Parkinson Disease Identification using Residual Networks and Optimum-Path Forest

Leandro A. Passos*⋆, Clayton R. Pereira⋆ †, Edmar R. S. Rezende ‡, Tiago J. Carvalho §,
Silke A. T. Weber ¶, Christian Hook ‖, João P. Papa †

*UFSCAR - Federal University of São Carlos - Department of Computing - São Carlos, Brazil
leandro.passosjr@dc.ufscar.br
†UNESP - São Paulo State University - School of Sciences - Bauru, Brazil
{*papa,clayton*}*@fc.unesp.br*
‡CTI Renato Archer - Campinas, Brazil
edmar.rezende@gmail.com
§IFSP - Federal Institute of São Paulo - Department of Computing - Campinas, Brazil
tiagojc@gmail.com
¶UNESP - São Paulo State University - Medical School - Botucatu, Brazil
silke@fmb.unesp.br
‖Ostbayerische Tech. Hochschule - Fakultät Informatik/Mathematik - Regensburg, Germany
christian.hook@oth-regensburg.de

***Abstract*—Known as one of the most significant neurodegenerative diseases of the central nervous system, Parkinson's disease (PD) has a combination of several symptoms, such as tremor, postural instability, loss of movements, depression, anxiety, and dementia, among others. For the medicine, to point an exam that can diagnose a patient with such illness is challenging due to the symptoms that are easily related to other diseases. Therefore, developing computational methods capable of identifying PD in its early stages has been of paramount importance in the scientific community. Thence, this paper proposes to use a deep neural network called ResNet-50 to learn the patterns and extract features from images draw by patients. Afterwards, the Optimum-Path Forest (OPF) classifier is employed to identify Parkinson's disease automatically, being the results compared against two well-known classifiers, i.e., Support Vector Machines and the Bayes, as well as the ones provided by ResNet-50 itself. The experiments showed promising results concerning OPF, reaching over 96% of identification rate.**

***Index Terms*—Parkinson's Disease, Residual Networks, Machine Learning.**

I. Introduction

Parkinson's disease (PD) is a neurological illness that mainly affects the person's movements causing several symptoms, such as slowness of movements, freezing of gait, and muscle stiffness, also developing tremor and depression, as well as changes in speech and writing skills [1]. PD is degenerative, chronic, and with no early diagnosis disease that should be treated appropriately; otherwise, it can make the person completely invalid due to deterioration of brain functions, leading the individual to death.

The English physician James Parkinson [2] was the one that first described PD, in 1817, as a primary disease of complex causes usually beginning after age 50. Researchers say that Parkinson's disease affects approximately 5 millions people worldwide. Even though it is not a contagious disease, it is estimated to reach nearly 10 million individuals until 2030.

Nowadays, one of the major problem related to PD encountered by medicine is the correct classification of the disease level. Such difficulty led researchers to deal with the problem using computational techniques, such as image processing, pattern recognition, and signal analysis, providing promising results [3]. Das et al. [4], for instance, presented a comparison among some classification techniques concerning PD diagnosis, achieving around 92.2% of classification accuracy employing Neural Networks. Spadotto et al. [5] introduced the Optimum-Path Forest (OPF) classifier [6], [7], in the context of automatic PD identification. Gharehchopogh et al. [8] used of Artificial Neural Networks with Multi-Layer Perceptron to diagnose the effects caused by Parkinson's Disease, and Spadotto et al. [9] also considered using a meta-heuristic-driven feature selection approaches aiming at recognizing such illness.

Recently, Pereira et al. [10] proposed to extract features from writing exams using image processing techniques, and later on Pereira et al. [11] introduced Convolutional Neural Networks to learn features from handwriting dynamics in the context of automatic PD identification. Further, Pereira et al. [12] introduced Restricted Boltzmann Machines for the same purpose.

