Neuroinformatics manuscript No.
(will be inserted by the editor)

A learning based framework for disease prediction from images of human-derived pluripotent stem cells of schizophrenia patients

Nickolas Fulaczyk · Jessica Di Re · Laura Stertz · Consuelo Walss-Bass · Fernanda Laezza · Demetrio Labate

Preprint date: May 8, 2022

Abstract Human induced pluripotent stem cells (hiPSCs) have been employed very successfully to identify molecular and cellular features of psychiatric disorders that would be impossible to discover in traditional postmortem studies. Despite the wealth of new available information though, there is still a critical need to establish quantifiable and accessible molecular markers that can be used to reveal the biological causality of the disease. In this paper, we introduce a new quantitative framework based on supervised learning to investigate structural alterations in the neuronal cytoskeleton of hiPSCs of schizophrenia (SCZ) patients. We show that, by using Support Vector Machines or selected Artificial Neural Networks trained on image-based features associated with somas of hiPSCs derived neurons, we can predict very reliably SCZ and healthy control cells. In addition, our method reveals that βIII tubulin and FGF12, two critical components of the cytoskeleton, are differentially regulated in SCZ and healthy control cells, upon perturbation by GSK3 inhibition.

Keywords Convolutional neural networks · fluorescence microscopy · PI3k/GSK3 pathway · human induced pluripotent stem cells · image processing · schizophrenia · statistical matrices

N. Fularczyk
Department of Mathematics
University of Houston
E-mail: nfularcz@Central.UH.EDU

J. Di Re
Department of Pharmacology & Toxicology
University of Texas Medical Branch
E-mail: jdire@utmb.edu

L. Stertz
Department of Psychiatry and Behavioral Sciences
UT Health
E-mail: Laura.Stertz@uth.tmc.edu

C. Walss-Bass
Department of Psychiatry and Behavioral Sciences
UT Health
E-mail: Consuelo.WalssBass@uth.tmc.edu

F. Laezza
Department of Pharmacology & Toxicology
University of Texas Medical Branch
E-mail: felaezza@utmb.edu

D. Labate
Department of Mathematics
University of Houston
E-mail: dlabate@math.uh.edu

1 Introduction

Human induced pluripotent stem cells (hiPSCs) have emerged as a very useful tool for studying a variety of human diseases allowing for the creation of accessible patient-derived cellular models. In the context of psychiatric disorders – that are notoriously heterogenous and complex – patient-derived hiPSCs have been used to establish molecular and cellular features of the disease that would have been otherwise impossible to reveal in traditional postmortem tissue samples [7, 11, 22]. As a result, by employing real time experimental perturbations in patient-derived live cells, researchers can identify disease-related cellular responses or *functional cellular endophenotypes*, hence expanding the diagnostic and therapeutic framework for psychiatric disorders.

Among psychiatric disorders, schizophrenia (SCZ) stands out for its complexity. Especially in this context, starting with the first seminal evaluation of cellular and molecular phenotypes of neural cells from SCZ patient-derived hiPSCs in [7], hiPSCs studies have been remarkably successful in unraveling cellular and molecular phenotypes that might be involved in the biological causality of SCZ [2]. For instance, reduced neurite outgrowth, decreased number of neurites,

reduced neural connectivity, synaptic disregulation and kinase signal abnormalities were found in hiPSC-derived neurons of SCZ patients [5, 33, 35] - observations that are consistent with animal models or postmortem studies. In addition, advanced DNA/RNA sequencing technologies at both population and single-cell levels in combination with patient derived hiPSCs have enabled the identification of molecular phenotypes relevant to the pathogenesis of SCZ [6]. Within this line of investigation, a recent study by some of the authors employed differential genomic profiling of hiPSC-derived neurons in healthy control vs. SCZ patients searching for gene expression alterations that could lead to mechanistic hypotheses of the disease [26]. This work revealed alterations in the phosphoinositide 3-kinase (PI3K) and glycogen synthase kinase 3 (GSK3) signaling pathways as a disease-related feature. Specifically, confocal imaging revealed that the expression levels of fibroblast growth factor 12 (FGF12), a protein associated with axonal transport, and the neuronal cytoskeleton component βIII Tubulin were differentially regulated in neurites of hiPSC-derived neurons of healthy control and SCZ upon perturbation by inhibitors of kinases in the PI3K/GSK3 pathway.

