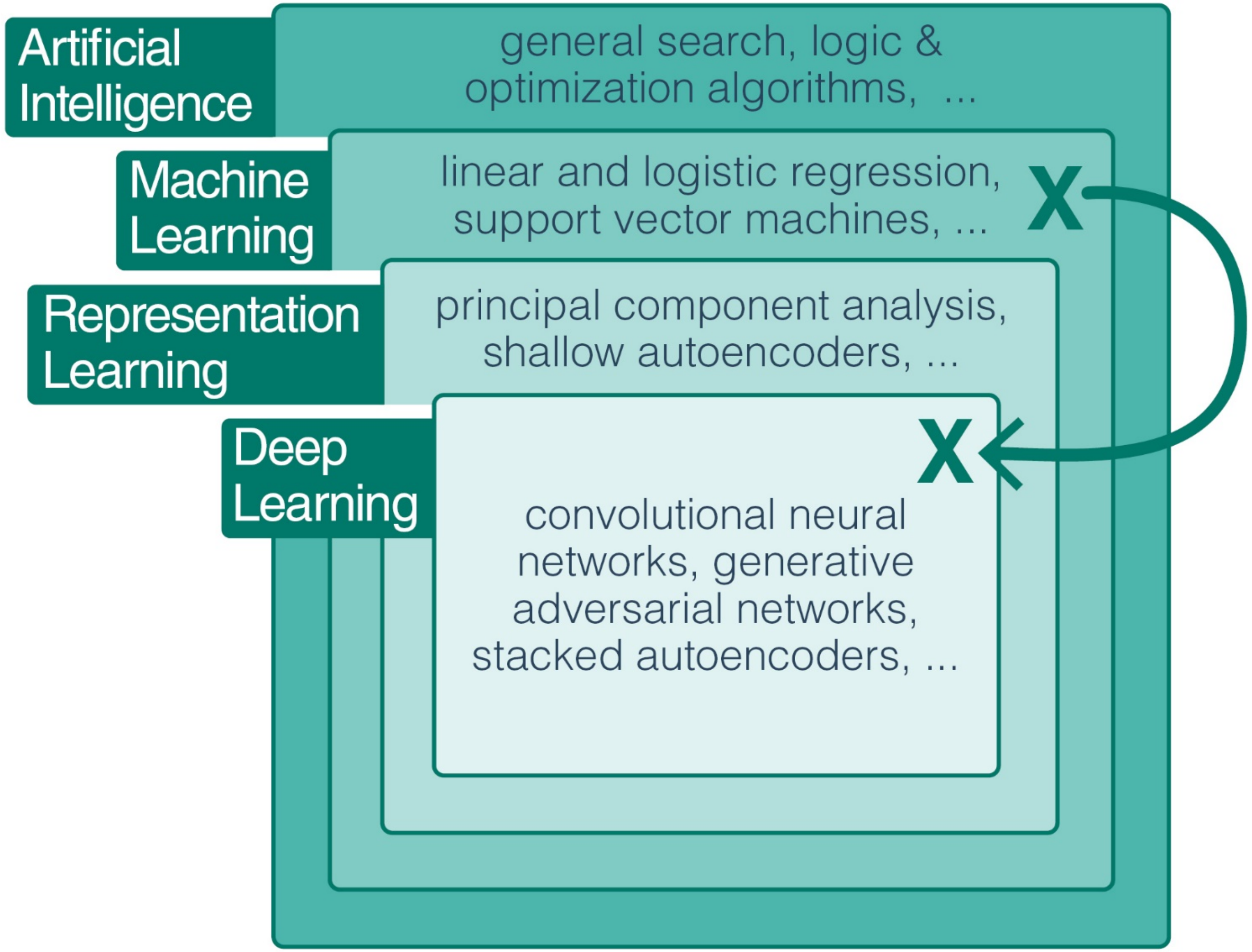
Artificial vs Human Intelligence



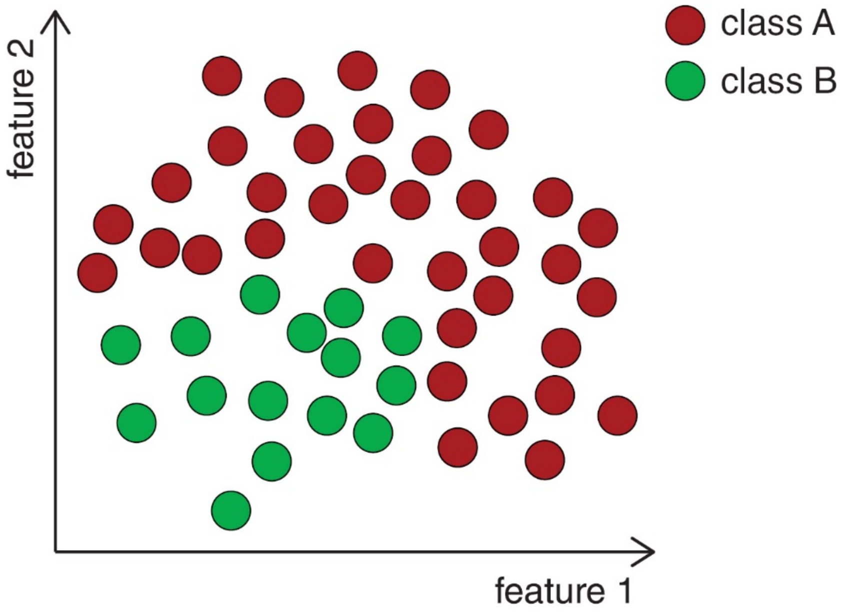
Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz & Hugo JWL Aerts

Artificial Intelligence in Radiology

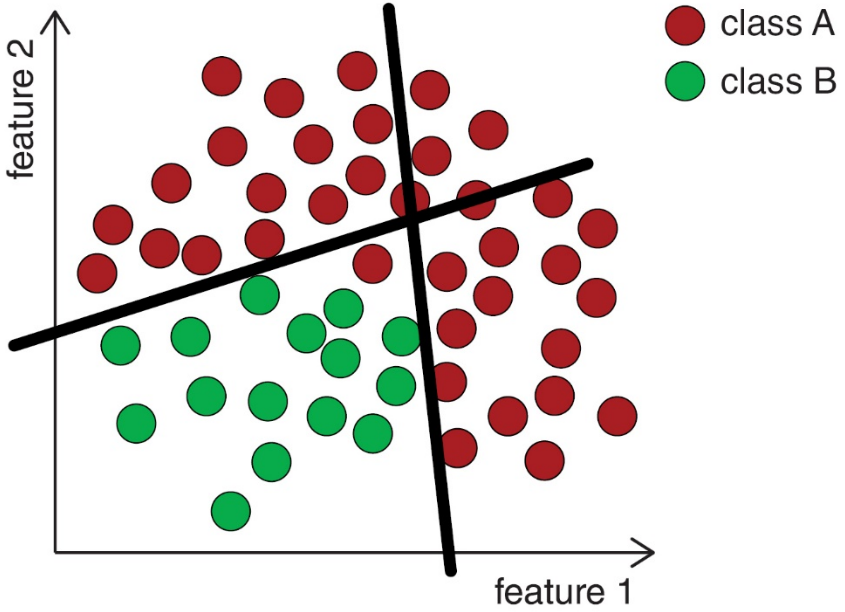
Nature Reviews Cancer - 2018


Artificial Intelligence
general search, logic & optimization algorithms, ...
Machine Learning
linear and logistic regression, support vector machines, ...
X
Representation Learning
principal component analysis, shallow autoencoders, ...
Deep Learning
X
convolutional neural networks, generative adversarial networks, stacked autoencoders, ...

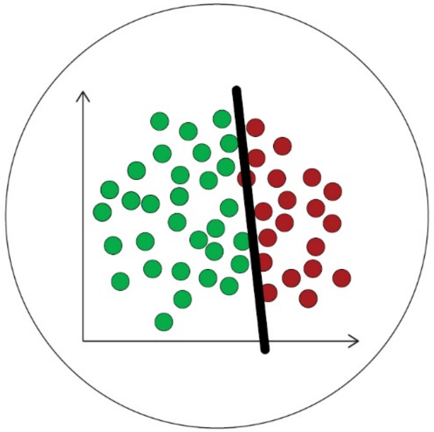
A Simple Neural Network

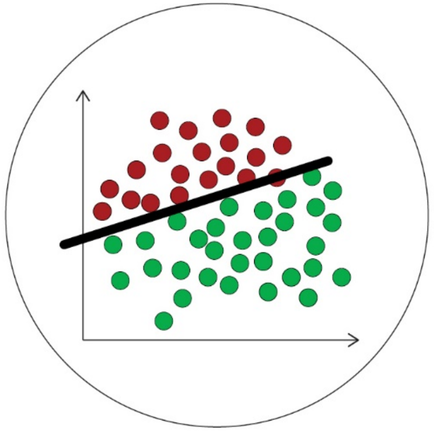


A Simple Neural Network



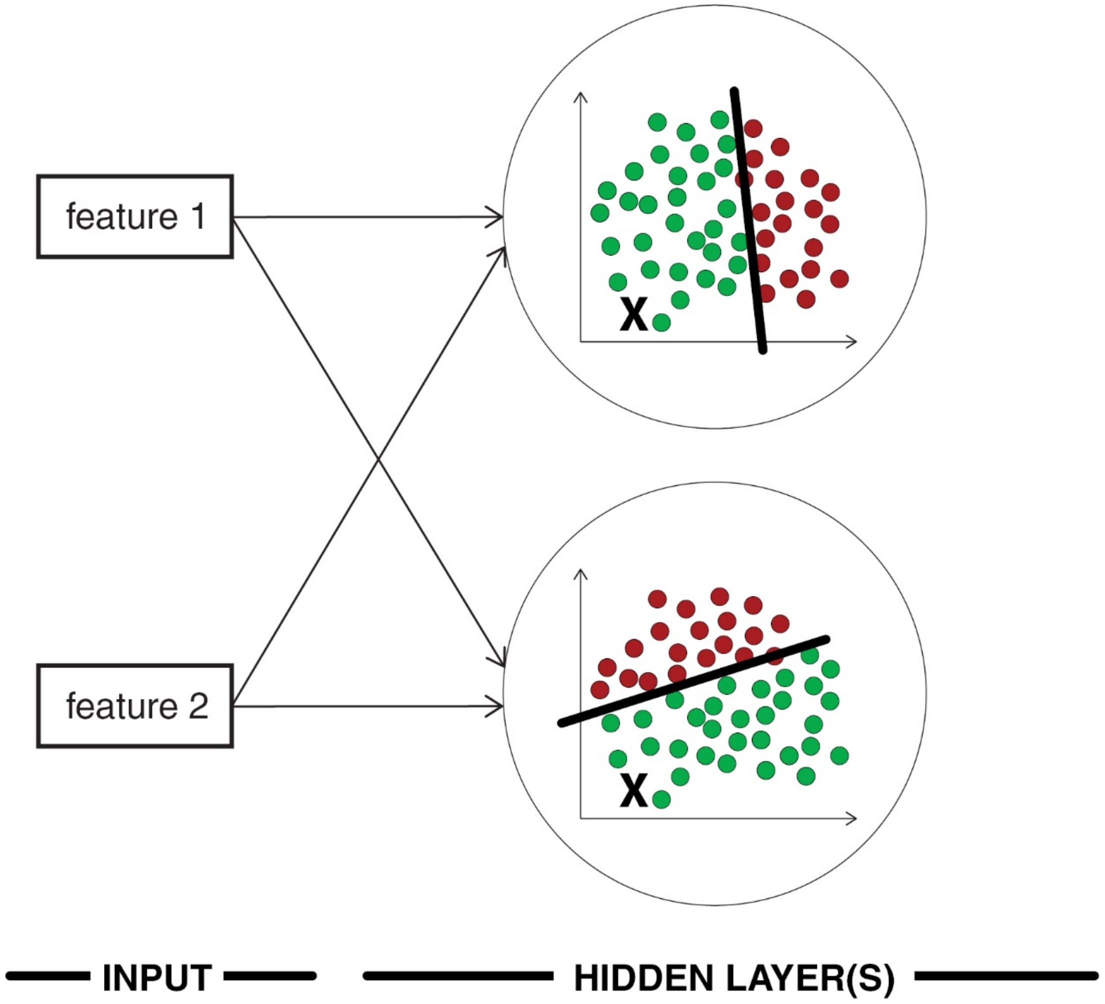
A Simple Neural Network



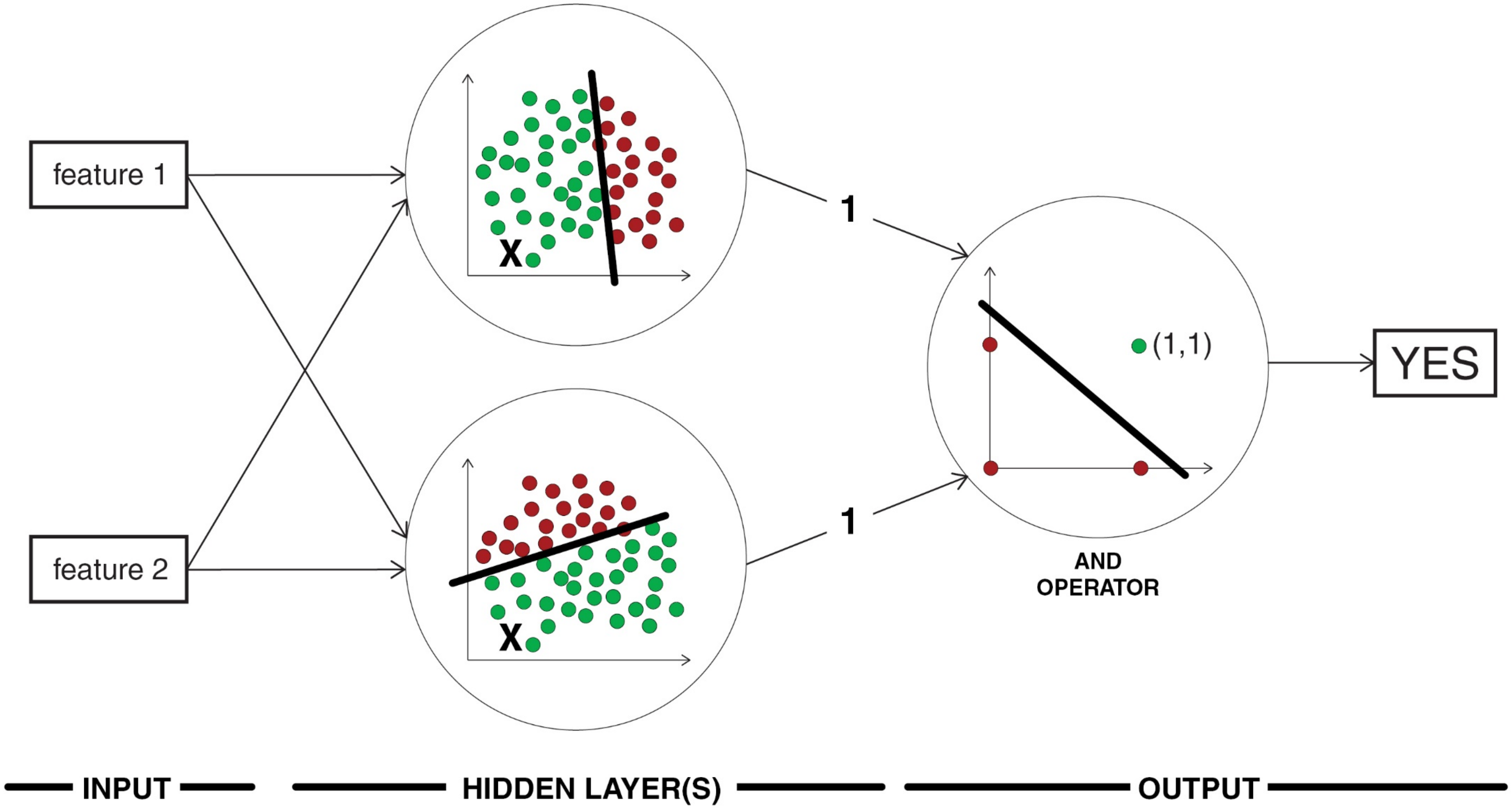


HIDDEN LAYER(S)

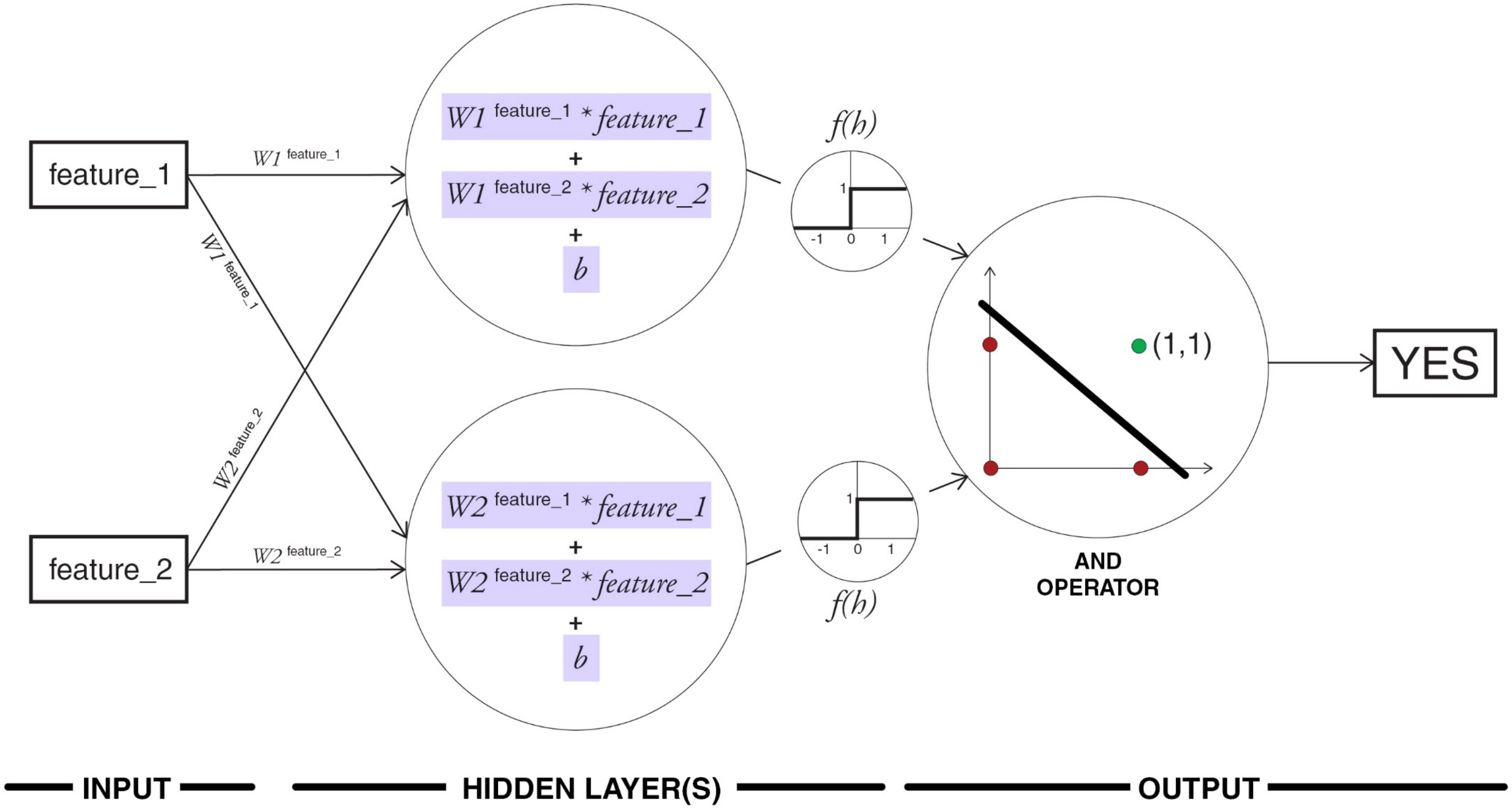
A Simple Neural Network



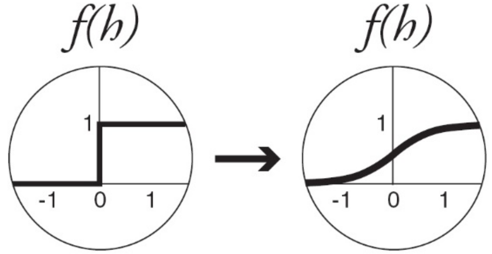
A Simple Neural Network



A Simple Neural Network



Backpropagation & Gradient Descent in Neural Networks



Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].

There have been many attempts to design self-organizing neural networks. The aim is to find a powerful synaptic modification rule that will allow an arbitrarily connected neural network to develop an internal structure that is appropriate for a particular task domain. The task is specified by giving the desired state vector of the output units for each state vector of the input units. If the input units are directly connected to the output units it is relatively easy to find learning rules that iteratively adjust the relative strengths of the connections so as to progressively reduce the difference between the actual and desired output vectors[2]. Learning becomes more interesting but more difficult when we introduce hidden units whose actual or desired states are not specified by the task. (In perceptrons, there are 'feature analysers' between the input and output that are not true hidden units because their input connections are fixed by hand, so their states are completely determined by the input vector: they do not learn representations.) The learning procedure must decide under what circumstances the hidden units should be active in order to help achieve the desired input–output behaviour. This amounts to deciding what these units should represent. We demonstrate that a general purpose and relatively simple procedure is powerful enough to construct appropriate internal representations.

The simplest form of the learning procedure is for layered networks which have a layer of input units at the bottom; any number of intermediate layers; and a layer of output units at the top. Connections within a layer or from higher to lower layers are forbidden, but connections can skip intermediate layers. An input vector is presented to the network by setting the states of the input units. Then the states of the units in each layer are determined by applying equations (1) and (2) to the connections coming from lower layers. All units within a layer have their states set in parallel, but different layers have their states set sequentially, starting at the bottom and working upwards until the states of the output units are determined.

The total input, x_j, to unit j is a linear function of the outputs, y_i, of the units that are connected to j and of the weights, w_{ji}, on these connections

$$x_j = \sum_i y_i w_{ji} \qquad (1)$$

Units can be given biases by introducing an extra input to each unit which always has a value of 1. The weight on this extra input is called the bias and is equivalent to a threshold of the opposite sign. It can be treated just like the other weights.

A unit has a real-valued output, y_j, which is a non-linear function of its total input

$$y_j = \frac{1}{1+e^{-x_j}} \qquad (2)$$

† To whom correspondence should be addressed

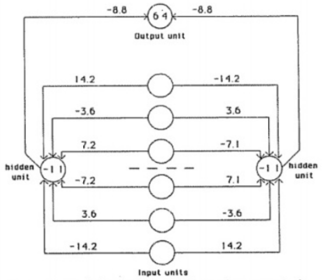


Fig. 1 A network that has learned to detect mirror symmetry in the input vector. The numbers on the arcs are weights and the numbers inside the nodes are biases. The learning required 1,425 sweeps through the set of 64 possible input vectors, with the weights being adjusted on the basis of the accumulated gradient after each sweep. The values of the parameters in equation (9) were $\varepsilon = 0.1$ and $\alpha = 0.9$. The initial weights were random and were uniformly distributed between −0.3 and 0.3. The key property of this solution is that for a given hidden unit, weights that are symmetric about the middle of the input vector are equal in magnitude and opposite in sign. So if a symmetrical pattern is presented, both hidden units will receive a net input of 0 from the input units, and, because the hidden units have a negative bias, both will be off. In this case the output unit, having a positive bias, will be on. Note that the weights on each side of the midpoint are in the ratio 1:2:4. This ensures that each of the eight patterns that can occur above the midpoint sends a unique activation sum to each hidden unit, so the only pattern below the midpoint that can exactly balance this sum is the symmetrical one. For all non-symmetrical patterns, both hidden units will receive non-zero activations from the input units. The two hidden units have identical patterns of weights but with opposite signs, so for every non-symmetric pattern one hidden unit will come on and suppress the output unit.

It is not necessary to use exactly the functions given in equations (1) and (2). Any input-output function which has a bounded derivative will do. However, the use of a linear function for combining the inputs to a unit before applying the nonlinearity greatly simplifies the learning procedure.

The aim is to find a set of weights that ensure that for each input vector the output vector produced by the network is the same as (or sufficiently close to) the desired output vector. If there is a fixed, finite set of input-output cases, the total error in the performance of the network with a particular set of weights can be computed by comparing the actual and desired output vectors for every case. The total error, E, is defined as

grad descent on E

$$E = \frac{1}{2}\sum_c \sum_j (y_{j,c} - d_{j,c})^2 \qquad (3)$$

$\frac{\partial E}{\partial w}$

$\sum$ case

where c is an index over cases (input-output pairs), j is an index over output units, y is the actual state of an output unit and d is its desired state. To minimize E by gradient descent it is necessary to compute the partial derivative ~~of E~~ with respect to each weight in the network. This is simply the sum of the partial derivatives for each of the input-output cases. For a given case, the partial derivatives of the error with respect to each weight are computed in two passes. We have already described the forward pass in which the units in each layer have their states determined by the input they receive from units in lower layers using equations (1) and (2). The backward pass which propagates derivatives from the top layer back to the bottom one is more complicated.

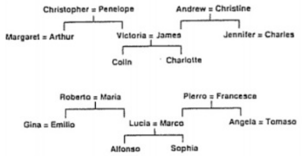


Fig. 2 Two isomorphic family trees. The information can be expressed as a set of triples of the form ⟨person 1⟩⟨relationship⟩⟨person 2⟩, where the possible relationships are {father, mother, husband, wife, son, daughter, uncle, aunt, brother, sister, nephew, niece}. A layered net can be said to 'know' these triples if it can produce the third term of each triple when given the first two. The first two terms are encoded by activating two of the input units, and the network must then complete the proposition by activating the output unit that represents the third term.

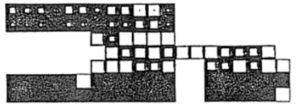


Fig. 3 Activity levels in a five-layer network after it has learned. The bottom layer has 24 input units on the left for representing ⟨person 1⟩ and 12 input units on the right for representing the relationship. The white squares inside these two groups show the activity levels of the units. There is one active unit in the first group representing Colin and one in the second group representing the relationship 'has-aunt'. Each of the two input groups is totally connected to its own group of 6 units in the second layer. These groups learn to encode people and relationships as distributed patterns of activity. The second layer is totally connected to the central layer of 12 units, and these are connected to the penultimate layer of 6 units. The activity in the penultimate layer must activate the correct output units, each of which stands for a particular ⟨person 2⟩. In this case, there are two correct answers (marked by black dots) because Colin has two aunts. Both the input units and the output units are laid out spatially with the English people in one row and the isomorphic Italians immediately below.