In this paper, we introduce deep residual networks in the context of Parkinson's disease identification, with special attention to the ResNet-50 [13] neural model. Furthermore, the data provided as output from ResNet-50 is used for classification purposes using the OPF, being the results compared against two well-known supervised classifiers, i.e., Support Vector Machines (SVM) and Bayes. Therefore, the contribu-

978-1-5386-4640-3/18/$31.00 ©2018 IEEE

tions of this paper are two-fold: (i) to introduce ResNet-50 in the context of Parkinson's Disease; and (ii) to foster the literature concerning PD early diagnosis.

The remainder of this paper is organized as follows. Section II presents the theoretical background related to Residual Networks, as well as the ResNet-50 architecture. Section III presents the dataset used in the experiments, and Section IV discusses the experimental results. Finally, Section V states conclusions and future works.

II. Theoretical Background

A. Residual Networks

Residual Networks (ResNets) [13] are a specific branch of deep neural networks (DNN) developed to deal with the degradation in optimization problems. To deal with such problem, He et al. [13] proposed a deep residual learning framework where sets of layers fit a residual map, in contrast with other architectures which hope that each layer fits the entire underlying map.

Let $H(x)$ be the final desired mapping, we have:

$$F(x) = H(x) - x, \tag{1}$$

where x represents the input image and $F(x)$ is the residual mapping. In a feed forward DNN, Equation 1 is treated as a shortcut connection [14]. In a more detailed way, we can rewrite Equation 1 as follows:

$$y_i = F(x_i, \{W_i\}) + x_i, \tag{2}$$

where y_i and x_i stand for the input and output at layer i^{th}, as well as W_i stands for the filter weights at the i^{th} layer. Finally, $F(x_i, \{W_i\})$ represents the residual mapping (also named "shortcut connection") to be learned.

Since these shortcuts need to be the same size of x, we can model them as *identity* or as a linear projection to match the dimensions. If this is not the case (e.g., when changing the input/output channels), we can perform a linear projection W_s by the shortcut connections to fit the output size:

$$y_i = F(x_i, \{W_i\}) + W_s x_i. \tag{3}$$

Figure 1 depicts an example of bottleneck blocks for ResNet-50, i.e., the blocks used to build the network, which are composed of convolution layers.

B. Transfer Learning Process

Yosinski et. al [15] defined transfer learning as the process of transferring the parameters from a model, trained for a specific task and dataset, to another with the same architecture but with a different dataset and task to solve.

In the DNN domain, solutions involving transfer learning usually make use of weights from previously trained networks (i.e., base model) as the starting point for the target network [16], [17]. Further, the new model can be fine-tuned using the well-known backpropagation learning algorithm. In a nutshell, one can replace the input and output layers of the base model by new ones from the target model and then train the network once again. However, if the dataset is not large enough and the number of parameters (i.e., layers and neurons) is quite considerable, fine-tuning the new model may lead to overfitting. Therefore, in this paper we froze the weights of the base model, i.e., we just kept the hidden layers and just replaced the input and output ones.

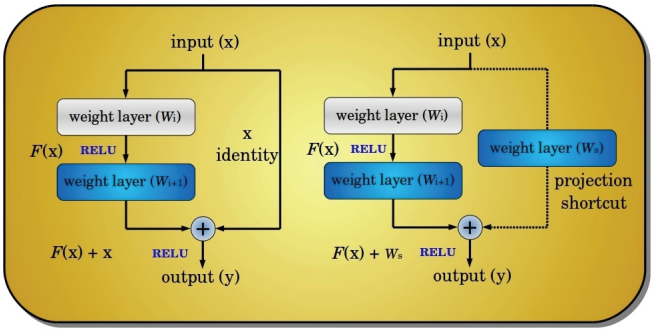

Fig. 1. Bottleneck Blocks for ResNet-50 (left: identity shortcut; right: projection shortcut with dimension of x different from shortcut output size).

Moreover, as mentioned earlier, despite the process of freezing layers may sound meaningless, many deep neural networks trained on natural images share a curious phenomenon: the first layers learn features that appear not to be specific to a particular task, but they usually generalize well to different situations. Features must eventually transition from coarser to finer layers of the network.

