Available online at **www.joac.info**

ISSN: 2278-1862

Journal of Applicable Chemistry

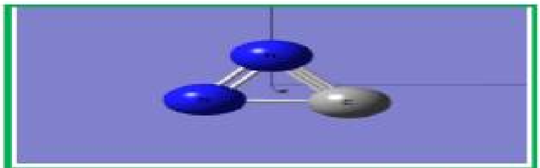
2020, 9 (3): 466-495

(International Peer Reviewed Journal)

New Chemistry News

$^{-}N=C=N^{-}$

New News of Chem (NNC)



ChemNewsNew (CNN)

Part 3

Artificial Intelligence (AI)

eXplainable AI (XAI)

Medical diagnosis

Data. Medical	- Uncertainty, unknown, incomplete, imbalanced, heterogeneous, noisy, dirty, erroneous, inaccurate, missing data - Probabilistic, fuzzy - Arbitrarily high-dimensional spaces

Dataset	China Acute Myocardial Infarction registry Tr: 9,619 ; Te: 9,125 patients
Task	In-hospital death in relation with clinical variables
Method (xAI)	XGBOOST
eXplainability of MachLrnMethod	New machine learning-based risk prediction model + Good discrimination ability + Offered individualized explanations on how clinical variables

	influenced the outcomes
FOM	<table><tr><td>Present method</td><td>0.899</td></tr><tr><td>random forest</td><td>0.861</td></tr><tr><td>logistic regression (LR) + top 15 variables</td><td>0.850</td></tr><tr><td>LR + L2 regularization</td><td>0.869</td></tr><tr><td>GRACE scores</td><td>0.810</td></tr><tr><td>89 variables</td><td>0.899 0.886-0.911 95% CI</td></tr><tr><td>12 variables</td><td>0.880 0.859-0.887 95% CI</td></tr></table>
Traditional statistical models	- Usually underestimate the complexity
Machine learning models	- Hard to interpret - Sensitive to the completeness of the input variables
Causal Inference	Engineering, 2020 https://doi.org/10.1016/j.eng.2019.08.016
KunKuang and Lian Li and ZhiGeng and Lei Xu and Kun Zhang and Beishui Liao and Huaxin Huang and Peng Ding and Wang Miao and Zhichao Jiang	

Task	Mammography images
Image size	3,000 x 3,000 pixels Mammography 300 x 300 pixels; ImageNet57 competition
Method (NN)	CNNs Transition from rule-based Computer Aided Detection systems to DeepLearning solutions
FOM	With embedded domain knowledge + Reduce diagnostic errors + Improve accuracy of radiologist + Helps in decision-making
Artificial intelligence in breast imaging	Clinical Radiology, 74 (2019) 357-366 https://doi.org/10.1016/j.crad.2019.02.006
E.P.V. Le and Y. Wang and Y. Huang and S. Hickman and F.J. Gilbert	

xAI	☞ Visual approach on therapeutic decision with more than two classes
interpretable models	o Rely on non-black box approach o Rule-based ones
Case Based Inference	☞ Similar cases can be used as examples for justifying the response of the system.

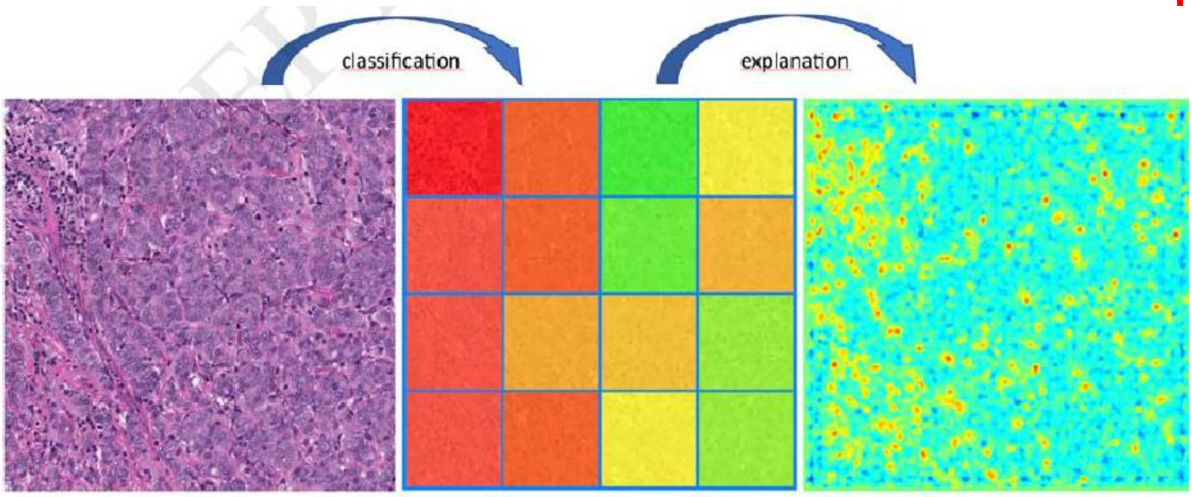
	+ This can be considered as an interpretable model. - Explanations, most CBR systems are limited to the display of the similar cases.

Dataset	☞ Breast Cancer Wisconsin (BCW) dataset2
Sample	☞ Digitized image of fine needle aspirate of breast mass
Data	☞ 683 cases; 9 dimensions; integer values ranging from 0 to 10 ☞ Classes: [benign or malignant]
Dataset	☞ Mammographic Mass (MM) dataset
Sample	☞ 830 cases; 2 numeric dimensions (age ; BI-RADS [Breast Imaging Reporting And Data System value]) ☞ 3 categorical dimensions (shape, margin, density of the mass) ☞ 2 classes (benign or malignant).
	☞
Dataset	o Breast Cancer (BC) dataset
Sample	o 286 cases o 4 numeric dimensions (age, tumor size, etc.), o 4 categorical dimensions (breast quadrant, etc.) o 2 classes (whether cancer is recurrent or not).
Simulated Dataset	o 4050 cases o 75 dimensions (22 Boolean, 14 integer and 39 nominal) o 4 classes, categories of treatment for breast cancer: [surgery, chemotherapy, radiotherapy,endocrine]
Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach	Artificial Intelligence in Medicine, 94 (2019) 42-53 https://doi.org/10.1016/j.artmed.2019.01.001
Jean-Baptiste Lamy and BoomadeviSekar and Gilles Guezennec and Jacques Bouaud and Brigitte Séroussi	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Discipline	o Ophthalmology
Goal	o To augment vision care o Improved efficiency of tools
Task	o To identify, localize and quantify ☞ Pathological features in macular and retinal disease
Understanding the advent of artificial intelligence in ophthalmology	Journal of Current Ophthalmology, 31 (2019) 115-117
Editorial	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI



- Left: original H&E (hematoxilin&Eosin) breast cancer image.
- Center: machine-learning-classification
- Right: Heatmap explaining classifier decisions with pixel-wise resolution

Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning	Seminars in Cancer Biology (2018), doi.org/10.1016/j.semcancer.2018.07.001

International Immuno-Oncology Biomarker Working Group
F. Klauschen, K.-R. Muller, A. Binder, M.Bockmayr, M. Hagele, P. Seegerer, S. Wienert, G. Pruneri, S.de Maria, S. Badve, S. Michiels, T.O. Nielsen, S. Adams, P.Savas, F. Symmans, S. Willis, T. Gruosso, M. Park, B.Haibe-Kains, B. Gallas, A.M. Thompson, I. Cree, C. Sotiriou,C. Solinas, M. Preusser, S.M. Hewitt, D. Rimm, G. Viale, S.Loi, S. Loibl, R. Salgado, C. Denkert