The backward pass starts by computing $\partial E/\partial y$ for each of the output units. Differentiating equation (3) for a particular case, c, and suppressing the index c gives

$$\partial E/\partial y_j = y_j - d_j \qquad (4)$$

We can then apply the chain rule to compute $\partial E/\partial x_j$

$$\partial E/\partial x_j = \partial E/\partial y_j \cdot dy_j/dx_j$$

Differentiating equation (2) to get the value of dy_j/dx_j and substituting gives

$$\partial E/\partial x_j = \partial E/\partial y_j \cdot y_j(1-y_j) \qquad (5)$$

This means that we know how a change in the total input x to an output unit will affect the error. But this total input is just a linear function of the states of the lower level units and it is also a linear function of the weights on the connections, so it is easy to compute how the error will be affected by changing these states and weights. For a weight w_{ji}, from i to j the derivative is

$$\partial E/\partial w_{ji} = \partial E/\partial x_j \cdot \partial x_j/\partial w_{ji}$$
$$= \partial E/\partial x_j \cdot y_i \qquad (6)$$

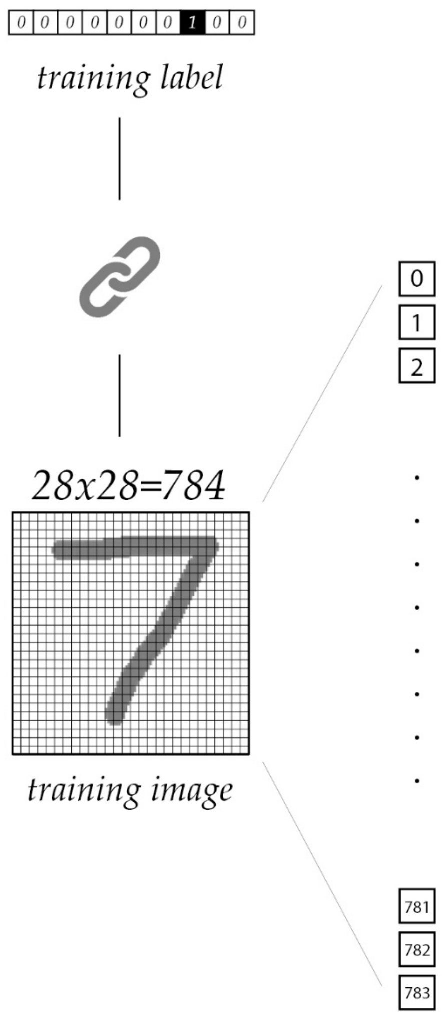
and for the output of the i^{th} unit the contribution to $\partial E/\partial y_i$

David E Rumelhart, Geoffrey E Hinton & Ronald J Williams

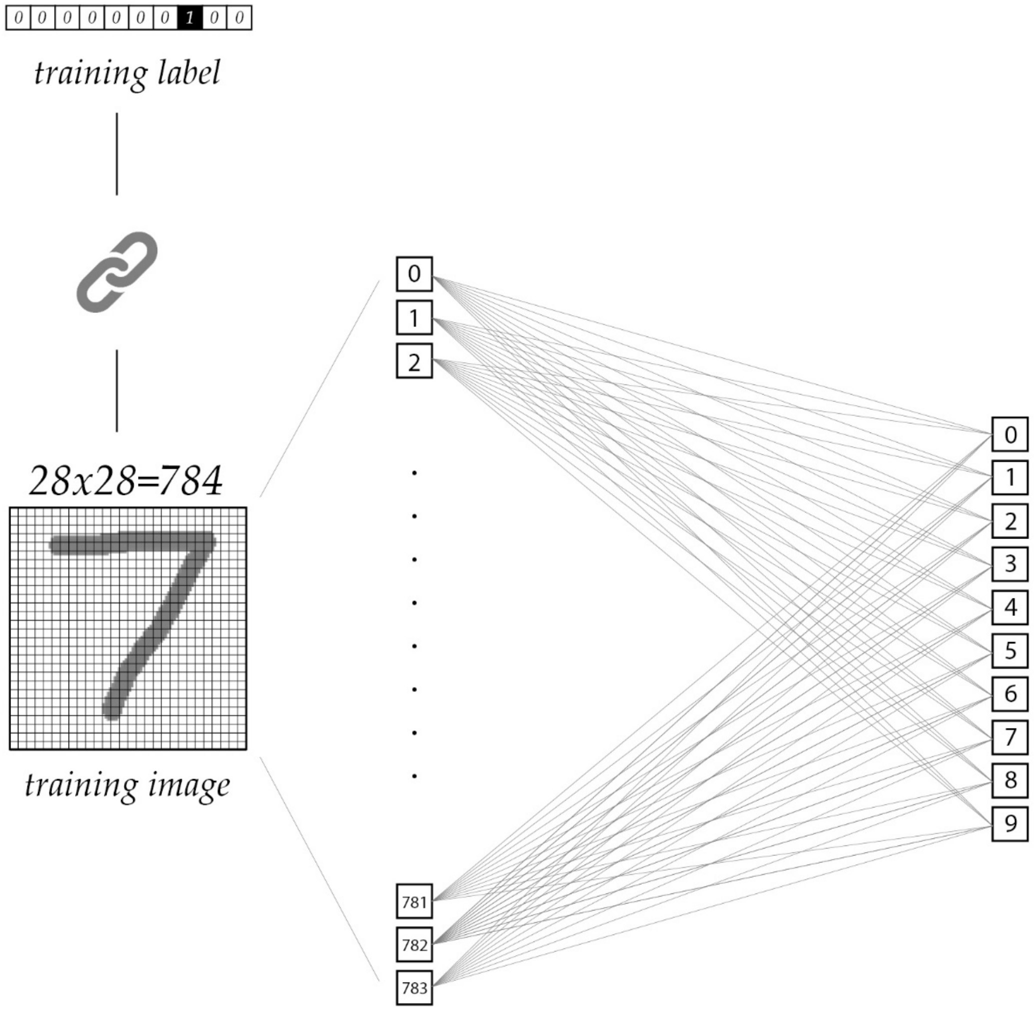
Learning Representations by Back-propagating Errors

Nature - 1986

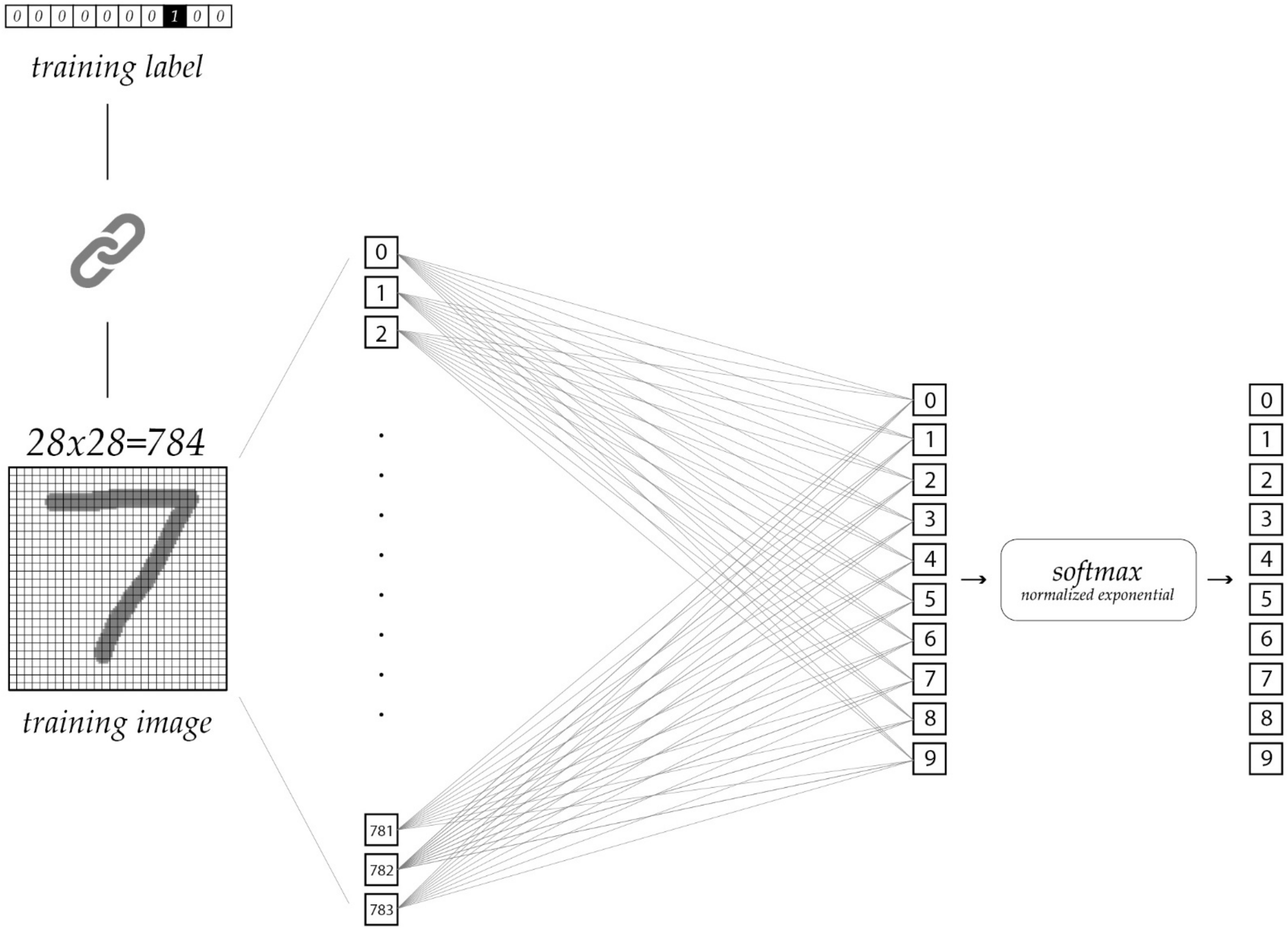
A Simple Neural Network



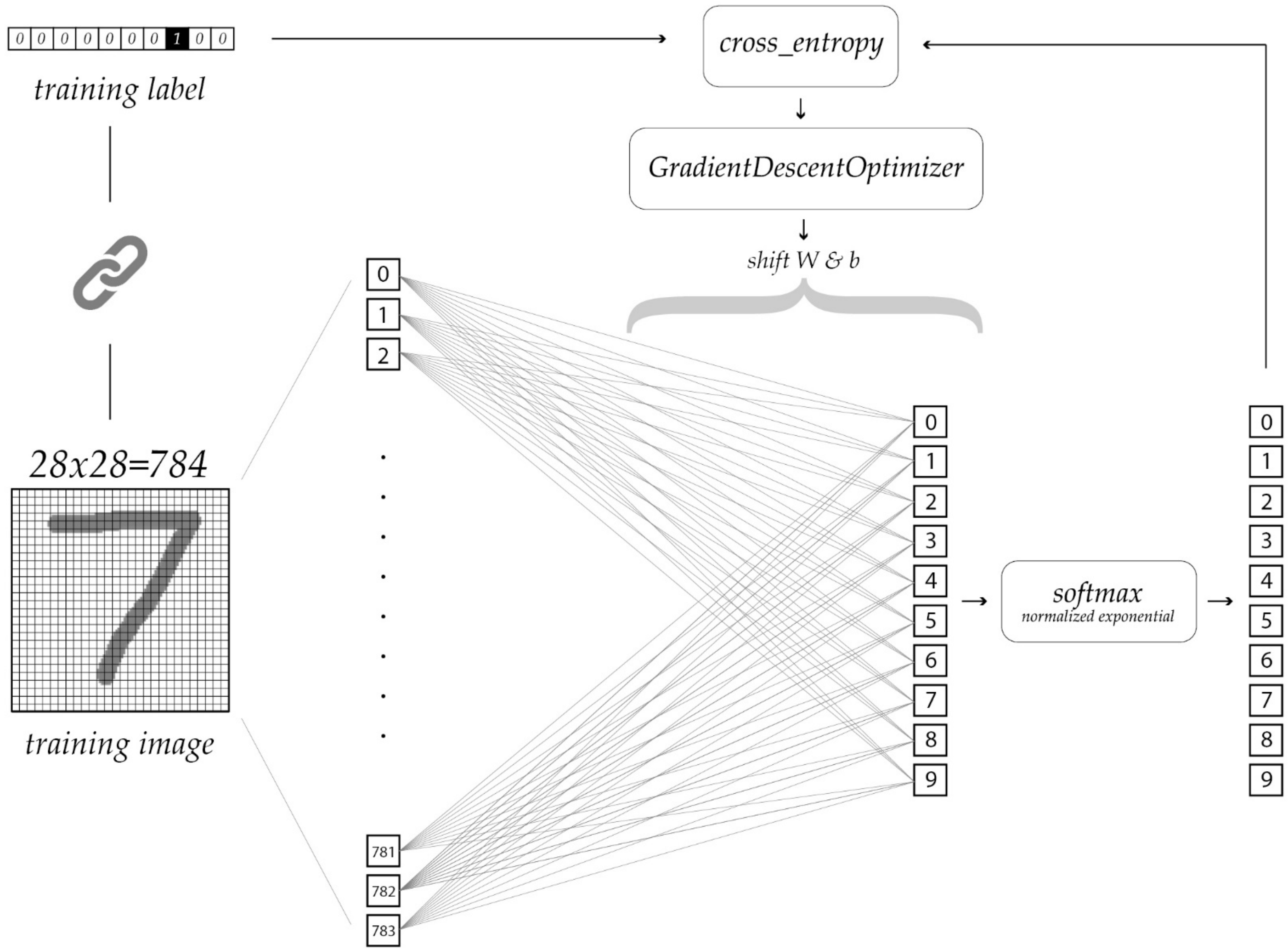
A Simple Neural Network



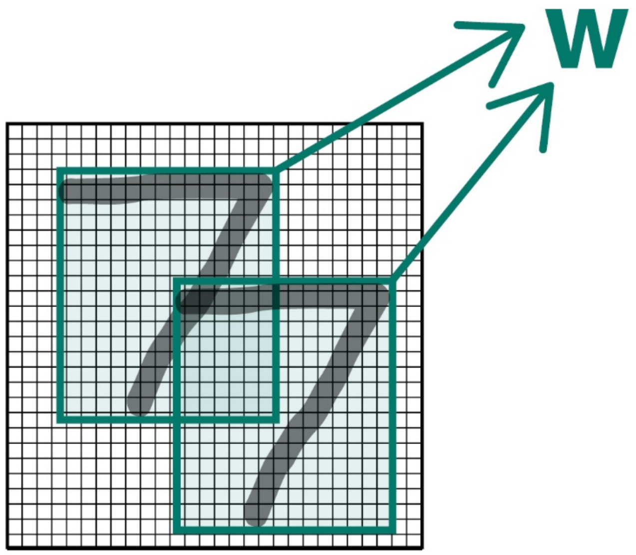
A Simple Neural Network



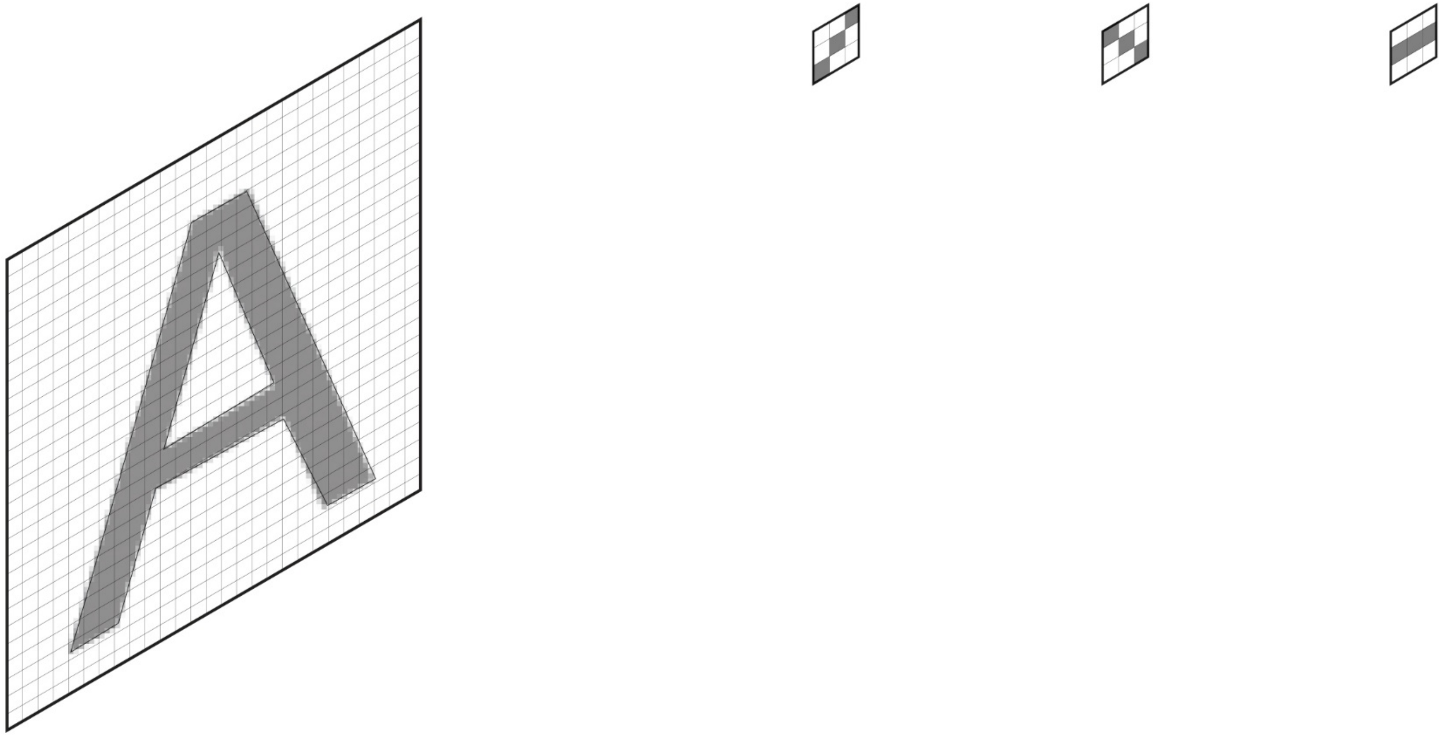
A Simple Neural Network



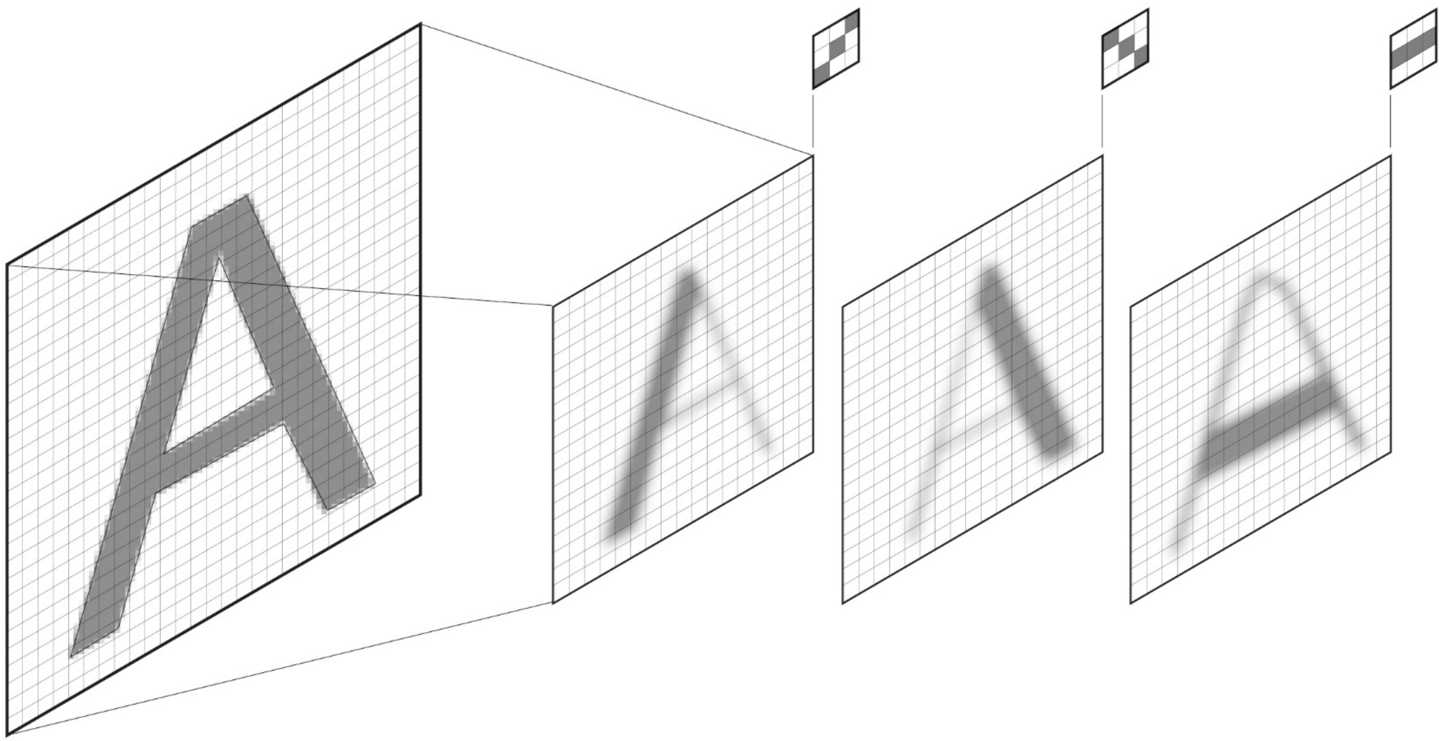
Weight Sharing



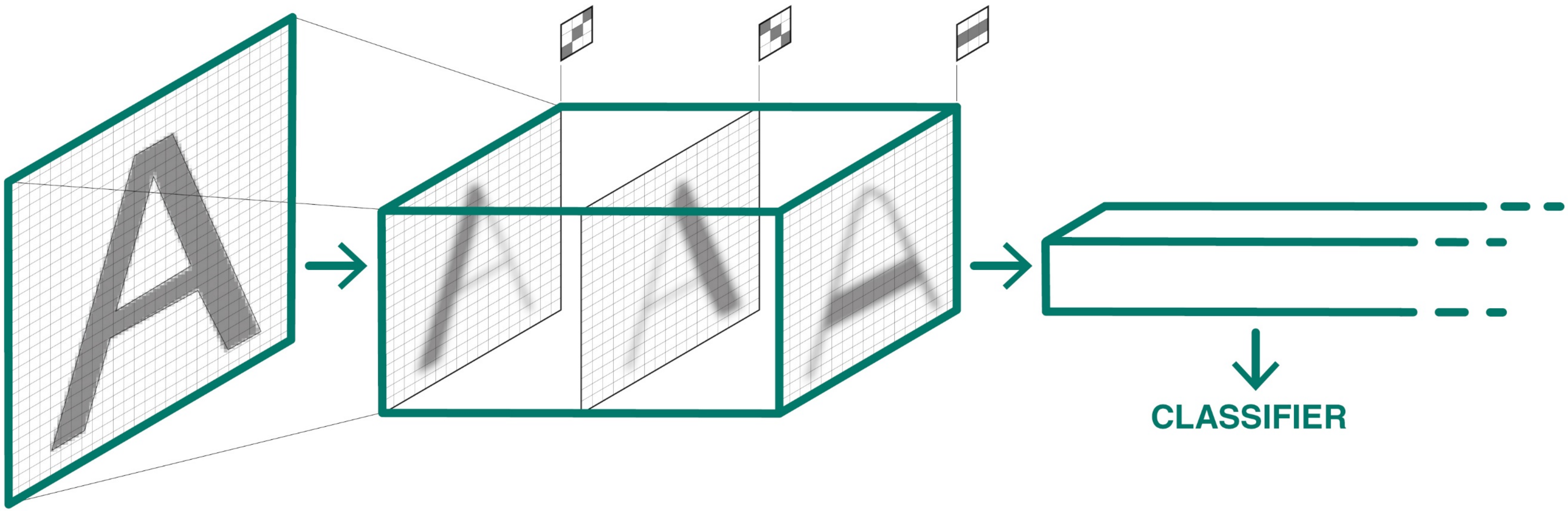
Convolutions



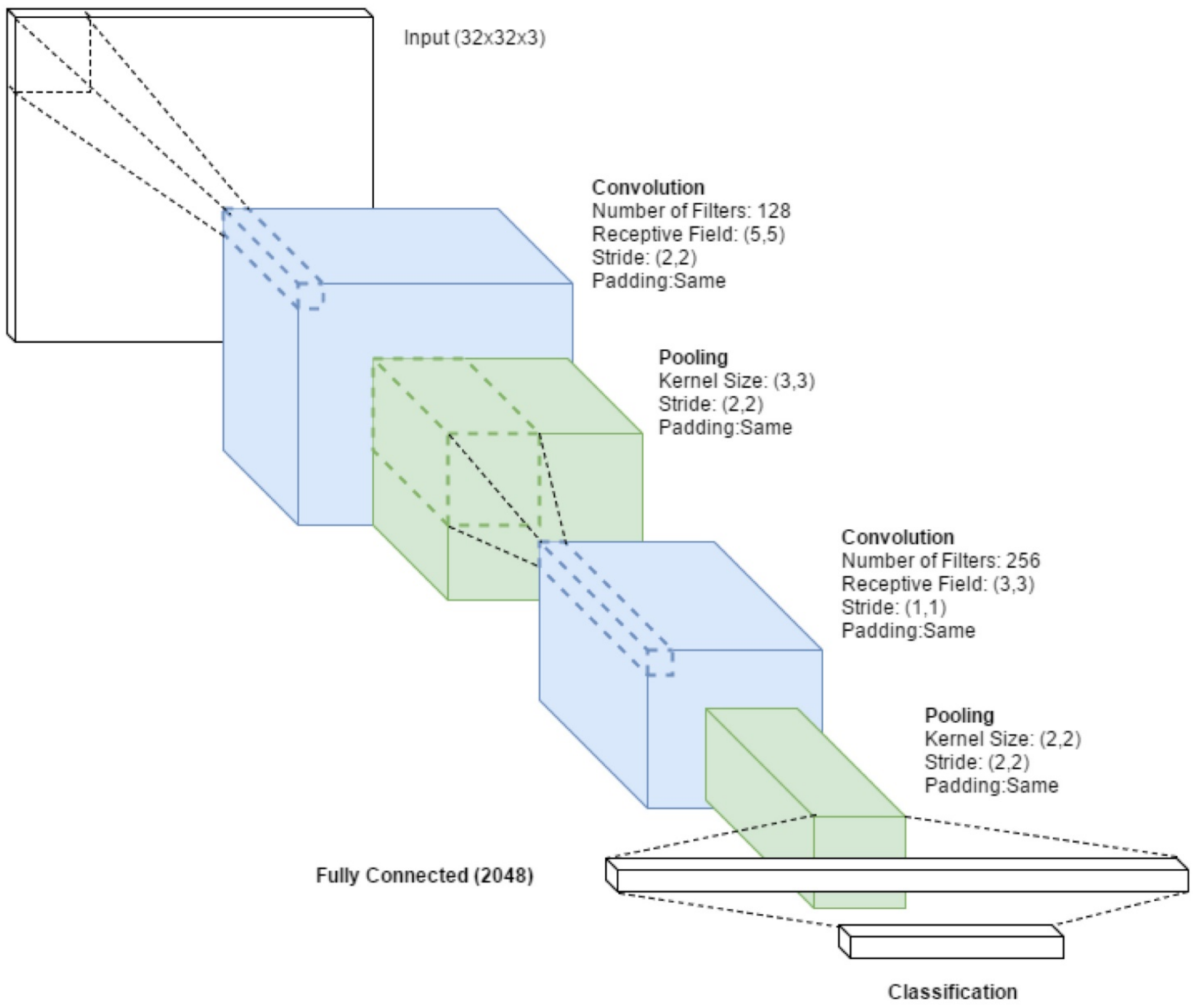
Convolutions



Convolutions



Convolutional Neural Networks



www.thushv.com/computer_vision/convolutional-neural-networks-mayor-of-the-visionville/

Neocognitron

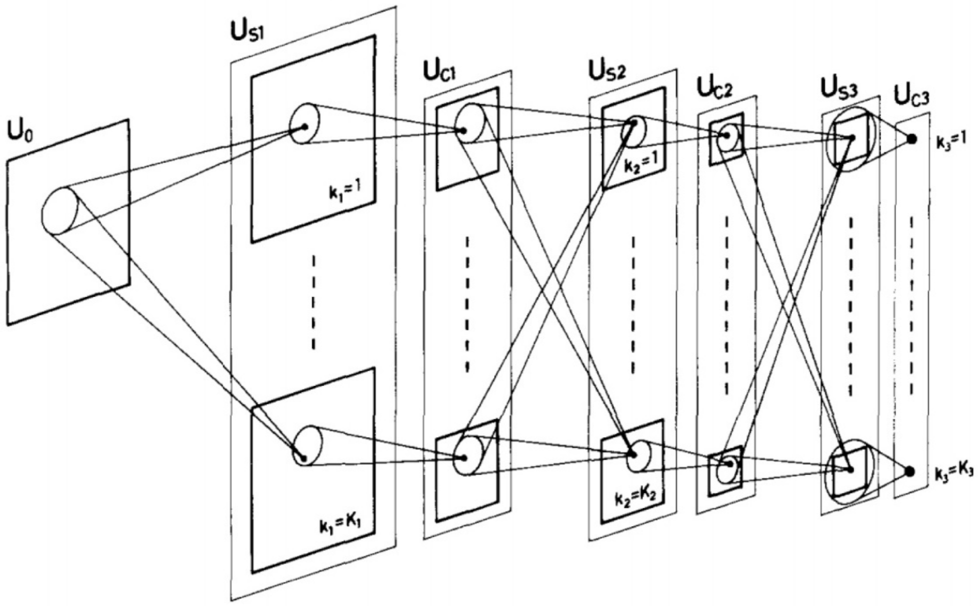


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

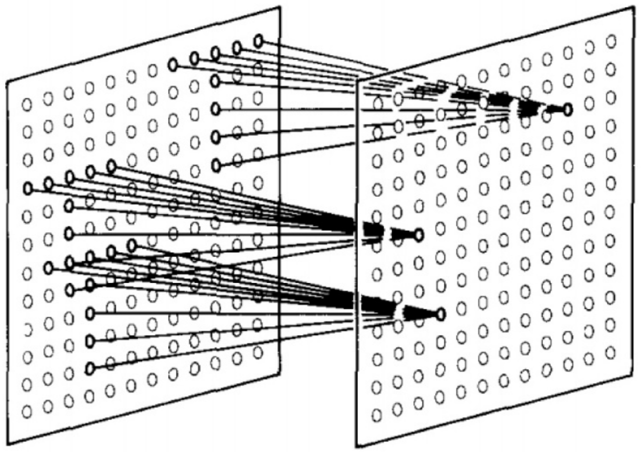


Fig. 3. Illustration showing the input interconnections to the cells within a single cell-plane

Kunihiko Fukushima

Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Biological Cybernetics - 1980

LeNet-5

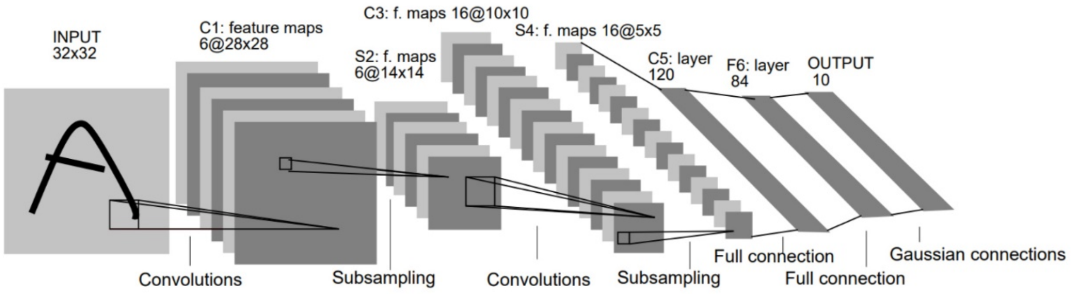


Fig. 1. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

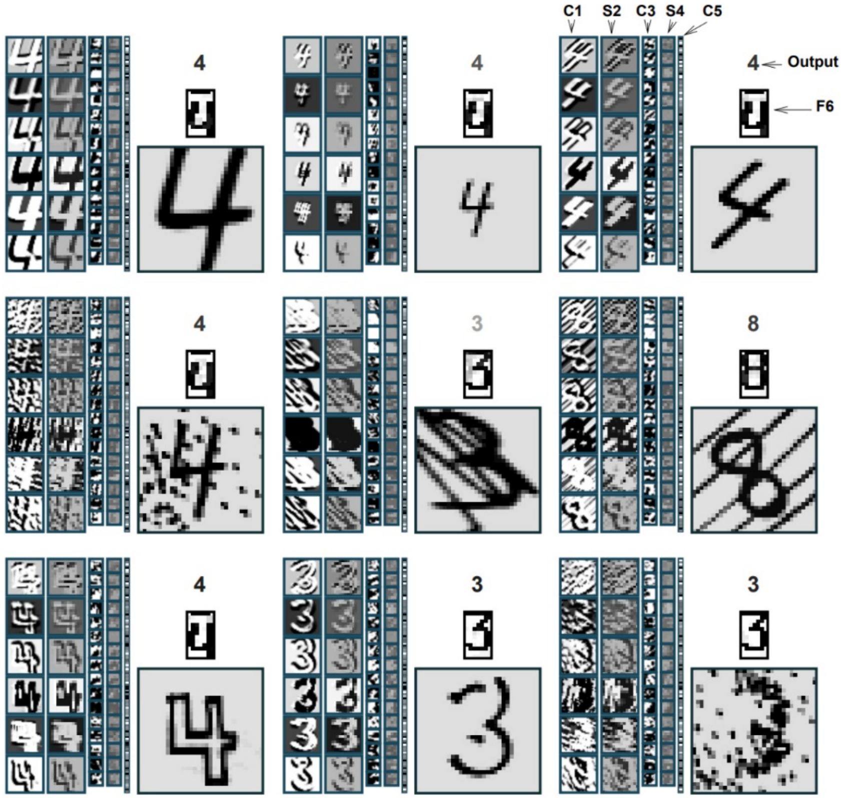


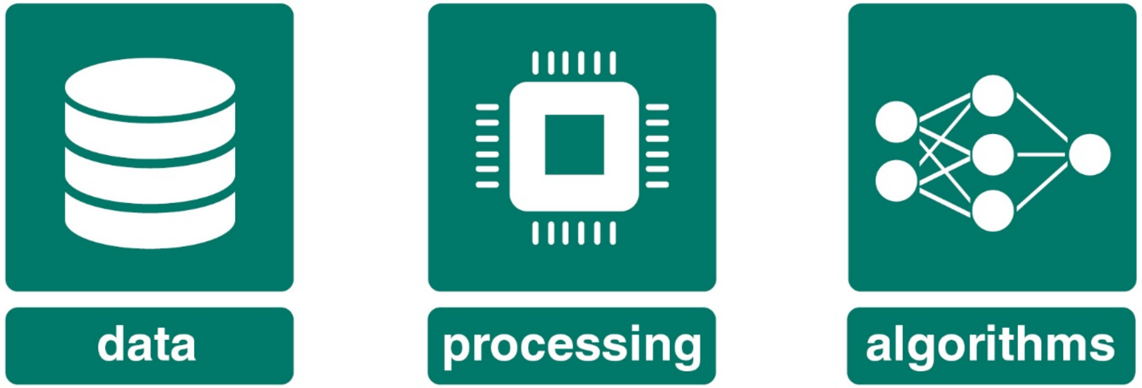
Fig. 4. Examples of unusual, distorted, and noisy characters correctly recognized by LeNet-5. The grey-level of the output label represents the penalty (lighter for higher penalties).

