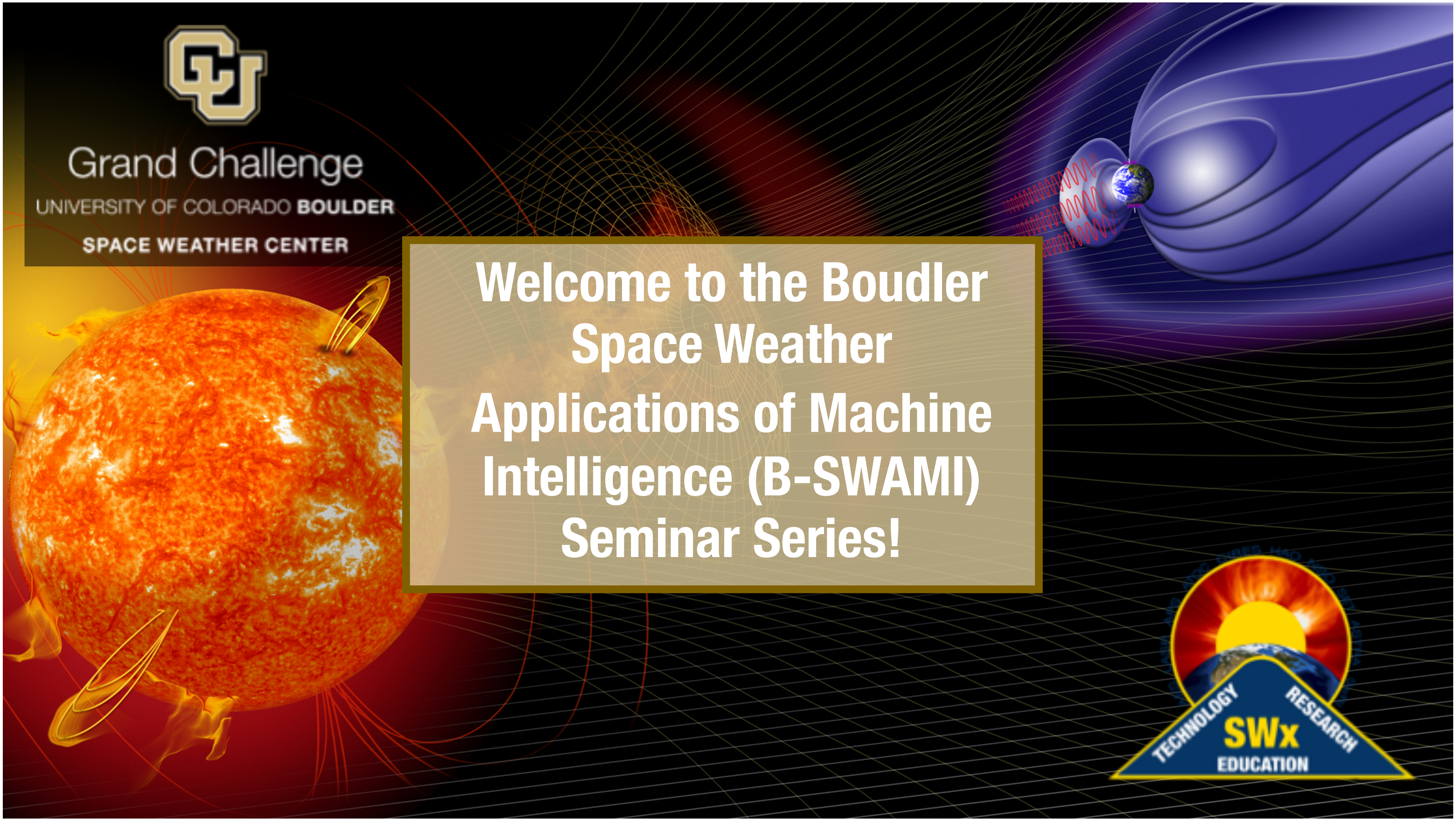

Grand Challenge
UNIVERSITY OF COLORADO BOULDER
SPACE WEATHER CENTER
Welcome to the Boudler Space Weather Applications of Machine Intelligence (B-SWAMI) Seminar Series!
TECHNOLOGY
RESEARCH
SWx
EDUCATION

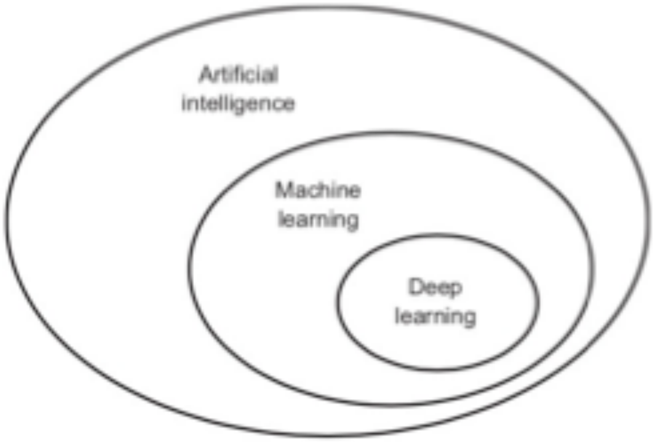
Outline

- "Machine Learning": Lexicon & Principles
- Deep Learning Examples
- Machine Learning in Space Weather Prediction
- Future Applications in SWx Forecasting
- Resources
- Open Discussion

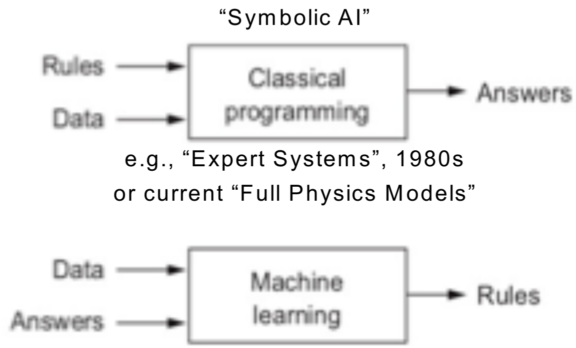
"Machine Learning": Lexicon & Principles I

- **Artificial Intelligence (AI):** the broadest term in use, referring to all attempts to automate human cognition.
- **Machine Learning (ML):** a sub-field of AI that aims to develop "programs" or "models" based on analysis of large data sets. "Learning" describes the automated process of setting parameters in the model based on "training data". New data are automatically processed.
- **Machine Intelligence*:** Alan Turing's preferred term for AI that is implemented specifically within an electronic computing machine ("computer"). Also makes for a much better acronym when combined with Boulder Space Weather…
- **Deep Learning (DL)**: a branch of ML in which "neural networks" with very many layers ("deep" networks) are trained as sequential geometric functions parameterized by learned "weights". Deep learning has recently (2012++) eclipsed all other ML techniques for complex problems.
- **Neural Network**: a connected network of layers, each consisting of "nodes" which are "activated" by their inputs to produce outputs. Activation function is necessarily non-linear. "Weights" for each node are the learned quantities.

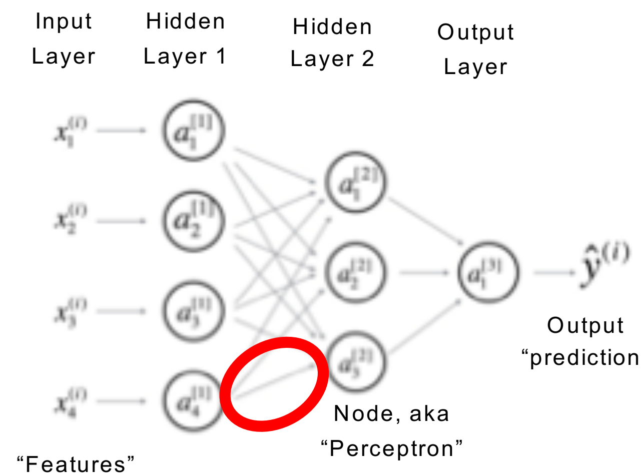
* A. M. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433-460.



