Investigating the Impact of Unsupervised Feature-Extraction from Multi-Wavelength Image Data for Photometric Classification of Stars, Galaxies and QSOs

Annika Lindh

Dublin Institute of Technology, Ireland

Abstract. Accurate classification of astronomical objects currently relies on spectroscopic data. Acquiring this data is time-consuming and expensive compared to photometric data. Hence, improving the accuracy of photometric classification could lead to far better coverage and faster classification pipelines. This paper investigates the benefit of using unsupervised feature-extraction from multi-wavelength image data for photometric classification of stars, galaxies and QSOs. An unsupervised Deep Belief Network is used, giving the model a higher level of interpretability thanks to its generative nature and layer-wise training. A Random Forest classifier is used to measure the contribution of the novel features compared to a set of more traditional baseline features.

1 Introduction

With the vast amounts of data gathered by today's astronomical surveys, it is no longer feasible to inspect the observations manually, making the role of Machine Learning (ML) in the field of Astronomy highly relevant at present [1,2]. However, incorporating this approach is an ongoing effort which Djorgovski et al. [3] describe as a change of culture in the field. Borne [4] also discusses this, and concludes that it is essential that the ML techniques be such that they are interpretable enough to allow for human astronomers to work in collaboration with them on the analysis. This would also allow for verification of their quality and would contribute more to the understanding of the underlying models of our universe than a black-box technique could do.

This paper takes these concerns into account while investigating a novel approach to the feature generation problem in the photometric classification of stars, galaxies and quasi-stellar objects (QSOs); the latter term referring to the highly active galactic nuclei believed to be caused by supermassive black holes [5]. The data source used is the Sloan Digital Sky Survey (SDSS) Data Release 12 (DR12) – a recent ground-based astronomical survey covering roughly a third of the sky [6].

The approach used in this paper is an unsupervised ML technique from the Deep Learning field called Deep Belief Networks (DBN) which has some interesting qualities when it comes to interpretability [7]. DBNs have been used successfully in Computer Vision tasks with unlabeled image data [8], which makes it a good candidate considering the vast amounts of this type of data available from recent surveys such as the SDSS.

The purpose of this paper is to provide an evaluation of the approach presented here. Conclusions drawn from this can be used either directly when designing the methodology for future classification tasks, or as a basis for future research in a similar direction.

It is outside the scope of this paper to investigate different classification algorithms in which the unsupervised features could be used, so this is left as a recommendation for future work. Further, while some exploration into different hyper-parameter settings was carried out during the development of the unsupervised model, extensive tuning was not a prioritized task, leaving some room for improvement on this aspect.

2 Photometric Classification

The golden record for object class labels in Astronomy is obtained by inspecting the object's spectra, i.e. the light emitted as a function of wavelength. From this, it is possible to analyze the physical qualities of the object that are highly relevant when determining its class [9]. A problem with using these spectroscopic measurements is that they are far more time-consuming to obtain than photometric measurements such as images and the parameters derived thereof [3]. In SDSS DR12, the number of objects with spectroscopic observations are roughly 1% of those with photometric observations [6]. Considering this, accurate automated classification using photometric data alone would increase both the coverage and speed of the classification task [1], [10].

While spectroscopic classification defines the golden record, photometric classification is an active area of research with promising results. The SDSS dataset in particular has become something of a standard for this type of research thanks to its publicly available dataset consisting of photometric observations of around 470 million objects, and spectroscopic measurements for around 5.3 million of these in SDSS DR12 [6], [11]. Current state-of-the-art on the three-way classification of stars, galaxies and QSOs in the SDSS dataset has reached rather high levels of accuracy, but still suffers from some problems, mainly with regards to the feature generation process and the interpretability of the models.