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

What do we need to build explainable AI systems for the medical domain?	arXiv:1712.09923v1[cs>AI 28Dec,2017

Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis, Douglas B. Kell

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Medical Care

Task	Phenotype diagnosis
Method	Multi-label gradient boosted tree (xgboost)
dataset MIMIC III	First 24 hours vitals of a patient in the ICU 833 extracted features (17 patient vital × 7 time windows × 7 statistics)

xAI	17 vital signs for explanation
SHAP	For attribution
LORE	For counterfactual rules
MOEA/D	For sensitivity analysis

Designing Theory-Driven User-Centric Explainable AI	In 2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)

Danding Wang, Qian Yang, Ashraf Abdul, Brian Y. Lim

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Method proposed	o Conceptual framework for building human-centered, decision-theory-driven XAI
XAI	✓ Mitigate common cognitive biases
Application	🕮 Medical diagnostic tool ☞ ICU ☞ Co-design exercise with clinicians.
Explainable tools	From philosophy, cognitive psychology,AI
Future	▪ Articulation of detailed design space of technical features of XAI ▪ Connecting methods with requirements of human reasoning, ▪ → Developers build more user-centric explainable AI-based systems
Designing Theory-Driven User-Centric Explainable AI. In 2019 CHI Conference on Human	Factors in Computing Systems Proceedings (CHI 2019) doi.org/10.1145/3290605.330083
Danding Wang, Qian Yang, Ashraf Abdul, Brian Y. Lim	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

DataSet	Medical Information Mart for Intensive Care (MIMIC-III)
Data Descriptor: MIMIC-III, a freelyaccessible critical care database	SCIENTIFIC DATA, 3:160035 DOI: 10.1038/sdata.2016.35
Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi& Roger G. Mark	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

xAI	Implementation of transparency/ traceability o For statistical black-box machine (deep)learning methods
AI	o Phenomenon of intelligence is very difficult to define ! AI itself esoteric term in engineering
Application	o Human explanation in histopathology
Future	o Go beyond explainable AI o Explainable medicine with causality
Causability and explainabilty of artificial intelligence in medicine	WIREs Data Mining KnowlDiscov. 2019;e1312. wires.wiley.com/dmkd 1 of 13 https://doi.org/10.1002/widm.1312
Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, Heimo Müller	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit (ICU): a retrospective study of high-frequency data in electronic patient records	www.thelancet.com/digital-health Published online March 12, 2020 https://doi.org/10.1016/S2589-7500(20)30056-X
Hans-Christian Thorsen-Meyer, Annelaura B Nielsen, Anna P Nielsen, Benjamin SkovKaas-Hansen, Palle Toft, Jens Schierbeck, Thomas Strøm, Piotr J Chmura, Marc Heimann, Lars Dybdahl, Lasse Spangsege, Patrick Hulsen, Kirstine Belling, SørenBrunak, Anders Perner	

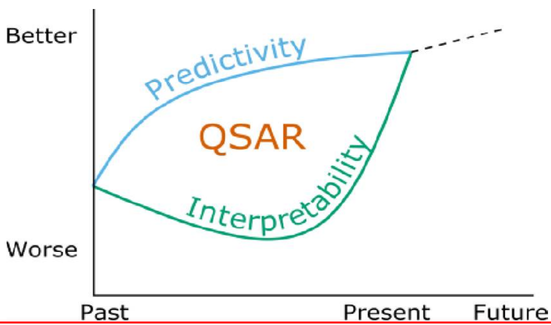
xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Protein folding

Computers; humans	! Computers are incredibly fast, accurate but stupid ! Humans are incredibly slow, inaccurate but brilliant, ! Together they are powerful beyond imagination
Datasets	o Protein Folding o Clustering of large high-dimensional gene expression data o Traveling Salesman Problem
Application	🕮 Integrative machine learning 🕮 Understanding intelligence
Intelligence	✓ What is it? Where is it? ✓ Solve intelligence – then everything else solved ✓ How real is AI?
Data; Knowledge	✓ Today is drowning in data ✓ Information overload ✓ A wealth of information creates a poverty of attention ✓ Yet, starving for knowledge
Future	✓ Multi-Task Learning to help to reduce catastrophic forgetting ✓ Multi-Agent Hybrid Systems making use of collective intelligenceand crowd-sourcing ✓ Transfer learning [learning to perform a task by exploiting knowledge acquired when solving previous tasks] ✓ Multi-Agent Hybrid Systems making use of collective intelligence and crowd-sourcing ✓ Automatic machinelearning (aML)
From Machine Learning to Explainable AI	IEEE DISA 2018 Conference, Kosice, August, 23, 2018 DOI: 10.1109/DISA.2018.8490530
Andreas Holzinger	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

QSAR (Structure actitivity relationships)

🕮 Evolution of the interpretation paradigm 🕮 Model → descriptors→ (structure)	
Interpretation of Quantitative Structure−Activity Relationship (QSAR) Models: Past, Present, and Future	J. Chem. Inf. Model. 2018 DOI:: 10.1021/acs.jcim.7b00274
Pavel Polishchuk	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Building of Robust and Interpretable QSAR Classification Models by Means of the Rivality Index	J. Chem. Inf. Model. 2019, 59, 2785−2804 DOI: 10.1021/acs.jcim.9b00264
Irene Luque Ruiz and Miguel ÁngelGómez-Nieto	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Interpretation of QSAR Models by Coloring Atoms According to Changes in Predicted Activity: How Robust Is It?	J. Chem. Inf. Model. 2019, 59, 1324−1337 DOI: 10.1021/acs.jcim.8b00825
Robert P. Sheridan	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Structural and physico-chemical interpretation (SPCI) of QSAR models and its comparison with MMP analysis	J Che Infor Model (2020)
Pavel G. Polishchuk, Oleg Tinkov, Tatiana Khristova, Ludmila Ognichenko, Anna Kosinskaya, Alexandre Varnek, and Victor Kuz'min	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Toxicology-Envronmental

Field	Environmental toxicology
Feature	o Model interpretability; Data interpretation
	o Organisation for Economic Co-operation and Development (OECD) o Five Principles for Quantitative StructureActivity Relationship (QSAR) validation

Machine Learning for Environmental Toxicology: A Call for Integration and Innovation	Environ. Sci. Technol. 2018, 52, 12953−12955
Thomas H. Miller, Matteo D. Gallidabino, James I. MacRae, Christer Hogstrand, Nicolas R. Bury, Leon P. Barron, Jason R. Snape, and Stewart F. Owen	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Bio-informatics

Supervised and Unsupervised Algorithms for,Bioinformatics and Data Science	Progress in Biophysics and Molecular Biology(2020) doi.org/10.1016/j.pbiomolbio.2019.11.012
Ayesha Sohaila;b, Fatima Arif	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI —

Climate and plant biology

Discipline	o Agriculture
Mega-Goal	**Socially oriented Sustainable Development Goals** o New crop ideotypes ☞ Water and nutrient use efficiency ☞ High food or net energy yield per hectare ☞ Carbonsequestration ☞ Optimized microbiome usage ☞ Disease resistance
Objective	▪ AI + decipherable decision-making process → ▪ Offers meaningful explanation to humans
Data	Large Data ▪ Multi-omics ▪ Imaging ▪ Ecophysiology ▪ Field-based data for large-scale population ▪ Plant omics ▪ Datasets of plant populations (genome, epigenome, transcriptome, proteome,metabolome, phytobiome, phenome)
Datasets	o Global exascale datasets o 12 major Elemental layers for soil o 48 light spectra (300 nm–780 nm) across 365 days o Calculated similarity indexes using the duo algorithm on summit supercomputer ▪ → Generated climate clusters globally at 1 km^2resolution
Resources	▪ 200-petaflop supercomputer
Goal	o Systems-level approach o To dissect biological mechanisms in plants\` o Exascale computing (from individual plant to global scale) o Advanced AI approaches to model climate type o Patterns/clustering across last 50 years o To predictfuture patterns

