ADAPTABILITY OF HYPERSPECTRAL UNMIXING FOR WATER MONITORING EVALUATED ON HYPSO SCENES

Randi Fjelltun Stensvand, Aria Alinejad, Joseph L. Garrett

Department of Engineering Cybernetics
Norwegian University of Science and Technology, 7034 Trondheim, Norway

ABSTRACT

Harmful algal blooms are a growing threat to activities ranging from aquaculture to fishing to recreation. Hyperspectral remote sensing is a potential tool for the detection and monitoring of algal blooms. However, developing algorithms that can detect a wide range of algal blooms and make their characteristics legible to scientists is challenging. Hyperspectral Unmixing enables scientists to jointly analyze the spatial and spectral content of an image. In this proceeding, we describe several experiments designed to test whether a hyperspectral unmixing model trained on one dataset can be adapted to others. The Entropic Descent Archetypal Analysis unmixing algorithm (EDAA), a simple state-of-the-art unmixing method, is used to develop an unmixing model for several scenes acquired by the HYPSO satellite constellation. An algorithm for adding new endmembers to the model, in order to adapt it to new scenes, is developed and tested. The performance of EDAA on these tests suffices to motivate further unmixing algorithm testing and development.

Index Terms— algae, online learning, archetypal analysis, cubesat constellations

1. INTRODUCTION

Harmful algal blooms are a threat to coastal ecosystems, recreation and aquaculture [1]. Hyperspectral imaging is a potential tool for detecting and monitoring algal blooms ([2], fig. 1). However, hyperspectral satellites collect more data than any team of scientists can manually peruse, e.g. well over a million pixels per day. Nonetheless, manual classification rules are inhibited by the large number of different algal species which can bloom, and the similarity of their spectral characteristics to other phenomena. The coastlines of the earth's continents are also long, and collecting sufficient water samples for supervised machine learning methods is generally not feasible, with most labeled datasets consisting of fewer than 10k pixels [3], in contrast to the billions of pixels included in most labeled computer vision datasets.

Thanks to the Research Council of Norway and industry partners for funding through under the projects SoS (Grant No. 337495), HYPSCI (Grant No. 325961), and SFI Harvest (Grant No. 309661).

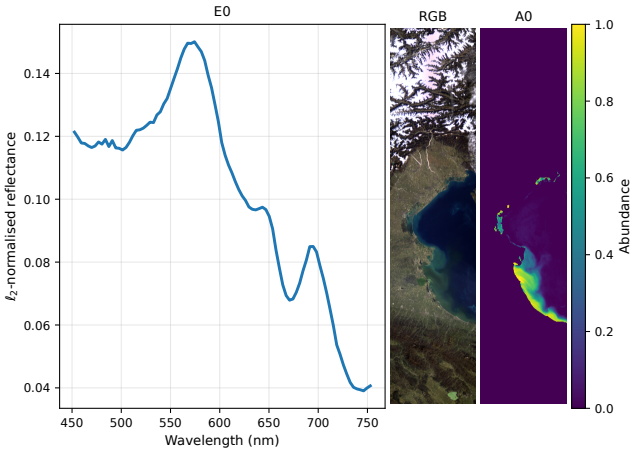


Fig. 1. Hyperspectral unmixing connects spatial and spectral patterns. Here, the spectrum of one endmember shows features indicative of an algal bloom: increased reflectance in the green bands (near 560nm) and evidence of chlorophyll-a absorption near 670nm. The abundances associated with this spectrum are shown spatially on an image of the northern Adriatic sea acquired by the HYPSO satellite.

Therefore machine learning methods which do not require extensive supervision are necessary to make hyperspectral signals legible to scientists. Initially, the HYPSO team found that Independent Component Analysis (ICA) could extract features which spatially appeared consistent with water quality parameters estimated by legacy band ratio algorithms [4]. It was thus expected that ICA could extract additional features with hyperspectral data as well. Unfortunately, the HYPSO team's collaborators from ocean science disciplines found the features extracted by ICA to be difficult to interpret, because of ambiguities regarding the connection between the scores and loadings produced by ICA and physical parameters describing the waters optical properties. Further investigations of general machine learning techniques did not uncover any unsupervised dimensionality reduction algorithms particularly well-suited to hyperspectral data [5].