Yann LeCun, Patrick Haffner, Léon Bottou & Yoshua Bengio

Object Recognition with Gradient Based Learning

Shape, Contour and Grouping in Computer Vision - 1999

Revival of Neural Networks



AlexNet

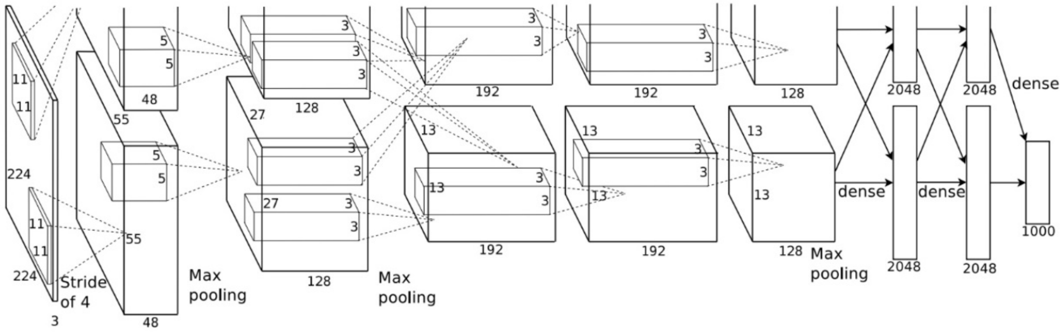


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

AlexNet @ ImageNet

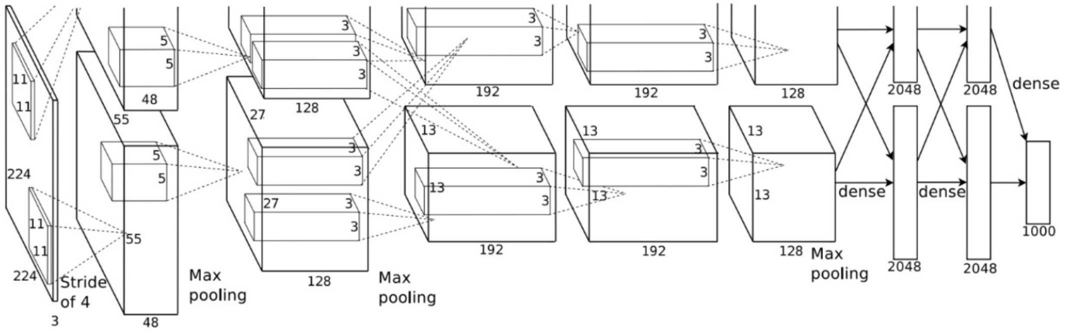
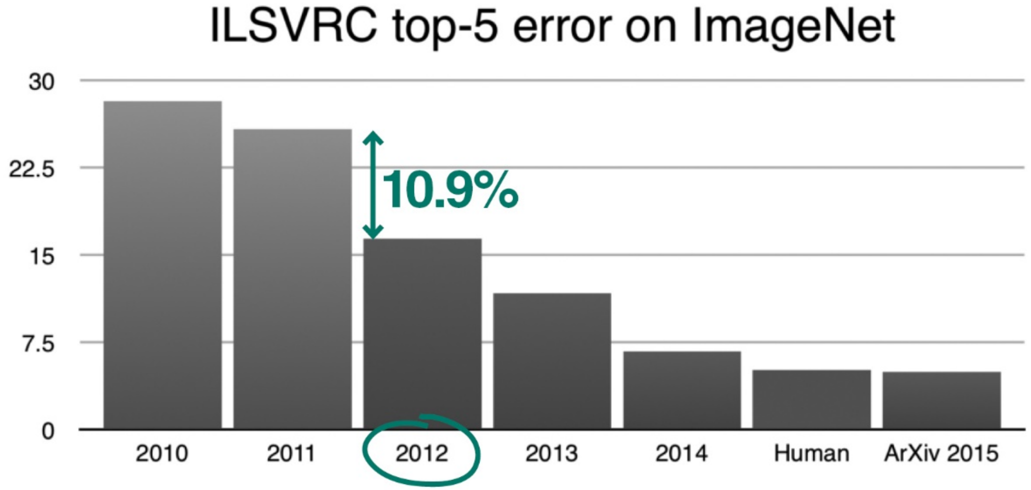
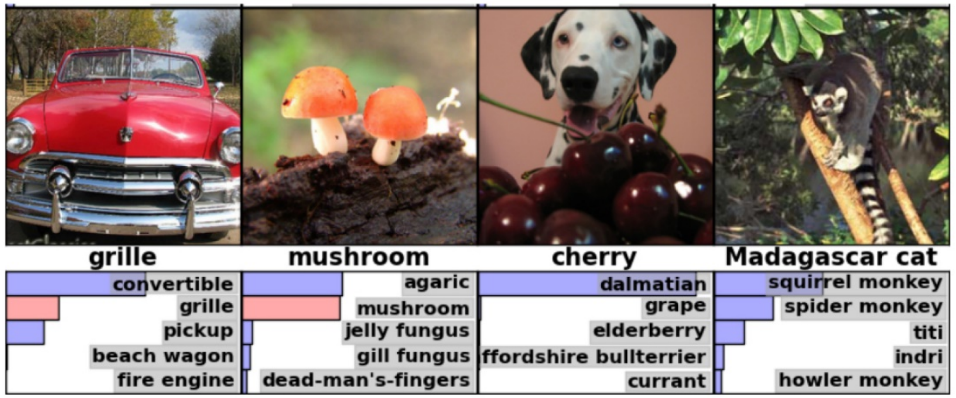


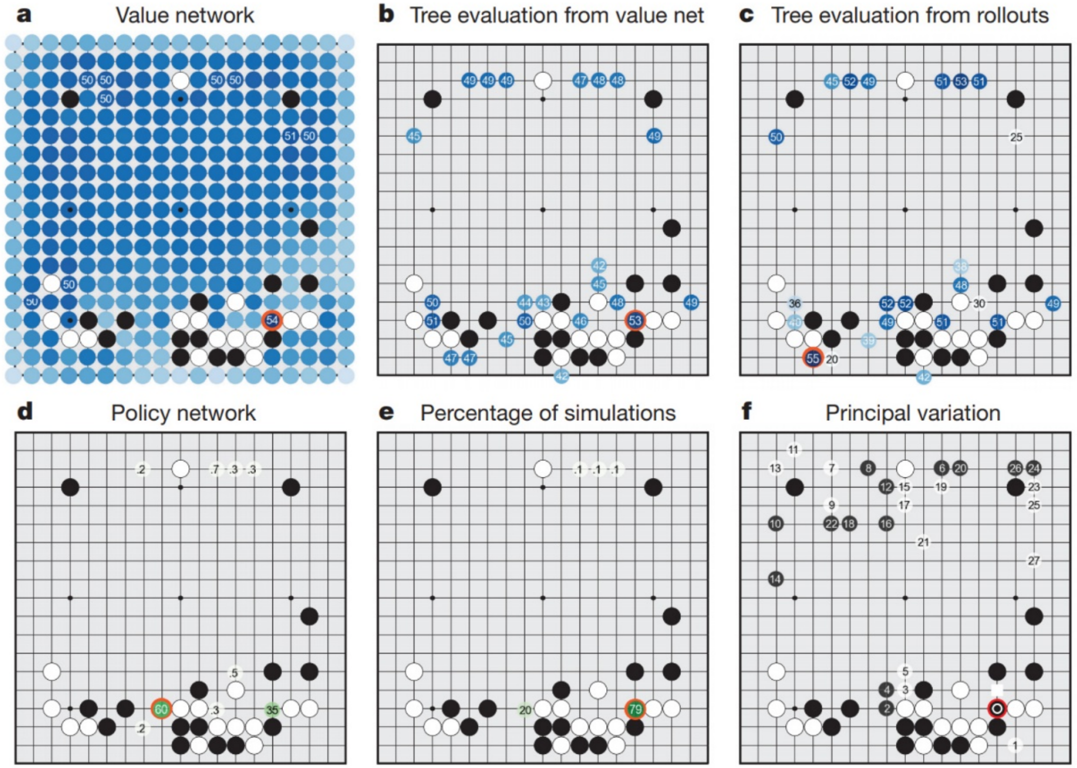
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.





devblogs.nvidia.com/mocha-jl-deep-learning-julia/

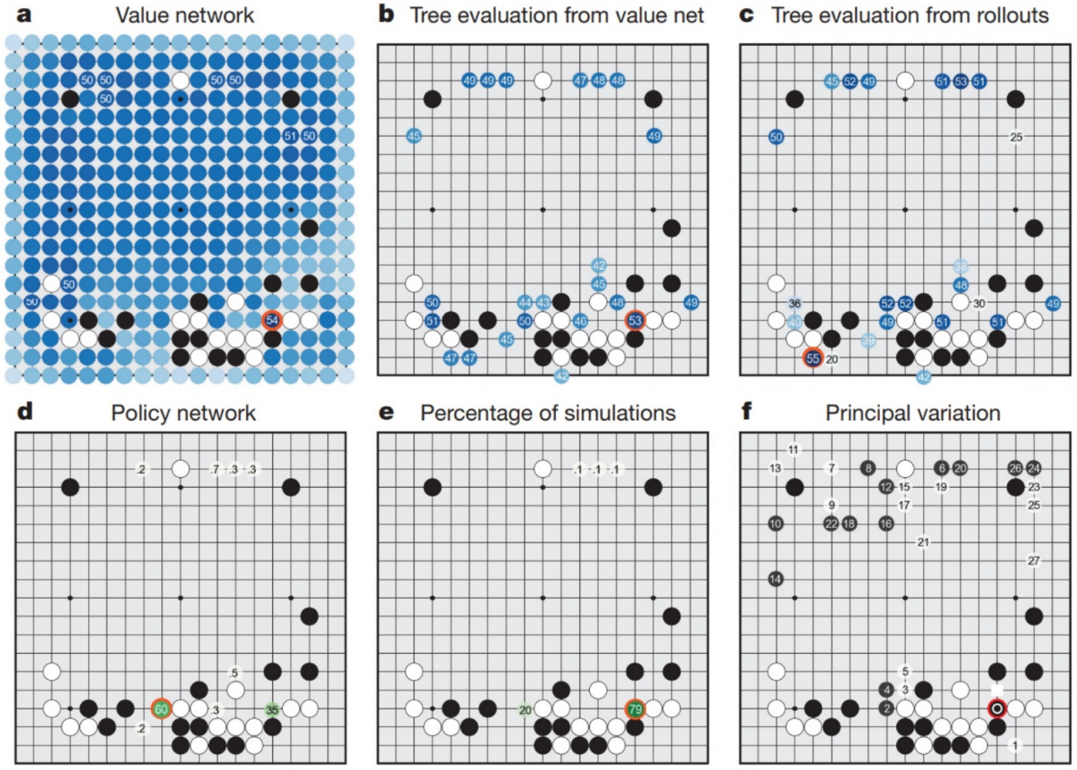
Alpha Go

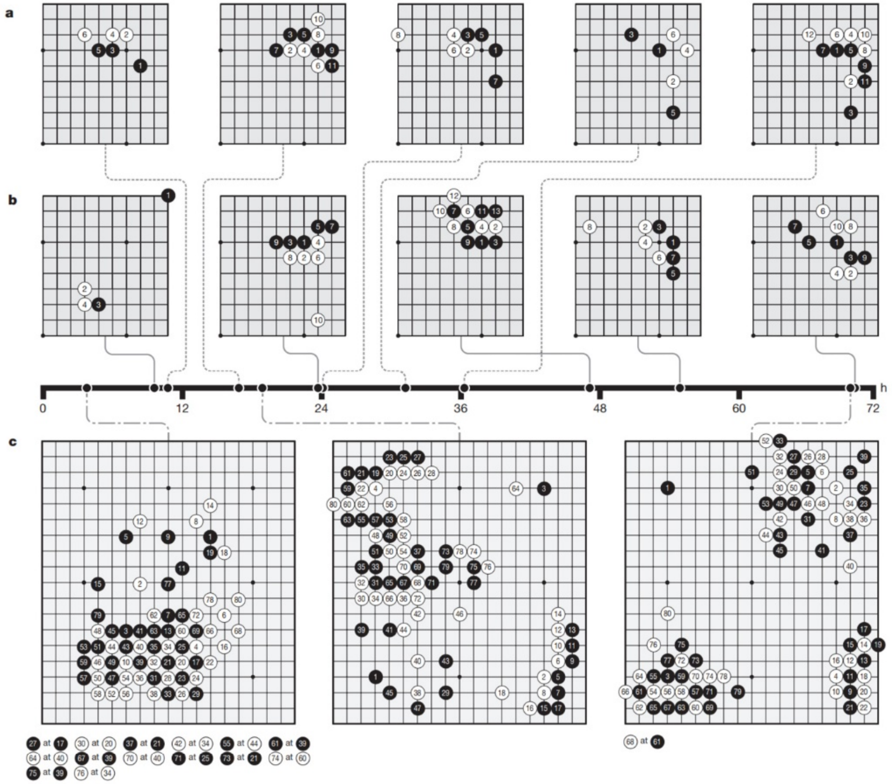


David Silver, Aja Huang, Chris J Maddison, et al.

Mastering the Game of Go with Deep Neural Networks and Tree Search
Nature - 2016

Alpha Go





David Silver, Aja Huang, Chris J Maddison, et al.

Mastering the Game of Go with Deep Neural Networks and Tree Search

Nature - 2016

David Silver, Julian Schrittwieser, Karen Simonyan, et al.

Mastering the Game of Go Without Human Knowledge

Nature - 2017

Open-Source Tools

houseroad Rename ZFNet to ZFNet-512 (#36)		Latest commit 3be4824 11 hours ago
bvlc_alexnet	Update bvlc_alexnet model	4 months ago
bvlc_googlenet	Add the value_info.json for the remaining of the models except style ...	3 months ago
bvlc_reference_caffenet	Add the value_info.json for the remaining of the models except style ...	3 months ago
bvlc_reference_rcnn_ilsvrc13	Add the value_info.json for the remaining of the models except style ...	3 months ago
densenet121	Add DenseNet-121 model	4 months ago
detectron	Add Detectron e2e_faster_rcnn_R-50-C4_2x model	3 months ago
inception_v1	Add Inception models	4 months ago
inception_v2	Add Inception models	4 months ago
resnet50	Add ResNet-50 model	4 months ago
scripts	Add Detectron e2e_faster_rcnn_R-50-C4_2x model	3 months ago
squeezenet	Correct SqueezeNet value_info to 227x227	3 months ago
style_transfer	Add other style transfer models	4 months ago
vgg19	Add VGG models	4 months ago
zfnet512	Rename ZFNet to ZFNet-512 (#36)	11 hours ago
.gitattributes	Remove squeezenet-specific lines from .gitattributes.	4 months ago
LICENSE	Add Apache 2.0 license	4 months ago
README.md	Update README to describe subdirectory access	3 months ago

Keras
DL4J
Caffe2
PYTORCH
torch
Caffe
Microsoft CNTK
TensorFlow
theano
mxnet

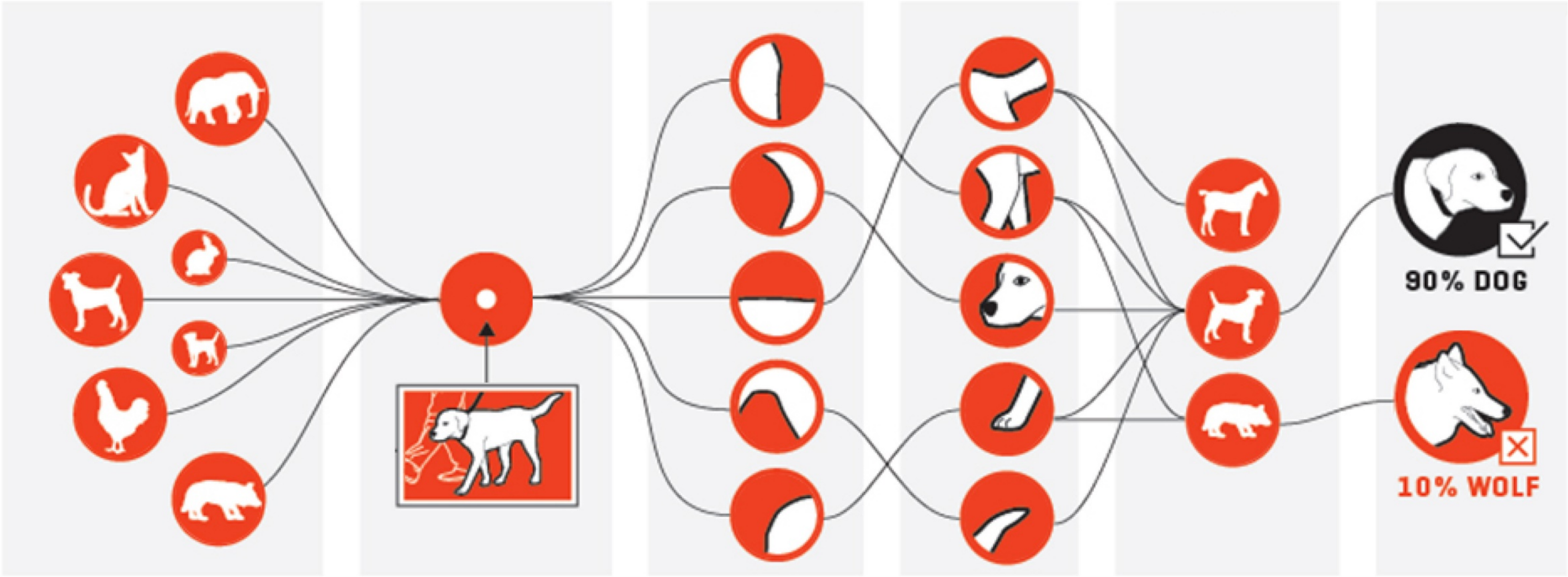
Yangqing Jia, Evan Shelhamer, Jeff Donahue, et al.

Caffe: Convolutional Architecture for Fast Feature Embedding
arxiv.org/abs/1408.5093

Misc.

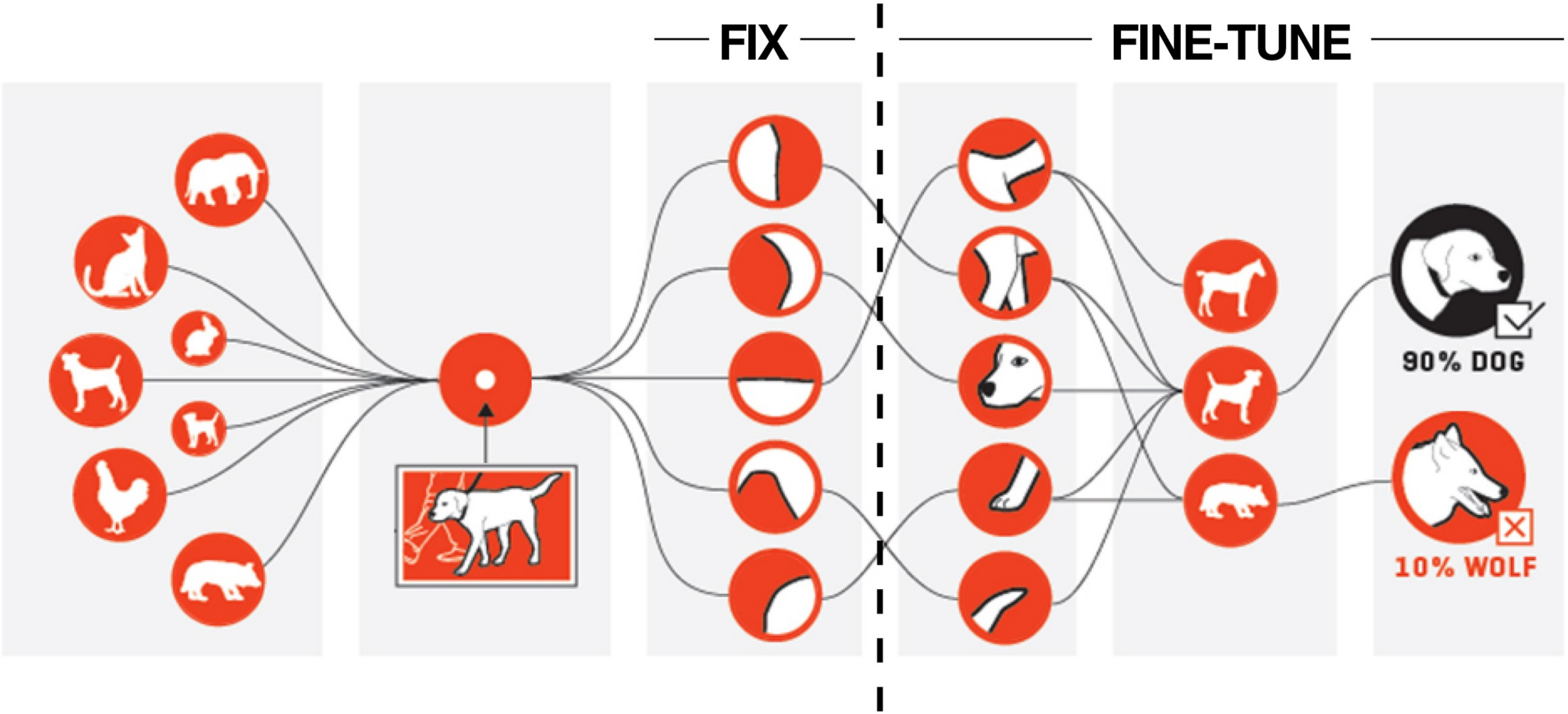
Open-Source Deep Learning Libraries
github.com

Transfer Learning



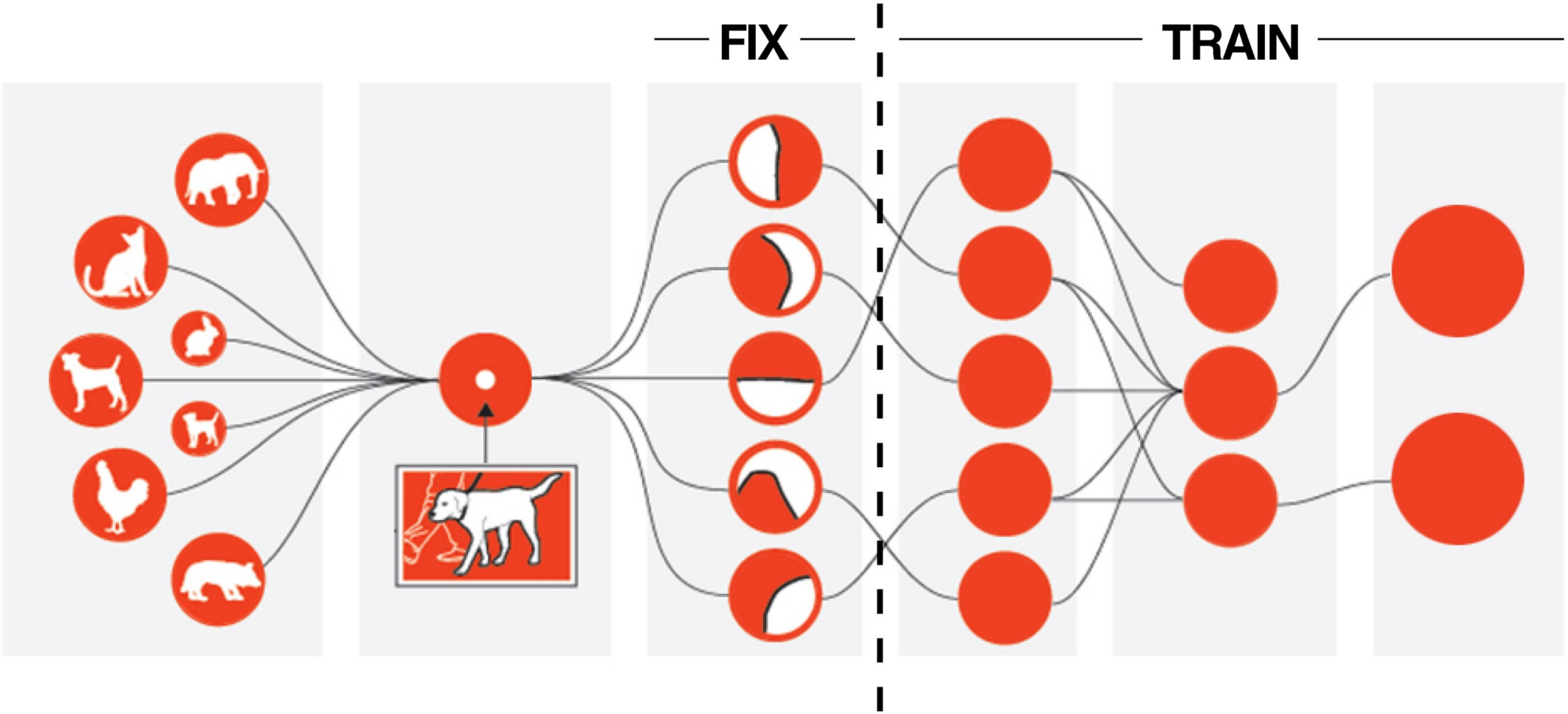
fortune.com/ai-artificial-intelligence-deep-machine-learning/

Transfer Learning



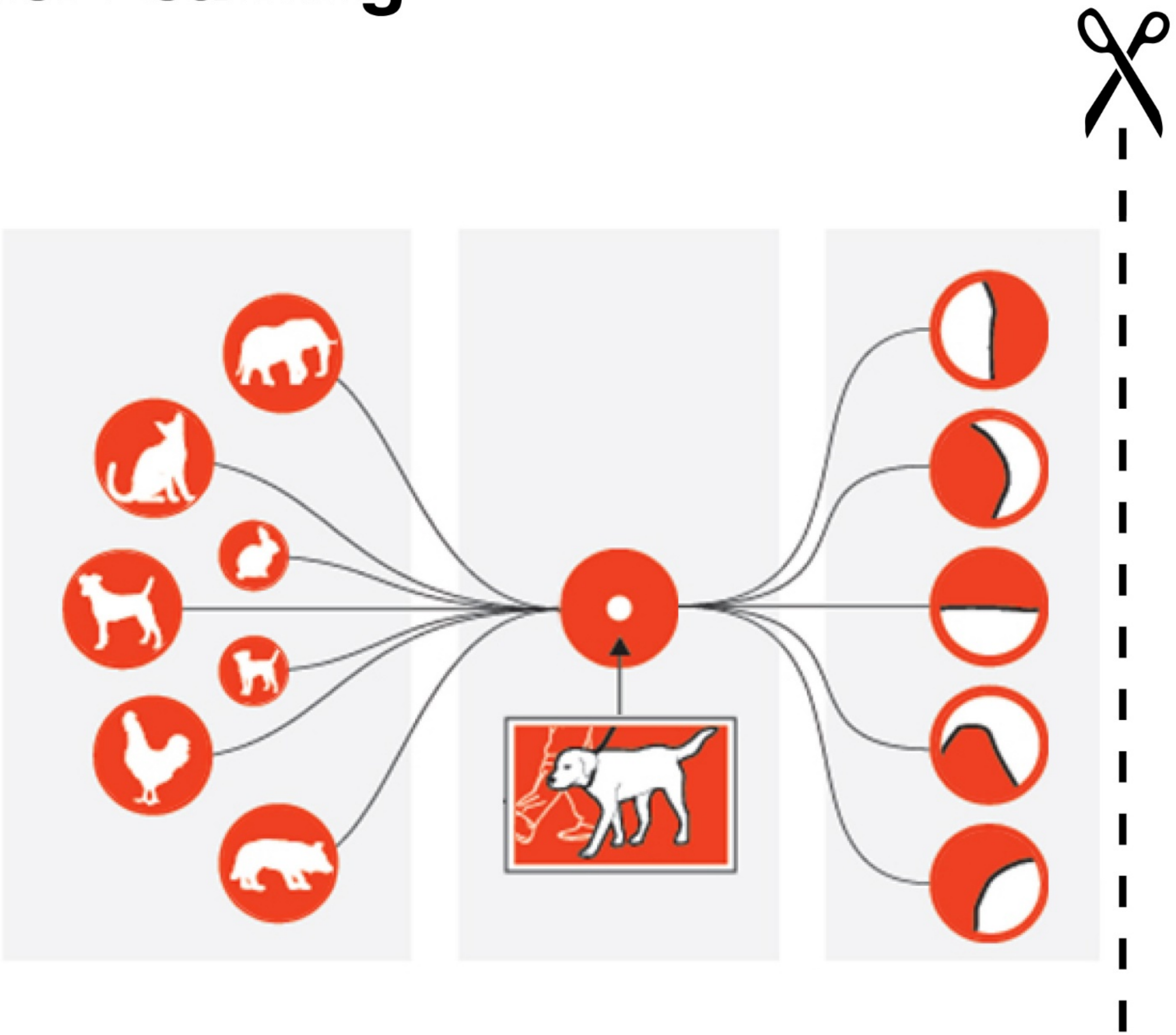
fortune.com/ai-artificial-intelligence-deep-machine-learning/

Transfer Learning



fortune.com/ai-artificial-intelligence-deep-machine-learning/

Transfer Learning



fortune.com/ai-artificial-intelligence-deep-machine-learning/

Deep Learning

Applications in Medical Imaging

Challenges

Early Logic and Statistical Pattern Recognition in Medicine

3 July 1959, Volume 130, Number 3366 SCIENCE

Reasoning Foundations of Medical Diagnosis

Symbolic logic, probability, and value theory aid our understanding of how physicians reason.

