



**HAL**  
open science

## Supervised diagnosis prediction from cortical sulci: toward the discovery of neurodevelopmental biomarkers in mental disorders

Pierre Auriiau, Antoine Grigis, Benoit Dufumier, Robin Louiset, Joel Chavas,  
Pietro Gori, Jean-François Mangin, Edouard Duchesnay

### ► To cite this version:

Pierre Auriiau, Antoine Grigis, Benoit Dufumier, Robin Louiset, Joel Chavas, et al.. Supervised diagnosis prediction from cortical sulci: toward the discovery of neurodevelopmental biomarkers in mental disorders. 21st IEEE International Symposium on Biomedical Imaging (ISBI 2024), IEEE, May 2024, Athènes, Greece. hal-04494994

**HAL Id: hal-04494994**

**<https://hal.science/hal-04494994>**

Submitted on 11 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SUPERVISED DIAGNOSIS PREDICTION FROM CORTICAL SULCI: TOWARD THE DISCOVERY OF NEURODEVELOPMENTAL BIOMARKERS IN MENTAL DISORDERS

*Pierre Auriiau<sup>1</sup>, Antoine Grigis<sup>1</sup>, Benoit Dufumier<sup>1</sup>, Robin Louiset<sup>1</sup>, Joël Chavas<sup>1</sup>,  
Pietro Gori<sup>2</sup>, Jean-François Mangin<sup>1</sup>, Edouard Duchesnay<sup>1</sup>*

<sup>1</sup> NeuroSpin, CEA Saclay, Université Paris-Saclay, France

<sup>2</sup> LTCI, TélécomParis, IPParis, France

## ABSTRACT

Recent advances in machine learning applied to structural magnetic resonance imaging (sMRI) may highlight abnormalities in brain anatomy associated with mental disorders. These disorders are multifactorial, resulting from a complex combination of neurodevelopmental and environmental factors. In particular, such factors are present in cortical sulci, whose shapes are determined very early in brain development and are a valuable proxy for capturing specifically the neurodevelopmental contribution of brain anatomy. This paper explores whether the shapes of cortical sulci can be used for diagnosis prediction using deep learning models. These models are applied to three mental disorders (autism spectrum disorder, bipolar disorder, and schizophrenia) in large multicentric datasets. We demonstrate that the neurodevelopmental underpinnings of these disorders can be captured with sMRI. Finally, we show the potential of visual explanations of models' decisions in discovering biomarkers for mental disorders.

**Index Terms**— deep learning, cortical sulci, psychiatric disorders, neurodevelopment

## 1. INTRODUCTION

Brain anatomy is an imprint of an individual's neurodevelopmental and environmental backgrounds. Brain anatomy abnormalities might feature either neurodevelopmental trajectories to psychiatric disorders or later deviations in non-pathological brain maturation that are harmful consequences of environmental interactions (excessive stress, abuse, etc.). Researching the disentanglement of these variability sources may offer essential insights concerning the individual prediction of clinical outcomes (e.g., treatment response), identification of prognostic brain signatures, and prevention of early negative influences on the developing child.

Machine Learning (ML) models offer a useful approach for more personalized diagnostic practices in psychiatry (i.e., precision psychiatry). Several models trained on structural MRI (sMRI) have highlighted abnormalities in brain anatomy that correlate with various mental disorders [1, 2]. Most

works that developed ML models use voxel-based morphometry (VBM) to warp the T1-weighted (T1w) images into a standard template to produce registered and comparable inputs for ML models. Unfortunately, these models cannot determine whether the abnormalities come from environmental or neurodevelopmental factors because VBM features retain information related to both factors [3, 4, 5].

To identify neurodevelopmental abnormalities, previous studies focused on folding patterns of cortical sulci, which are features of brain anatomy primarily impacted by early brain development [6]. Cortical sulci are a valuable proxy for capturing the neurodevelopmental contribution of brain anatomy because their shapes are set very early in brain development. Sulcus shapes are particularly relevant features as they are already known to keep crucial information about brain development and have already been linked to mental diseases [7, 8, 9].