We opted to employ ResNet-50 as the base model, which was pre-trained for object detection task on the ImageNet 2012 dataset [18] containing 1.28 million images from $1,000$ classes. Further, we considered all hidden layers and just changed the top layer by a different classification technique, as discussed later.

III. Methodology

In this section, we describe the dataset and the feature extraction procedure using DNN adopted in the work.

A. HandPD Dataset

This work employs the "HandPD"[1] dataset [19], which is composed of images from handwritten exams (Figure 2). This dataset comprises activities from specific tasks acquired during handwriting exams from healthy individuals (i.e., control group) and patients. We used data from spirals and meanders, summing up a total of 368 images, being 296 from patients and 72 images from the control group.

The main idea of using such images concerns that patients usually figure different levels of tremors, that are usually associated with the progress of the disease. Therefore, it is expected that subtle differences during the handwriting exams may be noticed by a machine learning algorithm, but not for the human eye.

[1]Available at http://wwwp.fc.unesp.br/ papa/pub/datasets/Handpd/

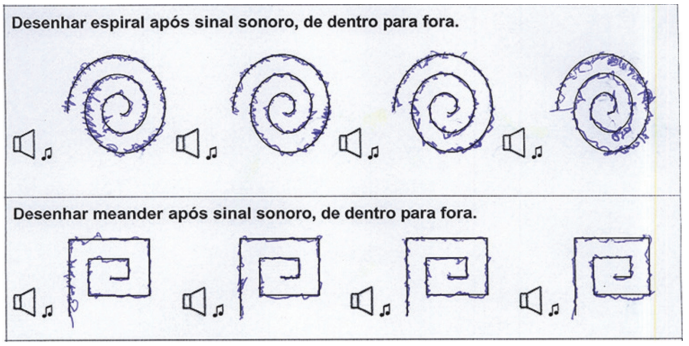

Fig. 2. Form used to assess the handwritten skills. Adapted from [19].

Further, the pre-processing step aimed at re-scaling the images to a resolution that is suitable to be used in the ResNet-50 architecture.

B. Feature Extraction Using Deep Neural Networks

This work employs a ResNet-50 architecture such that the input layer stands for the pixels from the input image ($240 \times 240 \times 3$), as well as the top-most layer concerns the output information, i.e., $2,048$ features. Once the ResNet-50 learning process is finished, the extracted features are mapped onto a 100-dimensional feature space using the Principal Component Analysis (PCA) technique.

The whole process outputs the features extracted from the dataset that will be further used to feed three different supervised learning techniques: SVM, OPF, and the Bayes classifier. Additionally, the recognition rates of those techniques will be compared against the ones obtained by ResNet-50 itself on its supervised phase.

In regard to SVM, we used a Radial Basis Function (RBF) kernel with parameters optimized with a grid search [20], and with respect to OPF we employed the LibOPF[2]. Concerning the Bayes classifier, we used our own implementation. Finally, to evaluate the techniques considered in this work, we compare the classification accuracy, the sensitivity, and specificity for each dataset.

Figure 3 depicts the transfer learning-based pipeline for feature extraction using ResNet-50 over HandPD dataset. Notice the learning does not require adjusting the filters, since the weights are transfered from the ones learned over ImageNet 2012 dataset [18].

IV. Experiments

This section presents the experimental results concerning ResNet-50, OPF, SVM, and the Bayes classifier to the task of Parkinson's Disease identification using features learned from handwritten images. The evaluation is taken upon a five-fold cross-validation, where each fold is composed of 80% of samples for training, and the remaining 20% are employed for testing purposes.

Tables I and II present the accuracy results concerning meander and spiral images, respectively. The most accurate accuracy results are in bold. Notice the values presented inside parentheses stand for the results concerning the original $2,048$ features extracted using ResNet-50 (i.e., without PCA), while the outside ones stand for the 100-dimensional vectors transformed through PCA. The main idea in considering these two scenarios is to evaluate the robustness of the techniques under lower and higher dimensional spaces. Notice the ResNet-50 was evaluated in one scenario oly (i.e., without PCA transformation).