Although patient-derived hiPSCs studies have associated SCZ to a multitude of genetic factors and intertwined endophenotypes, the molecular and cellular mechanisms underlying the disease are still poorly understood and there is an unmet need to identify markers based on measurable and easily accessible quantities, that could be used as biomarkers of the disease. Specifically, while results in the literature suggest that dysregulation of the kinase pathway may functionally affect the neuronal cytoskeleton in SCZ, there is currently no quantitative or predictive analysis of the relationship connecting measurable cytoskeleton alterations, kinase signal abnormalities and SCZ.

To address this outstanding need, in this paper we carry out a quantitative investigation of structural alterations in the neuronal cytoskeleton of patient-derived hiPSCs with the aim to identify accessible image-base molecular biomarkers that can be used to reliably predict SCZ in hiPSCs images. To do so, we have introduced a computational algorithm for the extraction of image-based features associated with the cytoskeleton of the somas of hiPSC-derived neurons that we have combined with a targeted supervised learning classifier. To provide both high accuracy and interpretability of disease prediction, we have developed two distinct methods: (i) our first method uses support vector machines (SVMs) that we trained on a targeted set of handcrafted features; (ii) our second method applies artificial neural networks, namely a convolutional neural network (CNN) or a vision transformer (ViT) and is fully data-based. The outcome of our supervised learning methods shows that structural alterations measured from soma images of hiPSC-derived neurons *at the single cell level* constitute a reliable predictor of disease status, with such alterations being correlated to disruption of the GSK3 signaling pathway. To our knowledge, this is the first result of this type in the context of HiPCSs studies of SCZ. By identifying image-based biomarkers of SCZ in hiPSC-derived neurons using a supervised classifier at single cell level, our approach offers a powerful avenue for elucidating cellular mechanisms that contribute to SCZ onset and progression.

2 Materials and Method

2.1 Materials

Cell preparation. Human iPSCs processed and analyzed in the present study derive from a previously published dataset [26]. Briefly, de-identified cell lines included in this study were generated from a Central Valley of Costa Rica SCZ multiplex family and included 6 individuals, namely 3 females and 3 males, of which 3 male siblings are affected with SCZ and all females are unaffected and served as healthy controls. Due to the disease history in the family, cell lines could not be matched for sex, which is an intrinsic limitation of the original and the present studies. All subjects were carefully characterized in previous studies, according with the Principles of the Declaration of Helsinki, and lymphoblastoid cell lines were already generated from each subject, as previously described in [26].

Immunocytochemistry. We describe below cell preparation recalling what was presented in a previous work [26]. After 24 hour treatment with 20 μm CHIR99021 (Tocris), a potent and selective GSK3 inhibitor or 0.5% DMSO (Vehicle) coverslips were fixed with 4% PFA/4% sucrose for 15 minutes, permeabilized using 0.25% Triton X-100 in PBS and blocked with 10% BSA for 30 minutes. Cells were stained using primary antibodies ms IgG1 FGF12 (NeuroMabs, catalog # 75-135, 1:200), ms IgG2a Nav1.2 (Neuromabs, catalog # 75-204, 1:500), Ck MAP2 (Invitrogen catalog #pa1-10005, 1:2000), and Rb βIII Tubulin (Abcam, catalog #ab18207; 1:2000) overnight at 4°C in 3% BSA. Isotype specific secondary antibodies ms IgG1 Alexa 647 (Life Technologies, 1:200), ms IgG2a Alexa 568 (Life Technologies, 1:200), rb Alexa 488 (Life Technologies, 1:200) Ck Dylight 405 (Jackson ImmunoResearch Laboratories, catalog # 103-475-155, 1:200) in 3% BSA were applied for 2 hours. Cells were washed and mounted using ProLong Gold Antifade Mountant (ThermoFisher, catalog # P36930).