Robert S. Ledley and Lee B. Lusted

The purpose of this article is to analyze the complicated reasoning processes inherent in medical diagnosis. The importance of this problem has received recent emphasis by the increasing interest in the use of electronic computers as an aid to medical diagnostic processes (1, 2). Before computers can be used effectively for such purposes, however, we need to know more about how the physician makes a medical diagnosis.

If a physician is asked, "How do you
make a medical diag
tion of the process m
"First, I obtain the
patient's history, phy
and laboratory tests.
the relative importan
signs and symptoms.
may be of first-orde
other data of less im
make a differential
the diseases which th
reasonably resemble.
disease after another from the list until it becomes apparent that the case can be fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a patient he often has a "feeling about the case." This "feeling," although hard to explain, may be a summation of his impressions concerning the way the data seem to fit together, the patient's relia-

be integrated by the physician with a large store of possible diseases. It is widely believed that errors in differential diagnosis result more frequently from errors of omission than from other sources. For instance, concerning such errors of omission, Clendening and Hashinger (3) say: "How to guard against incompleteness I do not know. But I do know that, in my judgment, the most brilliant diagnosticians of my acquaintance are the ones who do remember and consider the most possibilities."

Computers are especially suited to help the physician collect and process clinical information and remind him of diagnoses which he may have overlooked. In many cases computers may be as simple as a set of hand-sorted cards, whereas in other cases the use of a large-scale digital electronic computer may be indicated. There are other ways in which computers may serve the physician, and some of these are suggested in this paper. For example, medical students might find the computer an important aid in learning the methods of differential diagnosis. But to use the computer thus we must understand how the physician makes a medical diagnosis. This, then, brings us to the subject of our investigation: the reasoning foundations of medical diagnosis and treatment.

Medical diagnosis involves processes that can be systematically analyzed, as well as those characterized as "intangible." For instance, the reasoning foundations of medical diagnostic procedures are precisely analyzable and can be separated from certain considered intangible judgments and value decisions. Such a separation has several important advantages. First, systematization of the rea-

can be developed. However, a consideration of foundations is always essential as the first step in the development of practical applications.

The reasoning foundations of medical diagnosis and treatment can be most precisely investigated and described in terms of certain mathematical techniques. Before material to illustrate these techniques was selected, many of the *New England Journal of Medicine*

Dr. Ledley is a part-time member of the staff of the National Academy of Sciences-National Research Council, Washington, D.C., where he is principal investigator of the Survey and Monograph on Electronic Computers in Biology and Medicine. He is on the faculty of the electrical engineering department of George Washington University and mathematician at the Data Processing Systems Division of the National Bureau of Standards. Dr. Lusted is radiologist and associate professor at the University of Rochester School of Medicine, Rochester, N.Y.

"increasing interest in the use of electronic computers as aid to medical diagnostic processes"

VOL. 81 NO. 2 Radiology AUGUST 1963

a monthly journal devoted to clinical radiology and allied sciences

PUBLISHED BY THE RADIOLOGICAL SOCIETY OF NORTH AMERICA, INC.

The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer[1]

GWILYM S. LODWICK, M.D., THEODORE E. KEATS, M.D., and JOHN P. DORST, M.D.

THIS PAPER WILL DESCRIBE a concept of converting the visual images on roentgenograms into numerical sequences that can be manipulated and evaluated
by the digita
the results o
determine the
graphic findi
development
conveying ra
an electronic
possible expl
digital compu
diagnosis. T
vast numbers
from the phy
with its sto
the result of
of statistical
out these fu
accuracy ma
computer mo
communicati
mits high-sp
radiological data, is a logical approach to the control of a segment of exponentially expanding medical knowledge.

We have chosen to apply this concept to roentgenograms of lung cancer because, against a background of air density, the intimate details of the relationship between tumor and host may be faithfully reproduced roentgenographically. Par-

either resection of a segment of
pneumonectomy.

e-year survival data for this group of cases are shown in Table I. Less than 1 per cent of the total number were lost to follow-up. The absolute survival rate of 1.3 per cent for this highly malignant tumor is even lower than that

[1] From the Department of Radiology, University of Missouri School of Medicine, Columbia, Mo. (Drs. Lodwick and Keats), and the Department of Radiology, University of Iowa College of Medicine, Iowa City, Iowa. Dr. Dorst is now at the University of Cincinnati.

This investigation was supported in part by the James Picker Foundation on recommendation of the Committee on Radiology, National Academy of Sciences-National Research Council. Presented in part at the Forty-third Annual Meeting of the Radiological Society of North America, Chicago, Ill., Nov. 17–22, 1957. Submitted for publication in October 1962.

"a concept of converting the visual images on roentgenograms into numerical sequences... by the digital computer... to determine the significance of certain radiographic findings in lung cancer"

Robert S Ledley & Lee B Lusted

Reasoning Foundations of Medical Diagnosis

Science - 1959

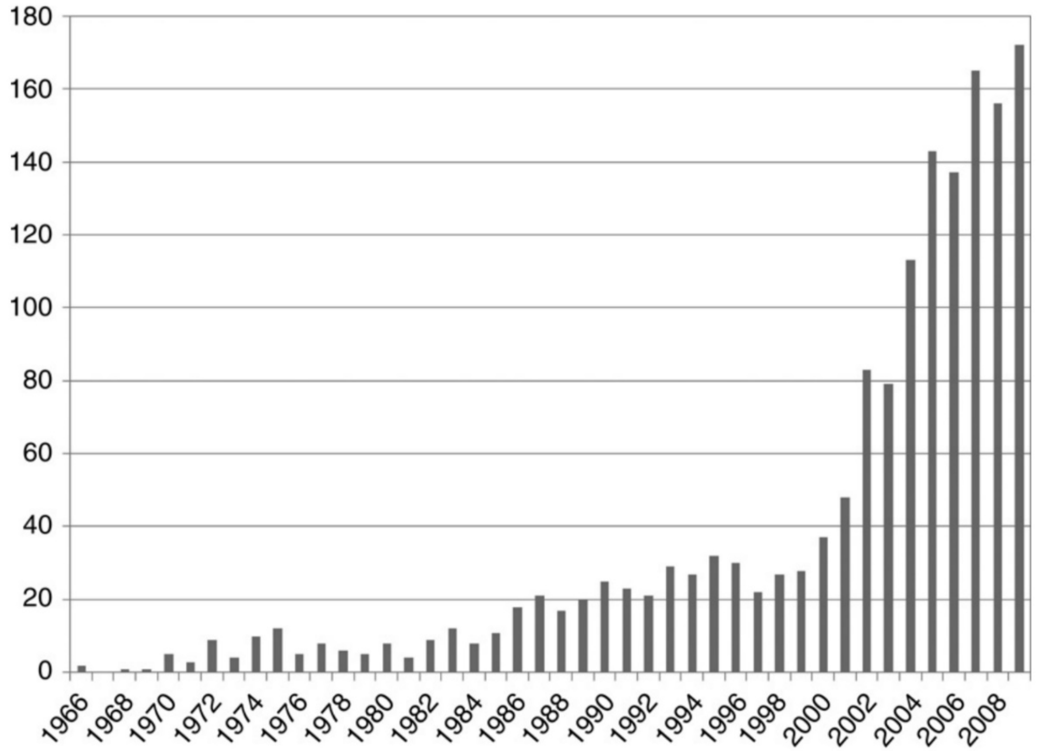
Gwilym S Lodwick, Theodore E Keats & John P Dorst

The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer

Radiology - 1963

Computer-Aided Diagnosis

number of publications on or related to CAD



CAD systems approved or cleared by FDA in the US

CAD Systems Approved or Cleared by the FDA in the United States

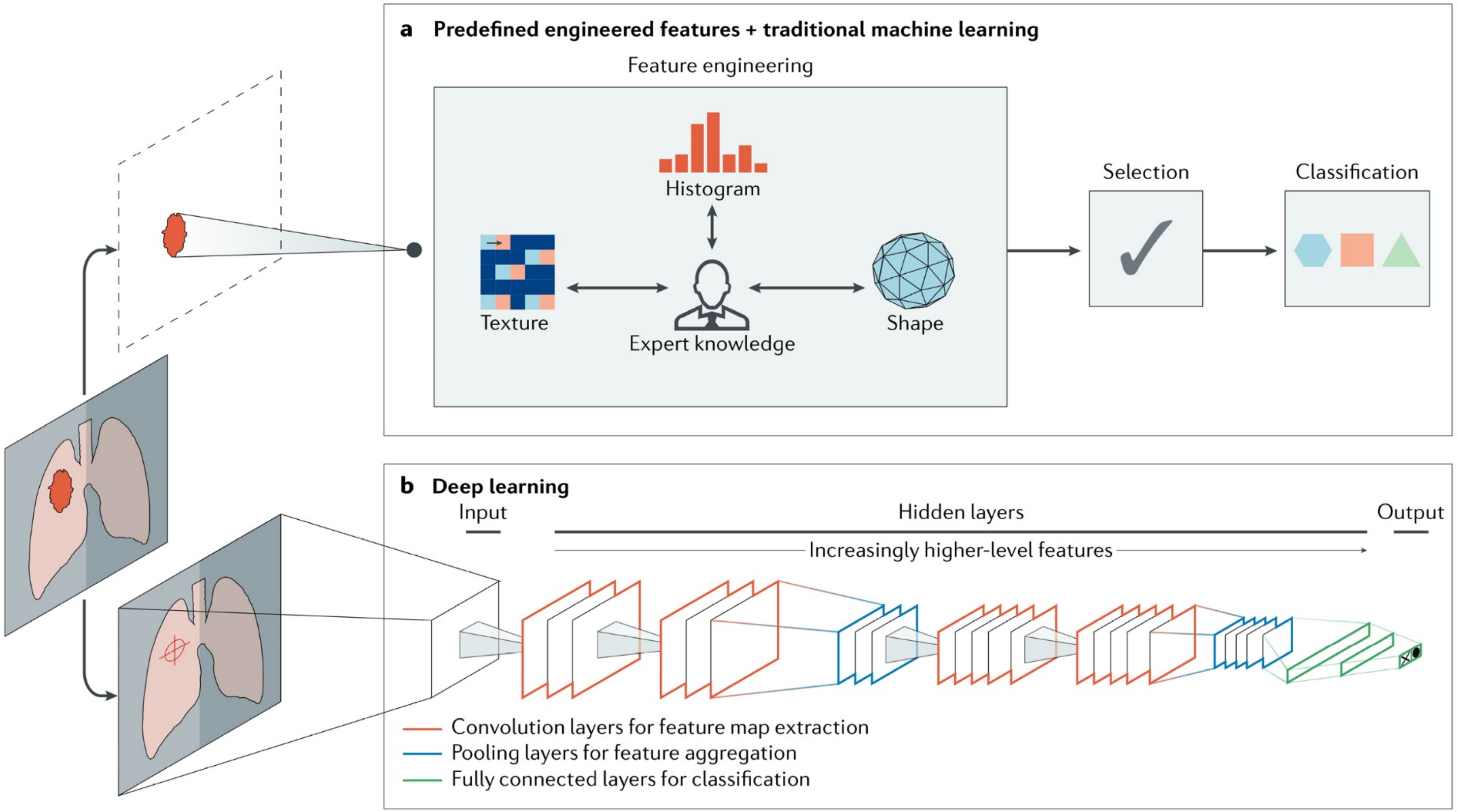
Name/Company	What It Does	Type of Approval	First and Last Date
Imagechecker/R2 Technology, Sunnyvale, Calif; Hologic, Bedford, Mass	Mass and microcalcification detection on mammograms	PMA	6/1998–9/2007
Logicon caries detection/GA Industries, Rancho Palos Verdes, Calif	Detection of caries on intraoral radiographs	PMA	9/1998–1/2007
Rapidscreen, Onguard/Riverain Medical, Miamisburg, Ohio	Nodule detection on chest radiographs	PMA	7/2001–8/2007
SecondLook/Icad, Nashua, NH,	Mass and microcalcification detection on mammograms	PMA	1/2002–10/2008
LungCare Nodule Enhanced Viewing/Siemens, Erlangen, Germany	Nodule detection and volumetry at chest CT	510(k)	11/2003
MedicLung/MedicSight, London, England	Nodule segmentation and viewing at chest CT	510(k)	12/2003
CT Colonography/General Electric, Fairfield, Conn	Detection of polyps at CT	510(k)	5/2004
Imagechecker-CT/R2 Technology, Sunnyvale, Calif	Detection of pulmonary embolism at chest CT	510(k)	6/2004
Lung CAR/MedicSight, London, England	Nodule detection and volumetry at chest CT	510(k)	7/2004
Colon Car/MedicSight, London, England	Detection of polyps at CT	510(k)	10/2004
Syngo Colonography/Siemens, Erlangen, Germany	Detection of polyps at CT	510(k)	10/2004
IQQA/EDDA, Princeton, NJ	Nodule detection on chest radiographs	510(k)	10/2004
Kodak Mammography CAD Engine/Carestream, Rochester, NY	Mass and microcalcification detection on mammograms	PMA	11/2004–3/2007
Advanced Lung Analysis 2/General Electric, Fairfield, Conn	Nodule detection and volumetry at chest CT	510(k)	11/2004
Syngo Lung CAD/Siemens, Erlangen, Germany	Nodule detection and volumetry at chest CT	510(k)	10/2006
ImageChecker CT CAD/Hologic, Bedford, Mass	Nodule detection and volumetry at chest CT	510(k)	12/2007

Bram van Ginneken, Cornelia M Schaefer-Prokop & Mathias Prokop

Computer-aided Diagnosis: How to Move from the Laboratory to the Clinic

Radiology - 2011

Deep Learning

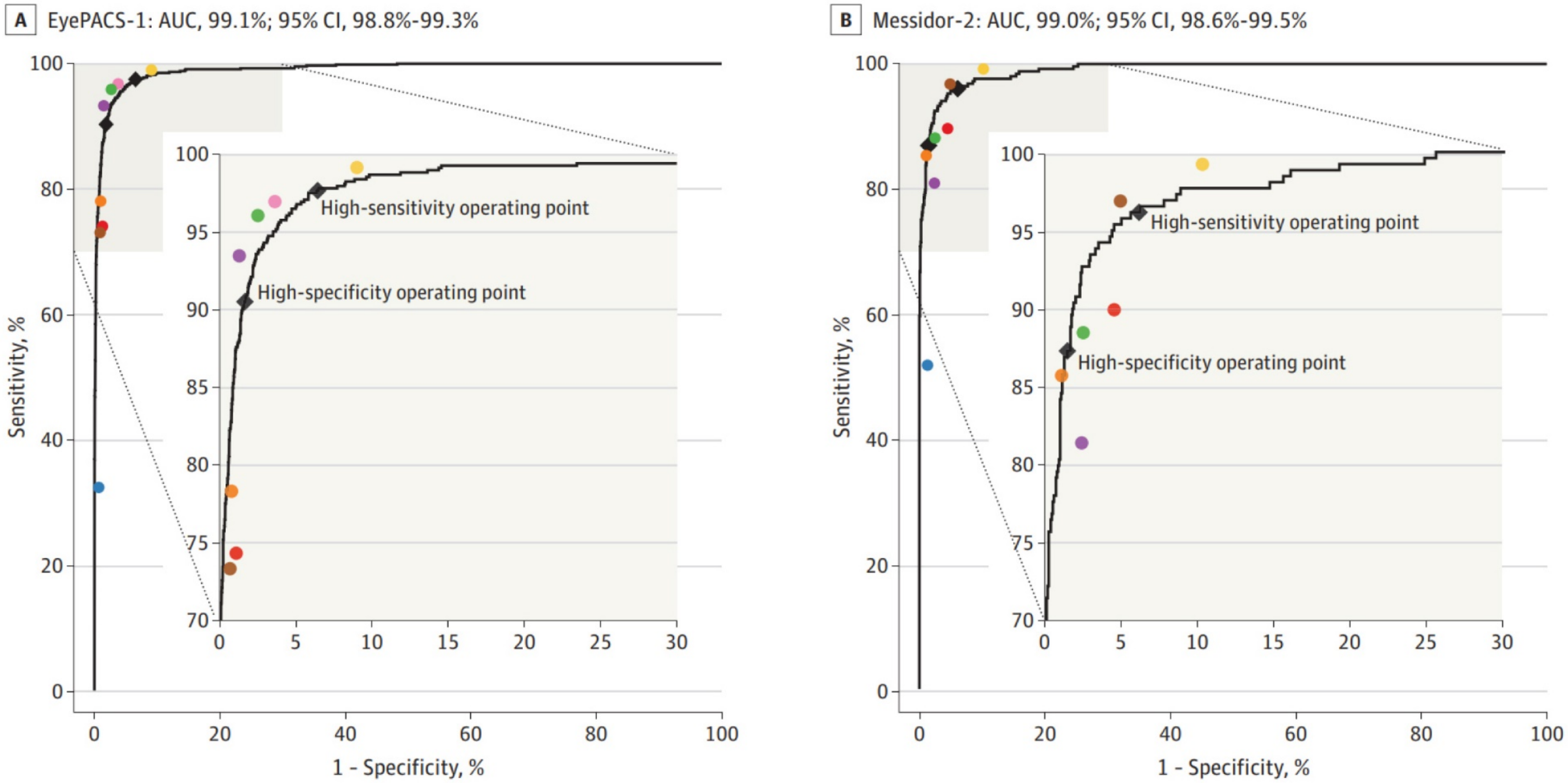


Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz & Hugo JWL Aerts

Artificial Intelligence in Radiology

Nature Reviews Cancer - 2018

Diabetic Retinopathy Detection

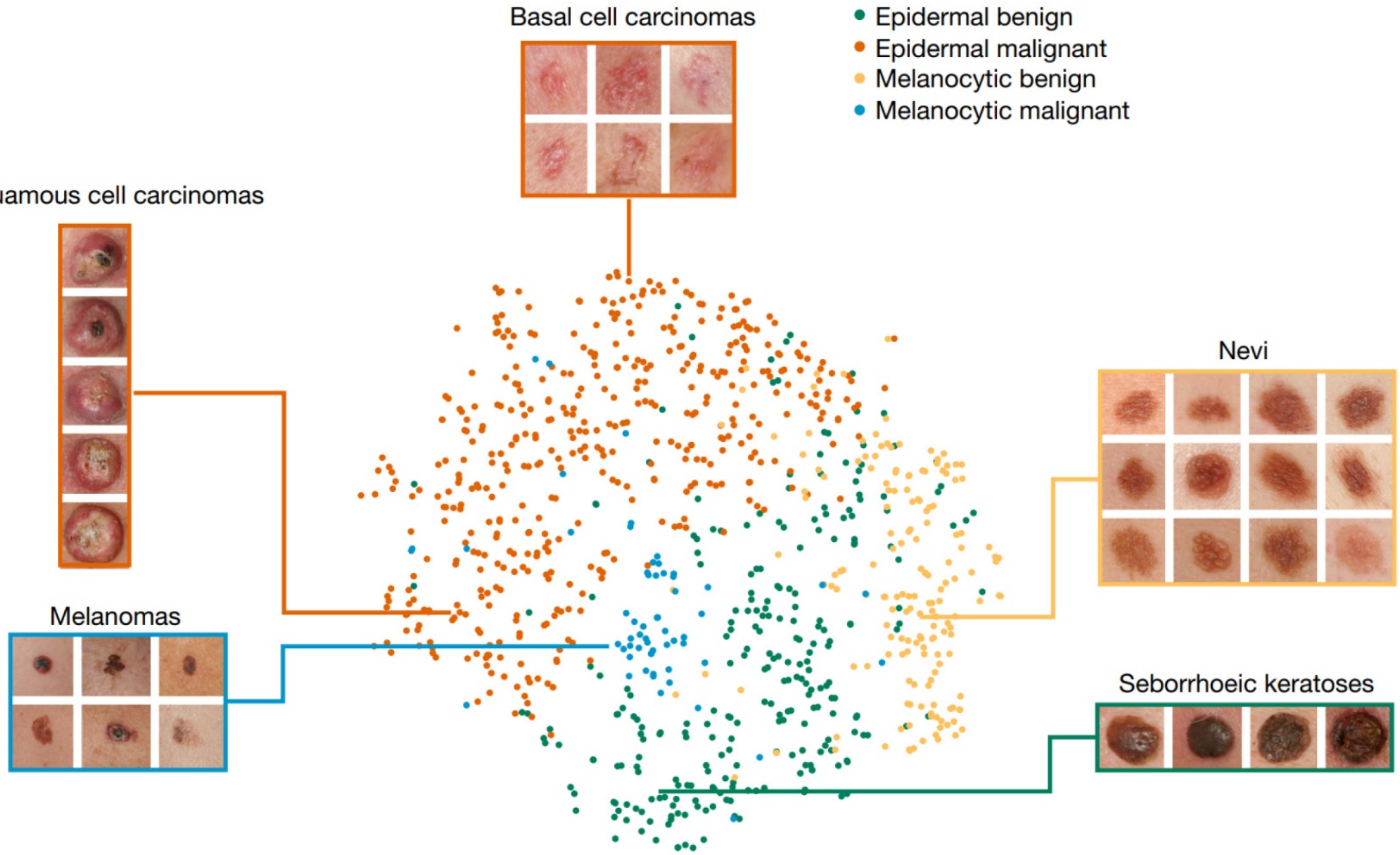


Varun Gulshan, Lily Peng, Marc Coram, et al.

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

The Journal of the American Medical Association (JAMA) - 2016

Skin Lesion Classification

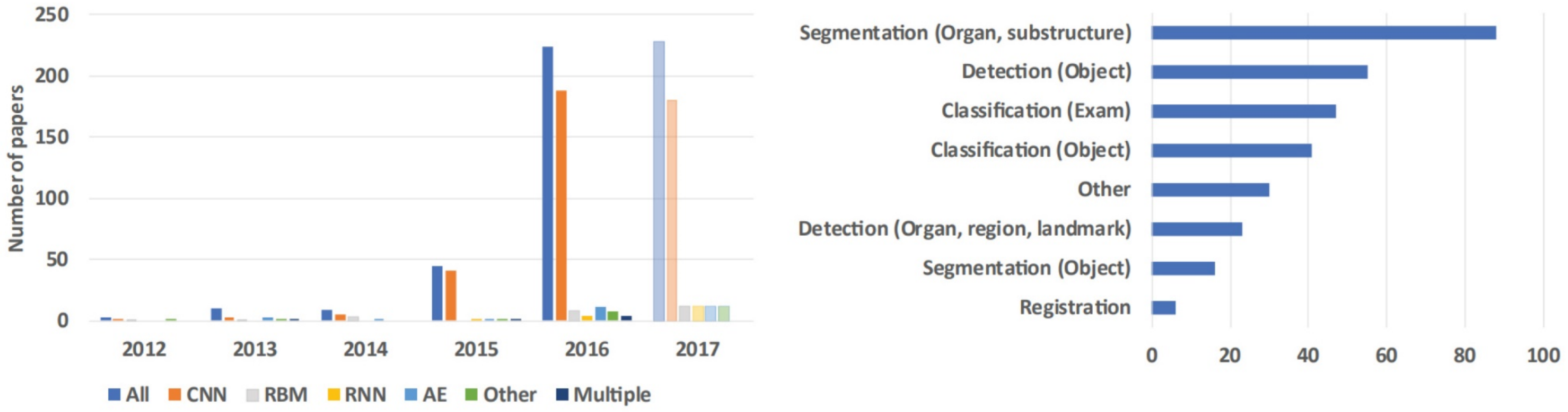


Andre Esteva, Brett Kuprel, Robert A Novoa, et al.

Dermatologist-level Classification of Skin Cancer with Deep Neural Networks

Nature - 2017

Deep Learning in Medical Imaging

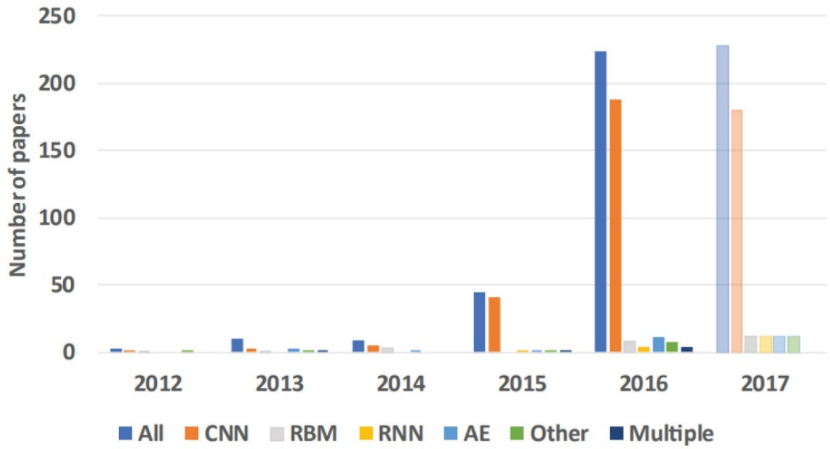


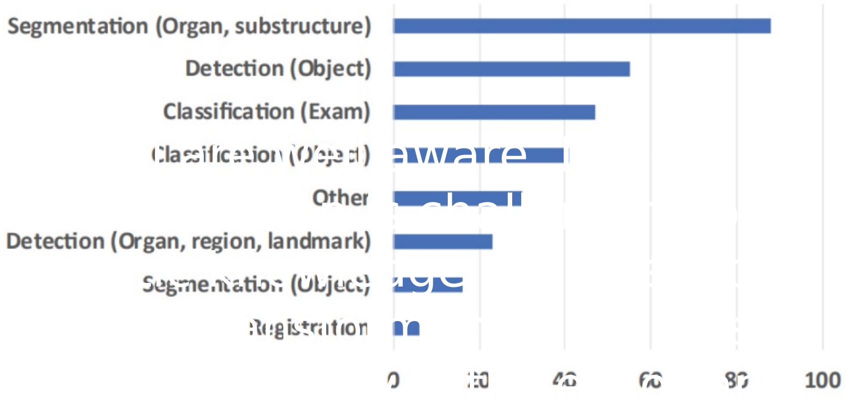
Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al.

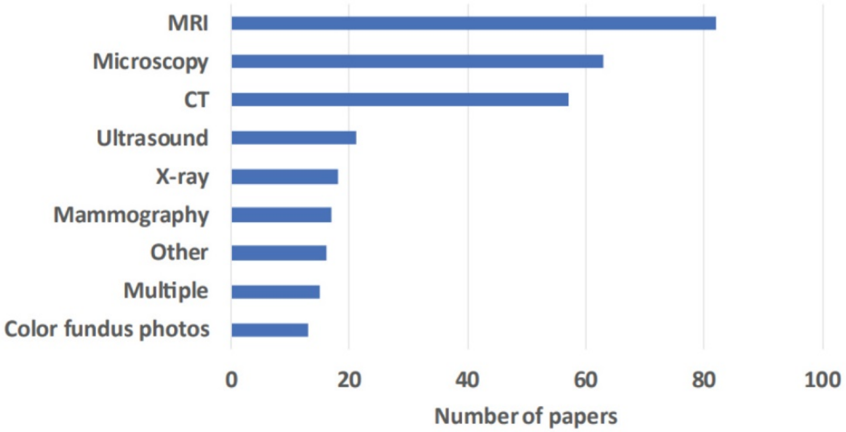
A Survey on Deep Learning in Medical Image Analysis

Medical Image Analysis - 2017

Deep Learning in Medical Imaging





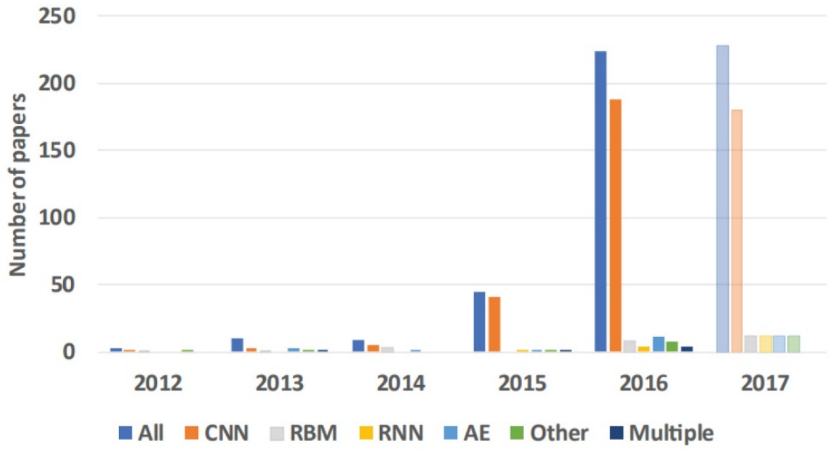


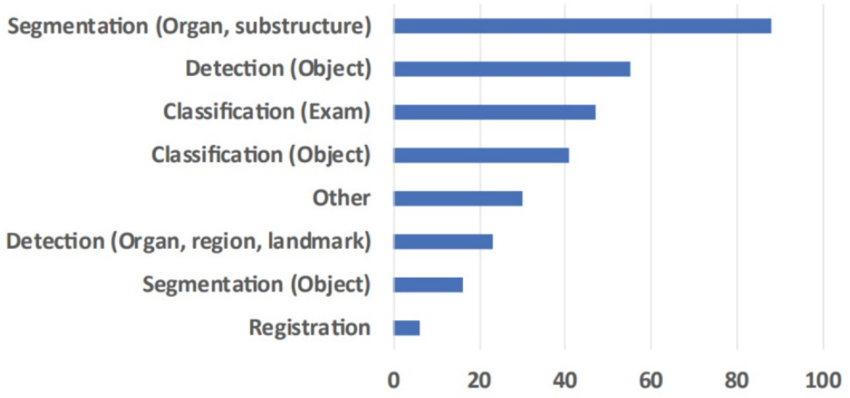
Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al.

