Literature Review of Deep Learning Applications in Astronomical Data Analysis

Engler, S.T., University of Hawaii, Department of Information and Computer Science, 12th November 2019

1. Introduction

Astronomy is undergoing a paradigm shift as the vast data collected using ground and space telescopes are now significantly larger than ever seen. Consequently, deep learning (DL) is beginning to have an impact in astronomy and astrophysics through the analysis of these large datasets. DL is categorized into two main types in astronomical data processing. The first is image processing through convolutional neural networks (CNNs), which has several astronomical applications, such as galaxy age categorization and stellar spectral classification. The second category is time series analysis, which is applied in eclipsing exoplanet identification through space telescope data, such as Kepler, TESS, and ground-based observations using the measurements of radial velocity extracted from spectral analysis. This document is based on the evaluation of parameters in systematic literature review [1], which is outlined in section 3. In this study, X papers were retrieved by manual search in four databases which are Google Scholar, University Hawaii Manoa's 'OneSearch', and Cornell University's 'arXiv'. Using keywords of 'deep learning', with 'astronomy' or 'astrophysics' only 65 journal articles contained both key words. From there, eight primary studies were finally included selected to represent different research subsets of astronomy. The primary studies are 88% journal articles, and 22% are conference proceedings. This literature review fills a gap in the literature in its scope of focusing specifically on DL as applied to astronomy and astrophysics.

Keywords: Deep Learning, Astronomy, Astrophysics, Literature review

2. Deep Learning

Deep learning is a subset of machine learning, which is a subset of artificial intelligence. DL involves the use of techniques to learn multiple levels of representation within a very large dataset. Predominantly, when people speak of DL, they are speaking about deep artificial neural networks rather than machine learning or artificial intelligence. [2] DL models require large-scale data sources in which traditional neural networks (NN) would encounter overfitting. Overfitting is when the NN outputs the same solution due to an over-training of the system. [3] DL has achieved outstanding successes in almost every scientific venture it has been applied to, and it is categorized into three different general approaches: supervised, reinforcement, and unsupervised learning. [2] Supervised learning uses labelled data that compare a loss value to a training data and iteratively modify the neural network parameters for better results. The types of supervised learning are deep neural network (DNN), CNN, and recurrent neural networks (RNN), which includes the long-term short memory (LSTM) and gated recurrent units (GRU). [2] Reinforcement learning (RL) is based on partially labelled datasets. The challenge with RL is that there is no solution to calculate a loss function from. If there are few enough parameters to tune, then parameter optimization can be performed through methods, such as cross entropy calculations and simulated annealing. [2] Unsupervised learning (UL) has no data labels, learns the important features of the dataset, and discovers structures within the data. [2] Google created "Deep Mind," which is the first unsupervised learning system. [4], [5]

Deep learning is built on neural networks, and the "deep" aspect did not appear until 2006.

[6] Initially, it was shown in 1943 that a perceptron could be used to solve such problems involving XOR logic. Subsequently, it was shown that a Turing machine could be created using neurons.[7] The building blocks of neural nets are the perceptron, which is a single neuron. The perceptron produced a solution provided one was able to guide towards a solution state. However, the perceptron was unable to solve XOR logical problems; hence, the research was stopped. In 1985, the backpropagation algorithm was then introduced, which created more usable NNs. [8]

After the introduction of back propagation, CNNs appeared in 1998 [7]. Subsequently, DNNs were introduced,[9] and the first full DNN appeared in 2012 [10]. The primary components of NNs are neurons, which are governed by activation functions (predominantly a sigmoid function). A NN is connected in multiple layers, with at least one hidden layer. As the NN performs computation, there are weights within the connections between the neurons, and the weights are found by iterating through the NN and making comparisons with the objective function (solution). The weights are adjusted using gradient decent, which is an optimization algorithm. Once a local minimum is found, the NN will use back propagation to push the new information back through the NN. For DNN, a stochastic gradient decent is often used to prevent the system from getting stuck in a local minimum. [11] The learning rate is the step size of the training. It can speed up the training process, cause the system to diverge if the steps are too large, or cause the system to converge quickly if the steps are of the right size. As the NN starts to converge, the learning rate often reduces at a constant or exponential decay. [2]