Hyperspectral unmixing (HU), which partitions an image into endmembers and abundances, has been developed specifically for imaging spectroscopy. A few previous studies ap-

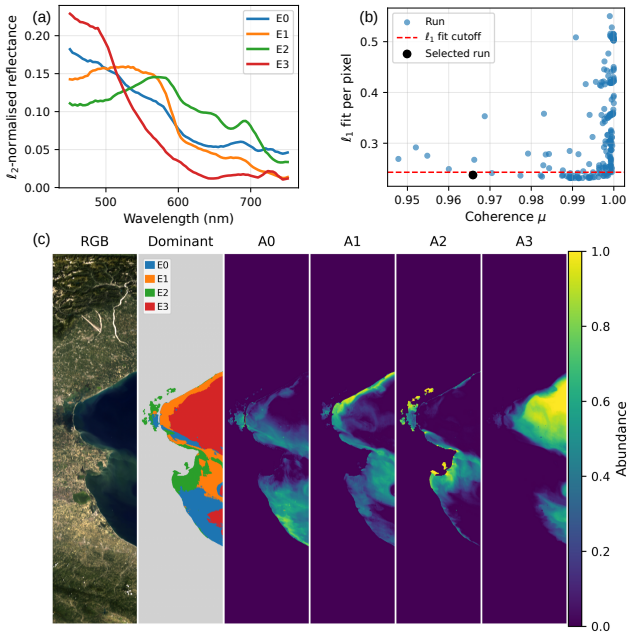


Fig. 2. The water pixels in a HYPSO image over the city of Venice, Italy are unmixed into 4 endmembers (a). Out of M=200 runs, one set of endmembers and abundances are selected according to their coherence and L1 residual (b). The endmember spectra which shows the clearest evidence of algae is spatially present in lagoons and the Po River plume (A2), while the endmember suggestive of clear water (A3) is further from the coast (c).

ply hyperspectral imaging to aquatic scenes [6, 2, 7, 8]. Two key aspects separate HU from other unsupervised machine learning techniques: (i) the abundances for a single pixel must sum to one and (ii) the abundances and endmembers are non-negative. These two properties resolve some of the ambiguities that plagued the aforementioned ICA. While our tests with ordinary non-negative matrix factorization (NMF), on which HU is historically built, showed middling performance [5], further tests that included the archetypal analysis constraint showed improved interpretability and predictive performance with on-water radiometry data [9]. Additionally, the recently-published Entropic Descent Archetypal Analysis technique (EDAA) enables a substantial acceleration of archetypal analysis.

While EDAA is a reasonably interpretable tool for making much of the information concealed in hyperspectral images legible to scientists, it still suffers some challenges. For example, the largest dataset that EDAA has been tested on consists of fewer than 100k pixels, while the HYPSO satellite constellation has collected a database of images consisting of over 2.5B pixels, and other hyperspectral satellites have accumulated much larger datasets. This proceeding will focus on how to adapt EDAA so that it can be used to analyze large, frequently updated hyperspectral datasets.

2. METHODOLOGY

Entropic Descent Archetypal Analysis addresses the non-convexity of the unmixing objective function by minimizing it repeatedly with different initial conditions, so that one can be selected based on its L1 error metric and endmember coherence [10, 11]. The coherence is the maximum pairwise spectral correlation between the estimated endmembers. By comparing these metrics for many different runs, it is possible to gain insight about how difficult a particular scene is to unmix (see fig. 2b). A recent comparison found that it was among the most accurate unmixing techniques [12], and our team's preliminary analysis determined that it was both much faster and more robust than other state-of-the-art unmixing techniques (see the appendix of [13]).

Once an EDAA model is created, the fully constrained least squares (FCLS) problem is solved with projected gradient descent in order to apply the model to new data. In this proceeding, an unmixing model refers to the dictionary of extracted endmembers.