AI, ML, and Deep Learning
Deep Learning with Python, Chollet 2018.



Symbolic AI Paradigm vs. ML Paradigm
Deep Learning with Python, Chollet 2018.

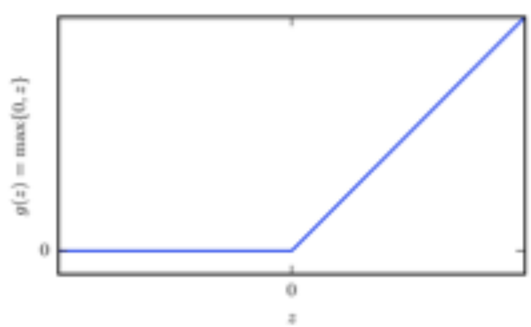


Fully Connected Neural Network with two "hidden" layers
Andrew Ng, deeplearning.ai

$$a_j^{[l]} = g^{[l]}(\Sigma_k W_{jk}^{[l]} a_k^{[l-1]} + b_j)$$

General Activation Equation for each Node
Andrew Ng, deeplearning.ai

g = non-linear activation function, e.g. tanh() or RELU()

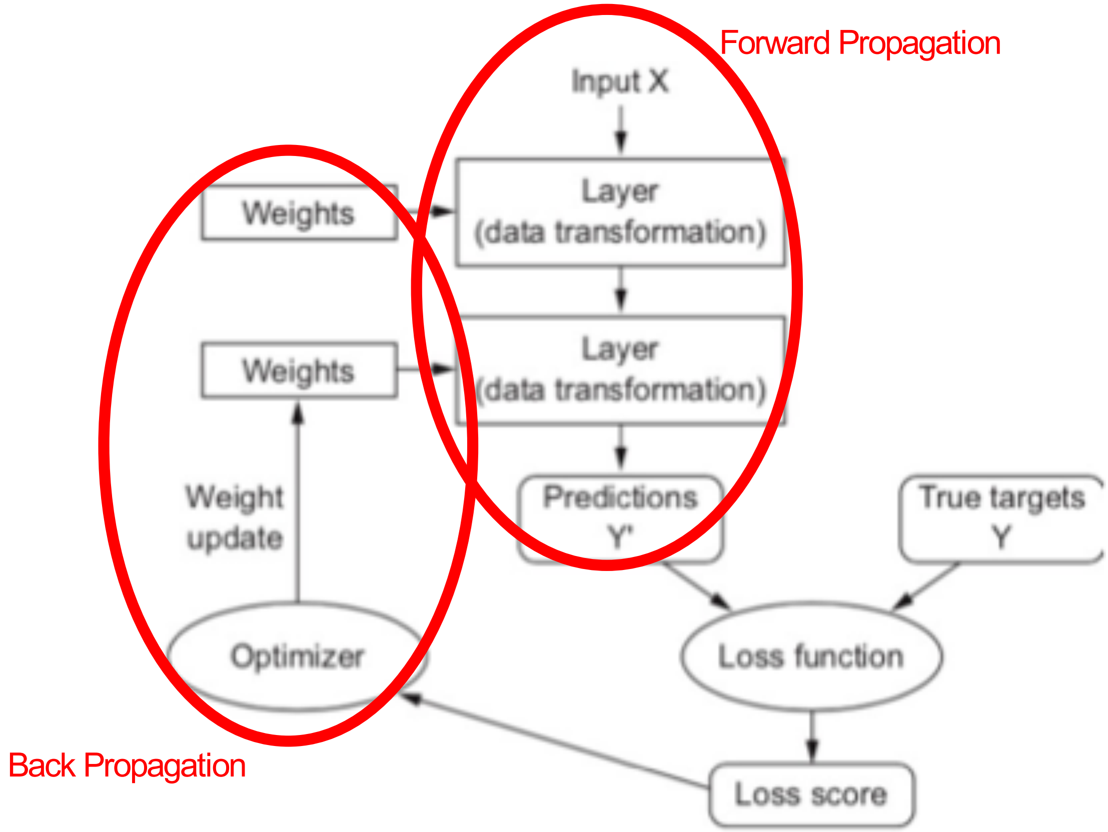


Rectified Linear Unit (RELU) Activation Function
Ian Goodfellow, www.deeplearningbook.org

Grand Challenge
UNIVERSITY OF COLORADO BOULDER
SPACE WEATHER CENTER

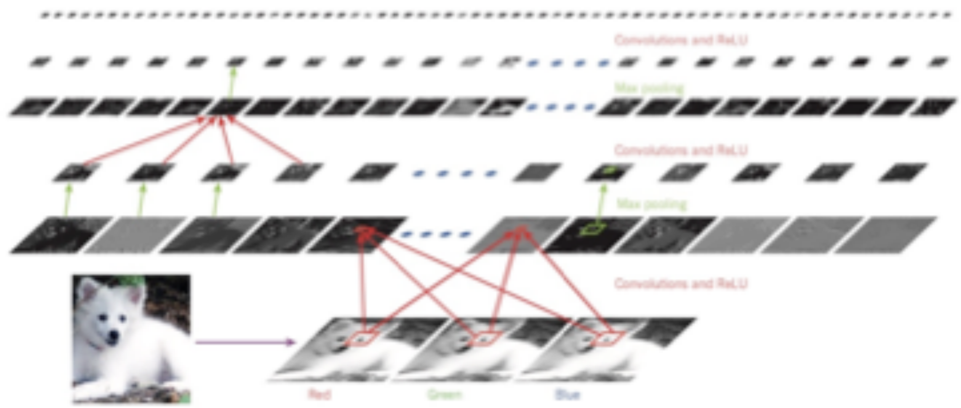
"Machine Learning": Lexicon & Principles II

- **Supervised Learning**: class of machine learning in which the user supplies inputs and desired output pairs (e.g. pre-classified examples) and the algorithm learns how to reproduce the desired outputs.
- **Unsupervised Learning**: class of machine learning in which only the input is provided to the algorithm and it performs segmentation or pattern recognition to determine outputs.
- **"Classical" Machine Learning Algorithms 1990s-Present**:
  - K Nearest Neighbors (KNN) – simplest classification and regression algorithm.
  - Support Vector Machine (SVM) – "hyperplane" determination for classification or regression.
  - Principal Component Analysis (PCA) – unsupervised analysis for common elements.
  - Decision Trees, including "Gradient Boosting" and Ensembles – "20 Questions" networks.
- **Deep Learning Algorithms ~2012-Present**:
  - Fully Connected Network (FCN) – earliest network architecture classification problems.
  - Convolutional Neural Network (CNN) – image and sequence analysis via convolution kernels.
  - Recurrent Neural Network (RNN) – sequence processing via "memory" of previous states.
  - Long Short-Term Memory (LSTM) network – more persistent/efficient memory architecture.
  - Reinforcement Learning – unsupervised learning strategy based on rewarding exploration or action.
  - Generative Adversarial Network (GAN) – generator trains to "fool" a discriminator agent.



Forward and Back Propagation in a NN

Deep Learning with Python, Chollet 2018.



Convolutional Neural Network Principals

Deep Learning, Lecun, Nature 2015.