2.1 Current State-of-the-Art

Brescia et al. [11] used a 2-layer Multi-Layer Perceptron with Quasi Newton Algorithm to simultaneously classify stars, galaxies and QSOs in the SDSS dataset, using only photometric features, with the spectroscopic class labels as the true labels. Their best result gave a precision of 93.82% for stars, 93.49% for galaxies and 86.90% for QSOs, with recall values of 86.40% for stars, 97.02% for galaxies and 90.49% for QSOs. For this model, the input features were the magnitude values that SDSS derives from the image data, representing the apparent strength of the source as seen from our planet. One problem with this feature set is that the magnitude values are dependent on the distance to the source, such that a closer source will have higher magnitude values than an identical source that is further away. While there is some correlation between the distance to an object and its class label due to cosmological reasons, the magnitude values themselves do not represent an accurate definition of the physical attributes of

the source. However, the *difference* in magnitudes from multiple wavelength bands of the same object does carry relevant information, essentially forming a very low-resolution spectra of the object [3], [12]. As such, it is possible that the classifier in [11] managed to find such correlations between differences in the magnitudes. There is also the risk, though, that it potentially overfits to irrelevant aspects of the individual magnitudes that might be correlated to the instrument quality or the target selection pipeline for the spectroscopic observations. This is a situation where a higher interpretability of the model would have been useful to investigate what the model based its decisions on.

In a separate experiment from the same paper, instead of using all the magnitude values directly, only one of them was used as-is, while the rest were used to generate a set of features from the differences between them [11]. These differences, referred to as *color values*, are commonly used in photometric classification tasks [13]. They are referred to by two-letter names based on the wavelength bands they compare, such as B-V or u-g; which wavelength bands are available depend on the photometric system that is used [12]. These color values should not be confused with the single-color terms used to describe subclasses of stars (e.g. red giants, white dwarves).

In this second experiment, 10 color values were used along with the magnitude value from the u-band and r-band respectively in two different models. The results from this experiment were slightly lower: for stars the precisions were 90.21% and 89.93% with recalls of 82.57% and 82.27%, for galaxies the precisions were 88.00% and 88.03% with recalls of 92.69% and 92.64%, and for QSOs the precisions were 85.56% and 85.60% with recall values of 87.83% and 87.77% [11]. It is difficult to say whether the results from the first experiment were higher due to the model's ability to find more complex correlations in the magnitude differences, or whether it was due to overfitting on aspects of the data that might not generalize to the full photometric population.

It should be mentioned that higher accuracies have been reported for specific classes under specific circumstances. In [14], QSO point-sources were separated from other point-sources to a precision of 96.96% and a recall of 99.01% for the QSO class. However, the sample selection resulted in a proportion of QSOs of 86.14% in contrast to 12.88% in the full dataset. While this might be useful for certain scenarios, the focus of the present work is on developing a model that generalizes well to all classes.

2.2 Gaps

Accurate object type classification still relies on spectroscopic measurements which are far less abundant than image data and its derived photometric parameters, but photometric classification results are reaching fairly high accuracies. Current state-of-the-art results have been obtained through the use of advanced ML algorithms. However, the interpretability of these techniques is often lacking, making it hard to say what they base their decisions on. For the methods to become more useful to astronomers, this needs to be addressed.

Another problem is the way features are derived from the original measurements. Often, this is done by handcrafted rules, increasing the risk of introducing a human bias. In the SDSS photometric pipeline, magnitude values are calculated based on assumptions about the nature of the object they measure, without actually knowing the nature

of that source; because of this, multiple values are derived for the same parameter, leaving it up to individual researchers which ones to use [15]. Coughlin et al. [16] argue for the use of methods that treat all sources in the same way, so that results can be compared more easily and biases be more readily quantified. Since the photometric parameters in SDSS are derived from the image data, using that data directly, in the same way for all objects, might be a step in the right direction.