Online learning refers to the capacity of machine learning methods to adapt their internal models in response to new data, after initial training. One example of online learning with hyperspectral data is On-the-fly-processing (OTFP) which can adapt a Principal Component-based model to new scenes as they are processed. Like OTFP, the EDAA-online update rule adds additional endmembers, as OTFP adds components, to the model when an error metric exceeds a certain threshold [14]. Unlike OTFP, large-error pixels are identified and reprocessed in their entirety (not only the residual). The process is:

1. FCLS calculates abundances
2. The pixelwise mean square error is calculated
3. Pixels which exceed the error threshold are selected
4. Those pixels are unmixed with EDAA to generate possible endmembers
5. The endmember which minimizes the coherence of the full model is added to the original endmembers
6. The scene is unmixed with the updated model

The extra steps of this procedure are required because archetypal analysis requires the original data for computation, e.g. the archetypes must be convex combinations of the data, not the residuals.

The HYPSO constellation consists of two 6U cubesats, each with a VIS-NIR hyperspectral imaging payload [15]. The primary objective of the HYPSO mission is to monitor water quality. This involves tasks such as providing an early warning system for harmful algal blooms and searching for signs of the crustacean *Calanus Finmarchicus* [16].

Preprocessing is applied before the tests to each top-of-atmosphere reflectance, also known as L1D, image. First, a smoothed dark pixel subtraction technique is used to mitigate the effect of the atmosphere over the whole scene [17]. The wavelengths between ≈400-750nm were retained for the

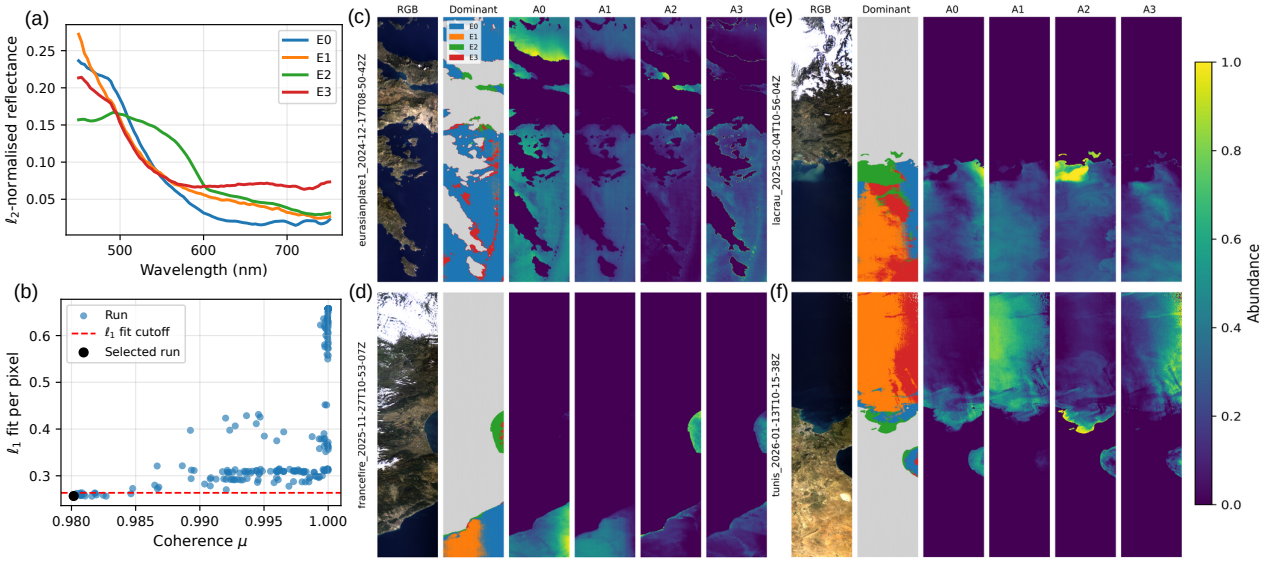


Fig. 3. Water pixels from four HYPSO scenes around the Mediterranean Sea are combined before unmixing. The endmember spectra (a) show somewhat less unique patterns than were recovered for a single scene, consistent with the larger coherence found in the model selection (b). The scenes show locations near Greece (c), the Pyrenees (d), the La Crau RadCalNet site (e), and Tunisia (f).

analysis, because that was the range over which the dark pixel subtraction performed adequately. Thus 88 of the original 120 bands are retained for both HYPSO-1 and -2 although their specific wavelengths vary slightly.