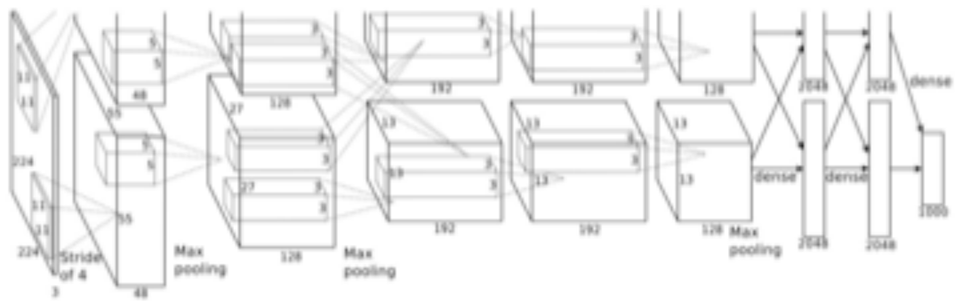


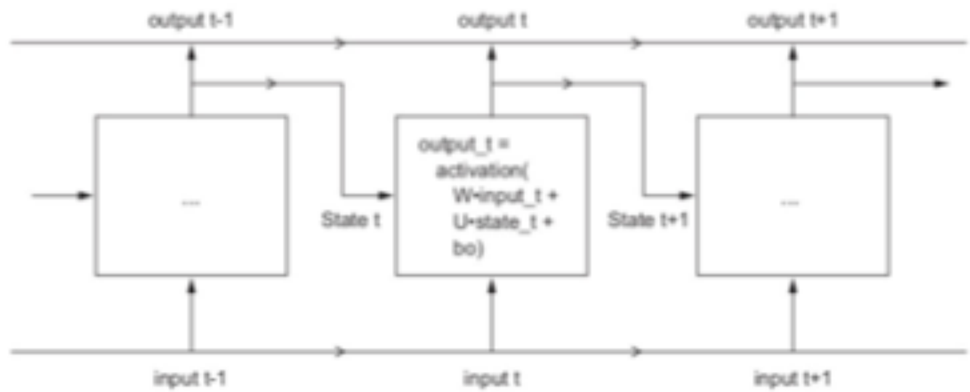
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input 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.

60 million parameters!

ImageNet Classification with Deep CNNs

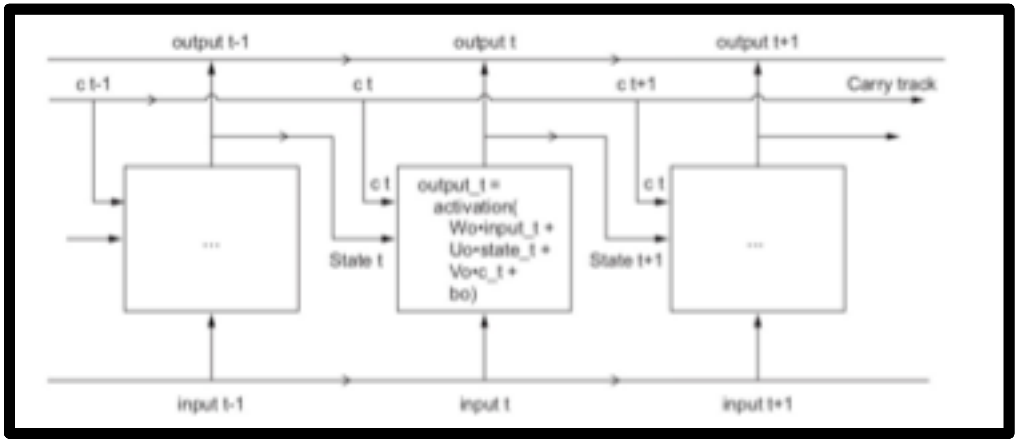
Krizhevsky, Sutskever, & Hinton, arXiv 2011

Grand Challenge
UNIVERSITY OF COLORADO BOULDER
SPACE WEATHER CENTER



RNN Network Timeline

Deep Learning with Python, Chollet 2018.



LSTM Network Timeline

Deep Learning with Python, Chollet 2018.

"Machine Learning": Lexicon & Principles III

- **Classification:** type of problem in which the goal is to identify an object as a member of a *pre-defined* class.
- **Object Detection:** sub-type of classification problem in which an object contained in an image is identified from a set of pre-defined classes.
- **Dataset Segmentation:**
  - **Training Set:** the subset of the available data used to train the network, i.e. optimize the weights until the loss function is below some acceptable threshold. Usually the majority of the available data (90—95%) in deep networks.
  - **Validation Set:** the subset of the data used to validate the current parameters and hyperparameters of a network. Cannot be used for Test Set since information from this set leaks into parameters and results in *overfitting*.
  - **Test Set:** a small subset of the available data used for final testing and error quantification. Modifying the network based on Test Set results turns the Test Set into a Validation Set and requires new data for final testing.
- **Overfitting:** optimizing parameters on validation data or on too small a training set – the network is tuned tightly to a small subset of the data and thus performs poorly on challenge data introduced in testing.
- **Class Imbalance**: condition of a training set in which one or more classes is rare and therefore the network learns to predict only the common class (e.g. X-class solar flares…)

Deep Learning Examples

Why has deep learning taken off in the last 10 years?

- Algorithms: backpropagation (1986), dropout, batch normalization, RELU activation
- “Big Data” availability
- Hardware (GPUs and now TPUs)
- Software: Tensorflow, Theano, Keras and Python support

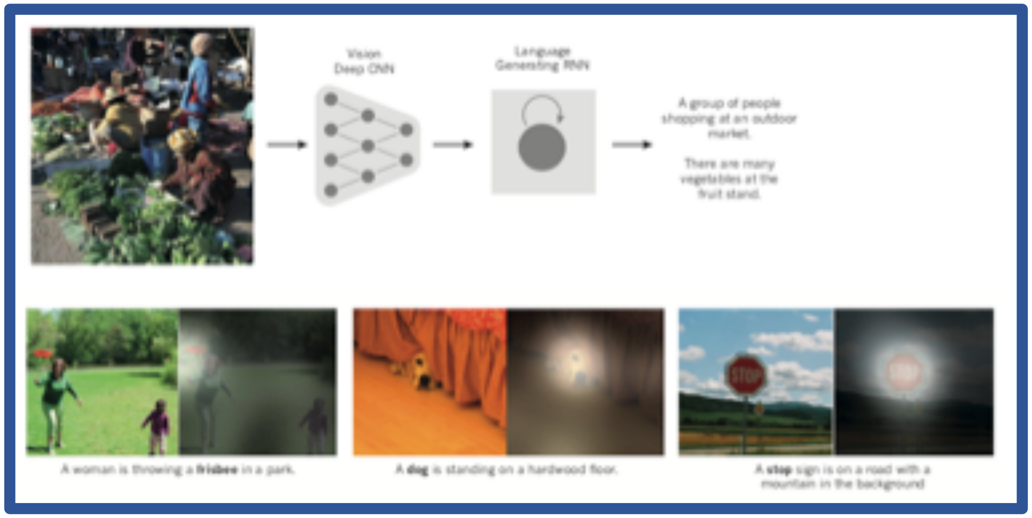
Amazing Examples to Date:

- Near human-level image classification
- ~~Near human-level speech recognition~~
- Near human-level handwriting transcription
- Near human-level language translation
- Highly accurate facial recognition
- Human-level driving ability
- Superhuman game playing ability
  - Learning Atari games from scratch
  - Defeat of the Go world champion
- “Creative” synthesis using GANs

- **Human-level** speech recognition
- **Human-like** speech synthesis

Google Now, Amazon Alexa

Google Duplex



Visual Attention and Caption Generation using CNN + RNN
(Xu et al., 2016)

What do you call a cat does it take to screw in a light bulb?
They could worry the banana.

What did the new ants say after a dog?
It was a pirate.

Why did the monsters change a lightbulb?
And a cow the cough.

What do you call a pastor cross the road?
He take the chicken.

Jokes generated by a text generating RNN
aiweirdness.com

Like humans, the system is imperfect...

Faces generated by a GAN (starting from random pixels)
www.miketyka.com

Machine Learning in Space Weather (Flare) Prediction
Background: The Language of Forecasting

Watch: "Something has been detected and *may* or *may not,* within some period of time, cause an event."