Further, there might be useful information in the image data that is not captured by the parameters from the current pipelines. Support for this has been demonstrated on the task of galaxy morphology classification in [17] and [18]. While [17] used a novel set of handcrafted rules as input features for a Multi-Layer Perceptron, [18] used a Convolutional Neural Network where the algorithm decides which details are relevant in describing the images, potentially reducing the human bias of the model. Additionally, thanks to the deep layer-wise training of the model on image data, they were able to visualize what was learned by the model, thus making it more interpretable. These qualities would be beneficial to the classification task in this paper as well, and their results lend some support to the approach of using unsupervised feature-extraction from astronomical image data, which is an approach that has not been well-explored when it comes to photometric classification of stars, galaxies and QSOs.

2.3 Deep Learning as a Potential Solution

Deep Learning is an area of ML that covers the use of multi-layered non-linear models [8]. It builds on previous research into Neural Networks which is a technique inspired by our understanding of the biological learning process in humans and other animals. The breakthrough that led to a surge of interest into Neural Networks in the early 2000s came from a combination of the rapid increase in processing power at a lower price scale, along with the work of Hinton, Osindero & Ten [7] who proposed a more efficient way of training Deep Neural Networks than the previous standard of using backpropagation [8]. Their approach was to pre-train a network through an unsupervised approach that made it possible to leverage large amounts of unlabeled data – a trait that is highly relevant in the Astronomy domain. The network was trained layer-by-layer so that increasingly sophisticated features could be found at each level. This not only improved the final result and the training speed – it also made the model more interpretable since the features learned in each layer could be visualized separately. Their motivation for using such an approach was that it should be possible to find a clearer link between what causes the image (i.e. the actual object being imaged) and its class label than there would be between individual pixel values and the class label of the object [19]; in practice, such a model learns the underlying features that cause the pixel values to take on certain distributions, which agrees well with the intention of the astronomical classification task.

To achieve the layer-wise training described above, a deep model was constructed by stacking a set of two-layer Restricted Boltzmann Machines (RBMs), each trained in isolation starting with the one closest to the input layer [7]. Each RBM is a network where all the units of one layer are connected to all the units of the next, with no connections between units within the same layer. The weights are symmetrically tied in

both directions: in a two-layer RBM, the weight matrix from the visible (input) layer to the hidden layer is the transpose of the weight matrix from the hidden layer to the visible. In addition to the weights, the two layers have their own set of bias units corresponding to the connected units of that layer. Training is carried out by using a learning technique called Contrastive Divergence (CD) which was introduced in [20] as an efficient approximation of maximum likelihood learning. As a final step, the weights of the RBMs can be unfolded into a single network where the symmetrical ties are dropped; this network can then be fine-tuned through a contrastive version of the wake-sleep algorithm that was introduced in [21]. This type of deep model is referred to in the literature as a DBN; however, the term is used somewhat ambiguously as it sometimes implies that the network includes a supervised softmax layer on top of the stack of RBMs, while at other times it refers to the fully unsupervised version, and yet other times it is used for a Multi-Layer Perceptron that has been pre-trained by CD and then fine-tuned by traditional backpropagation [8]. In this work, DBN refers to an unsupervised model trained by CD and fine-tuned in an unsupervised manner through the contrastive version of the wake-sleep algorithm.

2.4 Suitability of Using a DBN for Photometric Feature-Extraction

Feature-extraction by a DBN provides an unsupervised method that can be used directly on the image data. It is a data-driven method that avoids the use of handcrafted rules. Hence, the extracted features are chosen by the model to best represent the actual data, avoiding the human bias of handcrafted rules based on the researcher's assumptions about the data. Since it uses unsupervised training, it can potentially leverage the vast amounts of unlabeled data available from modern Astronomy surveys. Further, DBN is a generative model, meaning that what it learns is actually to generate new data samples from its understanding of the training data; this can be used to visualize what type of data the model has learnt to recognize. Additionally, thanks to its layer-wise training, it provides a straightforward way to visualize what sort of structures each unit is sensitive to in the images. These points address the gaps discussed here, while still providing a technique that has been shown capable of handling complex ML tasks.