Can exascale computing and explainable artificial intelligence applied to plant biology deliver on the United Nations sustainable development goals?	Current Opinion in Biotechnology, 61 (2020) 217-225https://doi.org/10.1016/j.copbio.2020.01.010
Jared Streich and Jonathon Romero and João Gabriel Felipe Machado Gazolla and David Kainer and Ashley Cliff and Erica Teixeira Prates and James B Brown and Sacha Khoury and Gerald A Tuskan and Michael Garvin and Daniel Jacobson and Antoine L Harfouche	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

Discipline	High yield cultivation; Climate; Environment
Task	! Predicting effects of expression of genes involved in plant growth = fn(changing water availability) ! Resistance to pests ! Ill-defined prediction targets

Tools	Next-Gen AI : [xAI + MachLrn + Deep NN + ...]
Implements	Automation of much of the analysis, but with human support/discretion in cycle
Big data	**Omics data** ☞ Heterogeneous ; high dimensional ☞ Derived from a wide range of experiments which yield different types of information
Data characteristics	- Noisy - Sparse, irregularly sampled - Collected under different conditions - Ambiguous time points

Accelerating Climate Resilient Plant Breeding by Applying Next-Generation Artificial Intelligence	Trends in Biotechnology, 37 (2019)1217-1235 https://doi.org/10.1016/j.tibtech.2019.05.007
Antoine L. Harfouche and Daniel A. Jacobson and David Kainer and Jonathon C. Romero and Antoine H. Harfouche and Giuseppe ScarasciaMugnozza and Menachem Moshelion and Gerald A. Tuskan and Joost J.B. Keurentjes and Arie Altman	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI xAI

Classification

Earlier successful approach	+ NNs with multiple hidden layers (deep neural networks) - More effective - More efficient
Limitation	- Not trivial to understand the way howthey derive their classification decisions
Method introduced	🕮 decompositional algorithm –DeepRED – o Able to extract rules from deep neural networks o Decision processes more comprehensible ▪ Ex: XOR function

	#attributes	#training ex.	#test ex.	NN structure	acc(training)	acc(test)
MNIST	784	12056	2195	784-10-5-2	99.6 %	98.8 %
letter	16	1239	438	16-40-30-26	96.9 %	97.3 %
artif-I	5	20000	10000	5-10-5-2	99.5 %	99.4 %
artif-II	5	3348	1652	5-10-5-2	99.4 %	99.0 %
XOR	8	150	106	8-8-4-4-2-2-2	100 %	100 %

🕮

DeepRED – Rule Extraction from Deep Neural Networks	Springer International Publishing Switzerland T. Calders et al. (Eds.): DS 2016, LNAI 9956, pp. 457–473, DOI: 10.1007/978-3-319-46307-0 29
Jan Ruben Zilke(B), EneldoLozaMenc´ıa, and Frederik Janssen	

Earlier successful approach	✓ Tree-based machine learning models: [random forests, decision trees; gradient boosted trees
Limitation	- No explanation of their predictions
Method introduced	Interpretability of tree-based models through three main contributions. 📖 Optimal explanations based on game theory 📖 Local feature interaction effects 📖 Local explanations of each prediction →Global Model Structure understanding
Local explanation methods	☞ Reporting decision path - Not helpful for most models(ex. Multiple trees) ☞ Assigning credit to each input featureby heuristic approach - Strongly biased based on tree depth ☞ Model-agnostic approaches - Executing the model for each explanation - Slow and suffer from sampling variability

Dataset	Chronic Renal Insufficiency Cohort (CRIC)
Patients	3,939 chronic kidney disease patients; 10,745 visits
Features	333 ; Electronic medical record dataset with 147,000 procedures and 2,185 features
Task	Classification End-stage renal disease within 4 yr or not

Dataset	National Health and Nutrition Examination Survey (NHANES) Epidemiologic Followup Study
Patients	14,407 individuals and 79 features
Task	Risk of death over 20 yr of followup

From local explanations to global understanding with explainable AI for trees	Nature Machine Intelligence, 56 (2020), 56–67
Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal and Su-In Lee	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Interpretable multiclass classification by MDL-based rule lists	Information Sciences xxx (xxxx) xxx; doi.org/10.1016/j.ins.2019.10.050
Hugo M. Proença, Matthijs van Leeuwen	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – Xai

Image Analysis

Caltech data set	- 9144 images - 102 classes (101 object classes and a "back-ground"class - Object classes: [human faces, leopards, motorbikes, binocular, brain, camera, etc.]	
	Dimension : 3,000 ; Tr:3060(30/class); Te:6084	
YaleB data	- 38 Persons (or classes) - 2414 Face Images - 64 Illumination Conditions - Images Resized to 24 ×21	
	- Dimension : 504 ; Tr:1216(32/class); Te:1198	
A group LASSO based sparse KNN classifier	Pattern Recognition Letters, 131 (2020) 227-233 https://doi.org/10.1016/j.patrec.2019.12.020	
Shuai Zheng and Chris Ding		

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Material Science

Explanation	Interpretable models - Material science - Small datasets
Task	Design and discovery - Of new materials with desired properties
Machine Lrn	- With Bootstrapped Projected Gradient Descent – BOPGD algorithm is constrained with Buckingham Pi theorem based dimensional analysis and scaling laws of relationships between different input descriptors(properties)
Positive features	+ Learn from Small data + Develop predictive models ▪ Accurate, computationally inexpensive physically interpretable.
Dataset	- 82 materials → classified into three different crystal structures,
Target property	- Predicting intrinsic dielectric breakdown (Fb) - Descriptors: Eight
Method	- PCA - Pairwise correlations
Machine learning constrained with dimensional analysis and scaling laws:Simple, transferable and interpretable models of materials from small datasets	Chem.Materials (2020) DOI: 10.1021/acs.chemmater.8b02837
Narendra Kumar, Padmini Rajagopalan, Praveen Pankajakshan, Arnab Bhattacharyya,Suchismita Sanyal, Janakiraman Balachandran, and Umesh V. Waghmare	

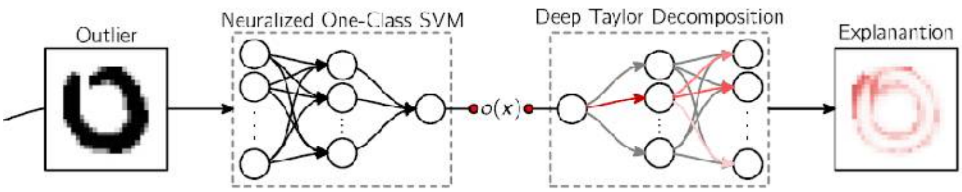
xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Explainable Machine Learning Algorithms To Predict Glass Transition Temperature	Acta Materialia (2020), doi: https://doi.org/10.1016/j.actamat.2020.01.047
EdesioAlcobac¸a, SauloMartielloMastelini, Tiago Botari, Bruno Almeida Pimentel, Daniel Roberto Cassar, Andr´e Carlos Ponce de Leon Ferreira de Carvalho, Edgar Dutra Zanotto	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – Xai

Deep Taylor Decomposition(DTD)

Deep Taylor decomposition (DTD)	o Quickly and reliably explain decisions in terms of input features o Basis: It leverages the model structure
Method	▪ OC (One class)-DTD
FOM	▪ Outperforms baseline procedures viz. ☞ Sensitivity analysis, distance to nearest neighbor, or edge detection ☞ Distance Decomposition, Gradient-Based, SHAP