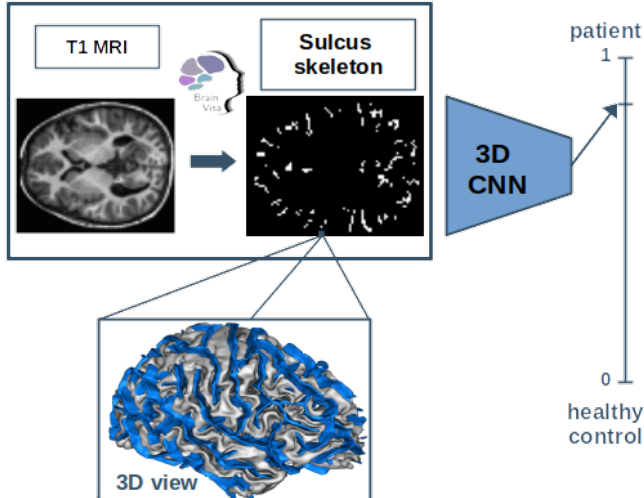
This article explores whether such worthwhile brain landmarks can be used for diagnosis prediction using deep learning models (see Fig. 1). Our model was tested on three mental disorders: autism spectrum disorder (ASD), bipolar disorder (BD) and schizophrenia (SCZ). We compared several training strategies and image preprocessing to address the sparsity and nature of the input data. Results show that cortical sulci contain sufficient information for reliable predictions. Finally, we showed the potential of feature importance techniques in highlighting cortical sulci potentially linked to mental disorders.

## 2. MATERIAL AND METHODS

### 2.1. Datasets

Three pooled clinical studies with sMRI were considered, one for each of the psychiatric pathologies: SCZ, BD, and ASD (for details, see Table 1). Each dataset consists of T1w images from patients and healthy controls (HC). Each image is processed through the Brainvisa/Morphologist pipeline<sup>1</sup>, which provides a sulcus skeleton for each individual. The sulcus skeleton is a binary image in which non-zeros vox-

<sup>1</sup><https://brainvisa.info>



**Fig. 1.** sMRI T1w images are pre-processed with the BrainVisa/Morphologist pipeline to obtain sulcus skeletons. Skeletons represent cortical folding. They are shown in blue in the 3D view. A 3D Convolutional Neural Network (CNN) is trained to predict the diagnosis based on these skeletons.

Dataset	# Subjects		# Sites	Pooled studies
	HC	PT		
SCZ	761	532	12	schizconnect-vip-prague cnp, candi, bsnip1
BD	695	469	15	biobd, bsnip1, cnp, candi
ASD	926	813	30	abide 1 & 2

**Table 1.** Number of healthy controls (HC), patients (PT) and sites in each considered SCZ, BD, and ASD dataset.

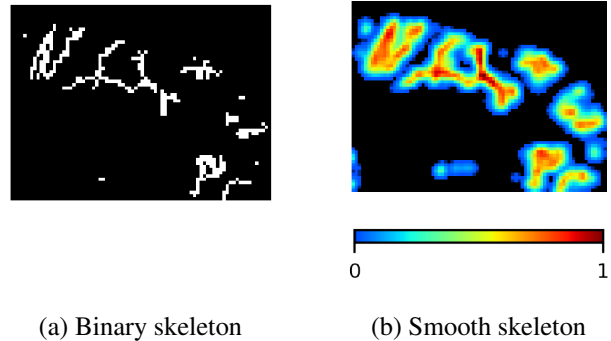
els form a negative cast of the brain’s folds [10, 11] (see Fig. 1). The ventricles are removed, and the resulting skeletons are affinely registered to the MNI template space with an isotropic resolution of 1.5 mm<sup>3</sup> using the DeepFolding toolbox<sup>2</sup>. A visual quality check was performed on the outlier skeletons detected by thresholding their size, calculated as the number of non-zero voxels in the image.