3 Methodology

3.1 Experiment Design

The classification task was to use photometric data to classify each astronomical object as either a star, galaxy or QSO according to its spectroscopic class label. To measure the advantage of using the features extracted by the unsupervised DBN, an experiment was designed where the result of two models could be compared. The *baseline model* uses the 10 color values introduced in section 2, which are easily derived from the existing dataset; this gives a baseline result which can already be obtained with little effort. The *DBN model* uses the novel features, in addition to the baseline features, to measure what added benefit they might provide. Both models use a Random

Forest (RF) classifier to assign the class labels. For the baseline model, this is the entire model; for the DBN model, the RF classifier is the final stage where the DBN features and the baseline features are used as inputs.

Each sample of input data for the DBN consisted of the pixel values from all 5 wavelength bands of one object with a resolution of 20 by 20 pixels each, making each sample 2000 pixels in total. Some preprocessing was necessary to ensure the images were cropped, centered and resized to the same dimensions. After this, the images were normalized in each wavelength band individually since the difference in signal strength between the bands is already covered by the color features; if a future experiment were to use the DBN features alone, the author recommends calibrating the images according to the information given in [22].

3.2 DBN Design

The DBN was constructed from two RBMs: the first with 2000 visible and 1000 hidden units, and the second with 1000 visible and 100 hidden units. The output from the DBN is thus 100 features, referred to here as the *DBN features*. Each layer was trained greedily in isolation, after which the symmetry of the generative and recognition weights were dropped and the RBMs were unfolded into a single network; the structure of these networks are shown in **Fig. 1**. The resulting network was globally fine-tuned through the contrastive wake-sleep algorithm where the generative weights were trained separately from the recognition weights [7], [21].
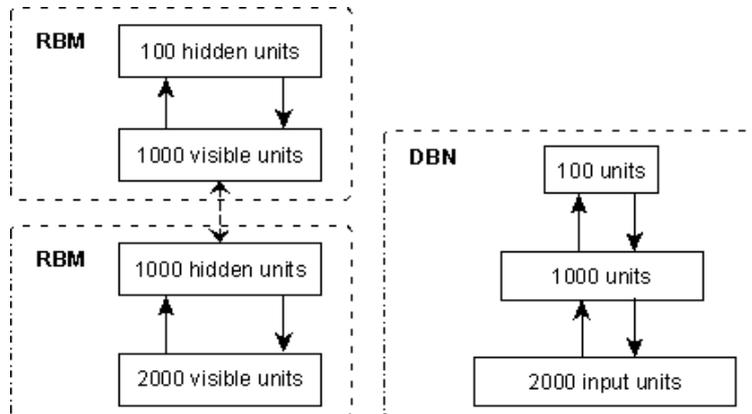
Fig. 1. The network used for unsupervised feature-extraction, with the RBM layers on the left and the combined DBN network on the right

3.3 Sample

The dataset for the experiments was acquired from the SDSS dataset by random sampling stratified on the spectroscopic class labels. **Table 1** shows the class label proportions and sizes of the full set of spectroscopic observations (which can contain duplicates and failed observations), the clean population (where objects must have valid

spectroscopic and photometric observations and not contain duplicates) and the actual sample. The criteria for the clean sample was based on SDSS flags only; no extra inspection was carried out on the images themselves. Specifically, only the *primary* photometric and spectroscopic observations were used. Spectroscopic observations were not allowed to have *zWarning* flags of *UNPLUGGED*, *BAD_TARGET* or *NODATA*. Photometric observations were required to have the *PHOTOMETRIC* flag set for its *calibStatus* in all wavelength bands, in addition to having the *clean* flag set. More details on the meaning of these flags can be found on the SDSS DR12 website[1] and the CasJobs[2] site where most of the numerical data can be downloaded from. The images were obtained separately through the DR12 Science Archive Server[3].