Warning: "Something has been detected and *will very likely*, within some period of time, cause an event."

Alert: "An event is in progress."

Current Capabilities

Event	Watch	Warning	Alert
Flare			X
Radiation Storm		X	X
Geomagnetic Storm	X	X	X

The ideal is to provide accurate, reliable, and timely **quantitative probabilities** for Watches and Warnings.

*As defined and used by, e.g., NOAA/SWPC.

Machine Learning in Space Weather (Flare) Prediction Background: Baseline Forecasts

- **Climatology Forecast:** the probability of an event occurring is the average of the probability over the relevant period.

For example, a climatology flare forecast would calculate the probability of an active flaring based on the probability of flaring all recorded active regions over, say, the past and current solar cycle.

If you can't do better than climatology, your method should be dropped.

- **Persistence Forecast**: things will stay just as they are right now, i.e., no flare is occurring now so that's the way it will stay.

Note that this is a *very accurate* forecast 90+% of the time. But it is also *useless* for high-impact episodic events like solar flares.

- **Recurrence Forecast**: the probability of an event occurring is the based on the probability of conditions returning.

This is the current operational method for forecasting coronal hole high-speed streams: the Sun rotates every 27 days so HSS events are predicted to return every 27 days.

Machine Learning in Space Weather (Flare) Prediction
Background: Current Operational Flare Forecasts

- **Classify a sunspot Active Region using the McIntosh System**

  Zurich Sunspot classification + penumbral development + density of medial spots

- **Use a Look-Up Table (LUT) based on 40+ years of flare statistics to find the probability of flaring at the M1—5 or X1 level within the next 24 hours for a given McIntosh class.**

  Flare levels are based on X-ray irradiance: A, B, C, M, X

  SWPC forecasts "R1—R2" = M1—M5 or "R3" = X1 "radio blackout" probabilities for each 24 hour period over 3-days.

- **Use human knowledge and prior skill biasing to modify probability**

  e.g. "Rapid flux emergence but McIntosh Class is not changing – increase the probability of R3 radio blackout in the next 24 hours from 50% to 75%."

Machine Learning in Space Weather (Flare) Prediction
Background: Contingency Tables

- Binary categorical tasks produce either “True/Positive” or “False/Negative” results.
  - “AR 10973 will flare in the next 24 hours” (P) or “AR 10973 will not flare in the next 24 hours” (N).

- Contingency Tables are used to create Skill Scores based on the relative numbers of
  - True Positive (TP)
  - False Positive (FP)
  - True Negative (TN)
  - False Negative (FN)

Observed

Forecast	P	N
P	**TP** 523	**FP** 67
N	**FN** 48	**TN** 1941

Skill Scores

$$\text{Skill} = \frac{A_{forecast} - A_{reference}}{A_{perfect} - A_{reference}}$$

Heidke Skill Score (HSS)

True Skill Score (TSS)

Accuracy

$$A_{forecast} = \frac{TP + TN}{N}$$

$$TSS = \frac{TP \times TN - FP \times FN}{P \times N} = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}.$$

Machine Learning in Space Weather (Flare) Prediction

Solar Phys (2013) 283:157–175
DOI 10.1007/s11207-011-9896-1

IMAGE PROCESSING IN THE PETABYTE ERA

Solar Flare Prediction Using Advanced Feature Extraction, Machine Learning, and Feature Selection

Omar W. Ahmed · Rami Qahwaji · Tufan Colak ·
Paul A. Higgins · Peter T. Gallagher ·
D. Shaun Bloomfield

Input “Feature Vector”

Table 1 SMART MF properties.

Property ID	Property	Description
v1	Type-Polarity	AR polarity (Unipolar/Multipolar)
v2	Type-Size	AR size (Big/Small)
v3	Type-Evolution	AR evolution (Emerging/Decaying)
v4	A	Area of the region
v5	Φ	Total unsigned magnetic flux of the region
v6	Φ_+	Total positive flux in the region
v7	Φ_-	Total negative flux in the region
v8	Φ_{IMB}	Flux imbalance fraction in the region
v9	$\Delta\Phi/\Delta t$	Flux emergence rate
v10	B_{MIN}	Minimum magnetic field in the region
v11	B_{MAX}	Maximum magnetic field in the region
v12	B_{MEAN}	Mean magnetic field in the region
v13	L_{NL}	Neutral-line length in the region
v14	L_{SG}	High-gradient neutral-line length in the region
v15	∇_{MAX}	Maximum gradient along the neutral line
v16	∇_{MEAN}	Mean gradient along the neutral line
v17	∇_{MEDIAN}	Median gradient along the neutral line
v18	R	Schrijver R value
v19	WL_{SG}	Falconer WL_{SG} value
v20	R^*	Schrijver R value with a lower threshold
v21	WL^*_{SG}	A modified version of WL_{SG}

SOHO/MDI 96 minute data: 1996 – 2010

21 magnetic field properties derived by manual definition

Cascade Correlation Neural Network with “several” hidden layers: shallow (Qahwaji & Colak 2006)

Binary classification problem: region is classified as **flaring** if it produced at least one C-, M-, or X-class flare in the following 24-hour period, and **non-flaring** if did not cause any C-, M-, or X-class flares in the ± 48-hour period around its observation time.

Table 5 Prediction measures achieved from applying machine learning with cross-validation on the segmented and operational datasets covering April 1996 – December 2010.

Association	Forecast-Verification Measures							
Method	MSE	TPR	FPR	TNR	FNR	FAR	ACC	HSS
Segmented	0.017	0.662	0.008	0.992	0.338	0.176	0.974	0.720
Operational	0.024	0.455	0.010	0.990	0.545	0.278	0.962	0.539

Machine Learning in Space Weather Prediction

The Astrophysical Journal, 798:135 (11pp), 2015 January 10 doi:10.1088/0004-637X/798/2/135

SOLAR FLARE PREDICTION USING *SDO*/HMI VECTOR MAGNETIC FIELD DATA WITH A MACHINE-LEARNING ALGORITHM

M. G. Bobra and S. Couvidat
W. W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305, USA; couvidat@stanford.edu
Received 2014 August 1; accepted 2014 November 1; published 2015 January 8