Each dataset was split into training, validation, and test sets. We used an iterative algorithm to stratify these splits on diagnosis, age, sex, and site [12]. The algorithm attempts to keep the histograms of the selected variables unchanged between the training and test sets. Note that the splits for all experiments are kept unchanged for fair comparisons.

## 2.2. Skeleton image pre-processing

Sulcus skeleton images contain binary values and are very sparse (i.e., less than 5% of the voxels are non-zero), which is

<sup>2</sup>[https://github.com/neurospin/deep\\_folding](https://github.com/neurospin/deep_folding)



**Fig. 2.** Zoom in on a sulcus skeleton a) with and b) without Gaussian smoothing.

not optimal for neural network training. To tackle this problem, we tried applying Gaussian smoothing to sulcus skeletons. The goal was to thicken the cortical sulci while preserving their shapes (e.g., we did not want to merge two parallel sulci). For this purpose, we chose a Gaussian kernel with a moderate full width at half maximum (FWHM) of 3.5mm and a small kernel size of 5x5x5 voxels (see Fig. 2). A visual quality check was performed to ensure that the sulcus shapes were preserved. Note that such a transformation converts skeleton images from binary to continuous values.

## 2.3. Network Training

Our goal is to build the best deep learning model to predict the clinical status of subjects in a supervised setting.

**Supervised learning:** First, we tried the straightforward approach by optimizing convolutional neural network (CNN) weights with a binary cross-entropy (BCE) loss. We benchmarked three well-known CNN architectures, namely AlexNet [13, 14], ResNet18 [15], and DenseNet121 [16]. While Alexnet consists of only a few convolutional layers, the other two architectures integrate skip connections to maintain information across layers.

**Weakly supervised learning:** Recent results show better stability with supervised contrastive learning over supervised learning with a BCE loss [17]. In contrastive learning, training attempts to bring positive pairs of augmented images together in the latent space and push negative pairs away [18]. In weakly supervised setups, labels are used to define pairs. In that case, positive pairs are images from the same class in the mini-batch.

We proposed to use the SupCon supervised contrastive model [17]. The SupCon loss is defined as:

$$L_{SupCon} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)}$$

where  $I$  is the set with all the samples,  $A(i)$  is the set of samples in the mini-batch that contains the sample  $i$ ,  $P(i)$  is the

set of positive samples in the mini-batch  $A(i)$  (samples from the same class as  $i$ ), and  $\tau$  is the temperature. Note that the temperature is an experimentally determined hyperparameter.

**Training settings:** When training the models, we used a grid search to find the best combination of hyperparameters: learning rate, scheduler, and temperature for weakly supervised learning models. For each set of hyperparameters, the model was trained three times with different random initializations. In this way, the epistemic variability was captured. An L2 penalty on the weights was added to the loss, ponderated by  $5 \cdot 10^{-5}$ . The models were trained for 50 epochs. For contrastive learning models, data augmentation was added as in [17]. In our case, we tried adding rotations, cutouts, and combinations of both.

**Evaluation settings:** We computed the area under the receiver operating characteristic curve (ROC AUC) of model predictions on test sets to evaluate model performance. For contrastive learning models, we trained a logistic regression from the latent representations of the training set to predict the diagnosis. We then evaluated the trained models on the test set and calculated the ROC AUC. To test if our results were above randomness, we computed the  $p_{value}$  from the U Mann-Whitney test between predicted probabilities of HC and those of patients.