Outlier explanation

Towards explaining anomalies: A deep Taylor decomposition of one-class models	Pattern Recognition 101 (2020) 107198 /doi.org/10.1016/j.patcog.2020.107198

Jacob Kauffmann , Klaus-Robert Müller, Grégoire Montavon

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

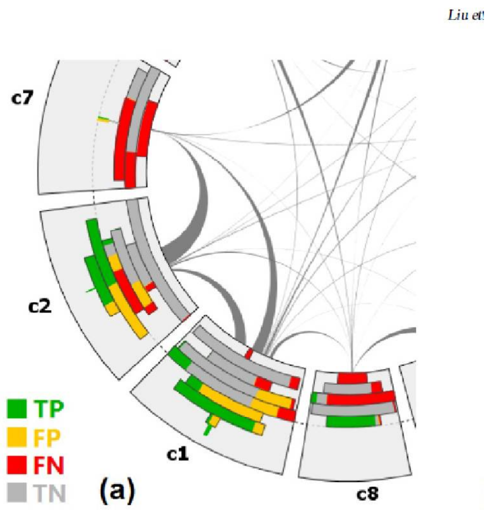
Deep Neural Networks	+ Gold standard in MachLrn - DNNS are black boxes due to their multilayer Nonlinear structure - Lack of transparency→Limiting interpretability - Prevents a human expert from being able to verify, understand reasoning of system
Method proposed	o Deep Taylor decomposition o Alg: backpropagating explanations from the output to the input layer + Explanation of classification decisions of a machine learning model in terms of input variables
Datasets	MNIST and ILSVRC
Explanation necessary	**Image classification** ☞ Indicate whether a test image belongs to a certain category or not ☞ Explain what structures (e.g. pixels in the image) were the basis for its decision - Sensitivity analysis ignores or overrepresents some of the relevant regions -
Explaining NonLinear Classification Decisions with Deep Taylor Decomposition, Pattern Recognition	Pattern Recognition 65,2017, 211-222 http://dx.doi.org/10.1016/j.patcog.2016.11.008

Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder,Wojciech Samek and Klaus-Robert Müller

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Computer science

xAI	🕮 Interactive visualization
	🕮 Understanding, diagnosis
	🕮 Creating explainable models

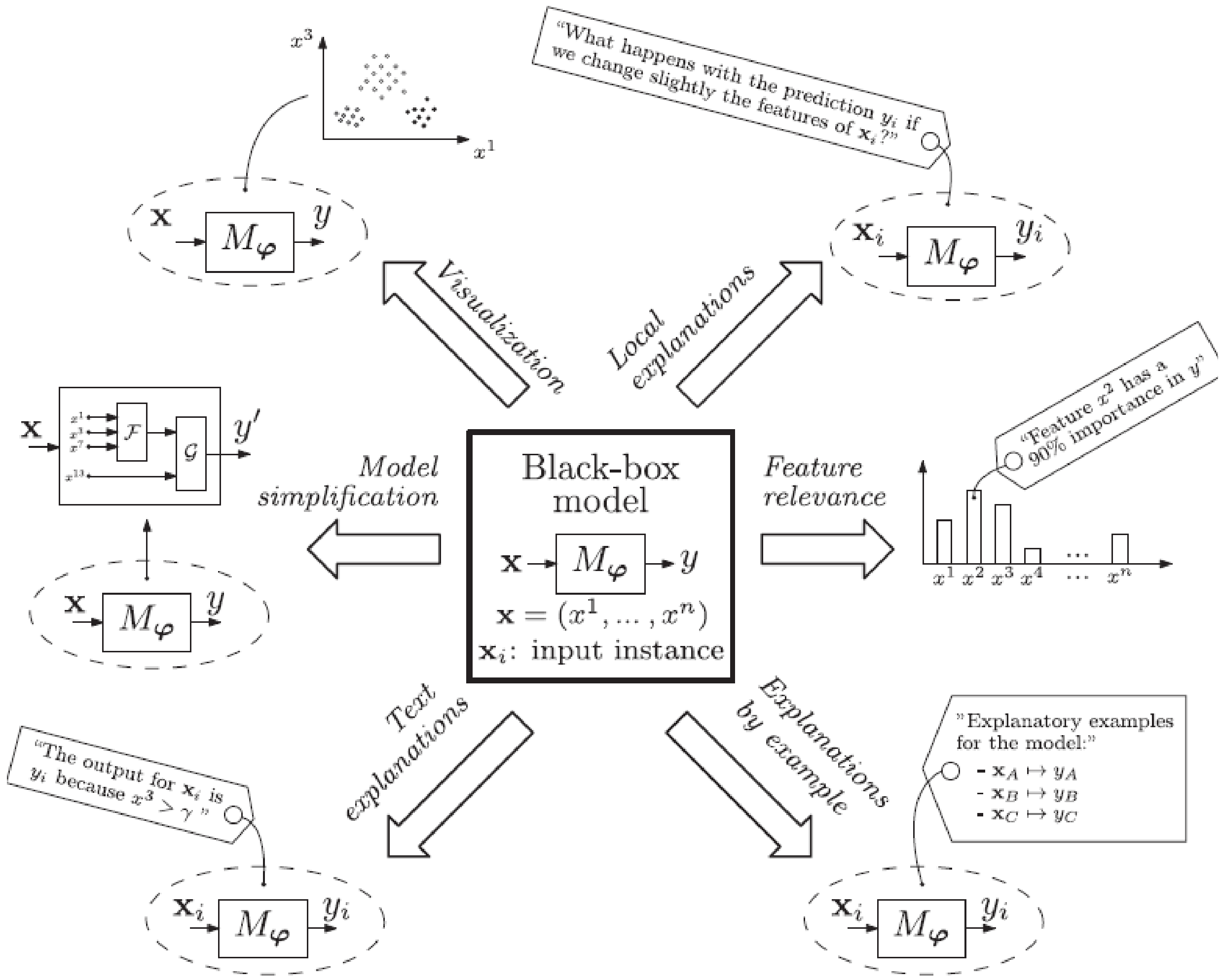


confusion wheel: A visual analytics tool – used by machine learning experts to diagnose model performance

Towards better analysis of machine learning models: A visual analytics perspective	Visual Informatics (2017), http://dx.doi.org/10.1016/j.visinf.2017.01.006
Liu, S.,et al.,	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Classical AI	Expert systems and rule based models	Review
Sub symbolic systems	Ensembles or Deep Neural Networks	
eXplainable AI (xAI)	Machine learning-explainability	
Methods	Data fusion ; workflows; explainability	
Responsible AI	Large-scale implementation of AI methods in real organizations Fairness, model explainability; accountability	



Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI	Information Fusion, 58 (2020) 82–115

Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez , Javier Del Ser, Adrien Bennetot, SihamTabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera

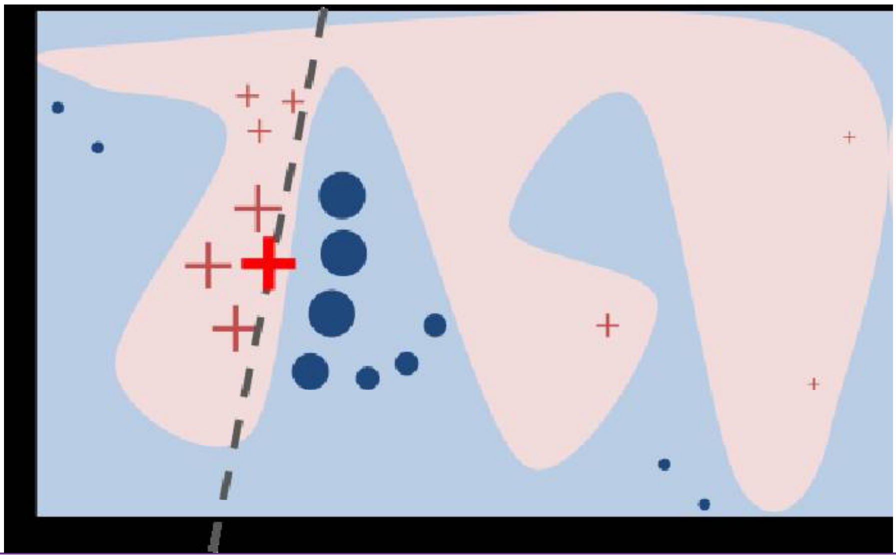
xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Explainable AI	+ Closer to explanation concept of outcome + Performance improvement is closer to concept of a benefit
Imperfect AI	Utilitarian benefit, empathy
Humans	! Capable of producing high-quality data that AI lacks ☞ Complex image recognition ☞ Speech recognition ☞ Translation in the field constructs ▶ Bias (conscious/unconscious)
AI	No bias (unless machine is a replica of human brain)

Egoistic and altruistic motivation: How to induce users' willingness to help for imperfect AI	Computers in Human Behavior, 101 (2019) 180-196 https://doi.org/10.1016/j.chb.2019.06.009
Yeonjoo Lee and Miyeon Ha and Sujeong Kwon and Yealin Shim and Jinwoo Kim	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Explanation technique	🕮 LIME[Local Interpretable Model-agnostic Explanations] 🕮 SP- [submodular pick] LIME 🕮 RP- [Random pick] LIME
Explainability	☞ Explains predictions of any model in an interpretable manner ☞ →Improving an untrustworthy classifier ☞ Identifying why a classifier should not be trusted
Humans	▪ Learn an interpretableModel locally around the prediction ▪ Explain the predictions of any classification



Dashed line: learned explanationlocally (but not globally) faithful
Bold red cross: Instance explained
Blue/pink background: black-box model's complex decision function (unknown to LIME)

"Why Should I Trust You?" Explaining the Predictions of Any Classifier	doi.org/10.1145/2939672.2939778 KDD 2016 San Francisco, CA, USA
Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Feature selectionmethod	☞ Informative Variable Identifier (IVI), o Identifying informative variables.
DataSets	o Non-linear Madelon Data o Digit Recognition Database MNIST o Synthetic linear classification problem with a binary output variable

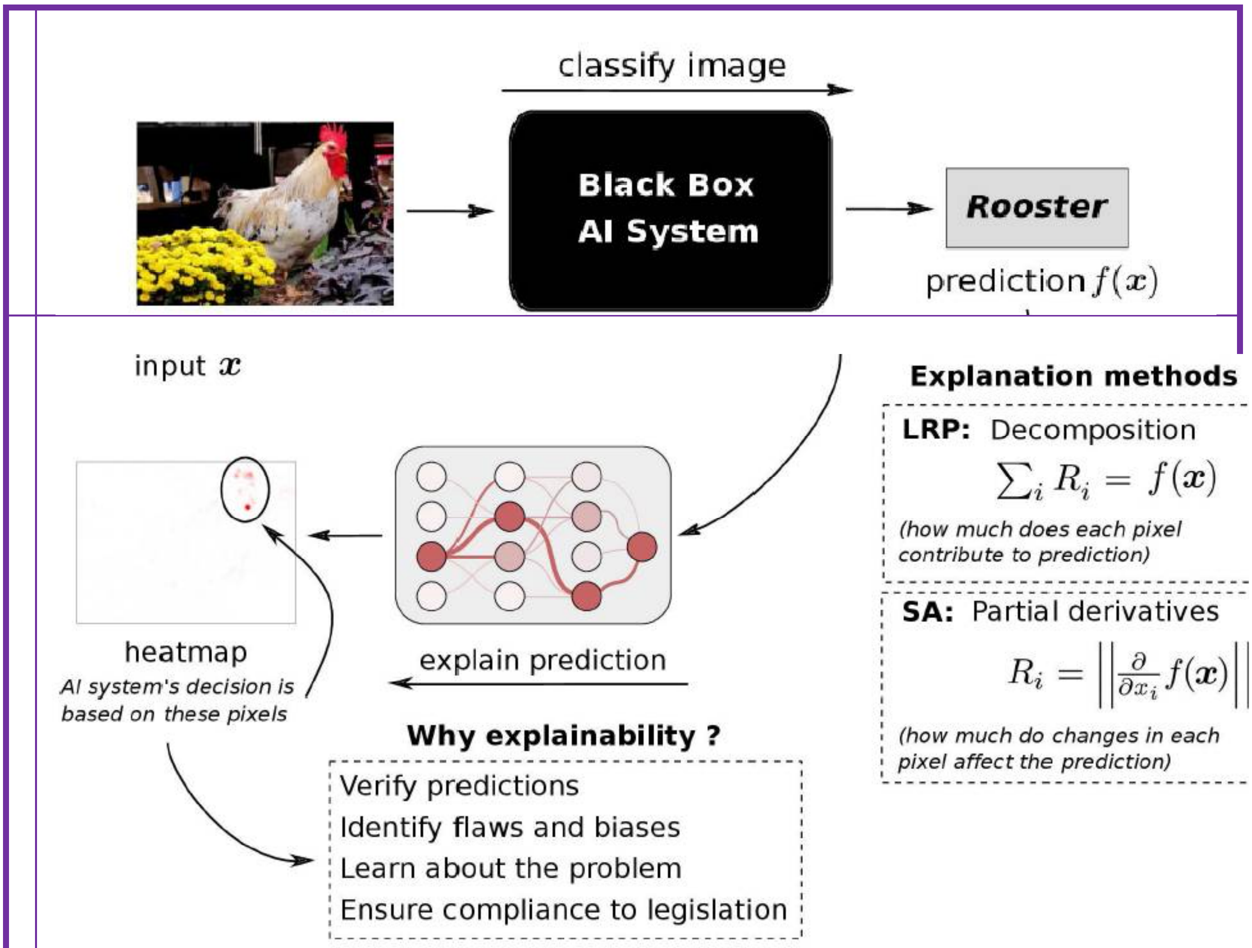
Informative variable identifier: Expanding interpretability in feature selection	Pattern Recognition, 98 (2020) 107077
Sergio Munoz-Romero, ArantzaGorostiaga, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, JoséLuisRojo-Álvarez	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – Xai

Explanation technique	🕮 SHAP (SHapley Additive exPlanations)
Explainability	☞ Interpreting predictions
Method	Additive Feature Attribution methods ▪ LIME ▪ DeepLIFT ▪ Layer-Wise Relevance Propagation ▪ Classic Shapley Value Estimation o Shapley regression values o Shapley sampling values o QuantitativeInput Influence
Basis	Additive feature attribution methods have an explanation model which is a linearfunction of binary variables
Limitation of classical methods	- Accuracy versus interpretability of model predictions **Remedy:**class of additive feature importance methods
Future methods	Faster model-type-specific estimation methods ! Make fewer assumptions ! Integrating work on estimation ! Interaction effects from game theory ! Defining new explanation model classes
A Unified Approach to Interpreting Model Predictions	31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA
Scott M. Lundberg, Su-In Lee	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Applications of xAI	o Medical domain - Wrong decisions of the system can be very harmful o Image classification o Sentimentanalysis, o Speech understanding o Strategic game playing
MachLrn	▪ Nested non-linear structure + Highly successful - Black-box manner [No informationabout what exactly makes system to arrive at decisions/predictions]
xAI	☞ Visualizing ☞ Explaining in text mode ☞ Interpreting deep learning models
Explanation tools	🕮 Sensitivity of the prediction with respect to changes in the input $R_i = \left\lVert \frac{\partial}{\partial x_i} f(\mathbf{x}) \right\rVert$ 🕮 Meaningfully decomposition of decision in terms of input variables 🕮 Layer-wise relevance propagation (LRP)