-	ResNet-50	OPF	SVM	Bayes
Mean Acc	90.76%	**96.31% (85.36%)**	89.59% (91.57%)	83.56% (84.72%)
Control Acc	65.27%	(88.89% 58.33%)	61.11% (70.83%)	46.57% (55.55%)
Patient Acc	96.96%	97.97% (91.89%)	95.61% (96.62%)	91.55% (91.89%)
Sensitivity	0.8393	0.8889 (0.6363)	0.7719 (0.8360)	0.5483 (0.6250)
Specificity	0.9198	0.9731 (0.9006)	0.9100 (0.9315)	0.9155 (0.8947)

TABLE I
MEAN ACCURACY RESULTS CONSIDERING MEANDER IMAGES.

Regarding meander images, one can notice the results obtained by OPF outperformed all techniques, achieving an accuracy of 96.31% against 90.76% obtained by ResNet-50, 89.59% by SVM, and 83.56% using the Bayes classifier. We can observe that OPF obtained considerably better results over the control class, i.e., healthy patients. Although the reader can notice results far below the reasonable for ResNet-50, SVM, and Bayes considering the control class, HandPD dataset figures some patients that were diagnosed later, which means at the time they were asked to do the drawings, they were at the early stage of the disease, and with good handwriting skills.

Table II presents the results concerning the spiral images. Once again, OPF outperformed the results obtained by the remaining techniques, achieving an even better accuracy, nearly 97%. One can observe from both experiments (i.e., meanders and spirals) that OPF benefit from lower dimensional feature spaces, but this behavior cannot be held for SVM. Since the latter works finer in high-dimensional feature spaces, it is expected to obtain better recognition rates. On the other hand, OPF is based on distance computations, which is known to be strongly affected by high-dimensional spaces.

-	ResNet-50	OPF	SVM	Bayes
Mean Acc	91.58%	**96.71% (87.77%)**	92.62% (95.38%)	80.16% (87.50%)
Control Acc	75.00%	88.89% (76.38%)	79.17% (83.33%)	62.50% (75.00%)
Patient Acc	90.54%	97.64% (90.54%)	95.94% (98.31%)	84.45% (90.54%)
Sensitivity	0.8060	0.9014 (0.6626)	0.8261 (0.9230)	0.6250 (0.6585)
Specificity	0.9402	0.9731 (0.9403)	0.9496 (0.9603)	0.8445 (0.9370)

TABLE II
MEAN ACCURACY RESULTS CONSIDERING SPIRAL IMAGES.

One can observe that OPF obtained promising sensitivity and specificity values under the PCA transformation experiment, showing to be a reliable method for identifying both healthy individuals and patients. The confusion matrices regarding meander and spiral images after PCA transformation (except for ResNet-50) are presented in Figures 4 and 5, respectively. One can observe that OPF has been consistently more accurate than other techniques regarding both classes and the two scenarios considered in this work, i.e., meanders and