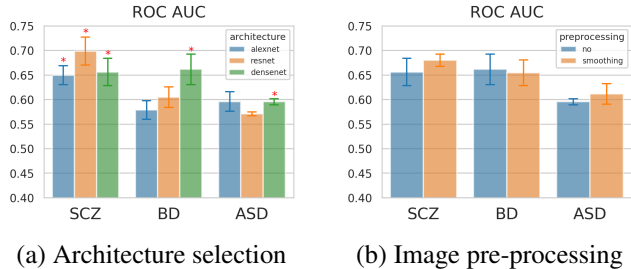
**Feature importance:** Perturbation-based visual explanations are one of the approaches to interpreting the decision-making process of deep learning models. These approaches indicate the importance of different voxels for the final decision taken by models. Among these approaches, we chose occlusion sensitivity [19]. Different regions of skeleton images were occluded, and the changes in the prediction scores were recorded. We used a patch size of  $16 \times 16 \times 16$  and a stride of 4 voxels. We chose a large patch size because skeletons were registered using a coarse affine transformation. The importance of each patch was measured by computing the difference between the binary cross entropy of predictions with and without occlusion, averaged over test set subjects.

The experiments’ reproducing codes are available here.

### 3. RESULTS

**Cortical folding is predictive of diagnosis:** All trained models can predict the diagnosis on the test set with a ROC AUC greater than 0.57 (see Fig. 3). Thus, cortical skeleton images contain sufficient information to predict the diagnosis of SCZ, BD, and ASD. In addition, the three classification tasks can be ordered by difficulty: SCZ vs HC, BD vs HC, and ASD vs HC. The same conclusion was reported in a similar VBM classification study [20]. In particular, the drop in performance on the ASD dataset can be attributed to the complexity of the pathology, which is known to have many different symptoms.

**Deeper models improve predictions:** The three benchmarked CNN architectures (see section 2.3) have close per-



**Fig. 3.** (a) Comparison of three CNN architectures: AlexNet, ResNet, and DenseNet for diagnosis prediction. (b) Comparison of models trained with smooth skeletons and binary skeletons for diagnosis prediction. In both cases, ROC AUC values are averaged on three training sets, error bars are standard deviations and red stars mean  $p_{val} < 0.05$ .

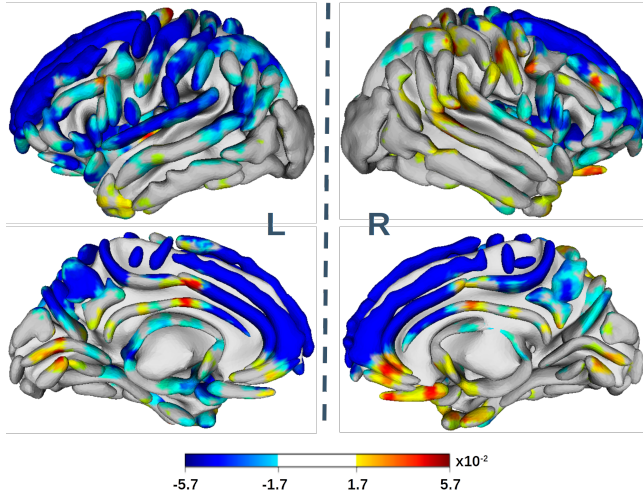
	SCZ	BD	ASD
BCE	0.656 $\pm$ 0.034	<b>0.661</b> $\pm$ 0.038	0.595 $\pm$ 0.008
SupCon	<b>0.675</b> $\pm$ 0.006	0.656 $\pm$ 0.044	<b>0.609</b> $\pm$ 0.001

**Table 2.** Comparison of models trained with BCE and SupCon loss. The models are trained to predict diagnosis on the SCZ, BD, and ASD datasets. Mean ROC AUC values and associated standard deviations are calculated from the three training iterations on the test set.

formances on the three classification tasks considered (see Fig. 3a). DenseNet is the only architecture that gives significant predictions (above chance  $p_{val} < 0.05$ ) on all test sets. Specifically, it can predict the diagnosis with an ROC AUC of 0.661 for BD vs HC, 0.656 for SCZ vs HC, and 0.595 for the ASD vs HC predictions. For the following experiments, we kept the DenseNet architecture.