Image acquisition. Confocal images were acquired with a Zeiss LSM-880 with Airy scan confocal microscope and a 63X oil immersion objective (1.4 NA). Multi-track acquisition was done with excitation lines at 405nm for Jackson Dylight 405 (MAP2 Channel), 488nm for Alexa 488 (βIII tubulin channel), 561nm for Alexa 568 (Nav1.2 channel) and 633nm for Alexa 647 (FGF12 channel). Z-stacks were

acquired every 0.43 μm with a frame size of 1024×1024 pixels and a scan speed of 7. The pinhole was set to 1AU for for the longest wavelength used in the acquisition (633nm) and the step size of the z-stack was kept at the optimal setting in the Zeiss software. Acquisition parameters, including photomultiplier gain and offset, were kept constant for each protein of interest.

Images used in this paper were projected using sum projections. For the later processing stage, region of interests delineating the boundaries of the somas were manually drawn on an overlay of the βIII tubulin and MAP2 image channels. A representative example showing the delineation of a soma boundary is reported in Fig. 1.

Dataset. Our dataset consists of 89 fluorescent image stacks, including a total of 340 cells, specifically, 65 HC-DMSO cells, 110 HC-CHIR cells, 67 SCZ-DMSO cells, and 98 SCZ-CHIR.

2.2 Method

We designed a supervised learning approach aimed at predicting the presence of disease (at the single cell level) in multi-channel fluorescent images of hiPSCs derived neurons and to quantify the impact of GSK3 inhibition on disease prediction and manifestation. To carry out such task, we considered two strategies:

1. a method using SVMs trained with a specially selected class of handcrafted features;
2. a fully data-driven approach implemented using a CNN or a ViT.

Before describing in detail our implementation of these two types of strategies, we briefly summarize general advantages and disadvantages of a SVM as compare to a CNN or ViT approach.

We recall that a SVM is a non-probabilistic algorithm that computes the optimal linear classifier [31]. Common to other machine learning algorithms, non-linear decisions can be implemented through an SVM using appropriate *kernel functions* (also called *filters*) as an embedding into a higher dimensional space (the feature space). A main advantage of SVM classifiers is that they can be trained with relatively few training samples, especially if feature selection (through appropriate kernel functions) is guided by domain-knowledge criteria to facilitate class discrimination.

Artificial neural networks such as CNNs [21] or ViTs [32] on the other hand are completely data-driven, that is, features are learned during training. These methods have gained increasingly more recognition in bioimaging applications in recent years, due to their remarkable performance in various classification tasks. As their multilayered architectures that may include a very large number of learnable parameters allow them significantly more flexibility than SVMs, they can learn more sophisticated classification models [20]. However, to learn satisfactory classification models, artificial neural networks typically need a very large number of annotated data for training - a requirement that may be unfeasible in some applications due to physical constraints or privacy issues. Another limitation of such methods concerns the interpretation of results. Since the classification model is entirely data-driven, it is usually very difficult to ascertain what data properties are responsible for class discrimination.

We refer to the excellent review by Lin et al. [16] for a more thorough comparison of handcrafted vs data-based features in biomedical imaging problems.

2.3 SVM approach

We developed an approach based on SVM to test the hypothesis that structural characteristics of the soma are altered in images of hiPSC-derived neurons of patients affected by SCZ. In addition, previous work published using these cells [26] showed that inhibition of GSK3 using CHIR99021 increased the fluorescent intensity of βIII tubulin in the neurons of HC cells.