2.1 Convolutional Neural Network

Fukushima was the first person to propose a CNN, which was able to recognize hand written numbers. [12] A CNN in its simplest form consists of a feature extractor and classification. It starts with a convolution, max-pooling, and classification layers. [3] A CNN takes advantage of the image inputs and allows the architecture to constrain it in an efficient way. When input is passed to the convolution layer, the output is an activation map that extracts features from the input data to pass further. The convolution layer consists of feature maps going through kernels that are passed into activation functions (such as the sigmoid, SoftMax, or rectified linear), which create feature maps. These feature maps can then be combined with multiple input feature maps. [2] The max pooling layer allows the network to process shape variations. This layer prevents the NN from having too many parameters to control, [3] allows the CNN to be less sensitive to the gradient problem, and produces a highly accurate and efficient NN. [3] Pooling layers further reduce the number of parameters. [3] The max-pooling layer is a form of subsampling, and it down-samples the output from the convolution layer. Each output layer is down-sampled into half the size of its incoming dimension. The down sample is obtained by averaging the neighboring values or taking the maximum value. [2] The classification layer scores the features extracted from the convolution layer. This information is backpropagated through the CNN with the convolutional operation on the filters and their previous layer. [2] The classification layer is often a feed-forward NN. [9] [13] The output of the CNN is a fully connected layer whose output is compared to the output layer for errors, and a mean square loss is found and backpropagated through the system. [3] CNNs are the first architecture to successfully solve into a deep network with multiple hierarchical layers. The number of parameters of the network are significantly reduced through the CNN topology by the sharing of parameters. This parameter sharing consequently creates equivariance in the system. [14]

A typical CNN for object detection in images is shown in Figure 1, and it is an important tool in several astronomical applications. The CNN consists of four layers. The image itself is the network input and is broken up into convolutional layers. [15] The convolutional layer is isolated from the previous layer to attach more weight to the localized structure. [16] The first layer extracts basic features, which are then combined by higher layers to create new features. Ultimately, a feature map is created, which acts as a filter to extract similar features within the images. Once the convolutional layer extracts the feature map, it enters into a pooling layer. The pooling layer extracts important features from the convolutional layer. This second set of feature maps is then passed through another convolutional layer to place the features within a fully connected layer, which consists of the entire image. In this manner, the feature maps are matched with identical items in the full image.

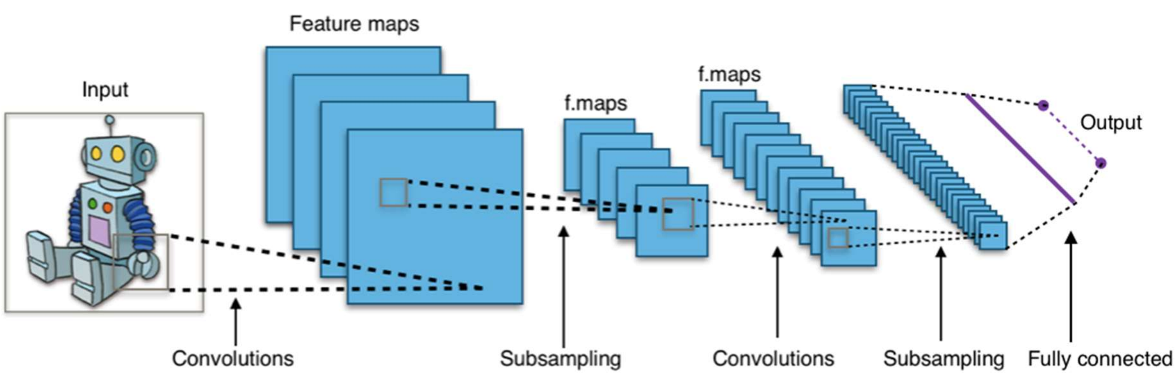

Figure 1: CNN. Source:

(https://en.wikipedia.org/wiki/Convolutional_neural_network)

2.2 Deep Recurrent Neural Network

The recurrent neural network is a NN where the output from a previous step is fed as input to the current step. The main and most important feature of an RNN is the hidden state, which remembers some information about a sequence. An RNN remembers every information through time and is easily applied to time-series data. Another form of RNNs is the long-short term memory RNN (LSTM-RNN), which passes its results to another copy of the network.