Table 1
SHARP Active Region Parameter Formulae

Keyword	Description	Formula	F-Score	Selection
TOTUSJH	Total unsigned current helicity	$H_{c_{total}} \propto \sum \lvert B_z \cdot J_z \rvert$	3560	Included
TOTBSQ	Total magnitude of Lorentz force	$F \propto \sum B^2$	3051	Included
TOTPOT	Total photospheric magnetic free energy density	$\rho_{tot} \propto \sum (\boldsymbol{B}^{Obs} - \boldsymbol{B}^{Pot})^2 dA$	2996	Included
TOTUSJZ	Total unsigned vertical current	$J_{z_{total}} = \sum \lvert J_z \rvert dA$	2733	Included
ABSNJZH	Absolute value of the net current helicity	$H_{c_{abs}} \propto \lvert \sum B_z \cdot J_z \rvert$	2618	Included
SAVNCPP	Sum of the modulus of the net current per polarity	$J_{z_{sum}} \propto \lvert \sum^{B_z^+} J_z dA \rvert + \lvert \sum^{B_z^-} J_z dA \rvert$	2448	Included
USFLUX	Total unsigned flux	$\Phi = \sum \lvert B_z \rvert dA$	2437	Included
AREA_ACR	Area of strong field pixels in the active region	Area $= \sum$ Pixels	2047	Included
TOTFZ	Sum of z-component of Lorentz force	$F_z \propto \sum (B_x^2 + B_y^2 - B_z^2) dA$	1371	Included
MEANPOT	Mean photospheric magnetic free energy	$\overline{\rho} \propto \frac{1}{N} \sum (\boldsymbol{B}^{Obs} - \boldsymbol{B}^{Pot})^2$	1064	Included
R_VALUE	Sum of flux near polarity inversion line	$\Phi = \sum \lvert B_{LoS} \rvert dA$ within R mask	1057	Included
EPSZ	Sum of z-component of normalized Lorentz force	$\delta F_z \propto \frac{\sum (B_x^2 + B_y^2 - B_z^2)}{\sum B^2}$	864.1	Included
SHRGT45	Fraction of Area with shear $> 45°$	Area with shear $> 45°$ / total area	740.8	Included
MEANSHR	Mean shear angle	$\overline{\Gamma} = \frac{1}{N} \sum \arccos \left(\frac{\boldsymbol{B}^{Obs} \cdot \boldsymbol{B}^{Pot}}{\lvert B^{Obs} \rvert \lvert B^{Pot} \rvert} \right)$	727.9	Discarded
MEANGAM	Mean angle of field from radial	$\overline{\gamma} = \frac{1}{N} \sum \arctan \left(\frac{B_h}{B_z} \right)$	573.3	Discarded
MEANGBT	Mean gradient of total field	$\overline{\lvert \nabla B_{tot} \rvert} = \frac{1}{N} \sum \sqrt{\left(\frac{\partial B}{\partial x} \right)^2 + \left(\frac{\partial B}{\partial y} \right)^2}$	192.3	Discarded
MEANGBZ	Mean gradient of vertical field	$\overline{\lvert \nabla B_z \rvert} = \frac{1}{N} \sum \sqrt{\left(\frac{\partial B_z}{\partial x} \right)^2 + \left(\frac{\partial B_z}{\partial y} \right)^2}$	88.40	Discarded
MEANGBH	Mean gradient of horizontal field	$\overline{\lvert \nabla B_h \rvert} = \frac{1}{N} \sum \sqrt{\left(\frac{\partial B_h}{\partial x} \right)^2 + \left(\frac{\partial B_h}{\partial y} \right)^2}$	79.40	Discarded
MEANJZH	Mean current helicity (B_z contribution)	$\overline{H_c} \propto \frac{1}{N} \sum B_z \cdot J_z$	46.73	Discarded
TOTFY	Sum of y-component of Lorentz force	$F_y \propto \sum B_y B_z dA$	28.92	Discarded
MEANJZD	Mean vertical current density	$\overline{J_z} \propto \frac{1}{N} \sum \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right)$	17.44	Discarded
MEANALP	Mean characteristic twist parameter, α	$\alpha_{total} \propto \frac{\sum J_z \cdot B_z}{\sum B_z^2}$	10.41	Discarded
TOTFX	Sum of x-component of Lorentz force	$F_x \propto -\sum B_x B_z dA$	6.147	Discarded
EPSY	Sum of y-component of normalized Lorentz force	$\delta F_y \propto \frac{-\sum B_y B_z}{\sum B^2}$	0.647	Discarded
EPSX	Sum of x-component of normalized Lorentz force	$\delta F_x \propto \frac{\sum B_x B_z}{\sum B^2}$	0.366	Discarded

In the solar-flare forecasting field, the two classes (non-flaring and flaring ARs) are strongly imbalanced: there are many more negative examples than positive ones, which reflects the fact that most ARs do not produce major flares in any given 24 or 48 hour period. **This class imbalance is a major issue for most ML algorithms**. Indeed, an ML classifier may strongly favor the majority class, and neglect the minority one. In other words, always predicting that an AR will not flare is likely to give very good results overall.

Li et al. (2007) and Yuan et al. (2010) used a soft margin SVM algorithm to forecast solar flares, demonstrating the fea- sibility of this approach. Here, we use the Scikit-Learn module implementation of **a soft margin SVM** in the Python programming language.

It is noteworthy that **using only the 4 highest-ranking parameters**— the total unsigned current helicity, total magnitude of the Lorentz force, total photospheric magnetic free energy density, and total unsigned flux— **gives roughly the same TSS score as the top 13 combined**.

Table 2
Flare Prediction Capabilities with 13 Features Compared to Other Studies, and with the SVM Parameters Selected to Achieve the Highest HSS_2

Metric	Segmented	Operational	Mason	Ahmed	Ahmed	Barnes	Bloomfield	Yu	Song
Time interval (no flare)	48h	24h	6h	48h	24h	24h	24h	48h	24h
Class-imbalance ratio	16.5	16.5	260	15.85	16.58	9.92	26.5	NA	2.23
Accuracy	0.973 ± 0.003	0.962 ± 0.004	0.694	0.975	0.963	0.922	0.830	0.825	0.873
Precision (positive)	0.797 ± 0.050	0.690 ± 0.049	0.008	0.877	0.740	NA	0.146	0.831	0.917
Precision (negative)	0.983 ± 0.003	0.978 ± 0.003	0.998	0.980	0.972	NA	NA	NA	0.860
Recall (positive)	0.714 ± 0.068	0.627 ± 0.049	0.617	0.677	0.523	NA	0.704	0.817	0.647
Recall (negative)	0.989 ± 0.003	0.983 ± 0.004	0.695	0.994	0.989	NA	NA	NA	0.974
f1 (positive)	0.751 ± 0.052	0.656 ± 0.035	0.015	0.764	0.613	NA	0.242	NA	0.758
f1 (negative)	0.986 ± 0.002	0.980 ± 0.002	0.819	0.987	0.989	NA	NA	NA	0.913
HSS_1	0.528 ± 0.062	0.342 ± 0.071	−78.9	0.581	0.339	0.153	NA	NA	0.588
HSS_2	0.737 ± 0.034	0.636 ± 0.037	0.008	0.751	0.594	NA	0.190	0.650	0.676
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
TSS	0.705 ± 0.067	0.610 ± 0.048	0.312	0.671	0.512	NA	0.539	0.650	0.620

Machine Learning in Space Weather Prediction

The Astrophysical Journal, 821:127 (7pp), 2016 April 20
doi:10.3847/0004-637X/821/2/127

PREDICTING CORONAL MASS EJECTIONS USING MACHINE LEARNING METHODS

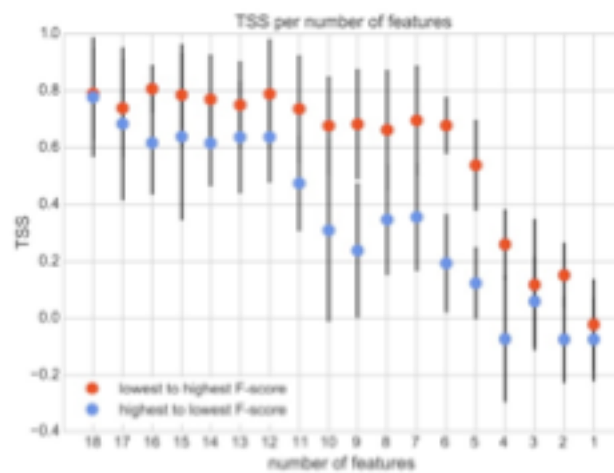
M. G. Bobra and S. Ilonidis
W.W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305, USA
Received 2015 December 14; accepted 2016 March 10; published 2016 April 21