Next, a water mask is generated for the image. First, EDAA is applied with p=2 endmembers. Then k-means clustering partitions the pixels into 3 clusters, which roughly correspond to clouds, land, and water [18]. Pixels retained in the water cluster, the darkest of the 3, were retained as candidate water pixels. A further refinement removed connected water regions with fewer than 100 pixels. The final water mask then selects the pixels for the subsequent EDAA analysis. The pixels are L2 normalized before processing.

Three experiments are performed to evaluate the suitability of EDAA for unmixing large datasets. First, the capacity of EDAA to resolve features in the ocean which suggest clear physical phenomena is tested. Next, the capacity of EDAA to unmix multiple hyperspectral scenes at once is tested on four HYPSO scenes collected around the Mediterranean (fig. 3). Finally, the capacity of EDAA models to ingest new images is tested on a new Mediterranean scene (fig. 4). The online learning procedure introduced above is tested on an additional scene.

In the **first experiment**, EDAA is applied to an image of Venice, Italy acquired by the HYPSO-1 satellite on September 4th, 2025. After segmentation, EDAA is applied to the water pixels to extract p=4 endmembers, with $M = 200$ runs, $T = 50$ outer iterations for each run, and $K_1 = K_2 = 30$ inner iterations for the endmember spectra and abundance matrices (retaining the original parameter names from [11]). The spatial distribution of the endmember abundances together with their spectra are inspected for evidence of chlorophyll and other optically active constituents (fig. 1).

Four different scenes from around the Mediterranean Sea are aggregated for the **second experiment**, which is designed to evaluate the applicability of EDAA to multiple scenes. One scene is from the HYPSO-1 satellite (fig. 3c), while the other three are from HYPSO-2. The water pixels are identified in all four scenes and they were jointly unmixed with EDAA.

The **final experiment** evaluates whether the update procedure described above can be successfully applied to an unmixing model. The Mediterranean model, developed in the second experiment, is applied to an image acquired by the HYPSO-2 satellite over the city of Venice (fig. 4). EDAA is applied to the 20% of pixels with the largest residuals to generate three new endmembers. The endmember which had the least correlation with the original endmbembers is added as a new endmember to the model, since it does not increase the original coherence, and the Venice scene is unmixed again with the new model. Finally, the updated model is applied to another image of Venice (used in the first experiment). The residuals with and without the new endmember are compared.

3. EXPERIMENTAL RESULTS

In all three of the experiments described above, EDAA performed well. In particular, its strategy of addressing the nonconvexity of the unmixing objective by using multiple runs