Explainable artificial intelligence: understanding, Visualizing and interpreting deep learning models	ITU Journal: ICT Discoveries, Special Issue No. 1, 13 Oct. 2017
Wojciech Samek, Thomas Wiegand, Klaus-Robert Müller	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – Xai

xAI	o Fuzzy linguistic modelling based approach
Intelligibility	o Machine learning models - Lack of intelligibility
Methods. Explanation	o LIME or SHAP o Make the predictions of ML transparent for humans o Still a gap to make enough intelligibility
	Intelligibility modes ▪ **Expert-2-Model:** Existing expert knowledge compatibility with the machine learning model ▪ **Expert-2-Expert:** Consolidation of knowledge from many experts in accordance with the model ▪ **Model-2-Expert:** Output of model explainers to humans ▪ **Feature-2-Expert:** Feature importance to humans

A fuzzy linguistic supported framework to increase Artificial Intelligence intelligibility for subject matter experts	7th International Conference on Information Technologyand Quantitative Management(ITQM 2019) Procedia computer science 162(2019)865-872
Juan Bernabé-Moreno, KarstenWildberger	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – Xai

Method	New visualization approach based on a Sensitivity Analysis
Application	🕮 To extract human understandable knowledge from supervised learning black box data miningmodels, Ex: NNs, SVMs. ensembles,including Random Forests
visualizations for SA	☞ Input pair importance ☞ Color matrix ☞ Variable effect characteristic surface
Datasets	o Bank direct marketing (classification) o Contraceptive method choice (classification) o Rise time of a servomechanism (regression) o White wine quality (regression)
Using sensitivity analysis and visualization techniques to open black box data mining models	Information Sciences, (2012) dx.doi.org/10.1016/j.ins.2012.10.039
Paulo Cortez, Mark J. Embrechts	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Evolution of eXplanation

- o Application fields
  - ☞ e-health, domestic robots, training
- o Necessary conditions for demanding explainability
- o Understandable explanations based on Social Science and psychological background
- o Platforms/architectures
  - ☞ BDI (Belief, Desires, and Intentions)
  - ☞ MDP(Markov Decision Process)
  - ☞ POSH (Parallel-rooted-ordered Slip-stack Hierarchical Action Selection),
  - ☞ STRIPS (Stanford Research Institute Problem Solver)
- o Explanatory granularity (Context; user-sensitive)
- o Explanation display
  - ✓ Expressive lights
  - ✓ Graphical. User interface
  - ✓ Natural language
- o Evaluation of explanation frame work
- o Future solutions for present xAI limitations

Explainable Agents and Robots: Results from a Systematic Literature Review	AAMAS 2019, May 13-17, 1078-1088, Montréal, Canada
SuleAnjomshoae, Amro Najjar, Davide Calvaresi, Kary Främling	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

xAI	🕮 Vital interdisciplinary research field ! XAI is not just a labcoat research field
Explainability	All aspects related to XAI ☞ Five W's o What, Who, When, Why, Where

	☞ How

Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)	IEEE, Access, 6(2018)52138-52160 DOI: 10.1109/ACCESS.2018.2870052
AMINA ADADI AND MOHAMMED BERRADA	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

ML vs Humans	ML algorithms performance exceed human level at times
Future	To integrate explanations into a larger optimization process →Improvementin performance of model or reduce its complexity.
xAI	Methods for visualizing, explaining and interpreting deep learning models
	Which one?? ! Predict right for the 'wrong' reason ! Predict wrong with right reasoning ! Evolution (natural/artificial) without explicit explanations
Towards Explainable Artificial Intelligence	Explainable AI, LNAI 11700, pp. 5–22, 2019 https://doi.org/10.1007/978-3-030-28954-6_1
Wojciech Samek, and Klaus-Robert Muller W. Samek et al. (Eds.)	

xAI — xAI — xAI — eXplainable Artificial Intelligence — xAI — xAI—xAI — Interpretable AI- IAI – IAI – xAI

AI procures	- Poor explainability
New method	o Exp-scalable method o Easily interpretable high-level summary of the relationship between entities
Data type	Dyadic datasets
FoM (Figure of Merit)	+ Explainability and accuracy + Extract relevant actionable information + Handles large datasets
A scalable decision-tree-based method to explain interactions in dyadic data	Decision Support Systems, 127 (2019) 113-141 https://doi.org/10.1016/j.dss.2019.
CarloartisEiras-Franco and Bertha Guijarro-Berdiñas and Amparo Alonso-Betanzos and Antonio Bahamonde", and Yealin Shim and Jinwoo Kim	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

xAI	Human knowledge role in explainable systems
New method	Neural Logic Networks (NLN); supervised incremental learning
Data set	Credit rating
Explanation	Tree method; NLN
Future scope	Fuzzy clustering and Bayesian models connection with NLN

Human Knowledge in Constructing AI Systems - Neural Logic Networks Approach towards an Explainable AI	Procedia Computer Science, 126 (2018) 1561-1570 https://doi.org/10.1016/j.procs.2018.08.129
Liya Ding	

xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Explanation	Basis ☞ Philosophy, psychology, social psychology, cognitive science
Article type	o Review
xAI	o Explainable AI scientists + Human computer Interface + … → Impressive results o Fool proof products not straight forward

Explanation in artificial intelligence: Insights from the social sciences	Artificial Intelligence, 267 (2019) 1-38 https://doi.org/10.1016/j.artint.2018.07.007
Tim Miller	

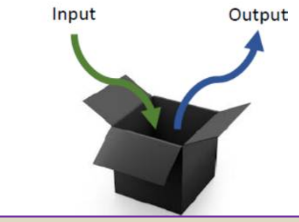
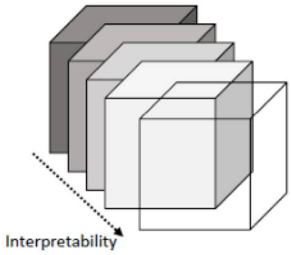
xAI—xAI—xAI—eXplainable Artificial Intelligence— xAI—xAI—xAI— Interpretable AI- IAI – IAI – xAI

Object orieted terminology (OOT)

Explainable AI [Machine Learning; Deep NN; Rule -Base; Robotics;]

DARPA goals	**Creation of technology for xAI** 🕮 New or modified machine learning techniques with embedded or external explanation interfaces/modules 🕮 Explainable models + Explanation approaches 🕮 Integrating state-of-the-art human-computer interaction (HCI) techniques (e.g., visualization, language understanding, language generation, dialog management) 🕮 Psychological assertions/theories of explanation for effective Interpretations

Autonomous systems of future	
☞ Perceive, learn, decide ☞ Act on their own	Civilian
Intelligent, autonomous, Symbiotic systems + Explainable AI	Department of Defence (DoD)
+ Help users to understand, appropriately trust, effectively manage AI systems	

[Black box; White box]	Computation: Intelligence [human; artificial], Learning; Machine learning; networks [Shallow; deep]	
Black box AI	- Employs complex opaque algorithms - Make no transparency for why a specific decision arrived - Not clear even to designers	
White box AI	Lighten up darkness ofcomplex black-box models	
	Opening black box ☞ Peeking Inside the Black-Box ☞ Whitening/increasing transparency or decreasing opaqueness ☞ Design of transparent deep models and deep learning modules ☞ Interfaces/modules for explanation	

xAI goals	**Development of new/modified machine learningtechniques with Transparent AI** + Actions should be easily understood by humans + Explainable models + Well-designed explanation interface + To work with existing old and new machine learning techniques to render them more explainable
	Interested in o New technology at the intersection of machine learning and HCI o Explaining machine learning models to end users o Interactive machine learning and visual analytics o Psychology of explanation
	DOD is not interested in XAI research 🕮 Unrelated to the specific issues of explainable AI 🕮 On effective explanation dialog Ex: user modeling, personalization, theory of mind