**Weakly supervised and supervised learning perform similarly:** Models were trained by optimizing a SupCon loss (see section 2.3) to make diagnosis predictions on the three datasets. We got comparable results with the three tested data augmentations, so we kept only cutouts to save computation time. The performances were compared with those obtained with a supervised model implementing a BCE loss (see Table 2). The Supcon loss performs better on SCZ and ASD but worse on BD. Thus, we kept a supervised model with binary BCE loss in the following experiments.

**Thickening skeletons does not help model training:** Gaussian smoothing was applied to the skeleton images (see section 2.2), and models were trained and evaluated with these newly transformed datasets. No data augmentation was performed at this stage, and the Gaussian smoothing parameters were fixed. We compared these results with those obtained without Gaussian smoothing (see Fig. 3b). The two settings gave close performances on the three test sets.



**Fig. 4.** Sulci importance on the model’s diagnosis prediction for subjects of the SCZ test set. Positive (red) values correspond to an increase in error when the region is occluded, revealing important regions for the prediction. Negative (blue) values highlight a decrease in error when the region is occluded, revealing regions that bring noise in the model.

We decided not to use Gaussian smoothing in the following experiments.

**Explainable artificial intelligence (XAI):** As deep learning based methods proliferate, there is a growing demand for their explainability, especially in high-stakes decision areas such as medical image analysis. As part of XAI, we computed occlusion sensitivity maps from the test set of the SCZ dataset (see section 2.3). To further improve interpretability, the resulting map was projected back onto a mean sulci template [21] (see Fig. 4). In this preliminary work, we can already see parts of cortical sulci brought out by the model. Such a feature importance technique can highlight part of sulci implicated in mental disorders.

#### 4. CONCLUSION AND FUTURE WORKS

We established a deep learning framework that can predict the diagnosis from cortical sulci for three mental disorders: ASD, BD and SCZ. Our results confirm a neurodevelopmental contribution to each of these mental disorders that can be observed in sMRI. Compared to state-of-the-art predictions, the predictions obtained are weaker but still significant. It could be explained as sulcal skeleton images contain less information than other features inferred from sMRI T1w images. Interestingly, the most straightforward approach works best. The best model results from supervised training that optimizes a binary cross-entropy loss with binary skeletons as inputs. Data transformations with Gaussian smoothing did not improve the prediction performance. Note that supervised contrastive learning does not help improve our predictions

over fully supervised learning, perhaps because classical data augmentations are not well suited to skeleton images. In future works, we would like to develop skeleton-specific data augmentations by drawing inspiration from data augmentations developed for brain sMRI [22].

All cortical sulcus deviations related to psychiatric disorders described in the literature are found between neurodevelopmental patients and other patients or HC [7, 9]. Based on this analysis, our model may correctly classify only neurodevelopmental patients among the whole patient population. In our work, patients cannot be split into two sub-groups because images come from aggregated studies that do not have the same phenotyping of patients. Our model predictions may have reached a glass ceiling with patients as a whole group. To further investigate this hypothesis, we would like to apply our model to a dataset from which we can split patients into a neurodevelopmental sub-group and compare the prediction performances of the two sub-groups of patients. On the other hand, we would like to try to improve our model predictions by taking advantage of a weakly supervised contrastive pre-training on a large dataset of HC, which has already been shown to improve significantly the performance and stability of the deep learning predictions [20].

Finally, the ability to explain the model is of fundamental importance in machine learning. Keeping a machine learning model as a “black box” is no longer an option, especially in neuroimaging. The proposed average maps allow us to highlight sulci associated with psychiatric disorders. However, the occlusion sensitivity approach fails to capture long-range dependencies between brain regions. A finer approach should be used to state cortical sulci involved in mental disorders, for instance, the SHAP approach [23]. In addition, our fixed-size patch approach may be too simple because a patch may occlude different brain areas across subjects. In future work, we would like to use more advanced techniques to define patches, for instance, by taking advantage of brain atlases.