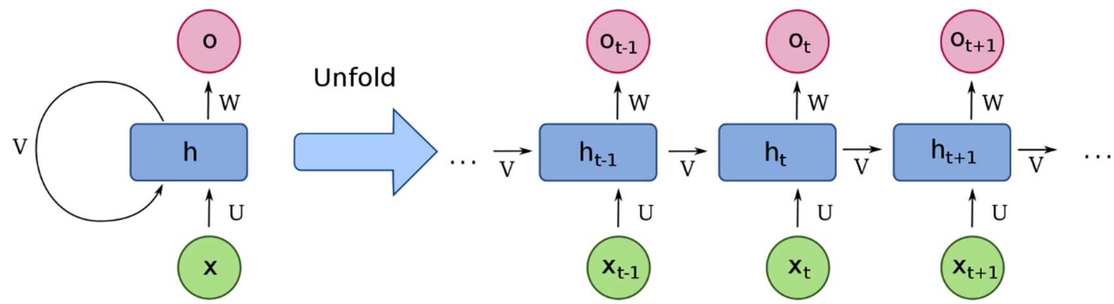

Figure 2: RNN Source:

(https://en.wikipedia.org/wiki/Recurrent_neural_network)

The network can be a long string of time series data with outputs of equal length. [17] RNNs by themselves are limited to the dependencies between each cell, which is called the long-term dependency problem. To overcome this, an LSTM cell is introduced to retain information indefinitely. [18] This is accomplished by using three different gates, which will let the cell read or write information according to parameters that open or close these gates. They are cells with logistic and linear units, where information gets into the cell when the "write" gate is open. The information stays in memory until the "keep" gate is opened. The "read" gate gives access. It is also possible to back-propagate through the memory cells. Each gate is a layer of neural nets governed by a sigmoid function. [19]

The LSTM is an architecture designed as an improvement on the RNN with the introduction of simple memory cells with a gating architecture. For each memory cell, the network computes the output of four gates: an update, input, forget, and output gates, as seen in Figure 3.

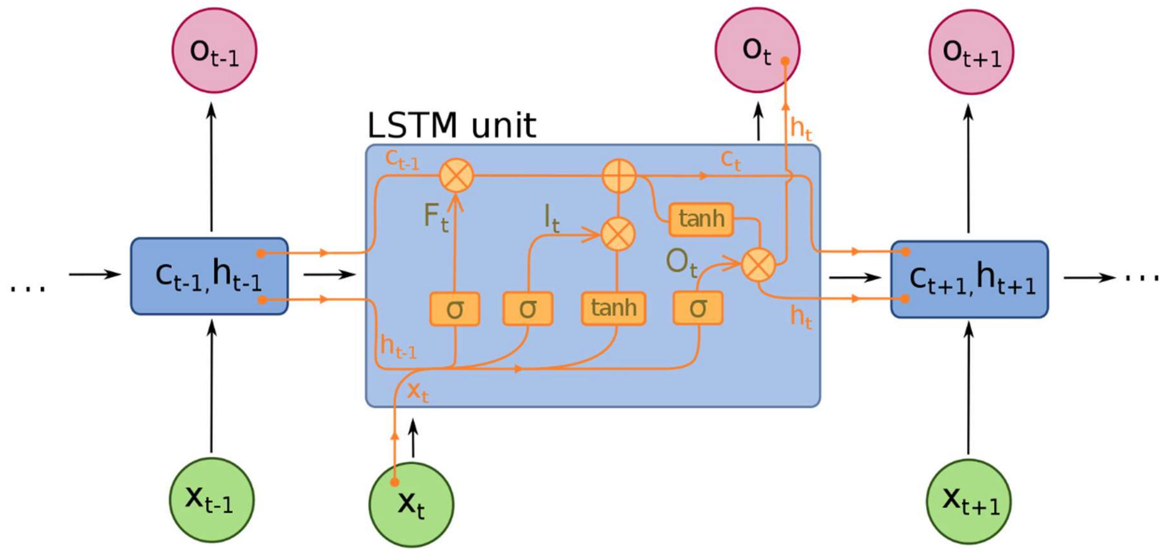

Figure 3: LSTM Cell

(https://en.wikipedia.org/wiki/Recurrent_neural_network)

3. Deep Learning Research in Astronomy and Astrophysics

A literature review provided by [1] outlines a way to analyze the effectiveness and quality of deep learning models using five features of the DL architecture:

1) Hierarchical layer – DL has multiple learning layers and is designed to extract low-level features to high-level features in a hierarchical structure. DL research should employ this carefully in classification applications.
2) High-level abstraction – Enables the extraction of data in unsupervised learning without

external input. The architectures should represent the algorithms through the composition of non-linear transformations.