Table 1
Features and F-scores

Keyword	Description	Formula	Scaling	Feature Ranking 24 hr	48 hr
MEANGBH	Mean gradient of horizontal field	$\overline{\lvert\nabla B_h\rvert} = \frac{1}{N}\sum\sqrt{\left(\frac{\partial B_h}{\partial x}\right)^2 + \left(\frac{\partial B_h}{\partial y}\right)^2}$	I	1	13
MEANJZH	Mean current helicity (B_z contribution)	$\overline{H_c} \propto \frac{1}{N}\sum B_z \cdot J_z$	I	2	2
MEANALP	Mean characteristic twist parameter, α	$\alpha_{total} \propto \frac{\sum J_z B_z}{\sum B_z^2}$	I	3	5
MEANGBT	Mean gradient of total field	$\overline{\lvert\nabla B_{tot}\rvert} = \frac{1}{N}\sum\sqrt{\left(\frac{\partial B}{\partial x}\right)^2 + \left(\frac{\partial B}{\partial y}\right)^2}$	I	4	18
MEANPOT	Mean photospheric magnetic free energy	$\overline{\rho} \propto \frac{1}{N}\sum (\mathbf{B}^{Obs} - \mathbf{B}^{Pot})^2$	I	5	12
MEANSHR	Mean shear angle	$\overline{\Gamma} = \frac{1}{N}\sum \arccos\left(\frac{\mathbf{B}^{Obs}\cdot\mathbf{B}^{Pot}}{\lvert\mathbf{B}^{Obs}\rvert\lvert\mathbf{B}^{Pot}\rvert}\right)$	I	6	9
SHRGT45	Fraction of Area with Shear >45°	Area with Shear >45°/Total Area	I	7	6
TOTPOT	Total photospheric magnetic free energy density	$\rho_{tot} \propto \sum (\mathbf{B}^{Obs} - \mathbf{B}^{Pot})^2 dA$	E	8	11
MEANJZD	Mean vertical current density	$\overline{J_z} \propto \frac{1}{N}\sum\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right)$	I	9	10
USFLUX	Total unsigned flux	$\Phi = \sum \lvert B_z\rvert dA$	E	10	3
MEANGAM	Mean angle of field from radial	$\overline{\gamma} = \frac{1}{N}\sum \arctan\left(\frac{B_h}{B_z}\right)$	I	11	14
TOTUSJZ	Total unsigned vertical current	$J_{z_{total}} = \sum \lvert J_z\rvert dA$	E	12	7
ABSNJZH	Absolute value of the net current helicity	$H_{c_{abs}} \propto \lvert\sum B_z \cdot J_z\rvert$	E	13	1
AREA_ACR	Area of strong field pixels in the active region	Area $= \sum$ Pixels	E	14	15
R_VALUE	Sum of flux near polarity inversion line	$\Phi = \sum \lvert B_{LoS}\rvert dA$ within R mask	E	15	17
TOTUSJH	Total unsigned current helicity	$H_{c_{total}} \propto \sum \lvert B_z \cdot J_z\rvert$	E	16	4
	Flare Class	$FC = CM$	neither	17	19
SAVNCPP	Sum of the modulus of the net current per polarity	$J_{z_{sum}} \propto \left\lvert\sum^{B_z^+} J_z dA\right\rvert + \left\lvert\sum^{B_z^-} J_z dA\right\rvert$	E	18	8
MEANGBZ	Mean gradient of vertical field	$\overline{\lvert\nabla B_z\rvert} = \frac{1}{N}\sum\sqrt{\left(\frac{\partial B_z}{\partial x}\right)^2 + \left(\frac{\partial B_z}{\partial y}\right)^2}$	I	19	16

…we use a (1) SVM and features derived from photospheric vector magnetic field data taken by the Solar Dynamics Observatory's (SDO) Helioseismic and Magnetic Imager (HMI) instrument **to forecast whether an active region that produces an M1.0-class flare or higher will also produce a CME**, and (2) a feature selection algorithm to determine which features distinguish these two populations.

Yashiro et al. (2005) showed that while more than 80% of X-class flares are associated with CMEs, this number drops as a power law with decreasing flare class. On average, about 60% of M-class flares produce a CME; even so, there is a great disparity across the M-class, where M1.0-class to M1.8-class flares are only ~44% likely to produce a CME. As such, we have **~6.5 times more events in the negative class than in the positive one**: our negative class includes 364 events, 230 of which are within the M1-range, whereas our positive class includes 56, where 7 are within the M1-range.



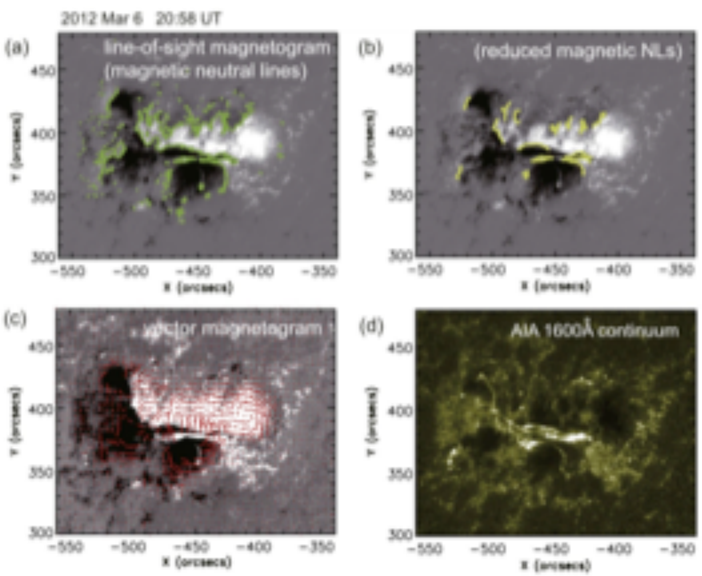
Grand Challenge
UNIVERSITY OF COLORADO BOULDER
SPACE WEATHER CENTER

Machine Learning in Space Weather Prediction

The Astrophysical Journal, 835:156 (16pp), 2017 February 1 doi:10.3847/1538-4357/835/2/156

Solar Flare Prediction Model with Three Machine-learning Algorithms using Ultraviolet Brightening and Vector Magnetograms

N. Nishizuka[1], K. Sugiura[2], Y. Kubo[1], M. Den[1], S. Watari[1], and M. Ishii[1]
[1] Applied Electromagnetic Research Institute, National Institute of Information and Communications Technology, 4-2-1, Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan; nishizuka.naoto@nict.go.jp
[2] Advanced Speech Translation Research and Development Promotion Center, National Institute of Information and Communications Technology, Japan
Received 2016 July 18; revised 2016 November 4; accepted 2016 November 4; published 2017 January 25



The procedures of our flare prediction model are as follows.

(i) observation data are downloaded from the web archives

(ii) ARs are detected from full-disk images of the line-of- sight magnetogram, and the ARs are tracked using their time evolution.

(iii) For each AR, features are calculated from multiwavelength observations, and **flare labels are attached to the solar feature database if an X/M-class flare occurs within 24hr after an image**.

(iv) Supervised machine learning is carried out with a 1 hr cadence to predict the maximum class of flares occurring in the following 24hr.

We used three machine-learning algorithms for comparison: the **SVM**, k-nearest neighbors (**k-NN**), and extremely randomized trees (**ERT**).

Table 4
The Prediction Results of X-class Flares and ≥M-class Flares, Neglecting Features of Previous Flare Activities

Algorithm	TP	FP	FN	TN	TSS
(a) X-class flares					
k-NN	136	16	15	54449	0.91 ± 0.02
SVM	130	23	21	54442	0.86 ± 0.02
ERT	87	4	49	54476	0.62 ± 0.03
(b) ≥M-class flares					
k-NN	1570	173	167	52706	0.904 ± 0.005
SVM	1501	759	236	52120	0.856 ± 0.009
ERT	1105	35	632	52844	0.63 ± 0.01

Note. The contingency tables of prediction results of X-class flares and ≥M-class flares, for the three machine-learning algorithms, k-NN, SVM, and ERT.