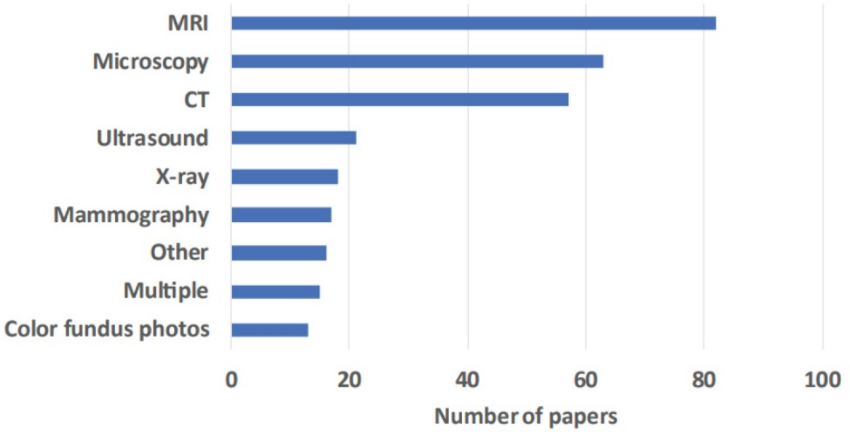
A Survey on Deep Learning in Medical Image Analysis

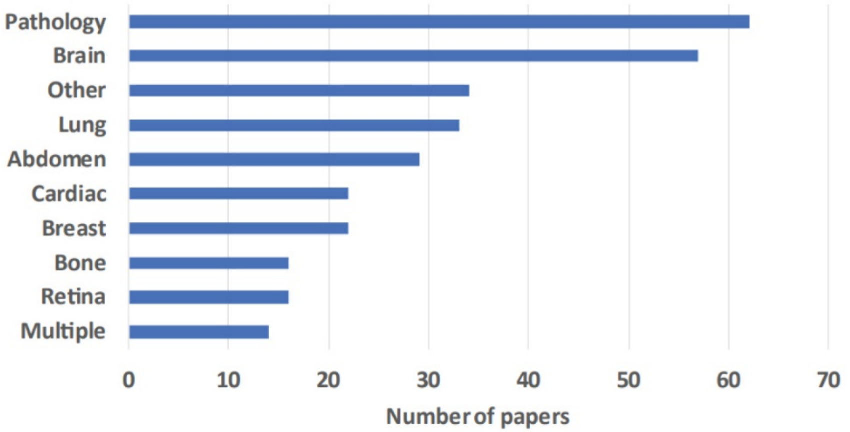
Medical Image Analysis - 2017

Deep Learning in Medical Imaging







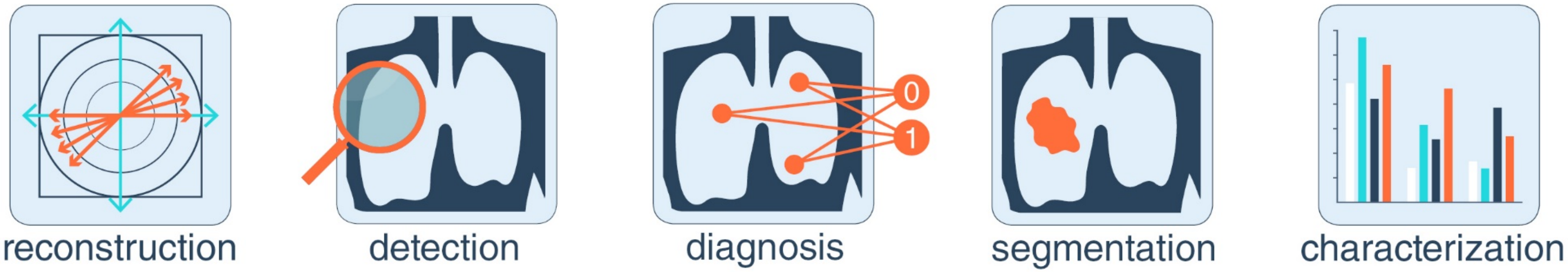


Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al.

A Survey on Deep Learning in Medical Image Analysis

Medical Image Analysis - 2017

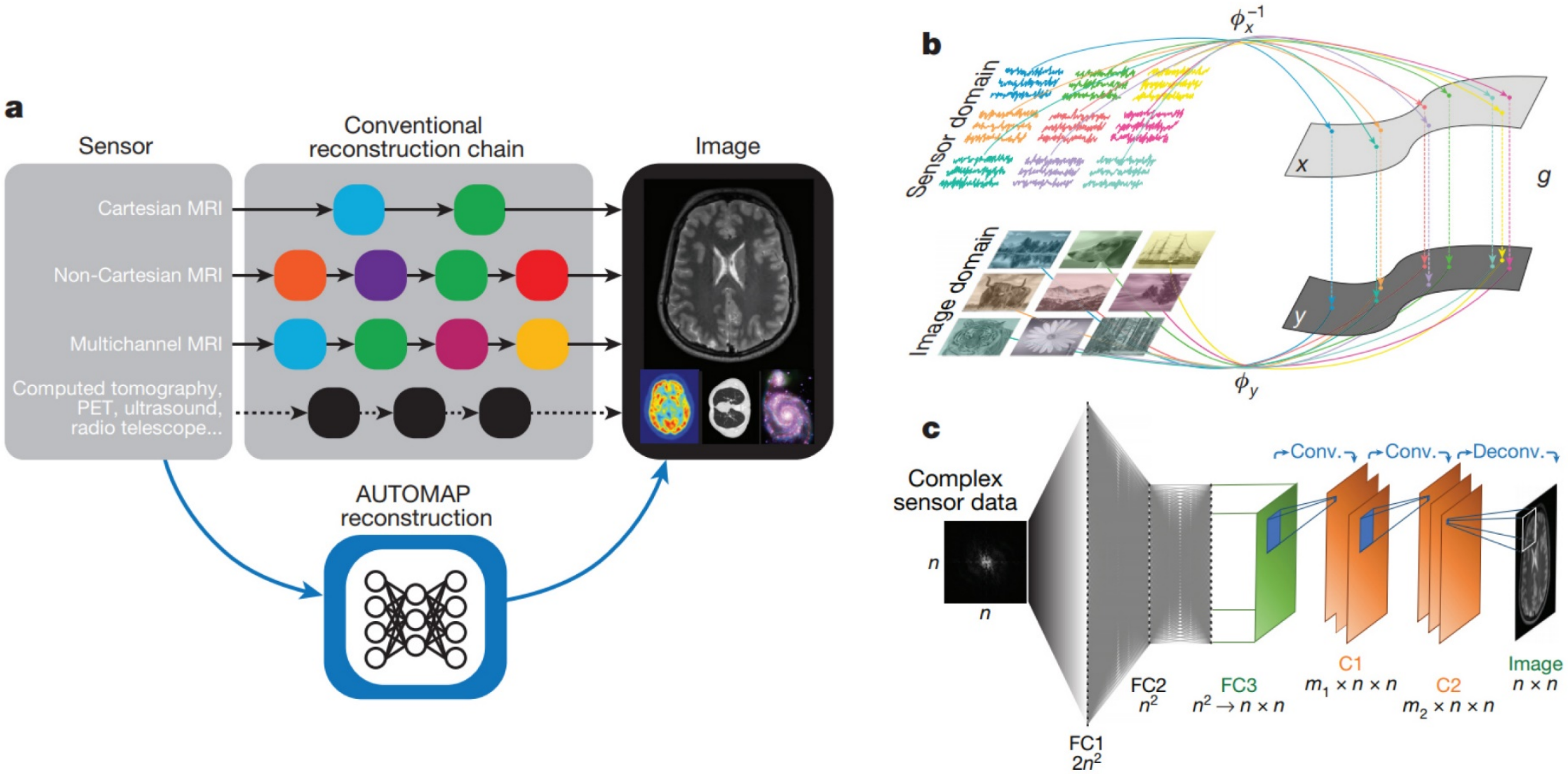
Artificial Intelligence in Radiology



Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz and Hugo JWL Aerts

Artificial Intelligence in Radiology
Nature Reviews Cancer - 2018

Reconstruction

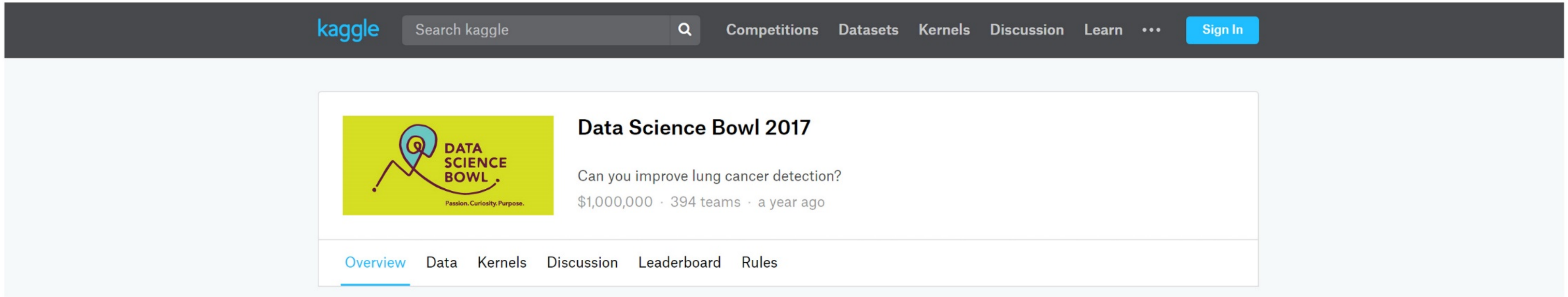


Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen & Matthew S Rosen

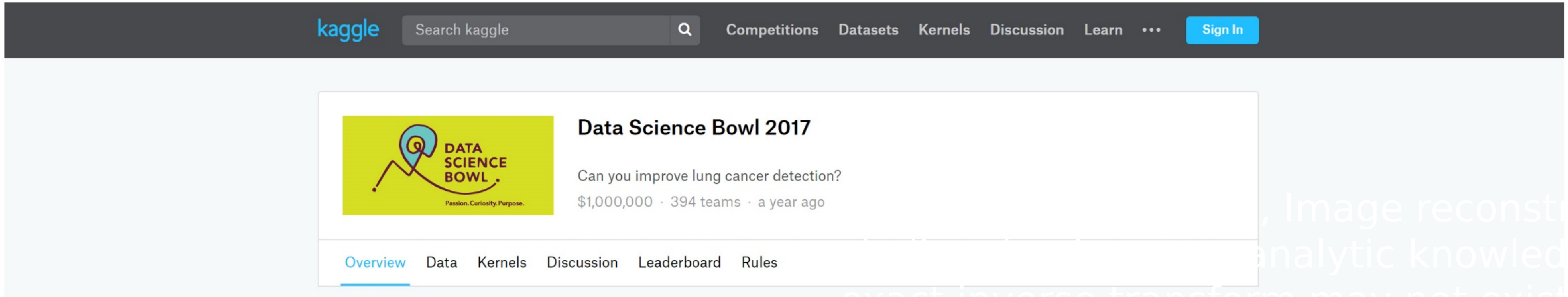
Image Reconstruction by Domain-transform Manifold Learning

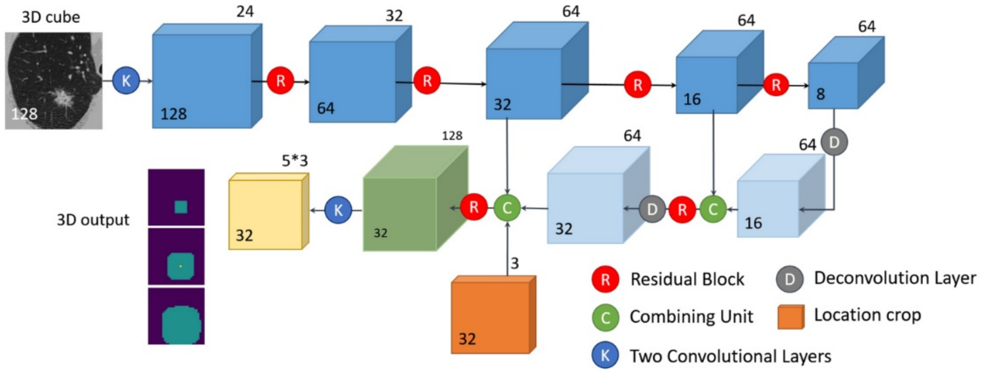
Nature - 2018

Detection & Diagnosis



Detection & Diagnosis



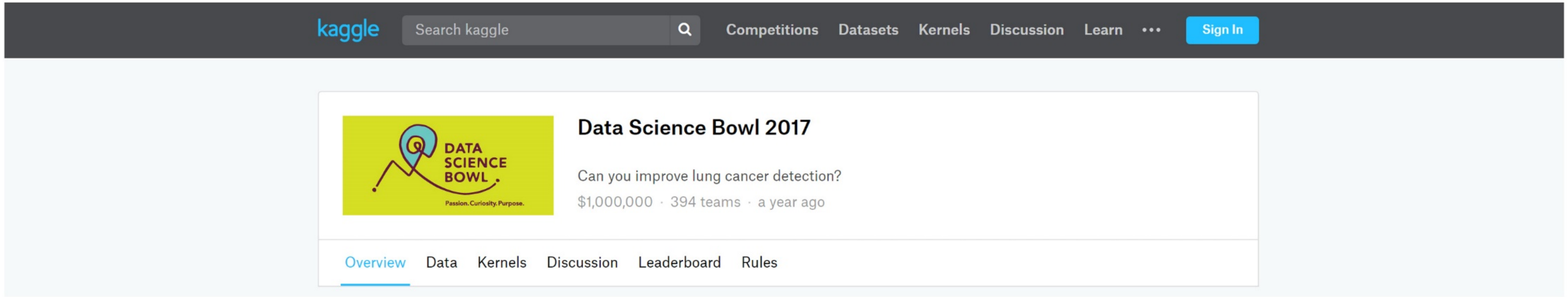


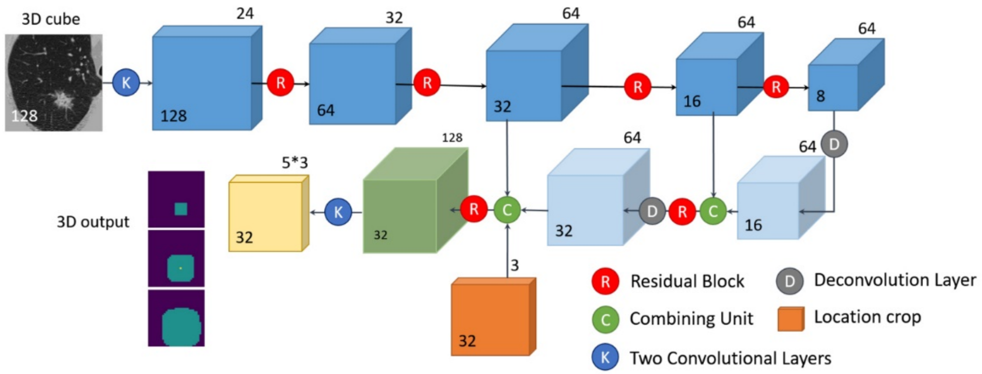
Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu & Sen Song

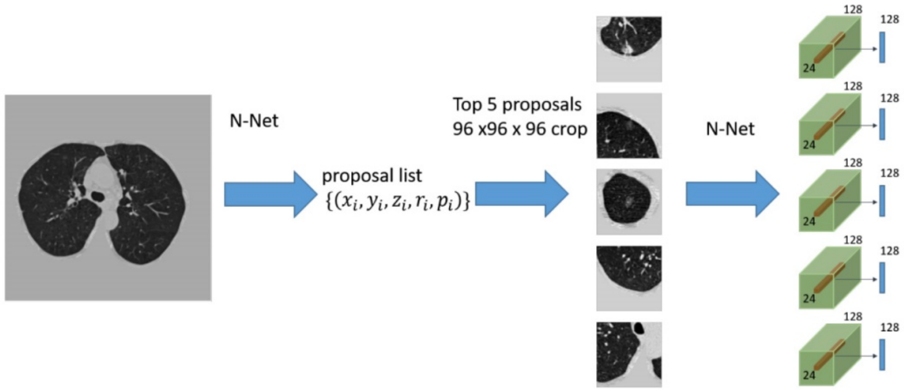
Evaluate the Malignancy of Pulmonary Nodules Using the 3D Deep Leaky Noisy-or Network

arxiv.org/abs/1711.08324

Detection & Diagnosis





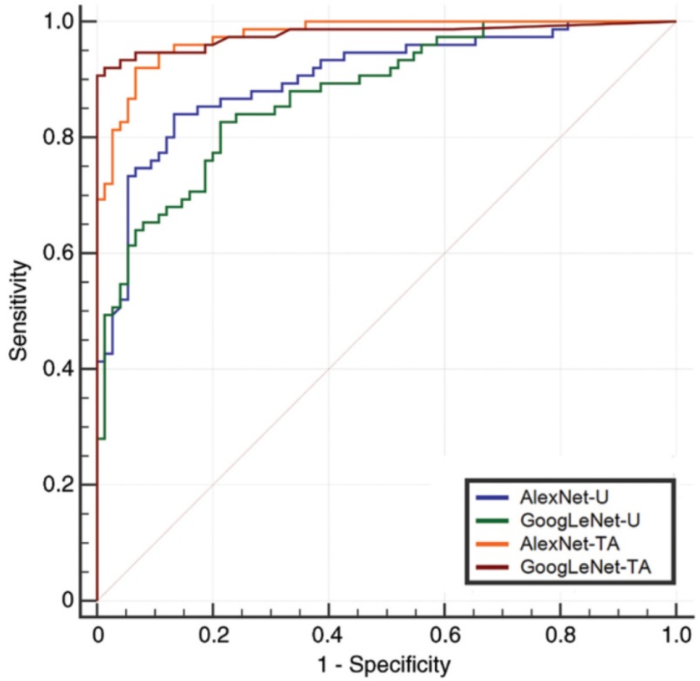


Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu & Sen Song

Evaluate the Malignancy of Pulmonary Nodules Using the 3D Deep Leaky Noisy-or Network

arxiv.org/abs/1711.08324

Detection & Diagnosis

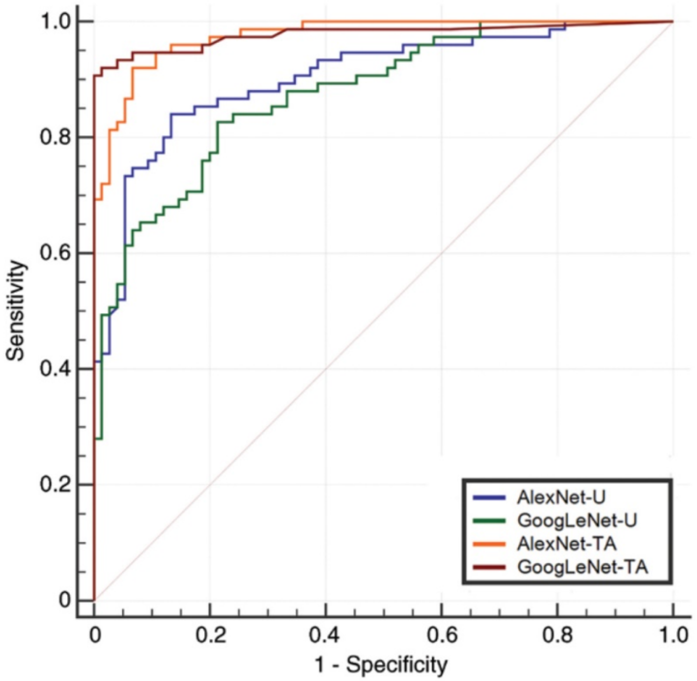


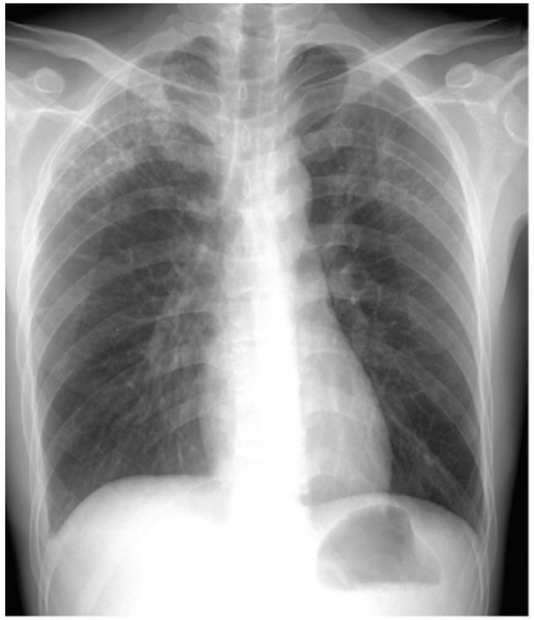
Paras Lakhani & Baskaran Sundaram

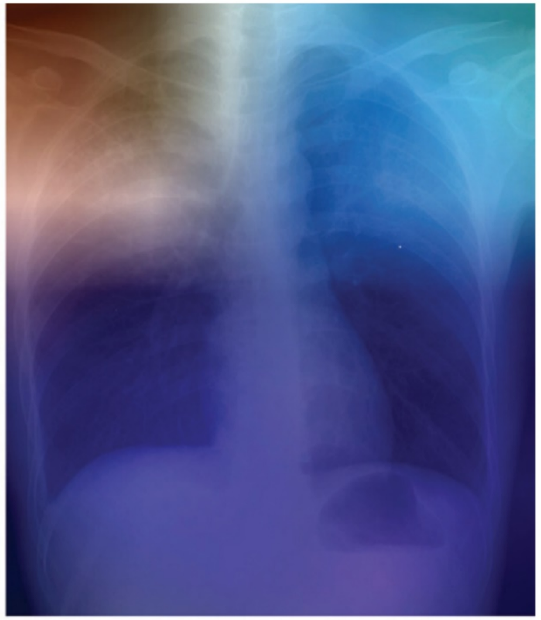
Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks

Radiology - 2017

Detection & Diagnosis







Paras Lakhani & Baskaran Sundaram

Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks
Radiology - 2017

Segmentation

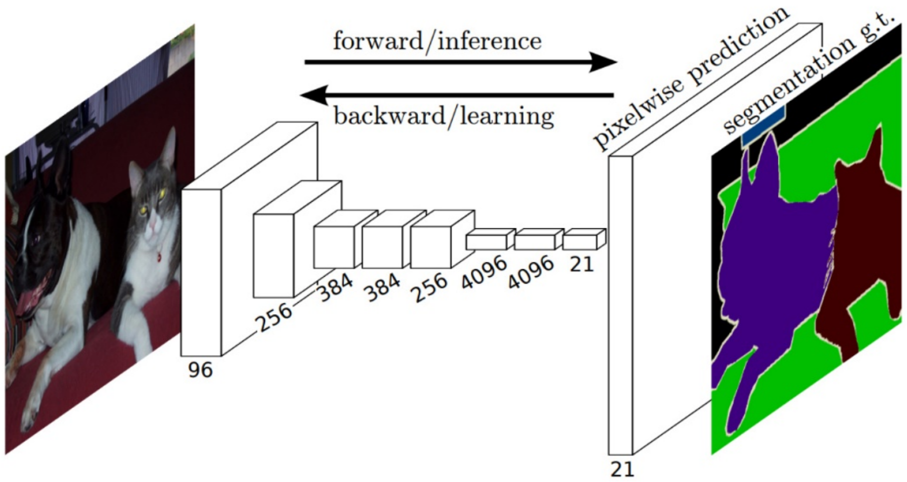


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

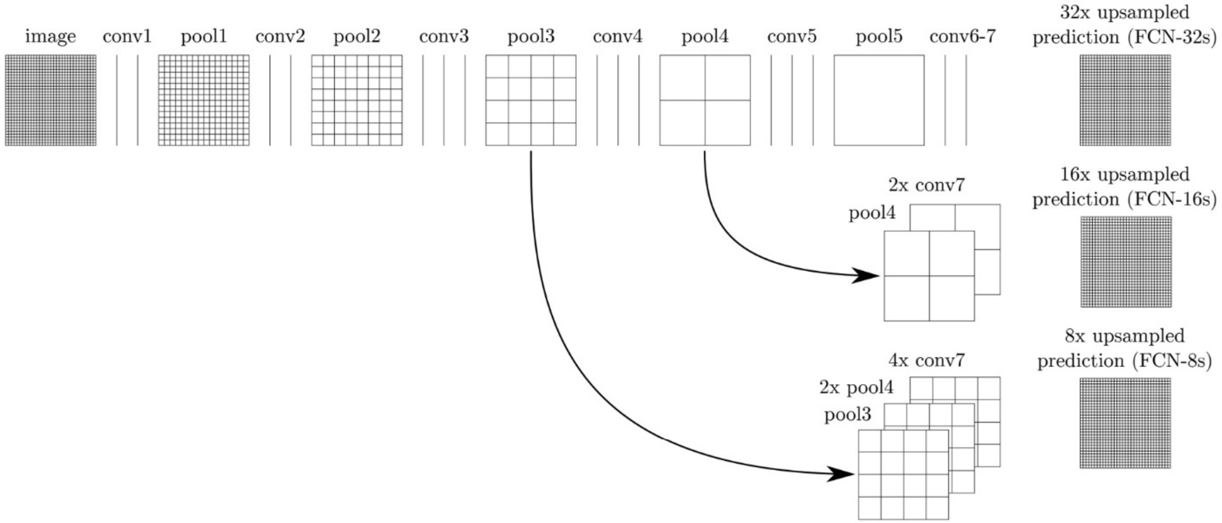


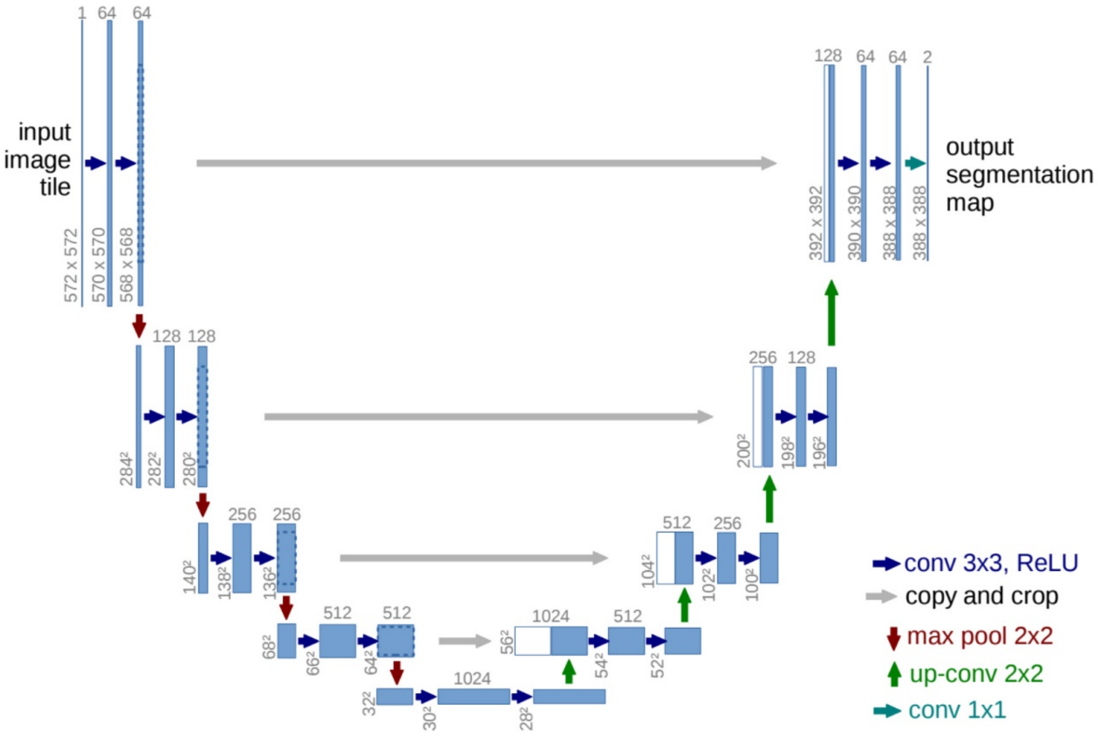
Figure 3. Our DAG nets learn to combine coarse, high layer information with fine, low layer information. Pooling and prediction layers are shown as grids that reveal relative spatial coarseness, while intermediate layers are shown as vertical lines. First row (FCN-32s): Our single-stream net, described in Section 4.1, upsamples stride 32 predictions back to pixels in a single step. Second row (FCN-16s): Combining predictions from both the final layer and the `pool4` layer, at stride 16, lets our net predict finer details, while retaining high-level semantic information. Third row (FCN-8s): Additional predictions from `pool3`, at stride 8, provide further precision.