3) Processing of high volume of data – How well does the DL method utilize large quantities of raw data for processing into results, and what is the amount of computer resources required?
4) Robust model – A useful measure of a DL learning model is its ability to be applied to different types of problems.
5) Overfitting – DL algorithms should not overfit the training dataset.

Some research articles on DL algorithms in the field of astronomy and astrophysics are reviewed using these five features.

New generation telescopes, such as the Atacama large millimeter array (ALMA), very large telescope (VLA), and sky survey telescope will produce enormous data that will require automatic processing in the astronomy community. It is not enough to only process the data, but the data needs to be capable of testing scientific questions based on the existing dataset. One research area in astronomy is the classification of astronomical sources. Classification tasks are well suited to CNN; however, there are challenges in dealing with bias, misclassified data, and portability of the CNN to other datasets it is not trained on.

Unsupervised classification relies on the statistical properties of objects while supervised classification is inherently biased through inconsistent human assessment. In the literature, [20] variable stars and exoplanet transits are classified using a recurrent-CNN, which uses the raw data to extract features. This enables data analysis directly from the raw data without the typically tedious data preparation conducted through human supervision. This study also simulates synthetic image sequences based on instrumentation and observations to adapt the training of the recurrent-CNN. Using this simulation for training allows the system to be fine-tuned to the data from the actual experiment using only a small number of real data points.

Table 1: DL methods typically applied to astronomical data.

	Convolutional Neural Network	**Recurrent Neural Network**	**Long-Short Term Memory (NN or RNN)**
Classification	• Stellar spectral type • Stellar photometric • Galactic morphological • Solar activity • Wide field detection	• Stellar photometric • Solar activity	• Stellar photometric
Regression		• Photometric redshifts • Spectral line parameter measurements	
Clustering	• Stellar spectral classification • Galaxy classification		
Time-Series Analysis		• Photometric planet detection • Variable star photometry	• Spectral planetary detection • Novel detection • Extrasolar planet detection

Table 2: Feature evaluation for deep learning for image sequence classification of astronomical events.[20]

DL Features	**Deep Learning for Image Sequence Classification of Astronomical Events**
Hierarchical layer	This study used a recurrent-CNN architecture with four layers of 21×21×64 data cubes followed by four layers of 11×11×64 and one layer of 6×6×64, then one layer of size 2304 followed by a 1024 layer. This is fed into an LSTM RNN of size 512. It then outputs data of size 7 for the classification. The architecture in this study is unusual as there are four identical layers. A good CNN should have multiple layers that are smaller than the previous layers to take advantage of the features. This model likely does not pick up nearly as many features as it could.
High-level abstraction	The data used in this study have a high level of abstraction and are suitable for this type of DL model.
Data volume processing	The volume of data for this DL architecture is questionable owing to the hierarchical layers of unusual sizes and the LSTM at the end of the system. It is likely a very slow model, but this is not discussed in the research article.
Robust model	This model should be fairly universal.
Overfitting	It is unlikely that this model would overfit due to the structure of the model and size of the dataset.

Unsupervised classification is seen in several papers based on the structural features of astronomical objects, such as galaxies. In [21] a CNN is used to classify galaxies into structural categories such as spheroid, irregular, point source, or unclassifiable using a Hubble super novae survey. Complexities in the research arose with the accuracy of category labels used in the training process of the CNN as the labels were selected by humans with a bias of personal conviction of the category the galaxy belonged to. To reduce this bias, the authors used five experts to classify 100 randomly selected galaxies to help fine-tune the CNN training.

Table 3: Feature evaluation for multiband galaxy morphologies for CLASH: A convolutional neural network transferred from CANDELS. [21]

DL Features	**Multiband Galaxy Morphologies for CLASH: A Convolutional Neural Network Transferred from CANDELS**
Hierarchical layer	The inception program uses a convolutional method that creates a computationally efficient program. For example, instead of creating a 5×5 grid with a filter, it creates two 3×3 grid with two filters, which is computationally more efficient. Ultimately, the program created a very large hierarchal structure through Inception and ConvNet code.
High-level abstraction	The data used in this study have a high level of abstraction and are suitable for this type of DL model.
Data volume processing	In theory, the inception code should create a highly efficient hierarchy, which would enable an effective data volume processing.