As remarked above, feature selection is critical for the success of such method, hence we defined appropriate kernel functions with the aim to capture the most relevant image characteristics of the soma region.

As shown in Fig. 1, the soma region of a representative image of hiPSC-derived neurons exhibits very different visual characteristics in the MAP2, βIII tubulin, Nav1.2 and FGF12 image channels. In the MAP2 (Fig. 1F) and βIII tubulin channel (Fig. 1G), the soma exhibits a pattern with recognizable filament-like structures. By contrast, in the FGF12 channel (Fig. 1I) the most remarkable characteristics is the presence of highly localized spots of high fluorescent intensity. The Nav1.2 channel (Fig. 1H) shows a combination of filament-like and spot-like characteristics.

Thus, to capture the micro-structures that are especially visible in the βIII tubulin and MAP2 image channels, we adopted a family of kernel functions associated to a class of statistical matrices called Gray-Level Dependence Matrices whose properties are described in the next section. Due to their sensitivity to complex patterns that may include in-homogeneous characteristics, we found that these kernel functions are very effective at capturing the image characteristics qualitatively observed in the different image channels and at discriminating SCZ from control cells.

2.3.1 Feature Selection.

To compute image-based features for our SVM approach we have chosen a class of statistical matrices derived from the pixel values in the soma region of the fluorescent images of hiPSC-derived neurons.

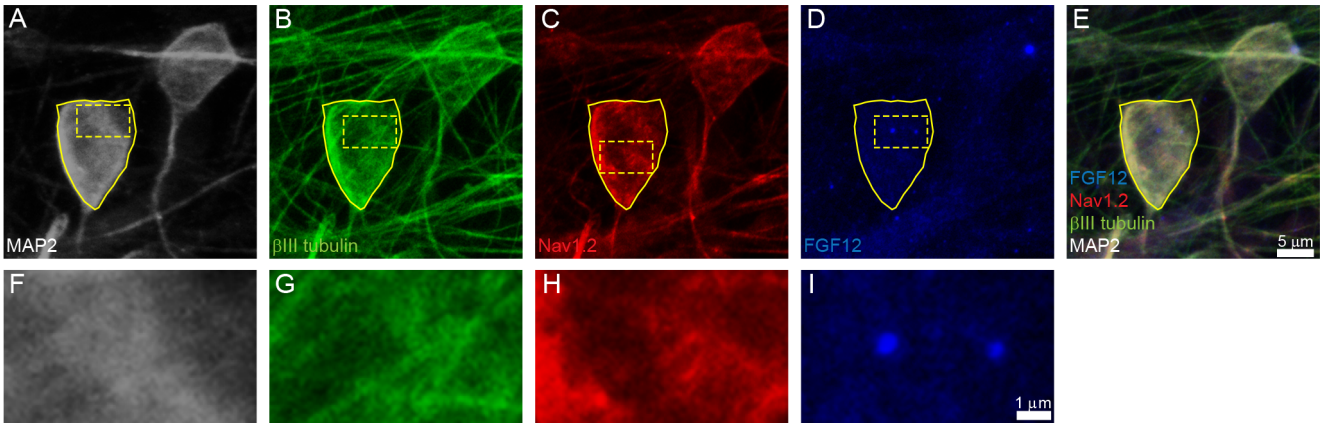

Fig. 1: Visual characteristics of a representative soma region (delineated in yellow) shown in the (A) MAP2, (B) βIII tubulin, (C) Nav1.2 and (D) FGF12 channels, with corresponding zoomed-in details (panels F-I) illustrating the distinct structural features of each image channel. Image overlay is shown in panel E.