Machine Learning in Space Weather Prediction

The Astrophysical Journal, 843:104 (14pp), 2017 July 10 https://doi.org/10.3847/1538-4357/aa789b

Predicting Solar Flares Using *SDO*/HMI Vector Magnetic Data Products and the Random Forest Algorithm

Chang Liu[1,2,3], Na Deng[1,2,3], Jason T. L. Wang[4], and Haimin Wang[1,2,3]
[1] Space Weather Research Laboratory, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA
[2] Big Bear Solar Observatory, New Jersey Institute of Technology, 40386 North Shore Lane, Big Bear City, CA 92314-9672, USA
[3] Center for Solar-Terrestrial Research, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA
[4] Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA
Received 2016 May 6; revised 2017 June 7; accepted 2017 June 7; published 2017 July 11

Table 1
Overview of AR Samples Using *SDO*/HMI Magnetic Parameters

SHARP Keyword	Formula	Unit	RF Importance	B Class ($n = 128$)	C Class ($n = 552$)	M Class ($n = 142$)	X Class ($n = 23$)
TOTUSJH	$H_{total} \propto \sum \lvert B_z \cdot J_z \rvert$	$10^2\ G^2\ m^{-1}$	37.4	4.8 ± 3.1	13.9 ± 9.9	27.7 ± 18.2	58.3 ± 40.0
TOTBSQ	$F \propto \sum B^2$	$10^{10}\ G^2$	17.9	1.0 ± 0.9	2.6 ± 1.9	4.6 ± 3.0	10.7 ± 8.6
TOTPOT	$\rho_{tot} \propto \sum (B^{Obs} - B^{Pot})^2 dA$	10^{23} ergs cm^{-1}	21.1	1.0 ± 1.4	2.7 ± 2.7	6.7 ± 5.7	19.6 ± 18.0
TOTUSJZ	$J_{total} = \sum \lvert J_z \rvert dA$	10^{12} A	50.6	9.5 ± 6.4	30.3 ± 21.4	53.9 ± 30.9	110.0 ± 73.4
ABSNJZH	$H_{abs} \propto \lvert \sum B_z \cdot J_z \rvert$	$10\ G^2\ m^{-1}$	19.9	6.1 ± 7.0	14.3 ± 17.0	39.2 ± 43.8	91.2 ± 63.6
SAVNCPP	$J_{sum} \propto \lvert \sum^{B_z^+} J_z dA \rvert + \lvert \sum^{B_z^-} J_z dA \rvert$	10^{12} A	24.6	2.7 ± 2.7	6.5 ± 6.4	15.8 ± 14.6	33.1 ± 24.0
USFLUX	$\Phi = \sum \lvert B_z \rvert dA$	10^{21} Mx	14.2	7.1 ± 5.5	19.9 ± 14.7	33.7 ± 21.0	72.2 ± 54.2
AREA_ACR	Area = $\sum$ Pixels	10^2 pixels	23.7	3.0 ± 2.4	8.2 ± 6.1	13.3 ± 7.7	29.2 ± 22.3
TOTFZ	$F_z \propto \sum (B_x^2 + B_y^2 - B_z^2) dA$	-10^{23} dyne	13.9	1.2 ± 1.3	2.7 ± 2.7	3.9 ± 3.7	6.1 ± 6.2
MEANPOT	$\bar{\rho} \propto \frac{1}{N} \sum (B^{Obs} - B^{Pot})^2$	10^3 ergs cm^{-3}	19.8	6.5 ± 5.8	5.9 ± 3.7	8.9 ± 4.2	12.1 ± 4.2
*R*_VALUE	$\Phi = \sum \lvert B_{LoS} \rvert dA$ within *R* mask	Mx	31.4	3.2 ± 0.7	3.8 ± 0.6	4.4 ± 0.5	4.9 ± 0.4
EPSZ	$\delta F_z \propto \frac{\sum (B_x^2 + B_y^2 - B_z^2)}{\sum B^2}$	-10^{-1}	15.4	2.1 ± 1.3	2.0 ± 1.3	1.7 ± 1.2	1.2 ± 1.1
SHRGT45	Area with shear >45°/Total Area	—	12.7	0.23 ± 0.17	0.27 ± 0.14	0.34 ± 0.13	0.40 ± 0.11

Notes.

13 features identified by Bobra & Couvidet (2015)

We resort to Random Forest, an inherent multiclass classifier, to perform flare prediction. RF is a general term for the random decision forests, an ensemble learning technique mainly for classifica- tion and regression tasks (Breiman 2001).

Table 4
RF Binary-class Flare Prediction Results (within 24 hr) Using 13 *SDO*/HMI Parameters and Comparison to Other Studies

Prediction ↓ Observation→	B/C Class ($n = 165$)	M/X Class ($n = 165$)
B/C Class	129.536 ± 4.025	41.722 ± 3.370
M/X Class	35.464 ± 4.025	123.278 ± 3.370
Recall: This work	0.785 ± 0.036	0.747 ± 0.030
Bloomfield et al. (2012)	N/A	0.704
Ahmed et al. (2013)	N/A	0.523
Bobra & Couvidat (2015)	N/A	0.832 ± 0.042
Nishizuka et al. (2017)	N/A	0.716
Precision: This work	0.756 ± 0.033	0.777 ± 0.033
Bloomfield et al. (2012)	N/A	0.146
Ahmed et al. (2013)	N/A	0.740
Bobra & Couvidat (2015)	N/A	0.417 ± 0.037
Nishizuka et al. (2017)	N/A	0.969
Accuracy: This work	0.766 ± 0.023	0.766 ± 0.021
Bloomfield et al. (2012)	N/A	0.830
Ahmed et al. (2013)	N/A	0.963
Bobra & Couvidat (2015)	N/A	0.924 ± 0.007
Nishizuka et al. (2017)	N/A	0.990
TSS: This work	0.532 ± 0.036	0.532 ± 0.030
Bloomfield et al. (2012)	N/A	0.539
Ahmed et al. (2013)	N/A	0.512
Bobra & Couvidat (2015)	N/A	0.761 ± 0.039
Nishizuka et al. (2017)	N/A	0.71 ± 0.002

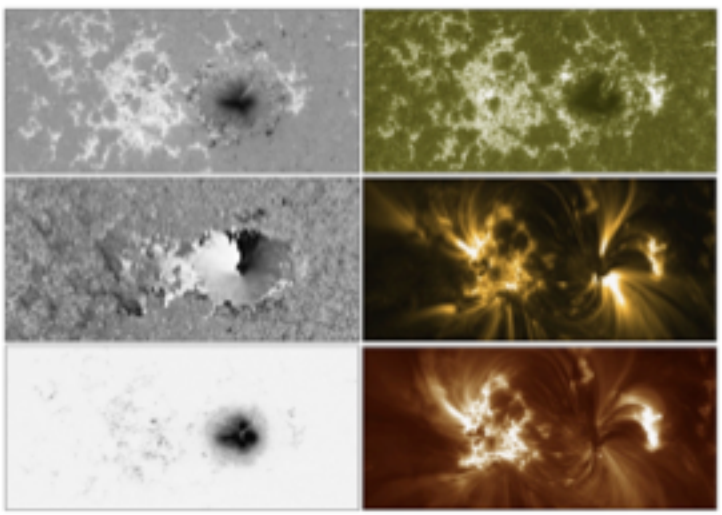
Grand Challenge
University of Colorado Boulder
Space Weather Center

Machine Learning in Space Weather Prediction

Solar Phys (2018) 293:48
https://doi.org/10.1007/s11207-018-1258-9

Flare Prediction Using Photospheric and Coronal Image Data

Eric Jonas[1] · Monica Bobra[2] · Vaishaal Shankar[1] ·
J. Todd Hoeksema[2] · Benjamin Recht[1]

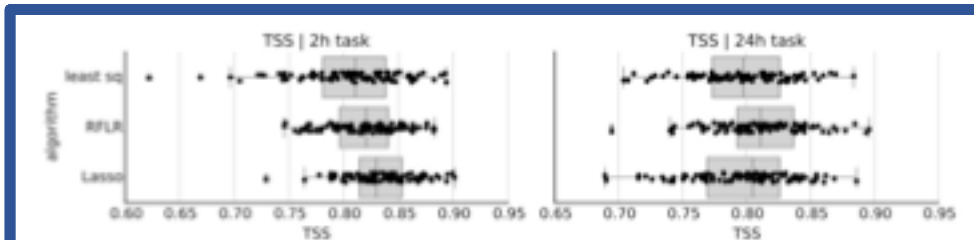


Such an approach may be useful since, at the present time, there are no physical models of flares available for real-time prediction.