Robust model	This model uses considerable specialized data in the code that is specific for the application. It is likely difficult to move these data into another set.
Overfitting	Overfitting is not a problem with this study.

A similar study by [22] classified galaxies through an online group of citizen-scientists as a form of creating a large categorized dataset. The researchers used this dataset to train and compared with real data using a quantile regression forest that provided a matrix of similarities between the image and training data. Another study [23] in unsupervised classification of galaxies was conducted using a large-scale dataset of over 30,000 images of galaxies for deep k-means clustering and combining the data with feature selection. Interestingly, the authors used only the observed data from photometer and spectroscopy, and they used no non-observational data. This approach was successful in identifying clusters of galaxies using different star formation histories. This study showed that DL was successful in categorizing data with strictly physical data and alleviated label bias in categorical applications.

Table 4: Feature evaluation for unsupervised classification of galaxies. I. Independent component analysis feature selection. [23]

DL Features	**Unsupervised Classification of Galaxies. I. Independent component analysis feature selection**
Hierarchical layer	Here, independent component analysis (ICA) is used to determine a set of derived independent components of the galaxies. K-means cluster analysis is then applied on nine independent components to

	obtain 10 distinct and homogeneous groups.
High-level abstraction	There is no high-level abstraction within k-means analysis.
Data volume processing	This approach has limited data processing owing to the limited computational processing of ICA.
Robust model	Once the ICA is conducted, the model is specific to the observational dataset. Hence, it is not universal.
Overfitting	There is some risk of overfitting.

Table 5: Feature evaluation for Identifying complex sources in large astronomical datasets using a coarse-grained complexity measure. [24]

DL Features	**Identifying Complex Sources in Large Astronomical Datasets Using a Coarse-grained Complexity Measure**
Hierarchical layer	A filter was applied to a 10×10 pixels and calibration with 256×256 radio image and 8-bit pixel intensity values. A 64×64 center of the image is cropped and filtered with window size *h*=1-.
High-level abstraction	There is very little high-level abstraction in this method.
Data volume processing	Data volume processing is good, but the data preparation required is excessive.
Robust model	There is little robustness in the model.
Overfitting	There is risk of overfitting depending on the calibration method.

Deep CNN has also been applied to astrophysical simulations of stellar asteroseismology data. [25] It enabled the group to interpolate between different stages of stellar evolution within the complex evolutionary models. The space telescopes, Kepler and TESS, which are both photometric measurements of stars over a large section of the sky, are predominantly time-series datasets extracted from telescopes hunting for exoplanets. There is an extremely large amount of data that requires the detection of transit events within these datasets.

Table 6: Feature evaluation for Deep learning applied to the asteroseismic modeling of stars with coherent oscillation modes. [25]

DL Features	**Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes**
Hierarchical layer	This study used a fully connected dense network with a 6-dimensional input vector, a 1-dimensional output vector, and two *n*-dimensional layers in-between. The layers have 6, *n*, *n*, and one neurons.
High-level abstraction	The architecture is used to train and predict oscillation frequencies within stellar models. The hierarchical structure is easy to modify for various other applications.
Data volume processing	Data are processed at very high volumes.
Universal model	The model can be universally applied.
Overfitting	Overfitting will be governed by the amount of data available.

One of the key points of interest in the hunt for exoplanets is the hunt for earth-like planets, and the detection of such transits is the presence of red (correlated) noises in the light curves. Researchers in [26] simulated the red noise through the application of hyperparameters and found that the noise can be filtered by RNN to extract data for earth-like planets. Research conducted by [27] expanded an exoplanet DNN that had a 95% accuracy. The work illustrates the importance of including expert knowledge in state-of-the-art DL models when applying them to scientific research problems that seek to identify weak signals in noisy data. This research paved the way for DL to be employed as a major tool in astronomy and the hunt for exoplanets.

Table 7: Feature evaluation for Shallow transits—deep learning.