Table 1. Spectroscopic class proportions in the sample and the relevant populations

	Sample n = 10,000	Clean population n = 3,140,923	Full population n = 3,537,411
STAR	23.97%	23.97 %	22.83 %
GALAXY	62.10%	62.10 %	64.29 %
QSO	13.93%	13.93 %	12.88 %

Due to resource limitations, the sample size is quite small compared to the total dataset, but as the experiments are meant to test the approach rather than draw conclusions about the class labels of the total dataset, this limitation should not be of great concern.

The sample was split into training and test sets by first splitting the dataset in half for use in the DBN and RF training each. The DBN data was split into 70% training and 30% cross-validation of the cost monitoring; the RF data was split into 40% for training, 30% for cross-validation of the number of trees for the RF and some hyperparameter tuning for the DBN, with 30% reserved for the final test results.

3.4 Evaluation

To assess the model's performance on the individual classes, the individual F_1 scores were used. For the overall performance, the macro-averaged F_1 score was used, calculated to prefer a model that generalizes well to all classes [23]. This was used since one of the strengths of unsupervised learning is that it can leverage unlabeled data, but by doing so, there's a risk of biasing the model towards data with a more common profile.

The final training and testing of the models was performed 100 times with different random number generator seeds to ensure that any differences between the models were not due to chance. The difference in means was tested for statistical significance with a one-tailed two-sample independent t-test in PSPP with a 95% confidence interval and a p-value cut-off of .05.

---

[1] http://www.sdss.org/dr12/

[2] http://cas.sdss.org/dr12/en/home.aspx

[3] http://data.sdss3.org/

3.5 Limitations

Since the absolute true labels of distant astronomical objects are unknown, the golden record will inevitably be an approximation based on current theories [3]. For this paper, the SDSS spectroscopic class labels were used, so any biases in the SDSS spectroscopic pipeline and target selection criteria are inherited here.

The strongest human bias of this work was the image preprocessing which unfortunately could not be avoided. Attempts have been made to keep this to a minimum by using threshold values relative to the measurements in each individual image.

4 Results

The DBN model performed better than the baseline model both overall and on each of the three classes. The differences in means were statistically significant ($p < .001$) when tested in a one-tailed independent two-sample t-test in PSPP with a confidence interval of 95% and no assumption of equal variance. **Table 2** shows the results along with the differences in means. The relative performance increase when adding the novel features is also given, calculated by taking the absolute performance increase divided by the means from the baseline model.

Table 2. Results from the baseline model and DBN model, the absolute difference between them, and the relative improvment of the DBN model compared to the baseline model

	baseline model	DBN model	Absolute increase	Relative increase
Macro-averaged F_1 score	0.8321	0.8682	0.0361	4.35%
F_1 score, STAR	0.8037	0.8664	0.0627	7.81%
F_1 score, GALAXY	0.9376	0.9582	0.0206	2.20%
F_1 score, QSO	0.7549	0.7801	0.0252	3.32%

5 Discussion

The results confirm the usefulness of unsupervised feature-extraction from image data in photometric classification of stars, galaxies and QSOs. The model generalizes fairly well to the different classes, but could potentially be improved on this aspect by training the classifier on a weighted sample to counter the class-imbalance.

Comparing to current state-of-the-art, the results reported in [11] can be converted into the same metrics used for this work, giving a macro-averaged F_1 score of 0.9135, with the F_1 score for stars at 0.8996, galaxies at 0.9522 and QSOs at 0.8866. The DBN model performs slightly better on the galaxy class but is otherwise below the state-of-the-art results, with mainly the QSO class lagging behind in performance. This leaves room for improvement, possibly by combining the classifier used in [11] and the novel features presented in this work.