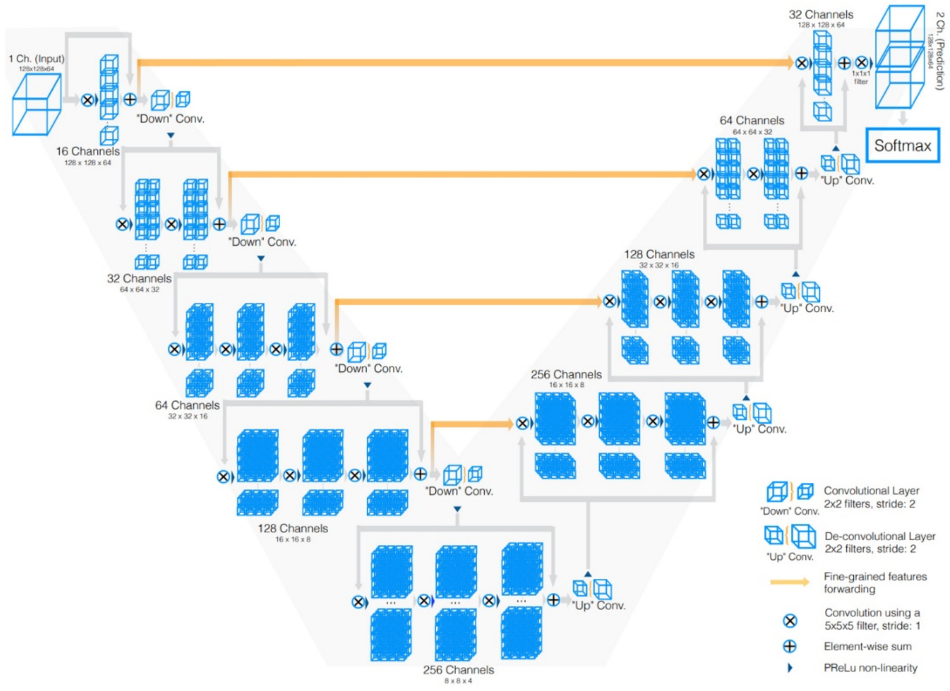
Jonathan Long, Evan Shelhamer & Trevor Darrell

Fully Convolutional Networks for Semantic Segmentation

Conference on Computer Vision and Pattern Recognition (CVPR) - 2015

Segmentation





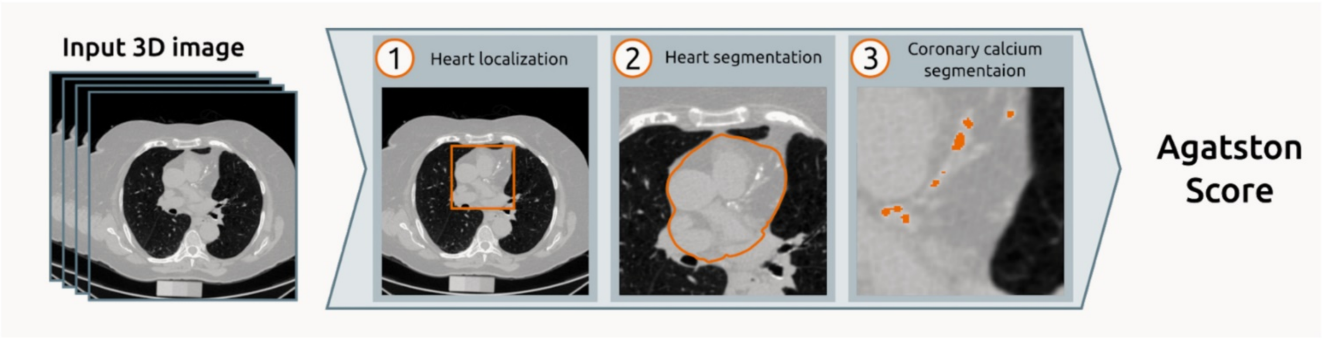
Olaf Ronneberger, Philipp Fischer & Thomas Brox

U-Net: Convolutional Networks for Biomedical Image Segmentation
MICCAI - 2015

Fausto Milletari, Nassir Navab & Seyed-Ahmad Ahmadi

V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation
arxiv.org/abs/1606.04797

Segmentation

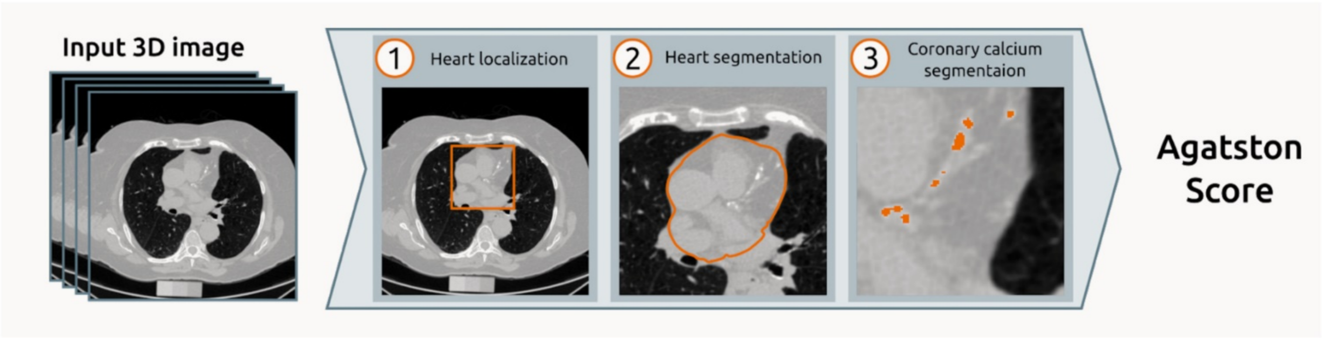


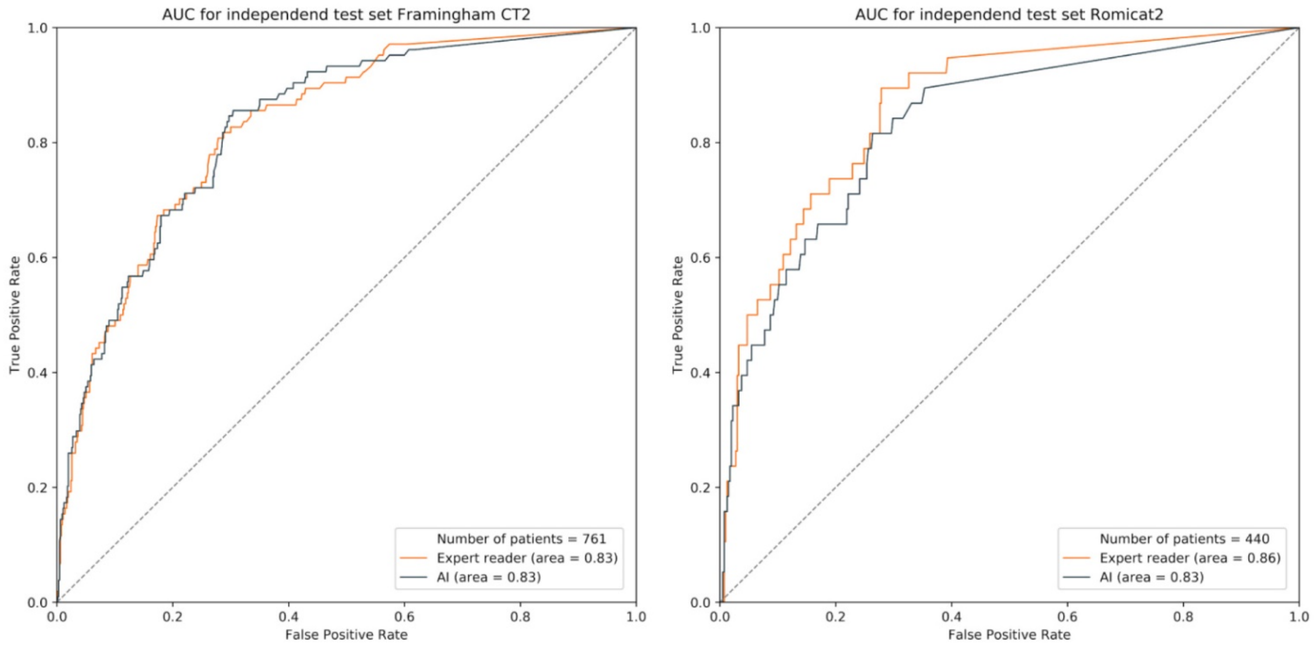
Roman Zeleznik, Parastou Eslami, Borek Foldyna, et al.

Deep Convolutional Neural Networks to Predict Cardiovascular Risk from Non-contrast Computed Tomography Images

Under Review - 2018

Segmentation



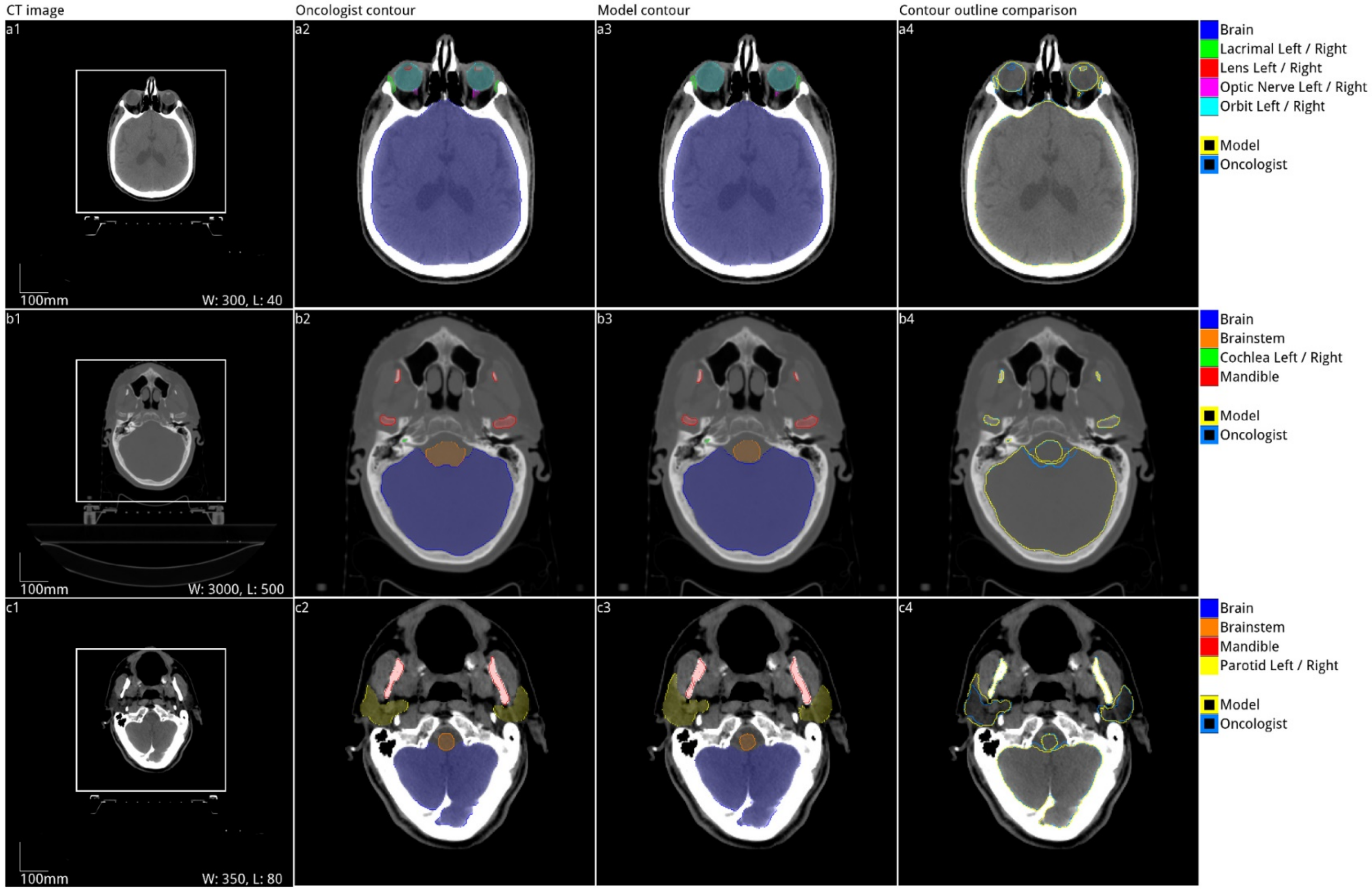


Roman Zeleznik, Parastou Eslami, Borek Foldyna, et al.

Deep Convolutional Neural Networks to Predict Cardiovascular Risk from Non-contrast Computed Tomography Images

Under Review - 2018

Segmentation for Radiotherapy



Stanislav Nikolov, Sam Blackwell, Ruheena Mendes, et al.

Deep Learning to Achieve Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy

Medical Image Computing & Computer Assisted Intervention (MICCAI) - 2018

Segmentation for Radiotherapy

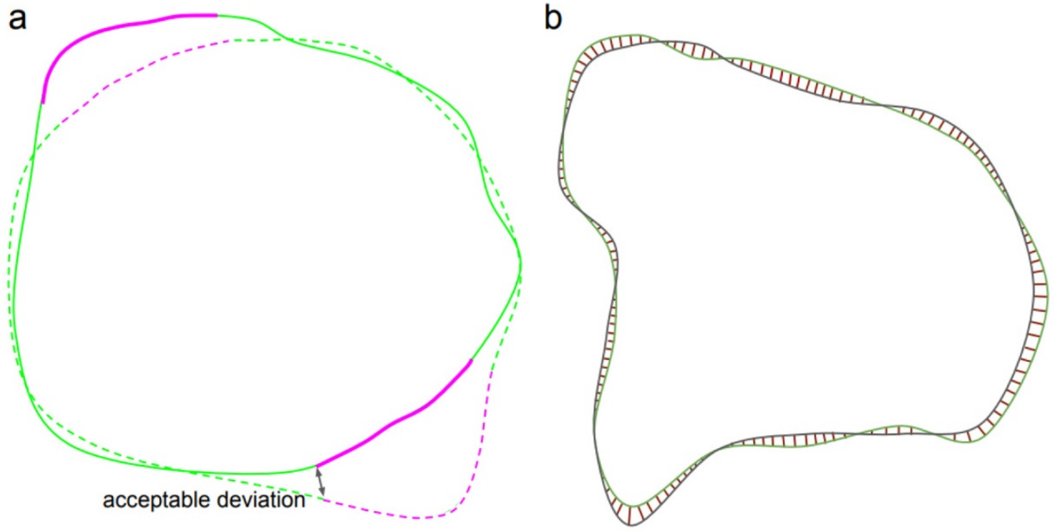
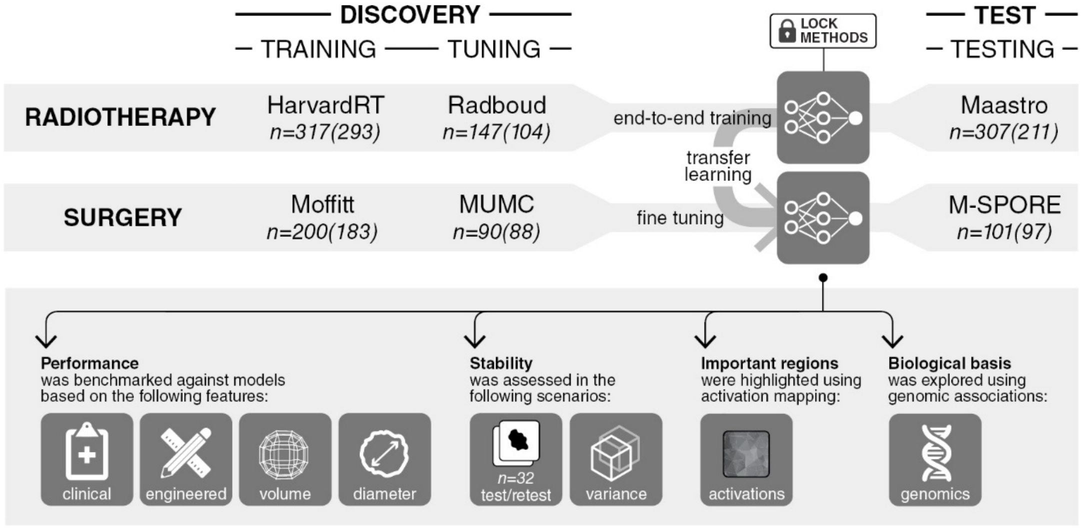
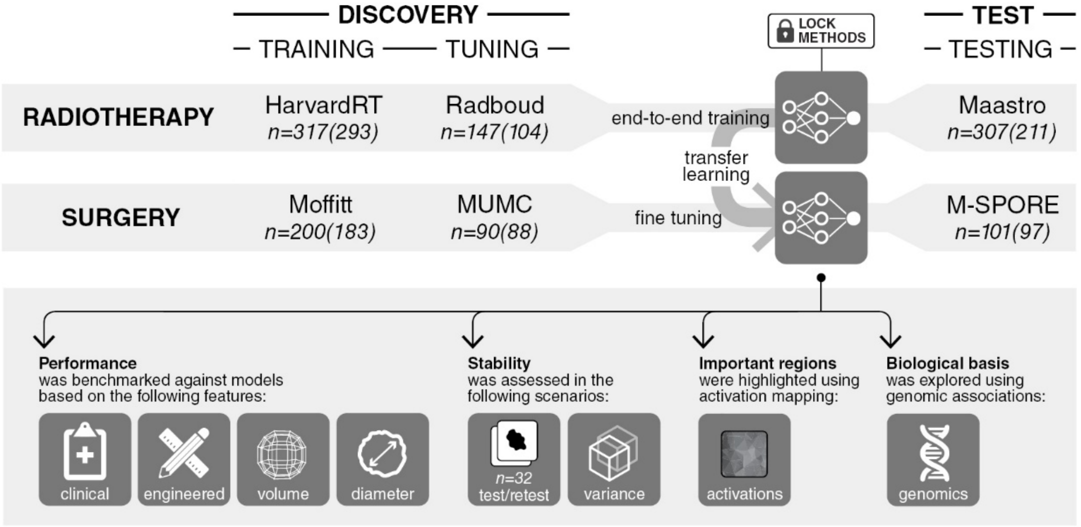


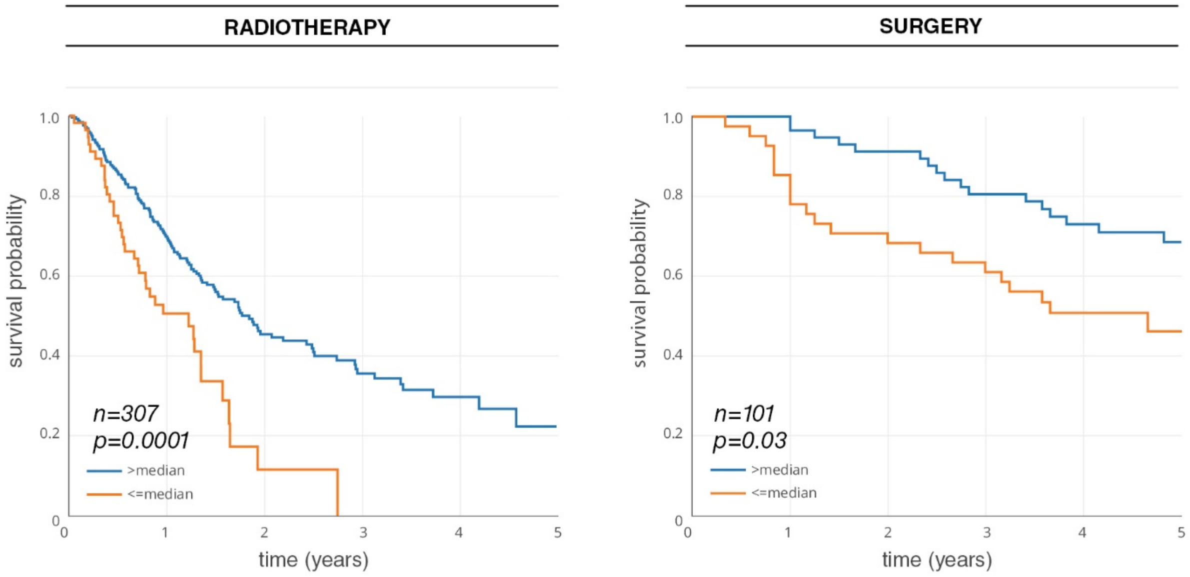
Figure 3 | Surface DSC performance metric. (**a**) Illustration of the computation of the surface DSC. Continuous line: predicted surface. Dashed line: ground truth surface. Black arrow: the maximum margin of deviation which may be tolerated without penalty, hereafter referred to by τ. Note that in our use case each OAR has an independently calculated value for τ. Green: acceptable surface parts (distance between surfaces $\leq \tau$). Pink: unacceptable regions of the surfaces (distance between surfaces $> \tau$). The proposed surface DSC metric reports the good surface parts compared to the total surface (sum of predicted surface area and ground truth surface area). (**b**) Illustration of the determination of the organ-specific tolerance. Green: segmentation of an organ by oncologist A. Black: segmentation by oncologist B. Red: distances between the surfaces. We defined the organ-specific tolerance as the 95th percentile of the distances collected across multiple segmentations from a subset of seven TCIA scans, where each segmentation was performed a radiographer arbitrated by an oncologist, neither of whom had seen the scan previously.

Characterization

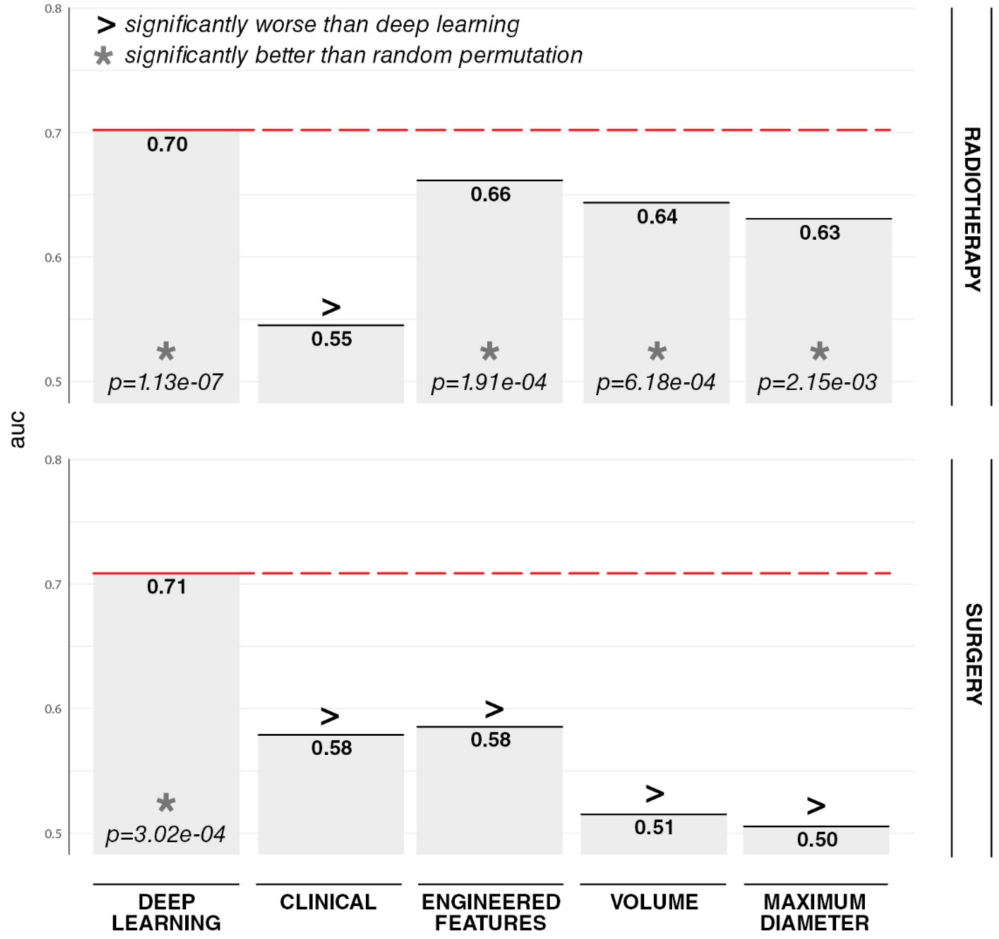


Characterization

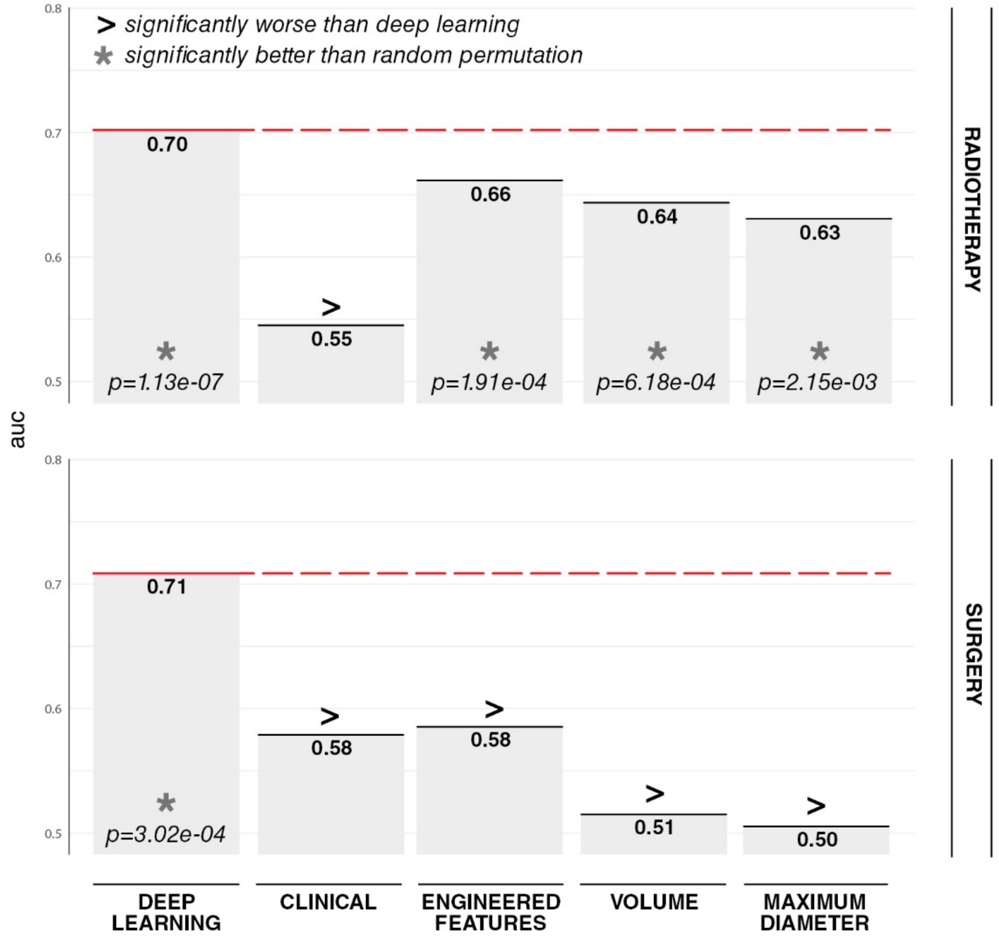


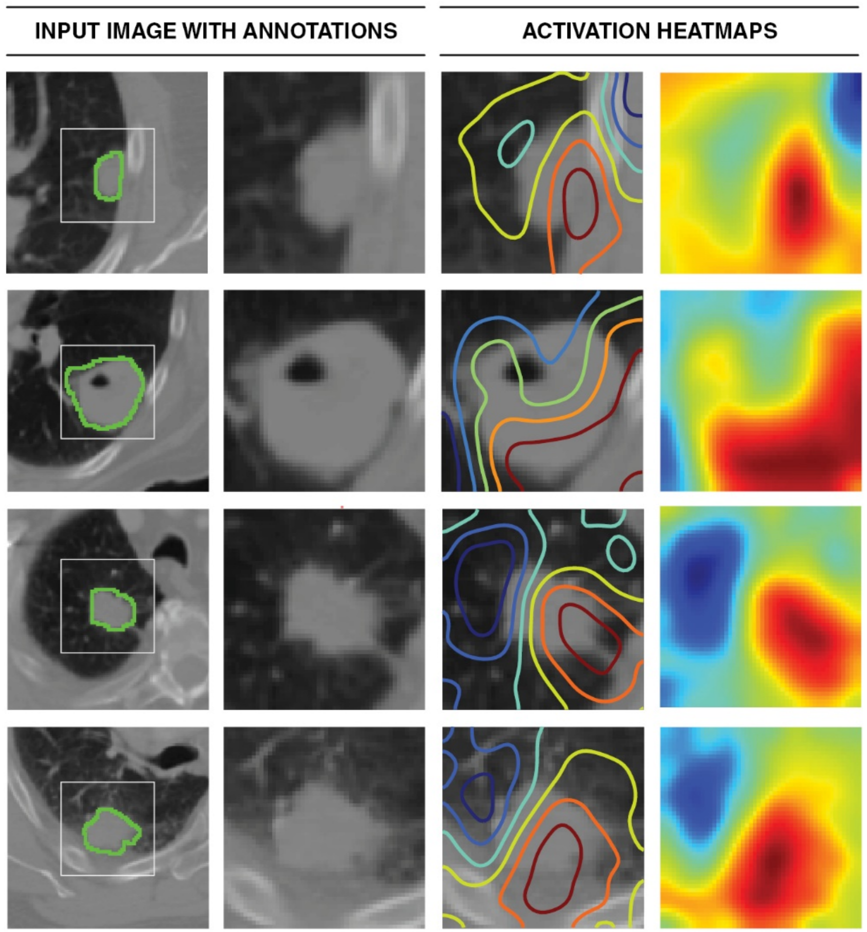


Characterization



Characterization



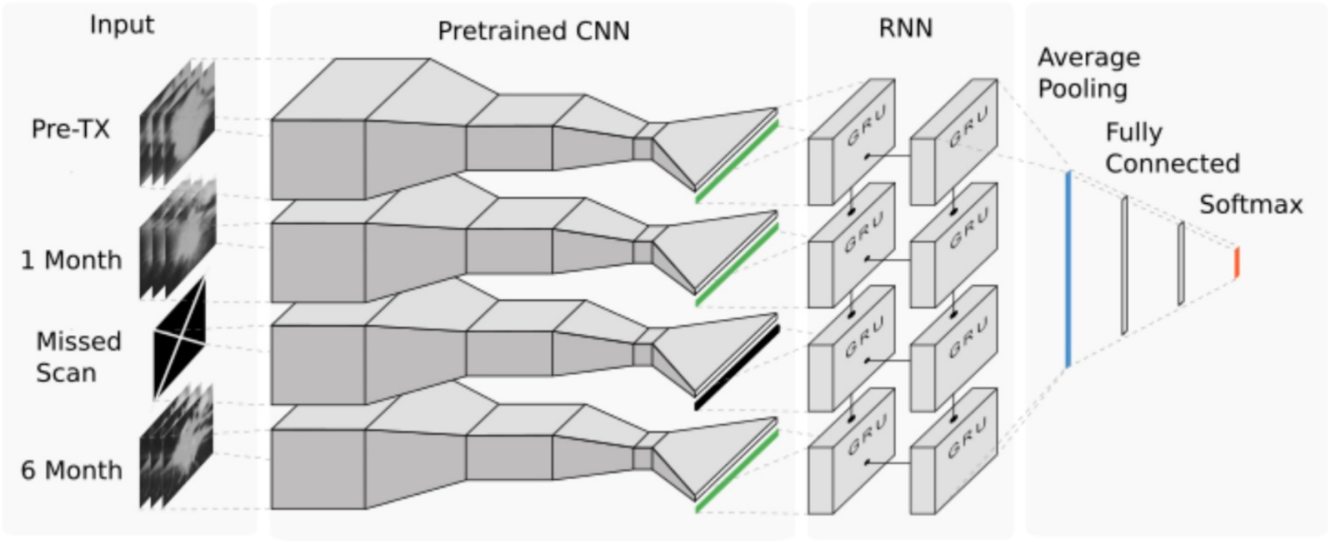


Ahmed Hosny, Chintan Parmar, Thibaud Coroller, et al.