[2]https://github.com/jppbsi/LibOPF

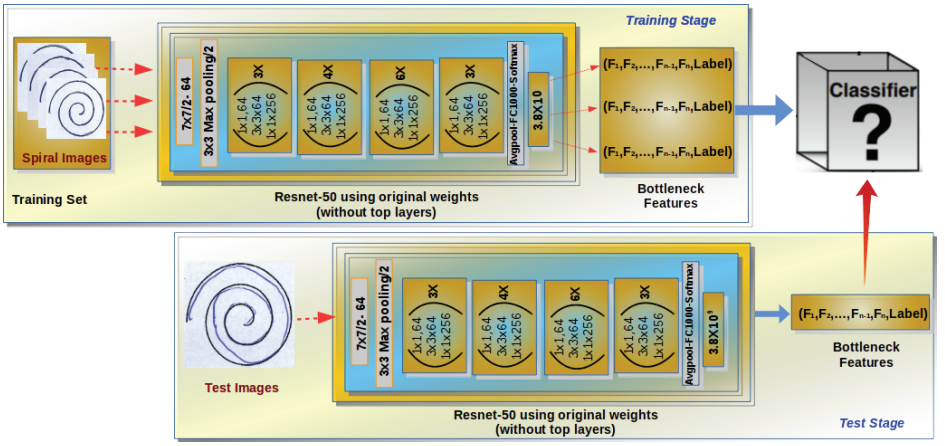

Fig. 3. Feature extraction of spiral images by means of transfer learning using ResNet-50. Inside the parenthesis we have the shapes each residual block, and outside we have the number of stacked blocks.

spirals. However, we did not observe any significant difference in terms of recognition rates between meanders and spirals.

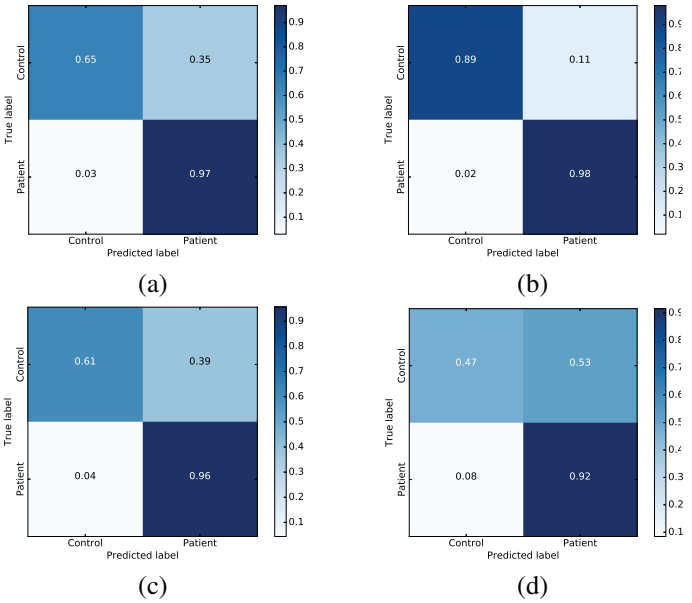

Fig. 4. Confusion matrices regarding meander images: (a) ResNet-50, (b) OPF, (c) SVM, and (d) Bayes.

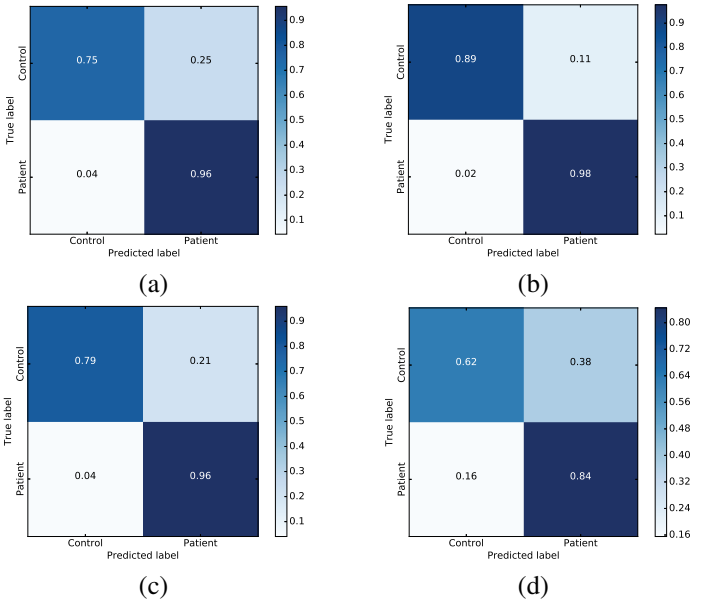

Fig. 5. Confusion matrices regarding spiral images: (a) ResNet-50, (b) OPF, (c) SVM, and (d) Bayes.

V. Conclusions and Future Works

We have presented the problem of automatic Parkinsons Disease identification using a transfer learning process applying residual networks. We considered two types of drawings that were used to feed a residual network for further supervised classification purposes.