For the interpretability of the model, however, a simpler classifier makes it possible to not only understand the extracted features but also how they were used in the classification task. This was done for the present work by examining the feature contribution values obtained from the RF classifier, averaged over the 100 training runs. The 10 color features were the top 10 contributors, while the total contribution of the 100 DBN features added up to 28.43%. A visualization of the top 10 DBN features is shown in **Fig. 2** where brighter areas represent a preference for the presence of signal, and darker areas represent the preference of an absence of signal. Since the images are normalized, they do not show where the zero-point of the weights are, but the contrast shows the type of shapes the model has learned to recognize. This visualization technique is possible regardless of the classifier, but the RF classifier can help by measuring their contributions.
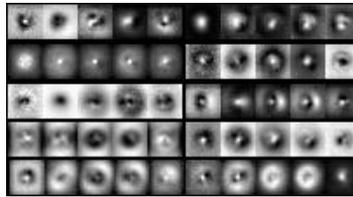
Fig. 2. Normalized visualizations of the top 10 contributing DBN features

Another way to provide some insight into what the model learned is by generating samples from the joint distribution learned by the model [7]. **Fig. 3** shows some of the actual data samples after the image preprocessing on the left, with samples generated from the model on the right-hand side, in both cases taken from different wavelength bands and different object types. The samples on either side are unrelated to each other and to the other samples in the same rows and columns.
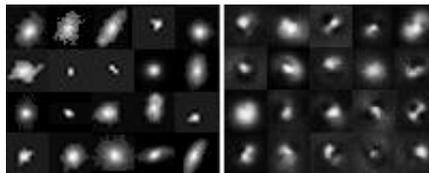
Fig. 3. Actual data samples after image preprocessing (left) and samples generated from the model's joint distribution (right). Individual images are 20x20 pixels and are taken from different wavelength bands and classes.

6   Conclusions

6.1   Contributions

The main contribution of this work is the finding that unsupervised feature-extraction from image data can provide a measurable advantage for photometric classification of stars, galaxies and QSOs. The approach makes it possible to leverage the large

amounts of unlabeled image data captured by modern Astronomy surveys, and it addresses the concern in the field regarding the interpretability of ML models. Additionally, thanks to the unsupervised training, the DBN features could be re-used even if the golden record for the class labels was expanded or re-evaluated.

As a side-product of this work, scalable implementations of the RBM and DBN algorithms have been constructed for running the experiments. These implementations are not tied to the task presented in this paper but can be applied to any domain with high-dimensionality (labeled or unlabeled) data. The source code has been made available online as free software under the GPLv3 license.[4]

6.2   Future Work

The approach presented in this paper could be used directly in future research with relevant classification objectives. The weakness in the QSO class could potentially be addressed by training the RF classifier on a weighted sample, or by using the DBN features in a more complex classifier such as the one from [11] at the expense of some of the model's interpretability.

A more specific recommendation regarding the quality of the DBN features would be to train the DBN with a sparsity target; sparser features take on more specific roles which improves their interpretability [24]. Potentially, this could also reduce the complexity of the DBN's learning, especially in the second layer and beyond where the features from previous layers would provide clearer building blocks.

A limitation to what can be concluded from this paper, which could be addressed by future research, is that it is not clear whether the image data from all five wavelength bands contributed to the results. More insight into this aspect could be provided by comparing these results to a model using only images from the wavelength band with the highest signal-to-noise ratio.

Finally, and perhaps the most interesting direction, would be for future work to investigate the usefulness of the DBN features for Transfer Learning. Previous work by [25] has shown that unsupervised feature-extraction can provide useful intermediate features for different classification tasks within the same domain. If the DBN features were to show promising results in this area, then building a database of such generalizable features could provide the Astronomy community with a set of ready-to-use lower-dimensionality features that would be more accessible than the full image data.

---

4   https://github.com/AnnikaLindh/DBNTensorFlow
5   https://github.com/blackecho/Deep-Learning-TensorFlow