Statistical matrices derived from gray-level pixel values have been used successfully in several image applications starting with the seminal work on Haralick on the *Gray-Level Co-occurrence* (GLC) Matrix and its applications in satellite image classification [12]. Among the most notable contributions to this area of research, we recall the work on the *Gray-Level Dependence* (GLD) Matrix, introduced for problems of texture classification [27], and the *Gray-Level Size Zone* (GLSZ) Matrix, proposed for applications in cell classification [28]. All these matrix-based features are part of the current Image Bio-marker Standardisation Initiative which aims to improve reproducibility and validation of radiomic features for applications in medical imaging [36]. The numerical implementation of such matrices is publicly available in the *py-Radiomics* toolbox [30].

Here is the definition of a 2D Gray-Level Dependence Matrix (GLDM) from [30] which we adapt to our setting.

Definition 1 *Let M denote a grayscale image and fix a pixel location p_1 in M with grayscale intensity i. We say that a pixel p_2 in M with the same grayscale intensity as p_1 is δ-dependent on the pixel p_1 if p_2 is in a δ-neighborhood centered at p_1. The (i, j) entry of a GLD matrix P is defined as the number of times a pixel of grayscale level i with j δ-dependent pixels appears in the image.*

We illustrate this definition using a 3×3 matrix, with $\delta = 1$ (this is the default value of the parameter which we adopt throughout this study):

$$M = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}, \quad GLD(M; \delta = 1) = \begin{pmatrix} 0 & 2 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}$$

The rows of the GLD matrix are associated with pixel values 1, 2 and 3 and the columns count the number of δ-dependencies. For instance, the (3,3) entry in the GLD matrix indicates that there are 2 pixels of grayscale level 3 which have 3 1-dependencies (specifically, the entries (1,2) and (1,3) in matrix M).

We computed the GLD matrices for all soma regions in our imaging set, where each individual image is a rectangular patch containing the soma and all pixel values outside the soma are set to 0. Since somas have different diameters, the corresponding rectangular patches containing the somas may differ in size and, similarly, the corresponding GLD Matrices may differ in size. Nonetheless, each GLD matrix is mapped to a feature vector of dimension 14, where each component is an appropriate weighted sum of matrix entries designed to capture local image patterns.

For example, let P be the GLD matrix for an input image M, N_g be the number of discrete intensity values in M, N_d be the number of discrete dependency sizes and $N_z = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j)$, be the number of dependency zones. Then the Small Dependence Emphasis (SDE), defined as

$$SDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2}}{N_z}$$

measures the homogeneity of local image patterns with smaller SDE values indicating a more homogeneous local structure.

Similarly, the other components of the feature vector quantify other statistical characteristics associated with local image patterns; we refer to [30] (and the related website [9]) for their definitions.

Using this approach, we mapped each soma image into a vector containing the fourteen radiomics features which we then used as input (as a csv file) for our supervised learning classifier, namely a Support Vector Machine (SVM).

2.3.2 SVM data classification

We applied our SVM classifier, trained on the image-based feature vectors computed above. to assign each cell (in the

test set) to one of four cell classes resulting from disease diagnosis and cell treatment – namely, disease (SCZ) vs. healthy control (HC) and GSK3 inhibition (GSK3i) vs. treatment control (vehicle). As remarked above, our features are selected to capture local image patterns including texture-like features such those we qualitatively observed in the βIII tubulin and MAP2 image channels.

In addition to using image features computed from a single fluorescent image channel, we also considered concatenated feature vectors obtained by combining different fluorescent image channels. Our rationale is that a perturbation (e.g., GSK3i) may affect different soma compartments resulting in structural alterations that may induce image phenotypes in different fluorescent image channels. Hence, by concatenating images features derived from multiple fluorescent image channels we combine information associated with different soma subcompartments that can possibly improve the performance of the classifier.

In our experiments below, we will report the classification performance obtained from a single fluorescent image channel as well as the classification performance obtained by concatenating the feature vectors computed from all 4 fluorescent image channels.

2.4 Artificial neural network approach

Artificial neural networks (ANNs) offer a platform for a supervised classification approach where feature selection is completely data-driven.