Deep Learning for Lung Cancer Prognostication: A Retrospective Multi-Cohort Radiomics Study

Under Review

Monitoring

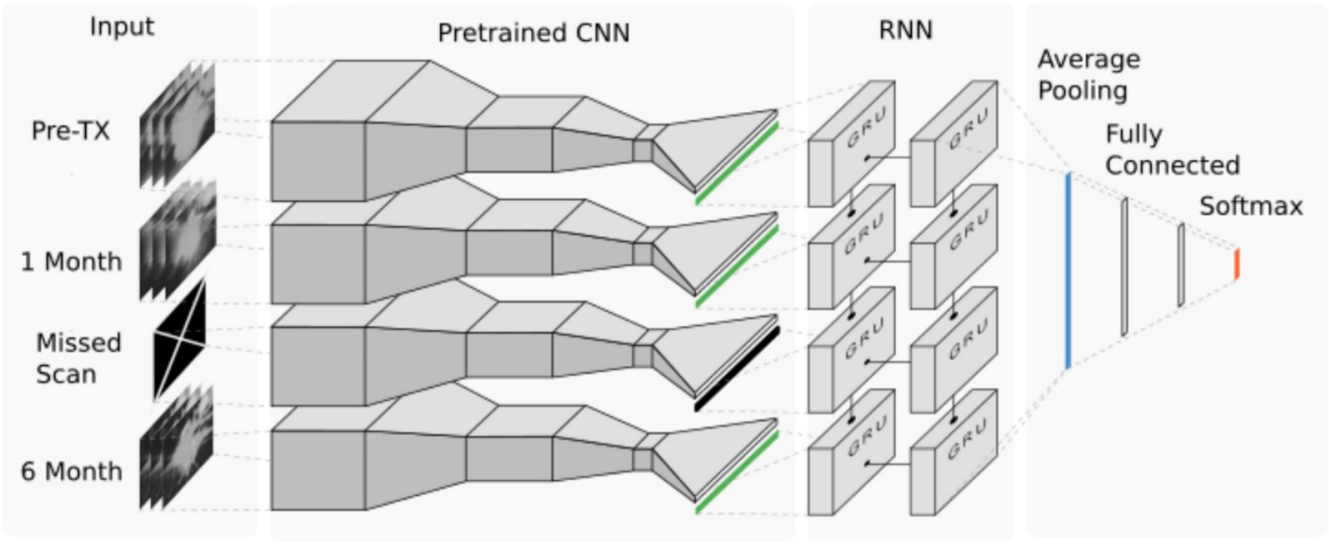


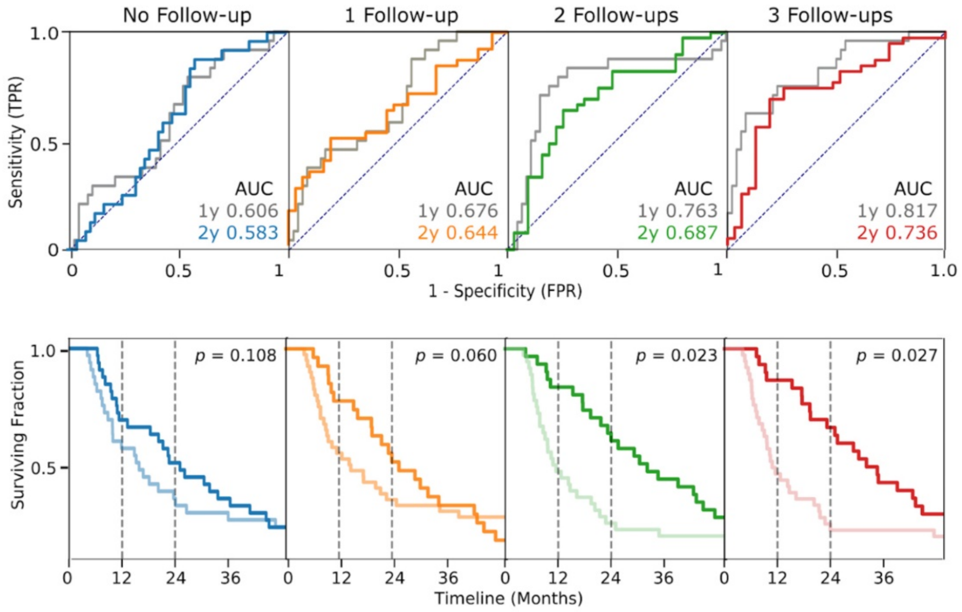
Yiwen Xu, Ahmed Hosny, Roman Zeleznik, et al.

Non-invasive Tracking of Lung Cancer Treatment Response Using Deep Learning-based Longitudinal Image Analysis

Under Review

Monitoring





Yiwen Xu, Ahmed Hosny, Roman Zeleznik, et al.

Non-invasive Tracking of Lung Cancer Treatment Response Using Deep Learning-based Longitudinal Image Analysis

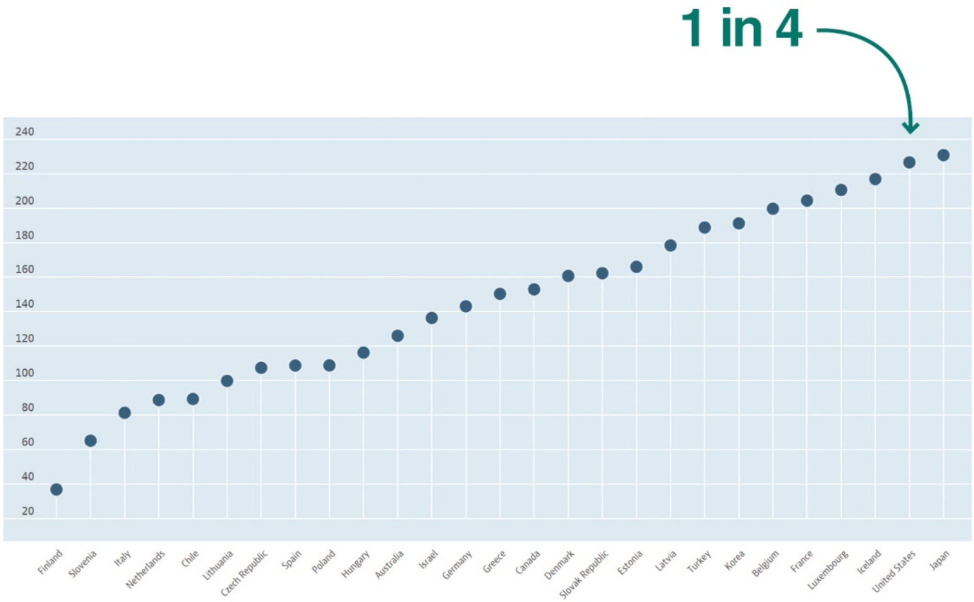
Under Review

Deep Learning

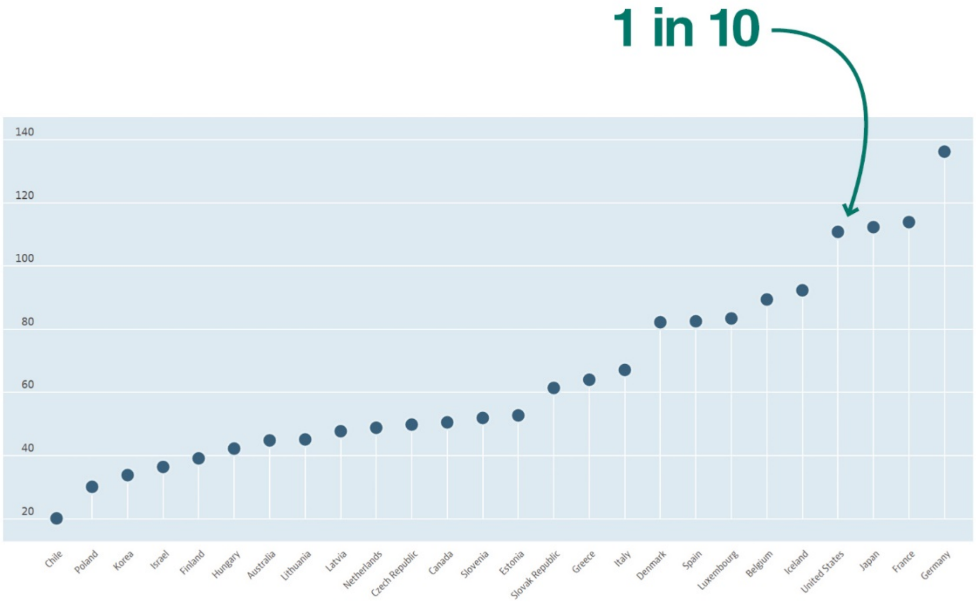
Applications in Medical Imaging

Challenges

Data, Data, and Data



CT exams per 1000 inhabitants (2017 or latest available)



MRI exams per 1000 inhabitants (2017 or latest available)

OECD

Health at a Glance 2017: OECD Indicators

The Organisation for Economic Co-operation and Development (OECD) - 2017

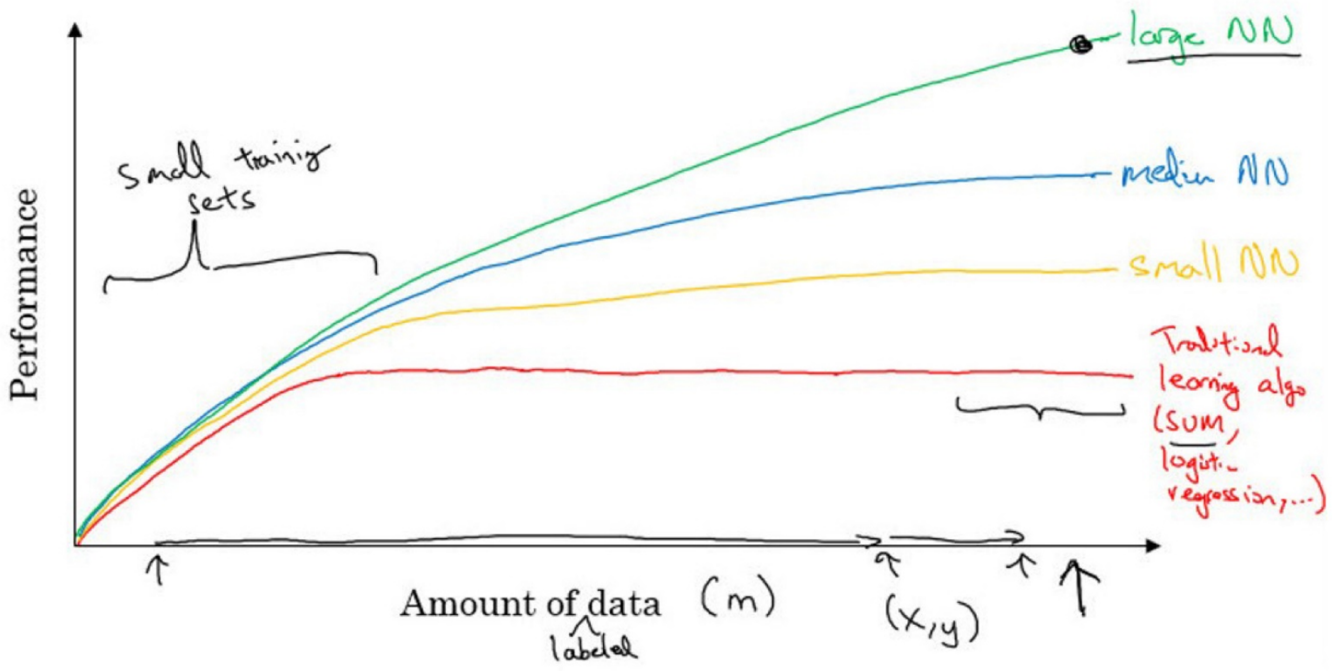
Data Curation

Claire Downs

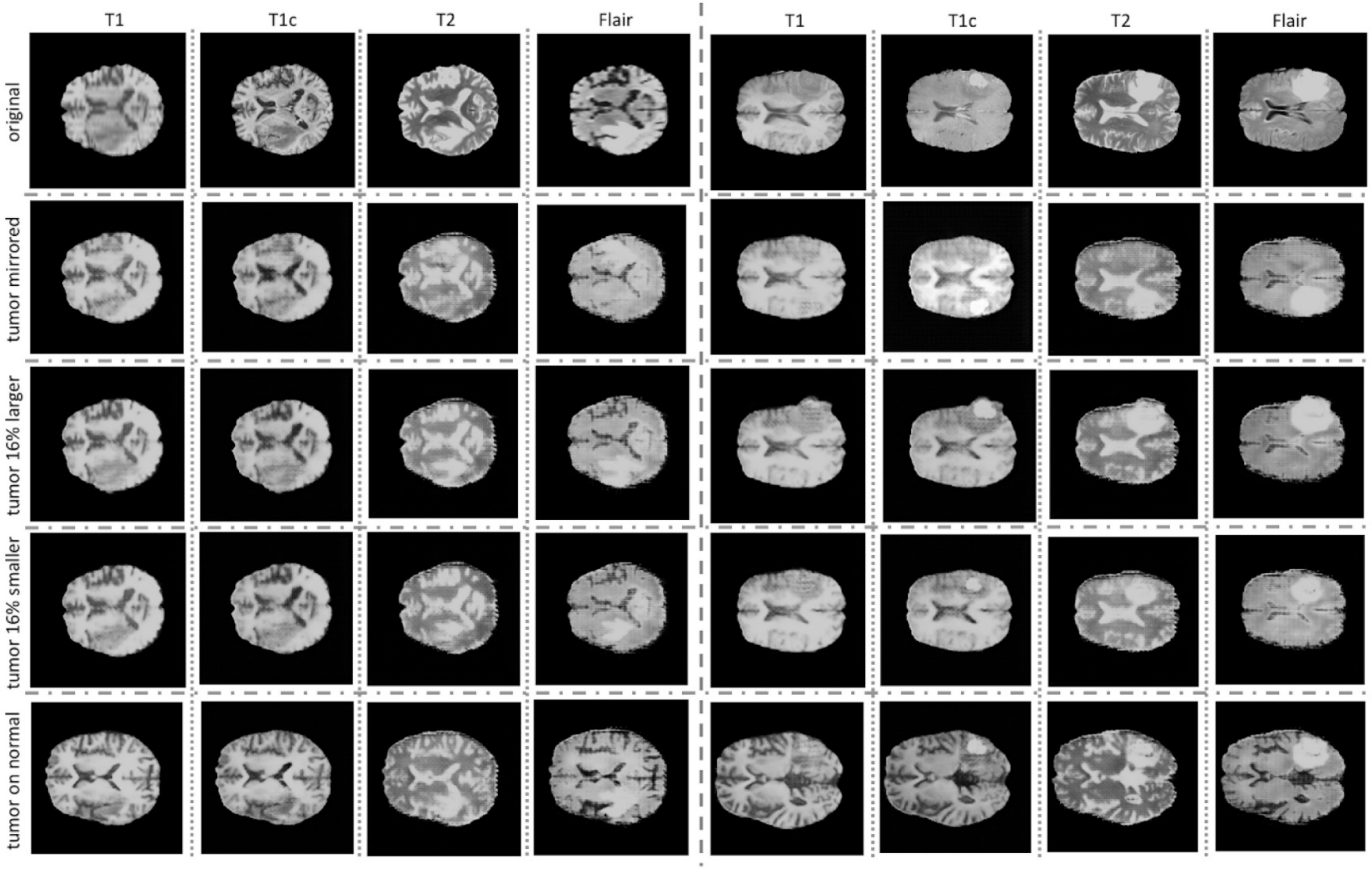
The Pros and Cons of Amazon Mechanical Turk

The Daily Dot

Scale vs Performance



Unsupervised Learning

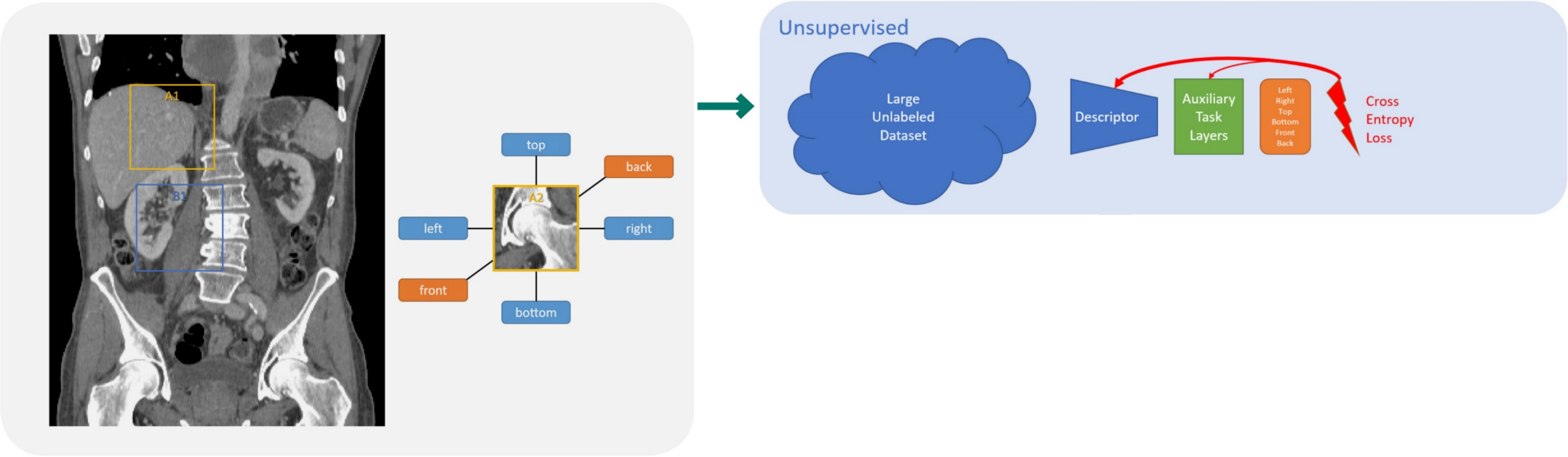


Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, et al.

Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks

Workshop on Simulation and Synthesis in Medical Imaging (SASHIMI) - 2018

Self-Supervised Learning

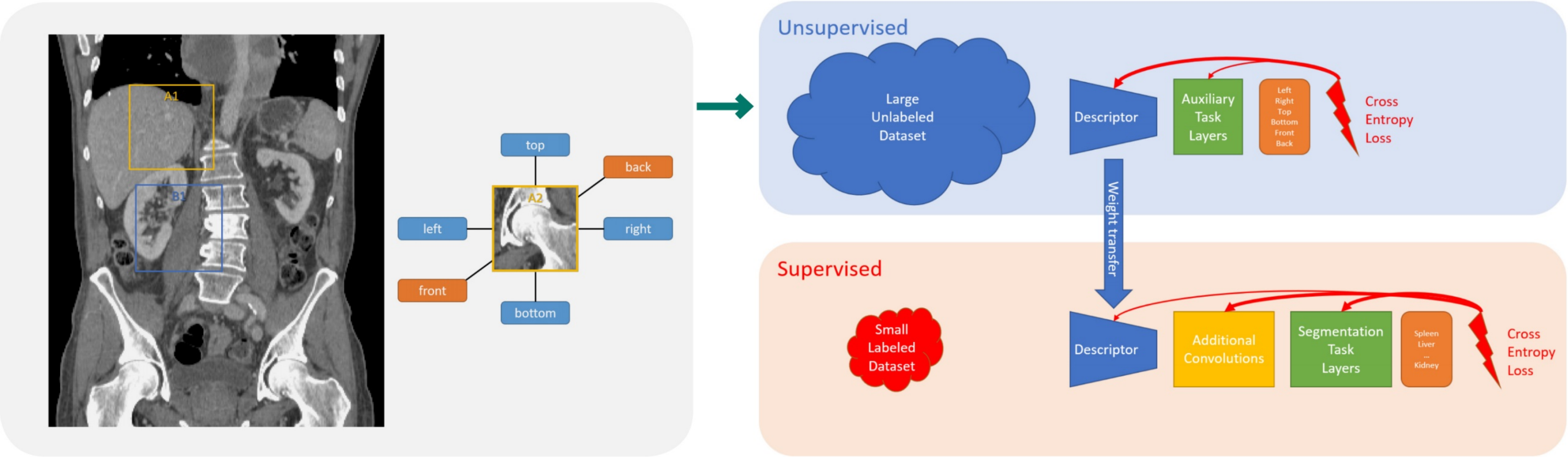


Max Blendowski, Hannes Nickisch & Mattias P Heinrich

Self-Supervised Convolutional Feature Training for Medical Volume Scans

semanticscholar.org

Self-Supervised Learning

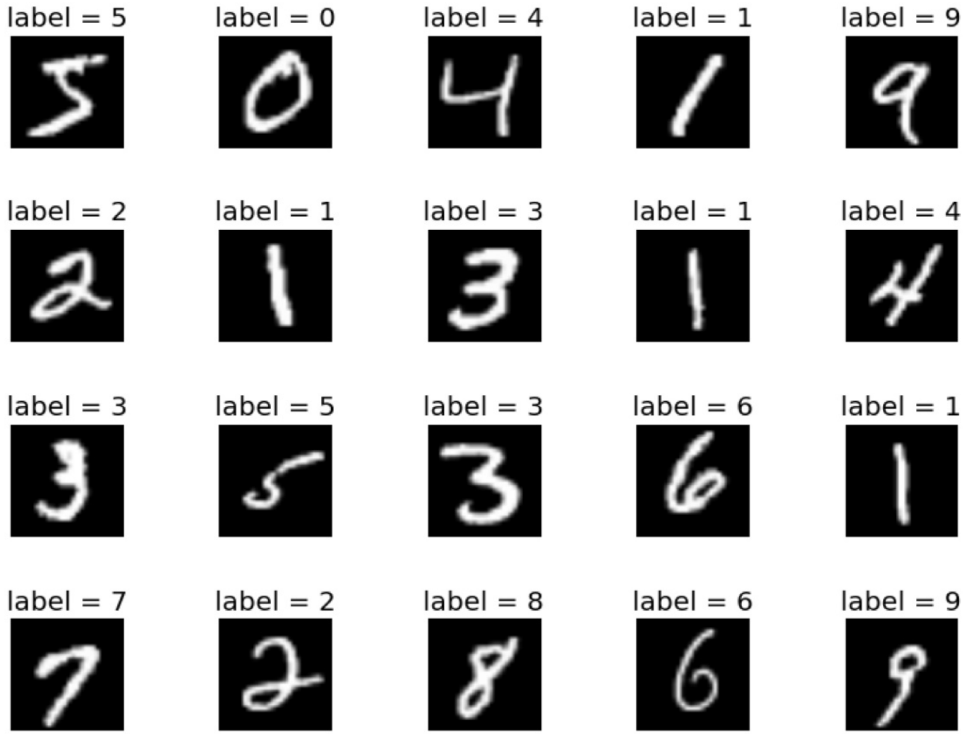


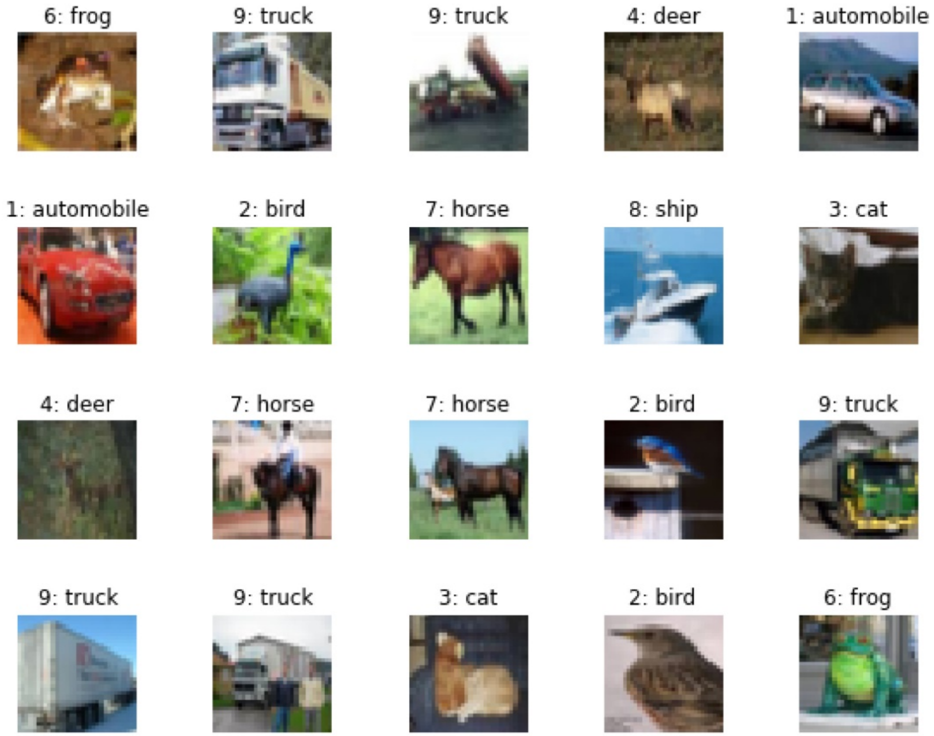
Max Blendowski, Hannes Nickisch & Mattias P Heinrich

Self-Supervised Convolutional Feature Training for Medical Volume Scans

semanticscholar.org

Benchmarking Datasets





Yann LeCun, Corinna Cortes & Christopher JC Burges

The MNIST Database of Handwritten Digits

yann.lecun.com/exdb/mnist & corochann.com

Alex Krizhevsky

The CIFAR-10 dataset

cs.toronto.edu/~kriz/cifar.html

Black-box Medicine

Davide Castelvecchi

Can we Open the Black Box of AI?

Nature - 2016

Interpretability

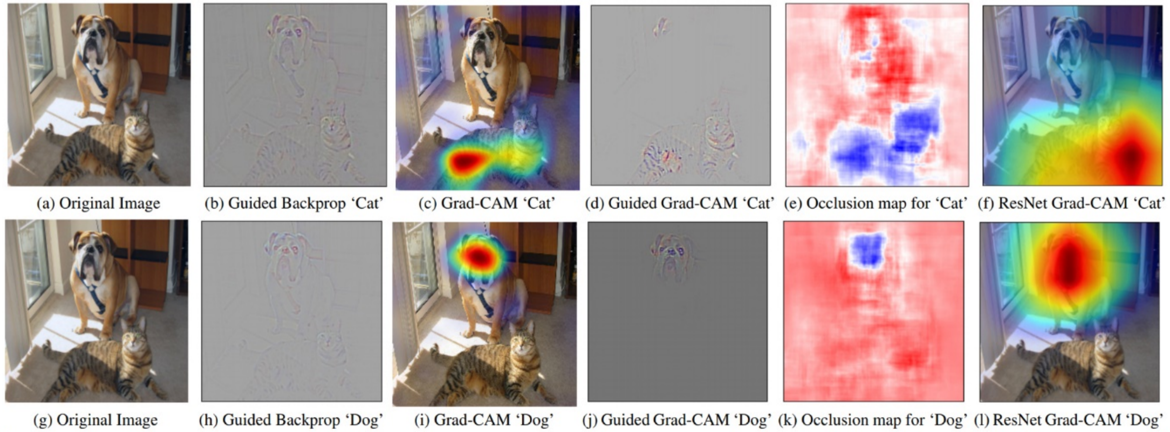


Figure 1: (a) Original image with a cat and a dog. (b-f) Support for the cat category according to various visualizations for VGG and ResNet. (b) Guided Backpropagation [46]: highlights all contributing features. (c, f) Grad-CAM (Ours): localizes class-discriminative regions, (d) Combining (b) and (c) gives Guided Grad-CAM, which gives high-resolution class-discriminative visualizations.Interestingly, the localizations achieved by our Grad-CAM technique, (c) are very similar to results from occlusion sensitivity (e), while being orders of magnitude cheaper to compute. (f, l) are Grad-CAM visualizations for ResNet-18 layer. Note that in (d, f, i, l), red regions corresponds to high score for class, while in (e, k), blue corresponds to evidence for the class. Figure best viewed in color.

Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, et al.

Grad-CAM:Visual Explanations from Deep Networks via Gradient-based Localization

IEEE International Conference on Computer Vision (ICCV) - 2017

Interpretability

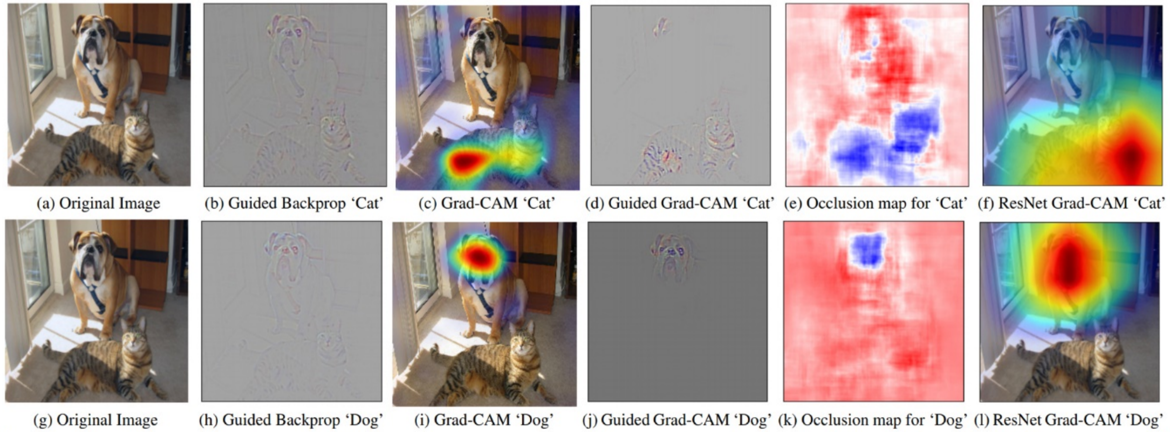
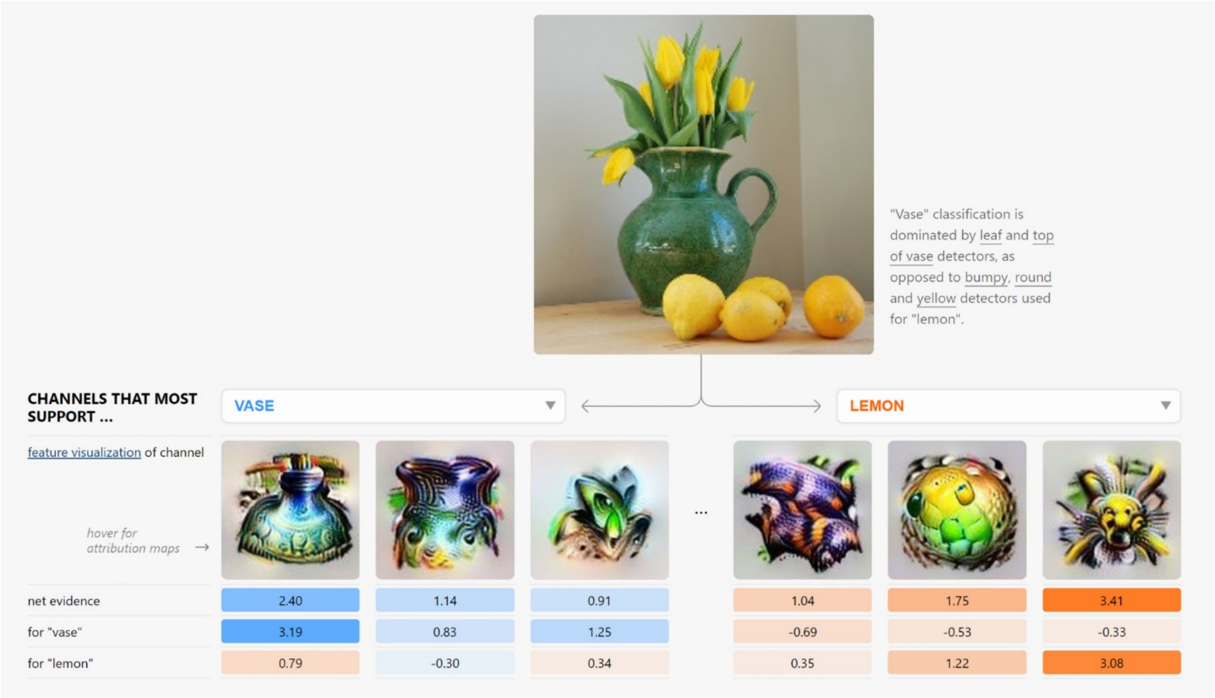


Figure 1: (a) Original image with a cat and a dog. (b-f) Support for the cat category according to various visualizations for VGG and ResNet. (b) Guided Backpropagation [46]: highlights all contributing features. (c, f) Grad-CAM (Ours): localizes class-discriminative regions, (d) Combining (b) and (c) gives Guided Grad-CAM, which gives high-resolution class-discriminative visualizations.Interestingly, the localizations achieved by our Grad-CAM technique, (c) are very similar to results from occlusion sensitivity (e), while being orders of magnitude cheaper to compute. (f, l) are Grad-CAM visualizations for ResNet-18 layer. Note that in (d, f, i, l), red regions corresponds to high score for class, while in (e, k), blue corresponds to evidence for the class. Figure best viewed in color.



Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, et al.

Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization
IEEE International Conference on Computer Vision (ICCV) - 2017

Chris Olah, Arvind Satyanarayan, Ian Johnson, et al.

The Building Blocks of Interpretability
distill.pub

Regulatory Aspects

Company	FDA Approval	Indication
Aidoc	August 2018	CT Brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT Stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI,CT) diagnosis
MaxQ-AI	January 2018	CT Brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

Eric Topol

FDA Approvals for AI in Medicine

twitter.com/EricTopol/status/1028642832171458563

Regulatory Aspects

Company	FDA Approval	Indication
Aidoc	August 2018	CT Brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT Stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI,CT) diagnosis
MaxQ-AI	January 2018	CT Brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

Eric Topol

FDA Approvals for AI in Medicine

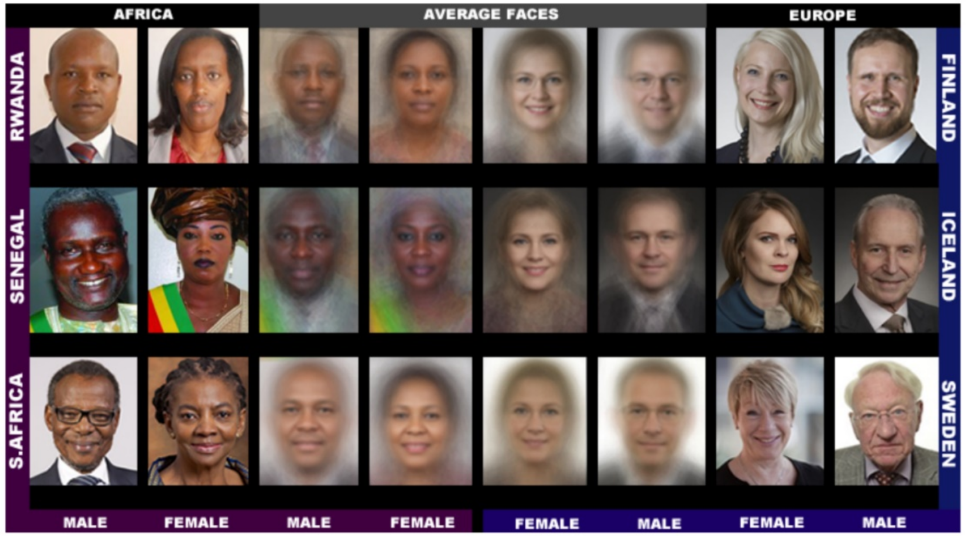
twitter.com/EricTopol/status/1028642832171458563

What is ground truth data?

Data as an active ingredient in models

Life-long learning models

Ethical Challenges

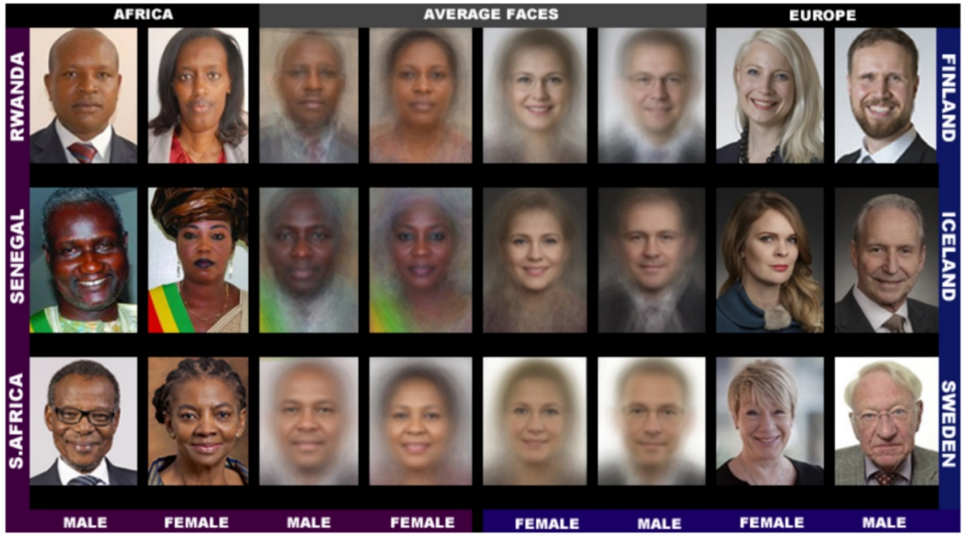


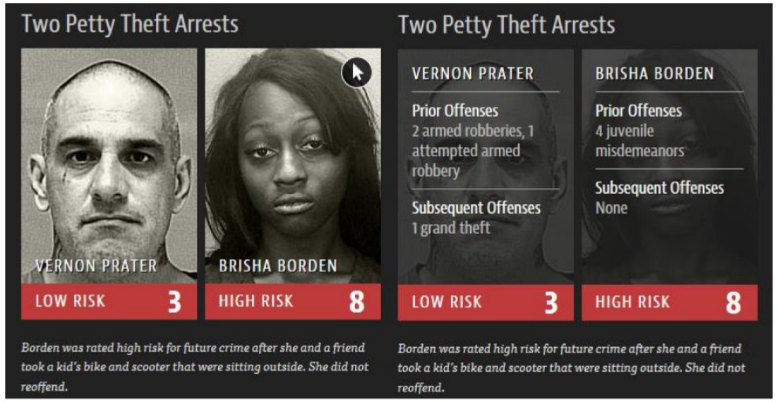
Joy Buolamwini & Timnit Gebru

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

Conference on Fairness, Accountability, and Transparency - 2018

Ethical Challenges





Joy Buolamwini & Timnit Gebru

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

Conference on Fairness, Accountability, and Transparency - 2018

Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner

Machine Bias

ProPublica - 2016

Ethical Breaches in Health Care

ico. Information Commissioner's Office

The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

Home | Your data matters | For organisations | Make a complaint | Action we've taken

About the ICO / News and events / News and blogs /

Royal Free - Google DeepMind trial failed to comply with data protection law

Date **03 July 2017**
Type **News**

The ICO has ruled the Royal Free NHS Foundation Trust failed to comply with the Data Protection Act when it provided patient details to Google DeepMind.

The Trust provided personal data of around 1.6 million patients as part of a trial to test an alert, diagnosis and detection system for acute kidney injury.

But an ICO investigation found several shortcomings in how the data was handled, including that patients were not adequately informed that their data would be used as part of the test.

The Trust has been asked to commit to changes ensuring it is acting in line with the law by signing an undertaking.

United Kingdom Information Commissioner's Office

Google DeepMind Trial Failed to Comply with Data Protection Law
ico.org.uk

Ethical Breaches in Health Care

ico.
Information Commissioner's Office

The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

Home | Your data matters | For organisations | Make a complaint | Action we've taken

About the ICO / News and events / News and blogs /

Royal Free - Google DeepMind trial failed to comply with data protection law

Date **03 July 2017**
Type **News**

The ICO has ruled the Royal Free NHS Foundation Trust failed to comply with the Data Protection Act when it provided patient details to Google DeepMind.

The Trust provided personal data of around 1.6 million patients as part of a trial to test an alert, diagnosis and detection system for acute kidney injury.

But an ICO investigation found several shortcomings in how the data was handled, including that patients were not adequately informed that their data would be used as part of the test.

The Trust has been asked to commit to changes ensuring it is acting in line with the law by signing an undertaking.

PROPUBLICA TOPICS ▾ SERIES ▾ NEWS APPS GET INVOLVED IMPACT ABOUT

Sloan Kettering's Cozy Deal With Start-Up Ignites a New Uproar

A for-profit venture with exclusive rights to use the cancer center's vast archive of tissue slides has generated concerns among pathologists at the hospital, as well as experts in nonprofit law and corporate governance.

by Charles Ornstein, ProPublica, and Katie Thomas, The New York Times, Sept. 20, 4:10 p.m. EDT

United Kingdom Information Commissioner's Office

Google DeepMind Trial Failed to Comply with Data Protection Law
ico.org.uk

Charles Ornstein & Katie Thomas

Sloan Kettering's Cozy Deal with Start-Up Ignites a New Uproar
propublica.org & nytimes.com

Ethical Challenges

Algorithms mirroring human bias

Unethical algorithms

Exacerbate tension between improving health and generating profit

Learned helplessness

Algorithm as third-party “actor” into the physician-patient relationship

Danton S. Char, Nigam H. Shah & David Magnus

Implementing Machine Learning in Health Care — Addressing Ethical Challenges

The New England Journal of Medicine (NEJM) - 2018

Security

NEWS

Employee error exposed Blue Cross patient data for 3 months

by Jessica Davis | September 21, 2018

An employee uploaded a file containing member information to a public-facing website in April, but officials did not discover

NEWS

Ransomware attack breaches 40,800 patient records in Hawaii

by Jessica Davis | September 13, 2018

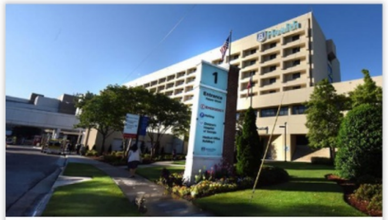
The Fetal Diagnostic Institute of the Pacific was able to restore data from backups, and with help from a cybersecurity firm wipe the

NEWS

Phishing attack breaches 38,000 patient records at Legacy Health

by Jessica Davis | August 22, 2018

The hackers went undetected for several weeks at the Portland, Oregon-based health system.



NEWS

417,000 Augusta University Health patient records breached nearly one year ago

by Jessica Davis | August 17, 2018

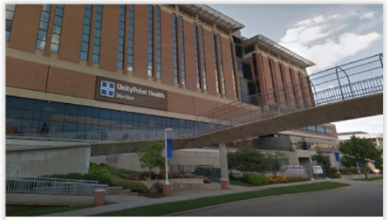
The Georgia provider was hit by two cyberattacks in September 2017, but did not explain when the breach was discovered.

NEWS

Canadian pharmacist fined for routinely accessing health records of acquaintances

by Lynne Minion | August 13, 2018

She snooped in the EHRs of nearly four dozen people over two years.



NEWS

1.4M records breached in UnityPoint Health phishing attack

by Jessica Davis | July 31, 2018

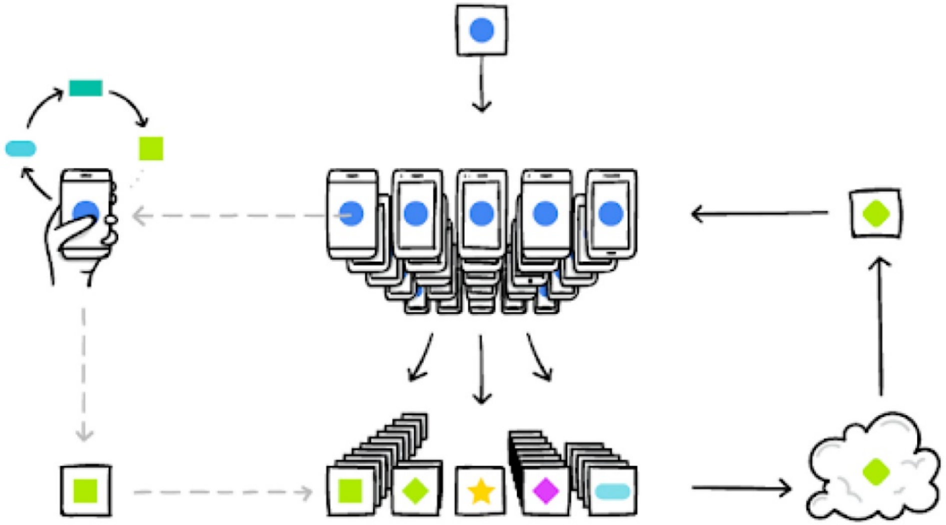
This is the second breach for the health system this year, and the biggest health data breach of 2018 in the U.S.

Healthcare IT News Staff

The Biggest Health Care Data Breaches of 2018 (so far)

healthcareitnews.com

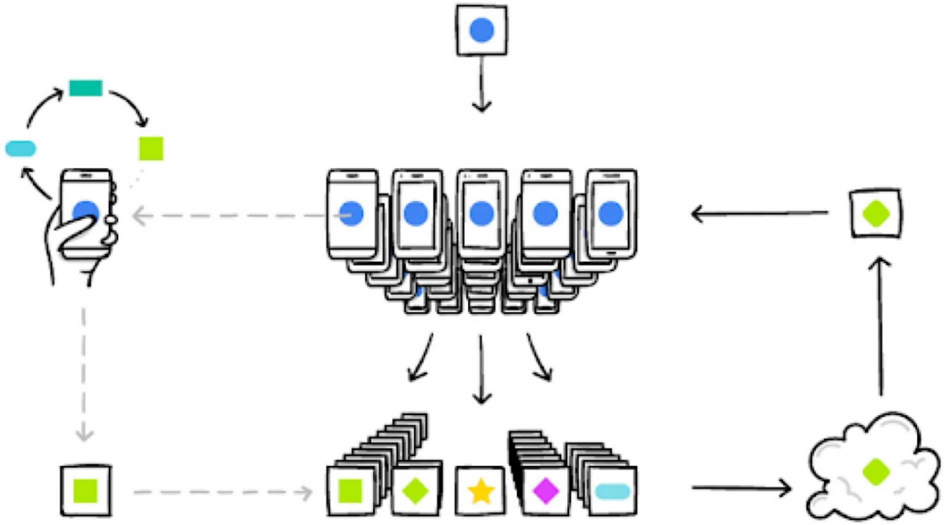
Security



Brendan McMahan, Eider Moore, Daniel Ramage, et al.

Communication-Efficient Learning of Deep Networks from Decentralized Data
20th International Conference on Artificial Intelligence and Statistics (AISTATS) - 2017

Security



Brendan McMahan, Eider Moore, Daniel Ramage, et al.

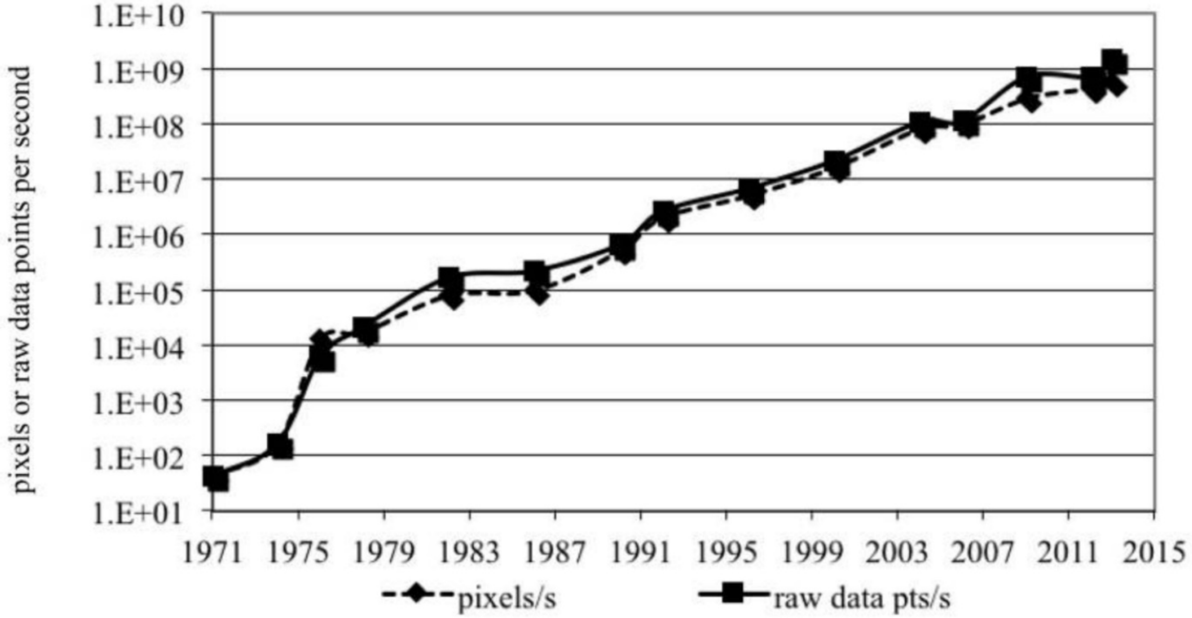
Communication-Efficient Learning of Deep Networks from Decentralized Data
20th International Conference on Artificial Intelligence and Statistics (AISTATS) - 2017

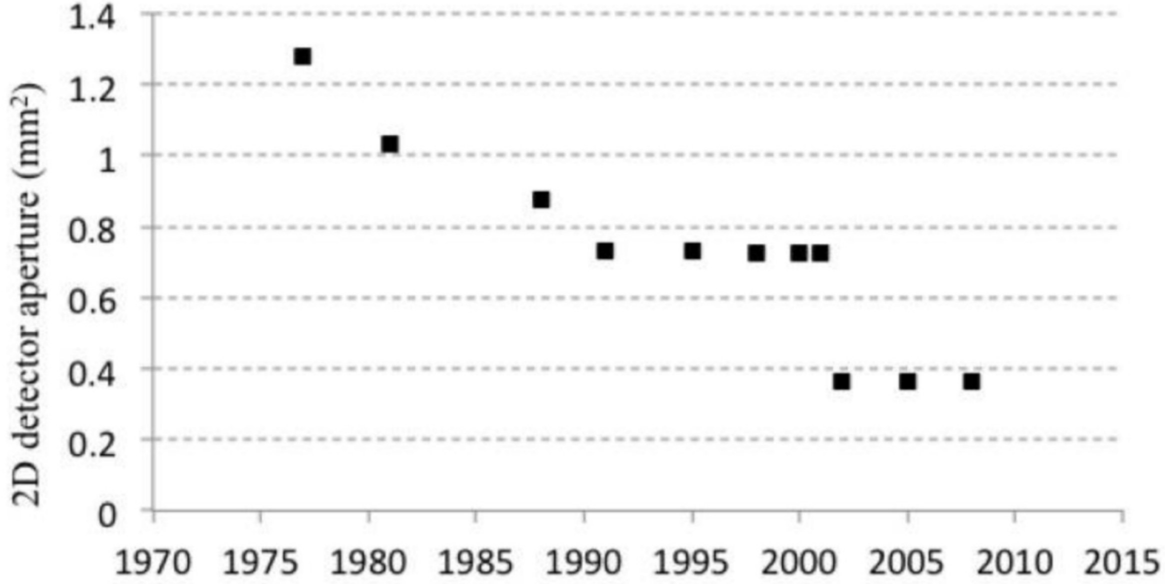
Nathan Dowlin, Ran Gilad-Bachrach, Kim Laine, et al.

CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy
International Conference on Machine Learning (ICML) - 2016

In Conclusion …

A Step in the Right Direction





Norbert J Pelc

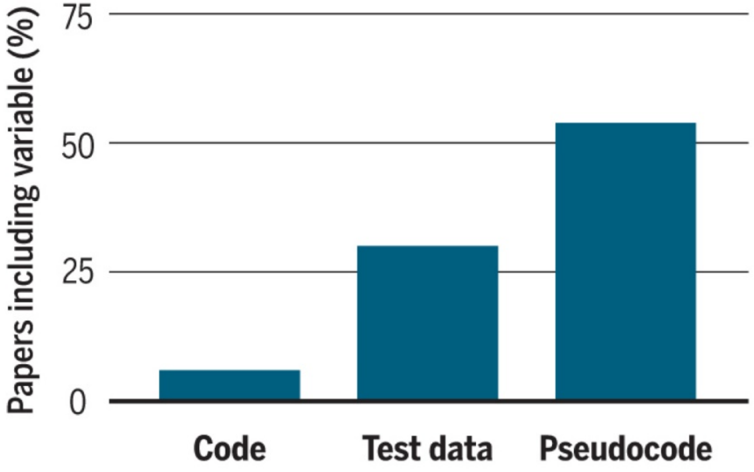
Recent and Future Directions in CT Imaging

Annals of Biomedical Engineering - 2014

Reproducibility

Code break

In a survey of 400 artificial intelligence papers presented at major conferences, just 6% included code for the papers' algorithms. Some 30% included test data, whereas 54% included pseudocode, a limited summary of an algorithm.

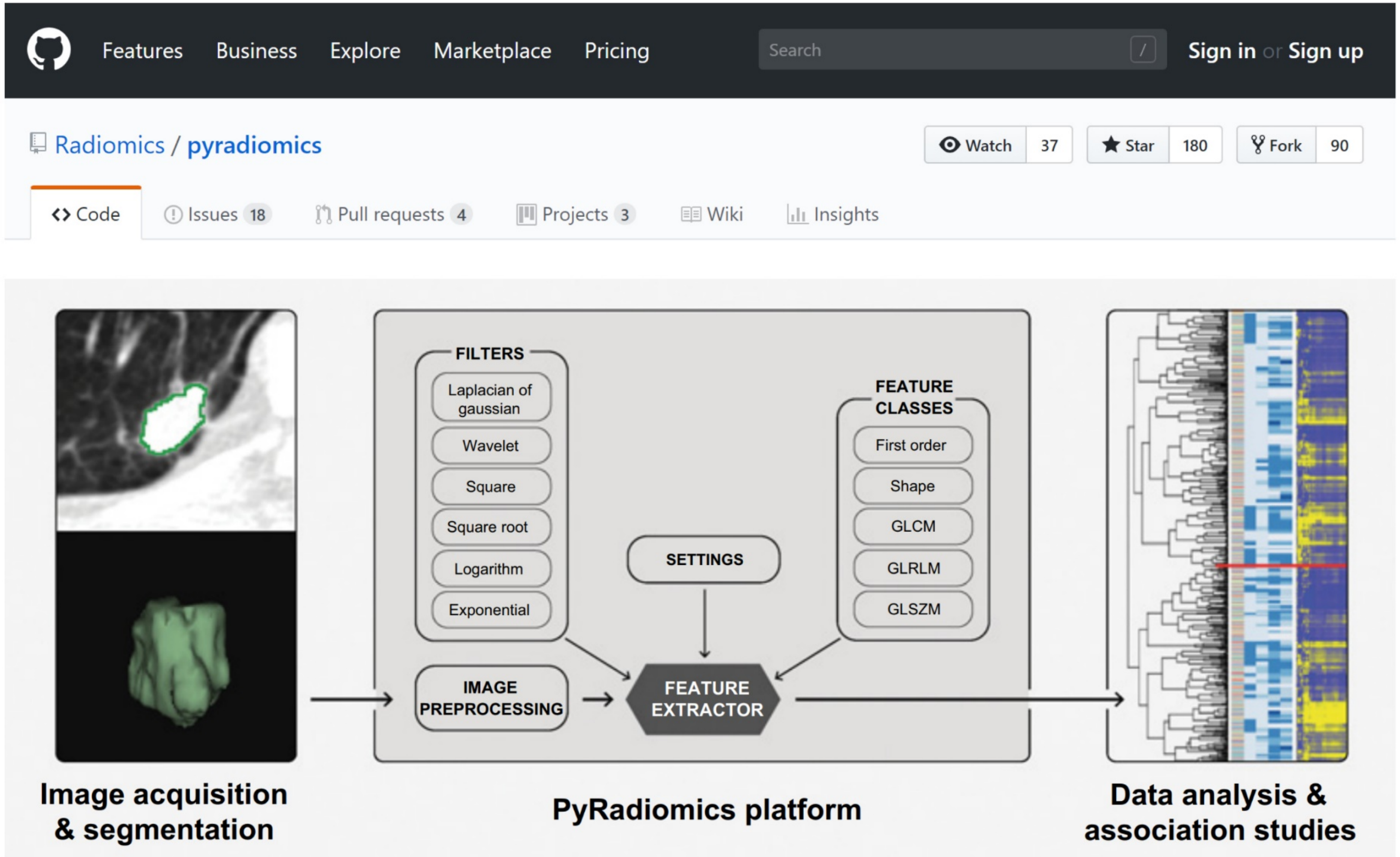


Matthew Hutson

Artificial Intelligence Faces Reproducibility Crisis

Science - 2018

PyRadiomics

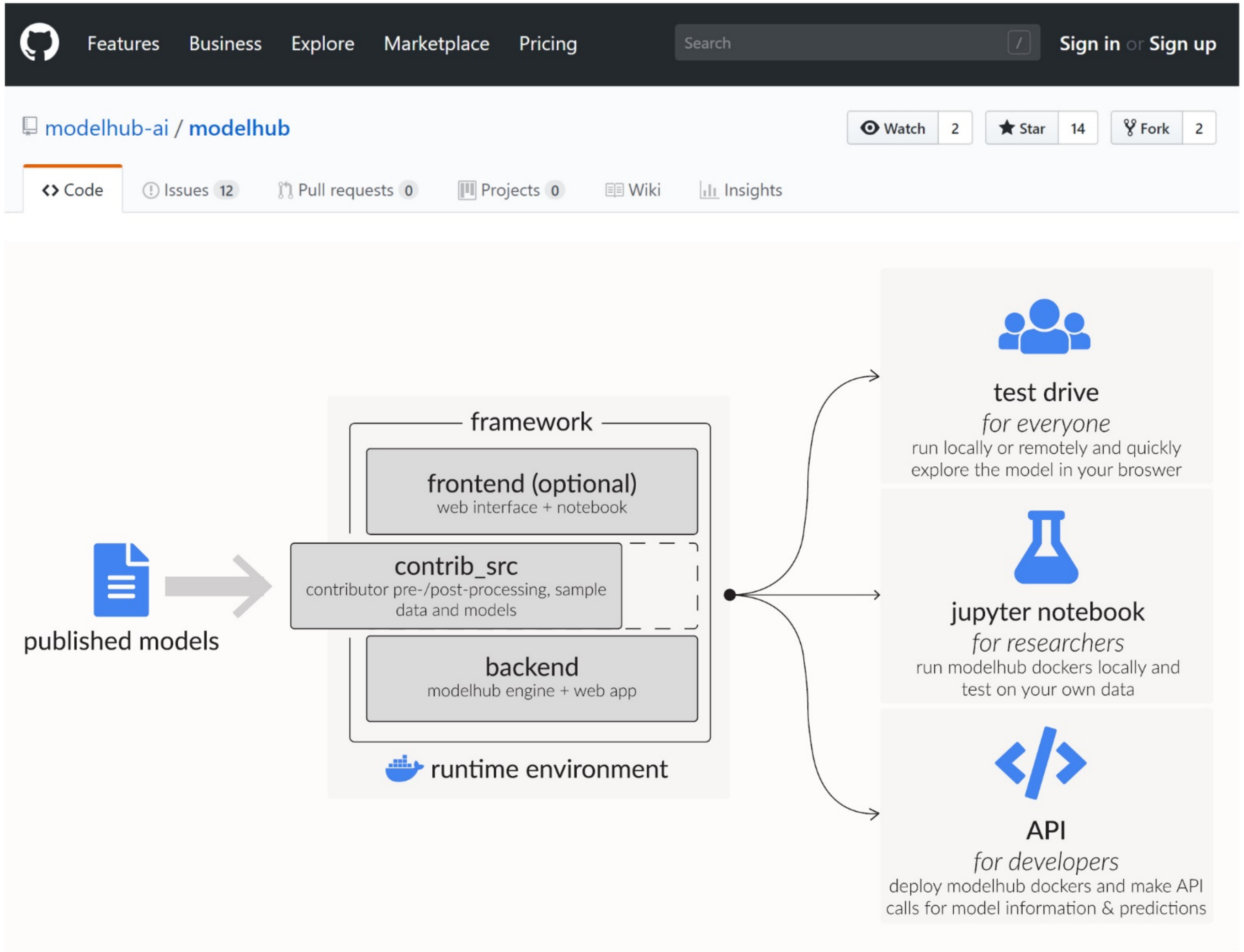


Joost JM van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, et al.

Computational Radiomics System to Decode the Radiographic Phenotype

Cancer Research - 2017

Modelhub



Ahmed Hosny, Michael Schwier, Andriy Y Fedorov and Hugo JWL Aerts

Modelhub: Plug & Predict Solutions for Reproducible AI Research

modelhub.ai

Mind The Hype

Advances in artificial intelligence (AI) will transform modern life by reshaping transportation, health, science, finance, and the military. To adapt public policy, we need to better anticipate these advances. Here we report the results from a large survey of machine learning researchers on their beliefs about progress in AI. **Researchers predict AI will outperform humans in many activities in the next ten years,** such as translating languages (by 2024), writing high-school essays (by 2026), **driving a truck (by 2027),** working in retail (by 2031), writing a bestselling book (by 2049), and **working as a surgeon (by 2053).** Researchers believe there is a 50% chance of AI outperforming humans in all tasks in 45 years and of **automating all human jobs in 120 years,** with Asian respondents expecting these dates much sooner than North Americans. These results will inform discussion amongst researchers and policymakers about anticipating and managing trends in AI.

Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang & Owain Evans

Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts
Journal of Artificial Intelligence Research - 2018

Jobs

Thomas H Davenport & Keith J Dreyer

AI Will Change Radiology, but it Won’t Replace Radiologists

Harvard Business Review

Thank you!