Desired Properties of xAI Systems	
- Informativeness - Low cognitive load - Usability - Fidelity - Robustness - Non-misleading - Interactivity /Conversational	- Accuracy - Interpretability - Responsiveness - Fairness - Privacy - Reliability - Robustness - Scalability

xAI workshops	
2017 IJCAI	Workshop on Explainable Artificial Intelligence5, and the
2018	Workshop on Explainable Smart Systems (EXSS)

XAI ANTITHESIS: EXPLAIN OR PREDICT
🕮 Simple and interpretable functions do not make the most accurate predictors
🕮 Accuracy requires more complex prediction methods
🕮 More complex the model, the more difficult it is to interpret

Software for explaining/interpreting black box models	
SHAP (Link)	SHapley Additive exPlanations github.com/slundberg/shap
ELI5	A library for debugging/inspecting machine learning classifiers and explaining their predictions github.com/TeamHGMemex/eli5
Skater	Python Library for Model Interpretation/Explanations github.com/datascienceinc/Skater
Yellowbrick	Visual analysis and diagnostic tools to facilitate machine learning model selection github.com/DistrictDataLabs/yellowbrick
Lucid	A collection of infrastructure and tools for research in neural network interpretability github.com/tensorflow/lucid
DeepExplain	perturbation and gradient-based attribution methods github.com/marcoancona/DeepExplain
iNNvestigate	A toolbox to iNNvestigate neural networks' predictions github.com/albermax/innvestigate

Explainability

Methods -- Explainability versus performance

- Highest performing (e.g., deep learning) methods are
  - Least explainable
- Most explainable (e.g., decision trees) are
  - Less accurate
- Which ??
  - Neither | either | No-choice

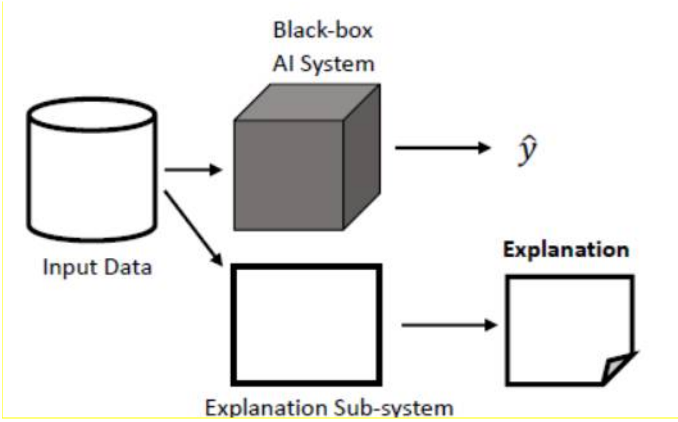
eXplainable high tech intelligent tools (xAI[xML;xI;xConsciousness]) of future

Extent of explainability

- Too Much, Too Little, or Just Right?
- Completeness or correctness (truth value)

Explainability Is it complete with?

- ! Humans alone
- ! Machines alone
- ! Humans and machines
- ! Human in the loop



Accuracy versus eXplainability

- Tradeoffs between "how smart an AI is" and "how transparent it is"
- Tradeoffs grow larger as AI systems increase in internal complexity

Explanation not mandatory --- But enhances credibility

Explanation Essential

Finance	Criminal Justice	Healthcare
• Credit		

Task	☞ Recommender system o online retail
Goal	☞ To show adverts, products o Social Media posts
Target users	☞ Right people at the right time
Necessary	☞ Accurate algorithms ☞ Commercially optimal approach ☞ Revenue optimization
Not essential	☞ eXplainability ('why'doesn't matter) ☞ Transparency

scoring • Loan approval • Insurance quotes	• People wrongly denied • Recidivism prediction • Unfair Police dispatch	☞ ICUs ☞ Critical Systems ☞ Diagnosis ☞ Med. Insurance
Big Data tasks	**Governess**	**Defense**

eXplainability standards
☞ Having human intelligence as the gold standard of AI ☞ Explaining capability of human decision makers
☞ Learning in defining semantic attributes, describing seed model, deciding layers and relation between layers, or verifying interpretations

Explainability	Narration of causal relationships of observed phenomena/model for I/O mapping or classification in a comprehensible manner through a linguisticdescription and visual display
	- Explanations cannot answer all queries of all users - No agreed definition of what an explanation is - No quantification/scale of comprehensibility of an explanation for humans (of different intellectual level)
Post-hoc explainability	A high complex uninterpretableblack-box model with high accuracy is developed. The model predicts outcome. It is explained in terms of readily available off-the-shelf interpretable knowledge Probes to reverse engineering process are used without altering or even knowing inner details of the black box model
Ante-hoc explainability	Explainability is included in the strucuture of work-flow during design itself So explanation is available for possible outcomes even before running the software

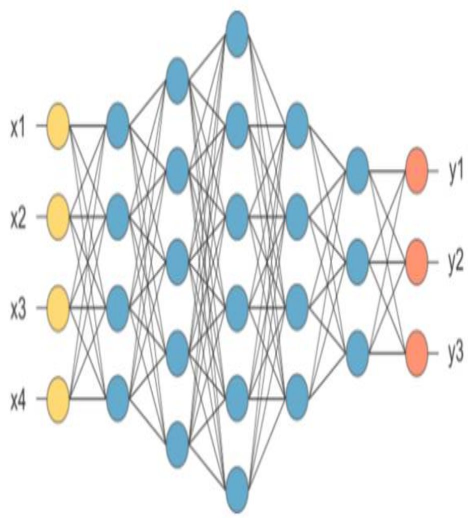
Agnostic explanation	This model approximates a black-box model locally in the neighborhood of any prediction of interest Dilating models even without knowledge of dataset
Causal explanations	The outcome or intermediate results explained from laws and conditions in a deductive way Hybridisation of machine learning → it develops a new dimenion in xAI systems

Transparency	The transformation of Input to output is clear
Understanding	🕮 Knowing context in which the facts appear In addition to ! Representation of facts ! Recognizing, perceiving ! Reproducing (stimulus–response on a physiological level) ! Content comprehension
Intelligibility to	[Scientific community [Developers; tool application scientists;] [non-experts, experts] [product designers, Engineers, data scientists] [Marketing personnel, business customers] [End users] Explainable [products; agents;] product user; public setting
Interpretability	Related to the model and notto the training data that is unknown
Accountability	For use of product by end users

Reasoning	o [Deductive; inductive; abductive]
Deductive	o "Top-down logic" : process of reasoning from premises to a conclusion
Inductive reasoning	o "Bottom-up logic" Reasoning from a single observation or instance to a probable explanation or generalization.
Abductive reasoning	o Reverse of deductive reasoning o Proceeds from an observation to the most likely explanation

Creation of Explainability modules from black-box-models or from scrap
Learning 🕮 Associating explanatory semantics with features of the model 🕮 Developing simpler models -- easier to explain 🕮 Proposing richer models that contain more explanatory content 🕮 Inferring approximate models -- purpose is only explanation

Explanators	
Decision Tree (DT) Decision Rules (DR)	Saliency Mask (SM) Saliency Map Sensitivity Analysis (SA) Partial Dependence Plot (PDP) Prototype Selection (PS) Activation Maximization (AM) Individual Conditional Expectation