References

1. Ž. Ivezić, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, Y. AlSayyad, ... H. Zhan, "LSST: from Science Drivers to Reference Design and Anticipated Data Products," *arXiv:0805.2366 [astro-ph]*, Aug. 2014.
2. A. M. Mickaelian, "Astronomical Surveys and Big Data," *arXiv:1511.07322 [astro-ph]*, Nov. 2015.
3. S. G. Djorgovski, A. A. Mahabal, A. J. Drake, M. J. Graham, and C. Donalek, "Sky Surveys," *arXiv:1203.5111 [astro-ph, physics:physics]*, pp. 223–281, 2013.
4. K. D. Borne, "Scientific Data Mining in Astronomy," *arXiv:0911.0505 [astro-ph, physics:physics]*, Nov. 2009.
5. D. Richstone, E. A. Ajhar, R. Bender, G. Bower, A. Dressler, S. M. Faber, … S. Tremaine, "Supermassive Black Holes and the Evolution of Galaxies," *arXiv:astro-ph/9810378*, Oct. 1998.
6. S. Alam, F. D. Albareti, C. A. Prieto, F. Anders, S. F. Anderson, B. H. Andrews, … H. Zou, "The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III," *The Astrophysical Journal Supplement Series*, vol. 219, no. 1, pp. 12–43, Jul. 2015.
7. G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
8. L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. e2, pp. 1–29, 2014.
9. K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, … Z. Zheng, "The Baryon Oscillation Spectroscopic Survey of SDSS-III," *The Astronomical Journal*, vol. 145, no. 10, pp. 1–41, Jan. 2013.

10. M. Brescia, S. Cavuoti, S. G. Djorgovski, C. Donalek, G. Longo, and M. Paolillo, "Extracting Knowledge From Massive Astronomical Data Sets," *arXiv:1109.2840 [astro-ph]*, pp. 31–45, 2012.

11. M. Brescia, S. Cavuoti, and G. Longo, "Automated physical classification in the SDSS DR10. A catalogue of candidate Quasars," *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 4, pp. 3893–3903, May 2015.

12. M. S. Bessell, "Standard Photometric Systems," *Annual Review of Astronomy and Astrophysics*, vol. 43, no. 1, pp. 293–336, Sep. 2005.

13. N. M. Ball and R. J. Brunner, "Data Mining and Machine Learning in Astronomy," *International Journal of Modern Physics D*, vol. 19, no. 7, pp. 1049–1106, Jul. 2010.

14. S. Abraham and N. S. Philip, "Photometric Determination of Quasar Candidates," presented at the Astronomical Data Analysis Software and Systems XIX, 2010, vol. 434, pp. 147–150.

15. C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton, S. Burles, F. J. Castander, … W. Zheng, "Sloan Digital Sky Survey: Early Data Release," *The Astronomical Journal*, vol. 123, pp. 485–548, Jan. 2002.

16. J. L. Coughlin, F. Mullally, S. E. Thompson, J. F. Rowe, C. J. Burke, D. W. Latham, … K. A. Zamudio, "Planetary Candidates Observed by Kepler. VII. The First Fully Uniform Catalog Based on The Entire 48 Month Dataset (Q1-Q17 DR24)," *arXiv:1512.06149 [astro-ph]*, Dec. 2015.

17. M. Banerji, O. Lahav, C. J. Lintott, F. B. Abdalla, K. Schawinski, S. P. Bamford, ... J. Vandenberg, "Galaxy Zoo: reproducing galaxy morphologies via machine learning," *MNRAS*, vol. 406, no. 1, pp. 342–353, Jul. 2010.

18. S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 2, pp. 1441–1459, Apr. 2015.

19. G. E. Hinton, "Learning to represent visual input," *Philos Trans R Soc Lond B Biol Sci*, vol. 365, no. 1537, pp. 177–184, Jan. 2010.

20. G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.

21. G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal, "The 'wake-sleep' algorithm for unsupervised neural networks," *Science*, vol. 268, no. 5214, pp. 1158–1161, May 1995.

22. J. A. Smith, D. L. Tucker, S. Kent, M. W. Richmond, M. Fukugita, T. Ichikawa, … D. G. York, "The u'g'r'i'z' Standard Star Network," *The Astronomical Journal*, vol. 123, no. 4, pp. 2121–2144, Apr. 2002.

23. M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

24. G. E. Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum*, vol. 9, no. 1, p. 926, 2010.

25. Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.