Our goal is to use past observations of an active region to predict its future flaring activity. We choose to model our problem as a **binary-classification task**: Will this active region produce an M- or X-class flare within the next T hours? For this study, we chose two values for T: 2 and 24.

There are many different metrics to assess the performance of a classification algorithm. These metrics are defined using four quantities: false positives (FPs), false negatives (FNs), true positives (TPs), and true negatives (TNs).

We focus on **linear classifiers**, which model the output $y_i \in \{-1, 1\}$ as a linear function of the input features…



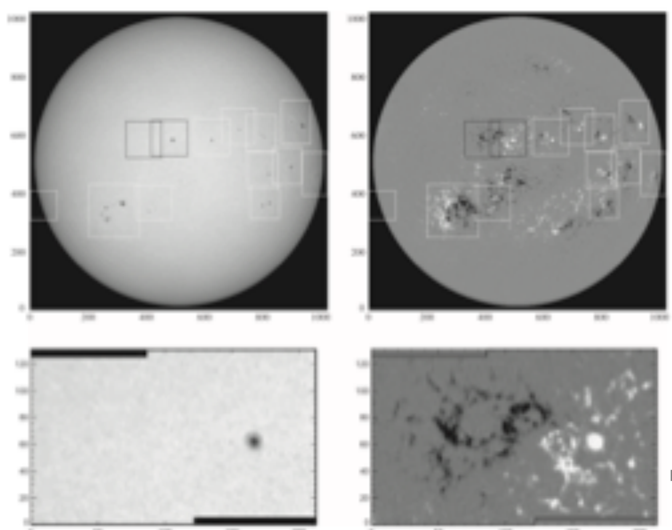
How to Really Compare Flare Predictions

The Astrophysical Journal, 829:89 (32pp), 2016 October 1 doi:10.3847/0004-637X/829/2/89

A COMPARISON OF FLARE FORECASTING METHODS. I. RESULTS FROM THE "ALL-CLEAR" WORKSHOP

G. Barnes[1], K. D. Leka[1], C. J. Schrijver[2], T. Colak[3], R. Qahwaji[3], O. W. Ashamari[3], Y. Yuan[4], J. Zhang[5], R. T. J. McAteer[6], D. S. Bloomfield[7,14], P. A. Higgins[7], P. T. Gallagher[7], D. A. Falconer[8,9,10], M. K. Georgoulis[11], M. S. Wheatland[12], C. Balch[13], T. Dunn[1], and E. L. Wagner[1]

Received 2016 May 10; revised 2016 June 15; accepted 2016 July 5; published 2016 September 26



The focus of the workshop was on "**all-clear**" forecasts, namely **predicting time intervals during which no flares occur that are over a given intensity** (as measured using the peak GOES 1–8 Å flux). For users of these forecasts, it can be useful to know when no event will occur because the cost of a missed event is much higher than the cost of a false alarm.

The data prepared and made available for the workshop participants constitute the basic level of data that was usable for the majority of methods compared. Some methods could make use of more sophisticated data or time series or a different wavelength, but **the goal for this particular comparison is to provide all methods with the same data**, so the only differences are in the methods, not in the input data.

The database prepared for the workshop is comprised of line-of-sight magnetic field data from the newest MDI calibration (Level 1.8) for the years 2000–2005 inclusive.

The Future

Major Challenges:

- Solar Eruption prediction (not “flares” *or* “CMEs” but eruptions that lead to flares *and* CMEs)
- Ionospheric scintillation prediction from GNSS data streams
- Geomagnetic storm intensity prediction based on L1 incoming data stream
- ??

Solar eruption prediction necessarily involves spatial and temporal pattern recognition. Therefore any successful system will be both a CNN capable of finding remote spatial correlations *and* an RNN capable of finding precursor signals **– *if they exist at all in current SDO/AIA and HMI data.***

For example: DL fails to predict the stock market because it is a fully stochastic system with insufficient prior information available in current datasets, i.e. *no discernible precursors*.

TREC strategy:

- CNN + RNN + Reinforcement Learning
- Data: all SDO AIA channels and HMI magnetogram (and dopplergram?) features at 45-720 sec cadence.

Conclusions

Deep Learning systems **are**

(a) tensor data transformation engines trained on **very** large datasets – *no physics involved*.
(b) outside any structured programming paradigm: data → model → pattern/prediction.
(c) characterized by **millions of parameters** that cannot be interpreted as physical quantities.

DL systems are **not** a form of **empirical physical models** (EPMs). EPMs are

(a) based on the laws of physics.
(b) within the structured programming paradigm: model + data → prediction.
(c) characterized by a few physical parameters ("NOx reaction rate", "Flux tube expansion", etc.).

DL systems will also **not** entirely replace **full physics models** (FPMs) since they offer little physical insight.

DL systems can be thought of as a form of **Super Observer** capable of finding patterns in data that no human could ever detect. By detecting these patterns (or failing to…), these Super Observers can inform future directions in FPM and EPM efforts.

Combined with Bayesian Probability Theory, DL networks will be capable of predicting Space Weather events that EPMs or FPMs cannot yet address because they are inadequate for predictive applications, e.g., Solar Eruptions.

Resources

General Machine Learning Books:
- *Introduction to Machine Learning with Python*, Andreas C. Müller & Sarah Guido, O'Reilly 2016.
- *Deep Learning with Python*, François Chollet, Manning 2018
- *Deep Learning*, Ian Goodfellow, Yoshua Bengio, & Aaron Courville, MIT Press 2018.
- *Machine Learning with Tensorflow*, Nishant Shukla & Kenneth Friklas, Manning 2018.

History of computing:
- Turing's Cathedral: the Origins of the Digital Universe, George Dyson, Vintage 2012.

Bayesian Probability Theory:
- *Probability Theory, the Logic of Science*, E. T. Jaynes, Cambridge, 2003.

Coursera:
- University of Michigan Data Science in Python Series: *Course 3 - Applied Machine Learning*
- Deep Learning, Andrew Ng (www.deeplearning.ai)

Papers online:
- Proc. Intl. Conf. on Learning Representations (ICLR)
- Conf. on Neural Information Processing Systems (NIPS)
- Many (most?) deep learning publications are published exclusively on arXiv

Space Weather:
- *Machine Learning Techniques for Space Weather*, E. Camporeale, S. Wing & J. Johnson eds., Elsevier 2018.

Resources (cont.)

Key Papers in Deep Learning:

- Original backprop paper: Rumelhart, Hinton, & Williams, Nature **323**, 533, 1986
- *Deep Learning*, Yann Lecun, Nature **521**, 436, 2015.
-

Other Papers in Applied Deep Learning:

- *DeepVel: Deep Learning for the estimation of horizontal velocities at the solar surface*, A. Asensio Ramos et al., A & A, **604**, A11, 2017.
- *Enhancing SDO/HMI Images Using Deep Learning*, C. J. Diaz Baso & A. A. Ramos, http://hmi.stanford.edu/hminuggets/?p=2552, 2018.
- *Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets*, Yunjie Liu et al., arXiv:1605.01156 2016.
- *Enabling large-scale viscoelastic calculations via neural network acceleration*, P. M. R. DeVries et al., GRL, **44**, 2017.

Software

- XGBoost library (gradient boosting machines)
- Scikit Learn
- Tensorflow
- Theano
- Keras (API that runs on top of TensorFlow or Theano)