#### 5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data under the local ethics policy.

#### 6. ACKNOWLEDGMENTS

The authors declare that they have no conflict of interest. This work was funded by French National Research Agency for the project Big2Small (Chair in AI, ANR-19-CHIA-0010-01), the project RHU-PsyCARE (French government’s “Investissements d’Avenir” program, ANR-18-RHUS-0014), and European Union’s Horizon 2020 for the project R-LiNK (H2020-SC1-2017, 754907). The authors acknowledge Denis Rivière and Clara Fischer for pre-processing raw T1w images.

## 7. REFERENCES

- [1] Thomas Wolfers et al., “From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics,” *Neuroscience & Biobehavioral Reviews*, vol. 57, pp. 328–349, 2015.
- [2] Anissa Abi-Dargham et al., “The search for imaging biomarkers in psychiatric disorders,” *Nature medicine*, vol. 22, no. 11, pp. 1248–1255, 2016.
- [3] Ronald J Janssen et al., “Making individual prognoses in psychiatry using neuroimaging and machine learning,” *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 3, no. 9, pp. 798–808, 2018.
- [4] Walter HL Pinaya et al., “Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study,” *Human brain mapping*, vol. 40, no. 3, pp. 944–954, 2019.
- [5] Mengjiao Hu et al., “Brain mri-based 3d convolutional neural networks for classification of schizophrenia and controls,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 1742–1745.
- [6] J Dubois et al., “Primary cortical folding in the human newborn: an early marker of later functional development,” *Brain*, vol. 131, no. 8, pp. 2028–2041, 2008.
- [7] G Auzias et al., “Atypical sulcal anatomy in young children with autism spectrum disorder,” *NeuroImage: Clinical*, vol. 4, pp. 593–603, 2014.
- [8] S Hotier et al., “Social cognition in autism is associated with the neurodevelopment of the posterior superior temporal sulcus,” *Acta Psychiatrica Scandinavica*, vol. 136, no. 5, pp. 517–525, 2017.
- [9] Colleen PE Rollins et al., “Evidence in cortical folding patterns for prenatal predispositions to hallucinations in schizophrenia,” *Translational Psychiatry*, vol. 10, no. 1, pp. 387, 2020.
- [10] Jean-François Mangin et al., “From 3d magnetic resonance images to structural representations of the cortex topography using topology preserving deformations,” *Journal of Mathematical imaging and Vision*, vol. 5, pp. 297–318, 1995.
- [11] Louise Guillon et al., “Identification of rare cortical folding patterns using unsupervised deep learning,” *arXiv preprint arXiv:2211.16213*, 2022.
- [12] Konstantinos Sechidis et al., “On the stratification of multi-label data,” in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III* 22. Springer, 2011, pp. 145–158.
- [13] Alex Krizhevsky et al., “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [14] Anees Abrol et al., “Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning,” *Nature communications*, vol. 12, no. 1, pp. 353, 2021.
- [15] Kaiming He et al., “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [16] Gao Huang et al., “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [17] Prannay Khosla et al., “Supervised contrastive learning,” *Advances in neural information processing systems*, vol. 33, pp. 18661–18673, 2020.
- [18] Ting Chen et al., “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.
- [19] Matthew D Zeiler et al., “Visualizing and understanding convolutional networks,” in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I* 13. Springer, 2014, pp. 818–833.
- [20] Benoit Dufumier, *Representation learning in neuroimaging: transferring from big healthy data to small clinical cohorts*, Ph.D. thesis, universit  Paris-Saclay, 2022.
- [21] Matthieu Perrot et al., “Cortical sulci recognition and spatial normalization,” *Medical image analysis*, vol. 15, no. 4, pp. 529–550, 2011.
- [22] Mehmet Saygın Seyfiođlu et al., “Brain-aware replacements for supervised contrastive learning in detection of alzheimer’s disease,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 461–470.
- [23] Scott M Lundberg et al., “A unified approach to interpreting model predictions,” *Advances in neural information processing systems*, vol. 30, 2017.