I. Feasibility study of deep learning to detect periodic transits of exoplanets. [26]

DL Features	**Shallow Transits—Deep Learning. I. Feasibility Study of Deep Learning to Detect Periodic Transits of Exoplanets**
Hierarchical layer	This project focused on a CNN with an input of 6144×2, which breaks into four hidden layers connected by 1527×40 followed by another four layers of size 377×32, 89×32, 17×32, and finally an output of 1×2.
High-level abstraction	The larger receptive field enables more global decisions at the earlier layers of the network.
Data volume processing	The data volume in this architecture should be sufficient.
Universal model	The model is fairly universal to images as input.
Overfitting	The object is not subject to overfitting.

Table 8: Feature evaluation for Identifying exoplanets with deep learning: A five-planet resonant chain around Kepler-80 and an eighth planet around Kepler-90. [28]

DL Features	**Identifying Exoplanets with Deep Learning: A Five-planet**

	Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90
Hierarchical layer	This CNN is broken into five multiple layers: a single input of 6144×1, four hidden 1527×40, 277×32, 89×2, and 17×32, as well as an output of 1×2.
High-level abstraction	The larger receptive field enables more global decisions at the earlier layers of the network.
Data volume processing	The data volume in this architecture should be sufficient.
Universal model	The model is fairly universal to images as input.
Overfitting	The object is not subject to overfitting.

In cosmology, deep learning has been applied in research by [29], which developed a pixel-based approach to implement convolutional and pooling layers on the spherical surface estimation of cosmological parameters from simulated maps of the cosmic microwave background (CMB). This showed the applicability of the CNN to CMB parameter estimation using a simple NN architecture consisting of four convolutional and pooling layers. The researchers were able to sample value temperature and polarization parameters directly from simulated maps using this method, which achieved better accuracy than simulations using Bayesian methods.

Table 9: Feature evaluation for Convolutional neural networks on the HEALPix sphere: a pixel-based algorithm and its application to CMB data analysis. [29]

DL Features	**Convolutional neural networks on the HEALPix sphere: a pixel- based algorithm and its application to CMB data analysis**

	[30]
Hierarchical layer	This group developed a pixel-based approach to implement convolutional on the spherical surface using a fully connected NN. The filter size swept across the image surface starts at 16 and reduces to 8, 4, 2, and 1, where it then hits a dropout layer followed by a fully connected 48-neuron layer. This is passed to an activation function and finally an output layer.
High-level abstraction	The design has the ability to be adjusted to almost any image classification problem.
Data volume processing	High-volume data can be processed on this system.
Universal model	The model is highly universal.
Overfitting	The model does not overfit.

4. Discussion

This literature review examined the origins of DL at the basic level. CNN and RNN were defined in the LSTM form. The review further examined where DL has been influencing astronomy. We found that DL is still at an early stage of various applications in astronomy. Further, we found that astronomers are considering DL as the main tool for processing data from the next generation of telescopes. The literature on DL in astronomy applications highlights transfer learning as a major concern for future research. [30] Transfer learning is the ability to take data from one type of data and apply it to another type. For example, taking photometric data from TESS and applying it to spectroscopic data, the authors [31] used a new approach to the time domain adaptation of the problem. The literature also reveals additional problems within the application of DL to astronomy. Learning algorithms are not fine-tuned to work with astronomical datasets as there is variation in the quality of astronomical data, and currently, there is no straightforward way to assess this within the DL algorithms. Specifically, astronomical measurements vary in terms of signal-to-noise ratios, measurement uncertainties, and training simulations having modelling uncertainties. The construction of classifiers that are able to handle situations, such as taking a ML model trained on one dataset and applying it to another, is of a high priority. Additionally, the new telescopes will be required to train as the data are being collected. The literature in [32] also calls for deep CNN that can present data in scientifically interpretable ways and are physically meaningful within the astrophysical models being tested.

References

[1] N. F. Hordri, A. Samar, S. S. Yuhaniz, and S. M. Shamsuddin, "A systematic literature review on features of deep learning in big data analytics," *International Journal of Advances in Soft Computing and its Applications*, vol. 9, no. 1, pp. 32–49, 2017.

[2] P. Hála, "Spectral classification using convolutional neural networks," Dec. 2014.

[3] M. Z. Alom *et al.*, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," *CoRR*, vol. abs/1803.0, Mar. 2018.

[4] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[5] I. Engineering and J. B. Raja, "SURVEY ON DEEP LEARNING ALGORITHMS," vol. 5, no. 1, pp. 38–43, 2019.