Evolution of transparency, interpretability and explainability of Model outcome		
Yester years	🕮 Blackbox model driven methods at core - No Explanation	
Now	🕮 Data driven models 🕮 Black box Explanation → transparent box model driven 🕮 Model specific implementation	
Tomorrow	o Transparent box (interpretation by design; Explanation embedded) o NLP interface---can also be used as black box at choice (of users in field operation)	
Future	! Agnostic models ! New models interpretable by design	

pDeep Explainability

Deep Explanation	o Operational details of deep NNs o Deconvolutional networks + Used to visualize the feature mapping output in layers of convolutional networks

eXplainablehybrid deep learning methods

- o Explainable features, explainable representations
- o Explanation generation facilities

Design choices fortransparent deep learning

- o Selection of training data, initial conditions, and training sequences
- o Architectural layers, loss functions, regularization, optimization

Categorization of models based on degree of explanation			
Method. Learning	Class of model		eXp. Scale
Bayesian belief NNs	Graphical	Models	3.5
Decision trees	Supervised unsupervised	Leaning	4
Logistic regression	Supervised unsupervised	Leaning	3
SVM	Supervised unsupervised	Leaning	2
k-means	Supervised unsupervised	Leaning	3
Random Forest/Boosting	Ensemble	Leaning	3
Q leaning	Reinforcement	Leaning	2
NNs	Deep	Learning	1
Hidden Markov Models	Natural Language process	Learning	3
			1: most difficult5: easiest

Evolution of AI during 1956-2020		
Generation	AI	Time period
First	▪ Symbolic expert system ▪ Shaky robots ▪ First order logic	1957-1970
Second	☞ Neural networks ☞ Probabilistic models (Statistical ; Bayesian] ☞ [GRNN, ProbNN; FuzzyNN] ☞ SVM, 🕮 learning [Mathematical; statistical; Fuzzy]	1980-2000
Third	▶ Neocognitron; Deep NN, Deep Learning ▶ Explainability of SLP, Stat models	1990; 2000-2020
Fourth	o Explainable (for hitherto existing black box models)	2016-
Fifth	! (Near) Realistic models for Real-life (micro- to mega) Phenomenon to control, communicate; command in ☞ Health, Environment, Defense, Governess, evolution (Hedge) ☞ Conscientious ; Consciousness	>2020

AI	[Comprehensive; understandable; Intelligible; Interpretable;] [Accurate AI; Responsible AI; [General AI; Super AI] [Accountable; Transparent; Fairness; Ethics;]
Intelligence	+ Accepted term (Psychology; Philosophy, Social Science) + Difficult todefine + Dependent on a wealth of different factors + Does not need a metal body to be a thread
Big data	o Coined by Cox and Ellsworth in 1997 o Originally referred to data being too big to fit into memory

	and processed by conventional means o Eight Vs – volume, velocity, variety, variability, visibility, value, veracity, vexing
Knowledge	☞ Processed and consolidated information or interpretations of the basic data, raw facts, observations from a particular point of view ;validated and is thought to be true
Knowledge distillation	▪ Compression method for training a small model to mimic a pretrained model or ensemble of models ▪ Used to transfer knowledge from cumbersome model to a small model
Data Mining (DM)	Data Mining (DM) aims to extract useful knowledge from raw data
Machine learning	**Methods:** Statistical and mathematical methods of increasing adaptability, complexity, goals/sub goals and utility **Computational facilities**: Computer hardware and software→ increased speed of computation and size of problem **Data:** Toy data sets to Big data; images/speech/hyphenated multisensory signals **Learning:**Learn important information, hidden patterns, associations from very large amount of data **State-of-art:** Machine learning has a niche in high performance computations
Deep networks	o Multilayered Neural networks o Neo-cognitron --- breakthrough in perception o Convolution NNs are of recent hype o Auto coders, decoders o Shallow NNs: If NN hidden layers restricted to two

Explanation	Response to a question; 🕮 Why did you do that? Why not something else? 🕮 When do you succeed? When do you fail? 🕮 When can I trust you? How do I correct an error?
Types	[Textual ; visual; graphical; dialectical]
Explanation about	☞ # Parts of model ☞ #Antecedents / consequents ☞ #Non-Zero weight (linear) ☞ Depth of tree (decision tree) 🕮 Model explanation: overall logic inside black box 🕮 Outcome: response for an instance input –local explanation [Lime; Anchors] 🕮 Model inspection [PDP; ICE; SHAP]
Local explanations	**Local explanations** ✓ Focus on data and provide individual explanations, ✓ Provide trust to model outcomes ✓ More faithful than global explanations ✓ LIME, onecovariate-out (LOCO)

LIME	Local interpretable model explanation ; Mathematical model: Fn(linear, x) + fn2(cubic,x) - Generate only 2^n different neighbors ☞ Remedy LioNets
LioNets	✓ Tried to interpret a neural network's prediction. o It is a **model-specific** outcome explanator Local interpretation of neural networks using penultimate layer coding
Global interpretability	☞ Focus on model and provide an understanding of the decision process ☞ Applications 🕮 Drugs prescription, diagnosis 🕮 Trends or a climatic change ☞ Global effect estimate is more helpful compared to many explanations for all possible idiosyncrasies

Explanation	Rules [symbolic] [If-then-else; [first order predicate; fuzzy; probabilistic; trees]] Graphic Textual [NLP; Human [expert; product-user; common-man[non-expert in all but potential consumer/user/propagator/promotor]
	[Input; process [work-flow; algorithm; output]
	Input: [Data [raw; pre-processed; generated from KIDs] [features] KIDs : [free/fixed parameters; data reduction/projection; dimension reduction; mapping to high dimensional space [SVM]}

Typical explainable Methods in research targets	
Linear trend in presence of outliers	Statistical parameters; graphical (ellipse) Residual analysis
Non-linear trend (polynomial – order zero to four)	Stat_Par; scatter diagrams-residual; Prior knowledge : data accuracy; If prior model available; y1 = f(x-square;par) Resid_1= yobs-y1; polyModel = f(resid_1); Explainable_1 + pseudo (explainable, black box) Predictability high
Stability constants	ML(l); ML(l)H(h); Symbolic; numerical; Parameters: Stat.beta; residual ; graphical Sensitivity analysis: errors (ingredient; data accuracy) Improvement: Experimental Design (ingredient concentration; number of experiments) Probes (GE Spectral, NMR)

Subtle differences between AI and xAI processing of Bigdata tasks		
Software Modules		
xAI	Integration of AI related technologies	(Challenging) Task
! Knowledge graph embedded Sequence Learning (using LSTMs)	o Deep Learning + Recurrent NN o semantics-augmented case-based reasoning o Natural Language Processing	o Airline caused delays o Globally 323,454 flights are delayed every year. o totaled 20.2 million minutes last year
! Knowledge graph embedded Random Forrest	o MachineLearning, o Reasoning, Natural Language Processing for building robust model	Accenture manages every year more than 80,000 opportunities and 35,000 contracts
Knowledge graph embedded Ensemble Learning	o Google Knowledge Graph, DBPedia (relational DataBase version of Wikipedia)	Predicting and explaining abnormally high employee expenses (Ex,:high accommodation price in 1000+ cities).
! Post-hoc explanation ! Local explanation ! Counterfactuals ! Interactive explanations	o Supervised learning o Binary classification	Loan applications
! Interactive explanations ! Multiplerepresentations	o Competing riskanalysis	Different treatments for Early invasive breast cancer

ACS.org ;sciencedirect.com : Information Source

R. Sambasiva Rao, School of Chemistry
Andhra University, Visakhapatnam
rsr.chem@gmail.com