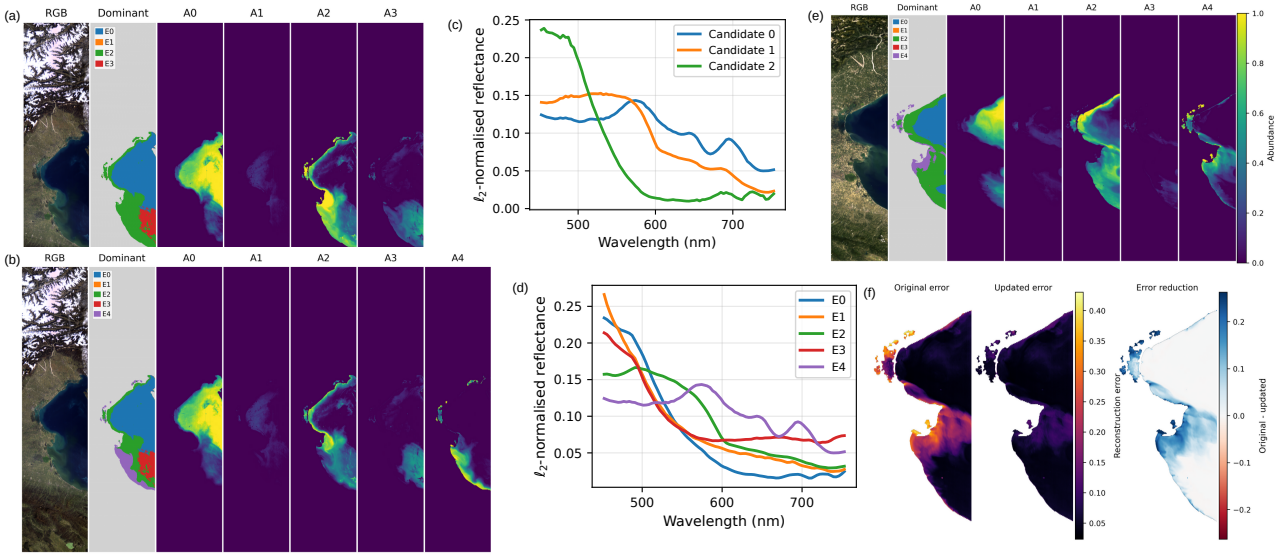


Fig. 4. The 4-endmember Mediterranean Unmixing Model applied to a Venice does not capture the local algal bloom (a). The pixels with the largest residual are selected and unmixed again to estimate a 5th endmember (b). The pixels in the error area are unmixed into three endmembers (c). Candidate endmember 0 is selected and added to the global model (d). The new 5-endmember model is tested on an additional Venice scene (e). The 5-endmember model shows a reduction in mean square error relative to the 4-endmember model on the test scene (f).

proved invaluable for visualizing and characterizing its performance. For all of the experiments, a few runs found much smaller minima of the objective function than the average, and these tended also to have lower coherence. Plots of the L1 error and coherence are shown in figs. 2b and 3b.

The **first experiment** clearly resolves an algal bloom in endmember #2 which furthermore shows evidence of chlorophyll absorption in its endmember spectrum. Endmember #3 in addition shows a nearly clear-water spectrum. The other two endmembers seem to be related to sediment in water, but their spectral features are more ambiguous. Despite this, the normalization strategy used by EDAA prevents the inversion of the recovered endmembers back into their original units, because all information about magnitude is lost, and thus inhibits interpretability. Even so, the first experiment showed that EDAA is capable of resolving spectra which display features characteristic of physical phenomena.

In the **second experiment**, EDAA clearly converges again. It achieves an L1 residual per pixel comparable to the first experiment, but its coherence was notably larger (indicating greater similarity among endmembers). It is more challenging to identify these endmember spectra with particular phenomena than in the first experiment, but endmember #2 does seem to recover some evidence of a river plume, although it is unclear whether this related colored dissolved organic matter (CDOM) or chlorophyll. The abundances from this experiment did seem to be moderately affected by an across-track artifact, though to a much lesser degree than observed for the techniques analyzed in [5]. Most importantly, the dark pixel subtraction atmospheric compensation seemed to standardize the data well enough for the endmembers to be distributed relatively evenly across the scenes.

The **final experiment** demonstrates that the model update rule functions as expected. The 4-endmember Mediterranean model does not initially identify the algal bloom in the new scene, and so the pixels over the algal bloom contain a relatively large residual. After the large-residual pixels are unmixed, two of the endmembers in the new model correlate with the endmembers of the original model, but one of them, which shows evidence of chlorophyll absorption, is not. After this endmember spectra is added to the original model, the algal bloom becomes much better resolved. When this new model is applied to an additional Venice scene, it achieves a much lower residual than the original model, suggesting that it has also spatially resolved a chlorophyll pattern in the water.

4. DISCUSSION

While these experiments point towards the usefulness of hyperspectral unmixing for aquatic imagery, they also suggest some limitations in the application of EDAA. First of all, it is readily apparent that EDAA extracts more detailed endmember spectra on the single-scene test than it does on the multi-scene test. Although the origin of this discrepancy was not directly analyzed here, we suspect that there may be inconsistencies between the images which are not eliminated by the simple atmospheric correction. These inconsistencies may be further coupled to imaging artifacts, such as the appar-

ent across-track artifact which afflicts experiment 2. Furthermore, the L1 residual vs. coherence selection plots produced by EDAA seem to suggest that there is an implicit structure to the unmixing objective function. Note in fig. 3b how the different runs seem to produce horizontal lines, in which runs with many different levels of coherence seem to produce the same L1 fit per pixel. Exploring what produces and characterizes these different lines may illuminate new structural aspects of the unmixing objective function.

Furthermore, the additional scenes may vary in different ways that are challenging for EDAA to capture. For example, EDAA seems to extract pixels with high chlorophyll and high suspended solids content fairly well, but if they vary in different proportions across scenes, they may be more challenging to separate. It could also be that the Mediterranean experiment simply includes more categories of spectra than can be captured by the model, although the very high level of coherence of the selected model suggests otherwise. Indeed, the similarity of endmembers #0,1, and 3–relative to their difference from endmember #2 (and endmember #4 extracted in the final experiment), suggests that there is an implicit hierarchical structure between the endmembers.