[6] V. Mnih *et al.*, “Playing Atari with Deep Reinforcement Learning,” in *NIPS Deep Learning Workshop 2013*, 2013.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, “Gradient-Based Learning Applied to Document Recognition,” *Proceedings of the IEEE*, no. November, pp. 1–46, 1998.

[8] S. Sardi, R. Vardi, A. Goldental, A. Sheinin, H. Uzan, and I. Kanter, “Adaptive nodes enrich nonlinear cooperative learning beyond traditional adaptation by links,” *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018.

[9] Y. Kali and M. Linn, “Reducing the Dimensionality of Data with Neural Networks,” *Science*, vol. 313, no. July, pp. 468–474, 2006.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[11] W. Xu, “Towards Optimal One Pass Large Scale Learning with Averaged Stochastic Gradient Descent,” pp. 1–19.

[12] K. Fukushima, “Neocognitron: A hierarchical neural network capable of visual pattern recognition,” *Neural Networks*, vol. 1, no. 2, pp. 119–130, Jan. 1988.

[13] V. (U of T. Nair and G. E. (U of T. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” *27th international conference on Machine Learning*, 2010.

[14] Z. Yang, Y. Yuan, Y. Wu, R. Salakhutdinov, and W. W. Cohen, “Review networks for caption generation,” *Advances in Neural Information Processing Systems*, no. Nips, pp. 2369–2377, 2016.

[15] and P. S. S. Gupta, Ankita, Gurunath Gurrala, “An online power system stability monitoring system using convolutional neural networks,” *IEEE Transactions on Power*

Systems, vol. 34, no. 2, pp. 864–872, 2018.

[16] L. Yann and B. Yoshua, "Convolutional Networks for Images, Speech, and Time-Series," vol. 4, no. April 2016, pp. 2571–2575, 1995.

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, pp. 3156–3164, 2015.

[19] W. Gerstner, "Coding properties of spiking neurons: Reverse and cross-correlations," *Neural Networks*, 2001.

[20] R. Carrasco-Davis *et al.*, "Deep Learning for Image Sequence Classification of Astronomical Events," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 108006, 2019.

[21] M. Pérez-Carrasco *et al.*, "Multiband Galaxy Morphologies for CLASH: A Convolutional Neural Network Transferred from CANDELS," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 108002, 2019.

[22] G. Longo, E. Merényi, and P. Tiňo, "Foreword to the Focus Issue on Machine Intelligence in Astronomy and Astrophysics," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 100101, 2019.

[23] T. Chattopadhyay, D. Fraix-Burnet, and S. Mondal, "Unsupervised Classification of Galaxies. I. Independent Component Analysis Feature Selection," *Publ. Astron. Soc. Pac.*, vol. 131, no. 1004, p. 108010, 2019.

[24] G. Segal, D. Parkinson, R. P. Norris, and J. Swan, "Identifying Complex Sources in Large

Astronomical Data Sets Using a Coarse-grained Complexity Measure," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 108007, 2019.

[25] M. Hon, D. Stello, and J. Yu, "Deep learning classification in asteroseismology," *Monthly Notices of the Royal Astronomical Society*, vol. 469, no. 4, pp. 4578–4583, 2017.

[26] S. Zucker and R. Giryes, "Shallow Transits—Deep Learning. I. Feasibility Study of Deep Learning to Detect Periodic Transits of Exoplanets," *The Astronomical Journal*, vol. 155, no. 4, 2018.

[27] L. Yu *et al.*, " Identifying Exoplanets with Deep Learning. III. Automated Triage and Vetting of TESS Candidates ," *The Astronomical Journal*, 2019.

[28] C. J. Shallue and A. Vanderburg, "Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90," *The Astronomical Journal*, vol. 155, no. 2, p. 94, 2018.

[29] N. Krachmalnicoff and M. Tomasi, "Convolutional neural networks on the HEALPix sphere: a pixel-based algorithm and its application to CMB data analysis," *Astronomy & Astrophysics*, vol. 628, p. A129, 2019.

[30] R. Vilalta, K. D. Gupta, D. Boumber, and M. M. Meskhi, "A General Approach to Domain Adaptation with Applications in Astronomy," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 108008, 2019.

[31] M. Ansdell *et al.*, "Scientific Domain Knowledge Improves Exoplanet Transit Classification with Deep Learning," no. Borucki 2016, pp. 3–10, 2018.

[32] L. Hendriks and C. Aerts, "Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 1004, p. 108001, 2019.