We conducted two distinct experiments to evaluate the robustness of the techniques with feature spaces under different dimensions, i.e., without PCA and after PCA transformation. We observed that the best results were obtained with OPF after PCA transformation in both datasets, outperforming the other techniques. Furthermore, the OPF obtained the best results with accuracies nearly to 97%. As far as we are concerned, these are the best results over HandPD dataset to date.

Regarding future works, we are interested in analyzing the images with different deep architectures, as well as the signals extracted during the handwritting exam.

Acknowledgments

The authors would like to thank FAPESP grants #2013/07375-0, #2014/16250-9, #2014/12236-1, #2015/25739-4 and #2016/21243-7, as well as Capes, and CNPq grants #306166/2014-3 and #307066/2017-7.

References

[1] Burke and E. Robert, "Evaluation of the braak staging scheme for parkinson's disease: Introduction to a panel presentation," *Movement Disorders*, vol. 25, no. S1, pp. S76–S77, 2010.

[2] J. Parkinson, *An Essay on the Shaking Palsy*. Whittingham and Rowland, 1817.
[3] B. Sakar, M. Isenkul, C. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 4, pp. 828–834, July 2013.
[4] R. Das, "A comparison of multiple classification methods for diagnosis of parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568 – 1572, 2010.
[5] A. A. Spadotto, R. C. Guido, J. P. Papa, and A. X. Falcão, "Parkinson's disease identification through optimum-path forest," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010, pp. 6087–6090.
[6] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.
[7] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares, "Efficient supervised optimum-path forest classification for large datasets," *Pattern Recognition*, vol. 45, no. 1, pp. 512–520, 2012.
[8] F. S. Gharehchopogh and P. Mohammadi, "Article: A case study of parkinsons disease diagnosis using artificial neural networks," *International Journal of Computer Applications*, vol. 73, no. 19, pp. 1–6, 2013.
[9] A. A. Spadotto, R. C. Guido, R. F. Carnevali, A. F. Pagnin, J. P. Papa, and A. X. Falcão, "Improving parkinson's disease identification through evolutionary-based feature selection," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010, pp. 7857–7860.
[10] C. R. Pereira, D. R. Pereira, F. A. Silva, J. P. Masieiro, S. A. T. Weber, C. Hook, and J. P. Papa, "A new computer vision-based approach to aid the diagnosis of parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 136, pp. 79–88, 2016.
[11] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *Proceedings of the 29th SIBGRAPI Conference on Graphics, Patterns and Images*, 2016, pp. 340–346.
[12] C. R. Pereira, L. A. Passos, R. R. Lopes, S. A. T. Weber, C. Hook, and J. P. Papa, "Parkinsons disease identification using restricted boltzmann machines," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2017, pp. 70–80.
[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016, pp. 770–778.
[14] C. Bishop, *Neural networks for pattern recognition*. Oxford University Press, USA, 1995.
[15] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Adv Neural Inf Process Syst*, 2014, pp. 3320–3328.
[16] E. Rezende, G. Ruppert, and T. Carvalho, "Detecting computer generated images with deep convolutional neural networks," in *Graphics, Patterns and Images (SIBGRAPI), 2017 30th SIBGRAPI Conference on*. IEEE, 2017, pp. 71–78.
[17] T. Carvalho, M. Teles, E. Rezende, F. Balieiro, and R. Sovat, "Exposing Computer Generated Images by Eyes Features Extraction via Transfer Learning of VGG19 CNN," in *Proceedings of the 16th International Conference On Machine Learning And Applications (ICMLA'17)*, december 2017.
[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *Springer IJCV*, vol. 115, no. 3, pp. 211–252, 2015.
[19] C. R. Pereira, D. R. Pereira, F. A. Silva, C. Hook, S. A. T. Weber, L. A. M. Pereira, and J. P. Papa, "A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements," in *IEEE 28th International Symposium on Computer-Based Medical Systems*, 2015, pp. 171–176.
[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.