The proposed update rule works well in this preliminary test. However, we expect that the update rule will need to be refined further before it is possible to create an unmixing model for the entire HYPSO constellation dataset. For example, the trade-off between model complexity and accuracy must be explored in more detail.

5. CONCLUSIONS

Unmixing partially solves the interpretability problem in hyperspectral data processing because the features are endmember spectra, which are quite similar to the original data. Likewise, the abundance sum constraint allows the output to be interpreted in terms of how much each endmember spectrum contributes to a particular pixel. However, many limitations to interpretability remain such as the relationship between the different endmembers, and the arbitrary selection of p. Moreover, EDAA normalizes the data, which prevents their inversion back into the units of the data, which in turn limits their capacity to be processed further.

Unmixing is not an inherently scalable technique, as most contemporary unmixing algorithms do not support batching. However, these experiments demonstrate the capacity of a simple rule for updating an unmixing model based on new data, without reconsidering the prior data. They, therefore, suggest that new, scalable learning strategies for unmixing are possible.

6. REFERENCES

[1] B. Karlson *et al.*, "Harmful algal blooms and their effects in coastal seas of Northern Europe," *Harmful Algae*, vol. 102, 2021.

[2] C. J. Legleiter *et al.*, "Spectral mixture analysis for surveillance of harmful algal blooms (SMASH): A field-, laboratory-, and satellite-based approach to identifying cyanobacteria genera from remotely sensed data," *Remote Sensing of Environment*, vol. 279, 2022.

[3] M. K. Lehmann *et al.*, "GLORIA - A globally representative hyperspectral in situ dataset for optical sensing of water quality," *Scientific Data*, vol. 10, 2023.

[4] C. Penne *et al.*, "Independent Component Analysis: A Tool for Algal Bloom Detection," in *13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2023.

[5] D. Lupu *et al.*, "Quick Unsupervised Hyperspectral Dimensionality Reduction for Earth Observation: A Comparison," *IEEE Transactions on Computational Imaging*, vol. 11, 2025.

[6] V. Kopačková and L. Hladíková, "Applying Spectral Unmixing to Determine Surface Water Parameters in a Mining Environment," *Remote Sensing*, vol. 6, 2014.

[7] E. Alfaro-Mejía *et al.*, "A blind convolutional deep autoencoder for spectral unmixing of hyperspectral images over waterbodies," *Frontiers in Earth Science*, vol. 11, 2023.

[8] P. Naik *et al.*, "Blind and endmember guided autoencoder model for unmixing the absorbance spectra of phytoplankton pigments," *Scientific Reports*, vol. 15, 2025.

[9] E. T. Øien, "Dimensionality Reduction of hyperspectral remote sensing data," Master's thesis, Norwegian University of Science and Technology, 2025.

[10] A. Cutler and L. Breiman, "Archetypal Analysis," *Technometrics*, vol. 36, 1994.

[11] A. Zouaoui *et al.*, "Entropic Descent Archetypal Analysis for Blind Hyperspectral Unmixing," *IEEE Transactions on Image Processing*, vol. 32, 2023.

[12] B. Rasti *et al.*, "Image Processing and Machine Learning for Hyperspectral Unmixing: An Overview and the HySUPP Python Package," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, 2024.

[13] J. L. Garrett *et al.*, "Hyperspectral Unmixing Hierarchies, 2026," arXiv:2604.16969 [cs] version: 1.

[14] R. Vitale *et al.*, "On-The-Fly Processing of continuous high-dimensional data streams," *Chemometrics and Intelligent Laboratory Systems*, vol. 161, 2017.

[15] A. Alinejad *et al.*, "In-orbit Intra-calibration of a Hyperspectral Satellite Constellation," *In preparation*," 2026.

[16] C. L. McCarry *et al.*, "Determination of zooplankton absorption spectra and their potential contribution to ocean color," *Optics Express*, vol. 32, 2024.

[17] P. S. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sensing of Environment*, vol. 24, 1988.

[18] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*," 2011.