CNNs are among the most widely used ANNs in image classification tasks, following their celebrated success in the ImageNet classification challenge [13]. A CNN consists of a sequence of convolutional and pooling layers, followed by a fully connected layer [14, 21]. In a CNN, the first section (consisting typically of multiple layers) performs feature selection starting from the input pixels while the last section performs a classification on top of the extracted features.

ViTs were introduced to improve vision processing tasks such as image recognition [32] and it was shown that they can achieve greater performance than CNNs on image classification [10]. Their key idea consists in breaking down input images into patches and then analyze their relationships. This is in contrast with a CNN where input pixels are simply mapped into feature vectors. A ViT is implemented using a more complex network architecture than a CNN and requires typically a larger amount of training samples.

2.4.1 Selection of network architectures

For our CNN, we implemented a standard 3-layer architecture as illustrated in Fig. 2. It consists of three convolutional units where each unit includes a layer of 3×3 convolutional filters, Group Normalization, dropout with probability one-half and a Rectified Linear Unit (ReLU). We set the number of convolutional filters in each layer to increase by a multiple of 2, starting with 16 filters in the first layer, followed by 32 and 64 filters in the second and third layer, respectively. The output layer is a linear layer that takes as input the output of the last convolutional unit and applies a linear transformation to generate a 2-component feature vector whose size corresponds to the number of classes, e.g., 2 classes if we carry out a binary classification.

Our network was implemented using PyTorch by adapting the code used by one of the authors in [24].

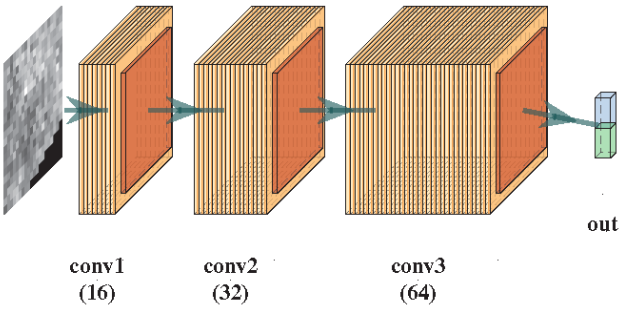

Fig. 2: 3-layer Convolutional Neural Networks with 16, 32 and 64 nodes in layer 1, layer 2 and layer 3, respectively.

We adopted the ViT from the TIMM package [34] which is implemented in PyTorch. The network architecture includes 12 alternating Transformer encoder's which can be thought of as blocks; a layer norm precedes each block while a residual connection follows each block; each Transformer encoder contains 6 self-attention heads followed by a multilayer perceptron [10]. This implementation takes individual soma images as input and processes them into patches of size 16x16 pixels. We set the hyper-parameters to the software defaults.

2.4.2 Data preparation and training

We applied our CNN approach to image patches extracted from the soma regions. Since somas from different cells may exhibit a variability in size, this approach allows us to classify somas based solely on their cytoskeleton structures and not on their shape and size. This way, we can have a fair comparison with our SVM approach that we trained on features extracted from the soma independently of its size.

For the selection of the patch size, we tested several options and found the best classification performance to occur for size 16x16 pixels. We explain this heuristic result with the observation that this patch size is large enough to include informative structural characteristics as shown in Fig. 1, that

might otherwise be missed by smaller patches. By allowing a small overlap between contiguous patches, we can extract a relatively large number of them from our image database, so that we can effectively train our CNN.

To automatize patch extraction, we designed an automated routine that extracts rectangular regions of size 16×16 pixels from the soma region of each image. This routine extracts patches sequentially by rows allowing a 8 pixel slide, i.e., there is a 8 pixel overlap between a patch and the next one. In addition, we discarded any patch that includes more than 10% of pixel with 0 intensity. This procedure removes patches that are either dominated by background (hence, uninformative) or patches that including a significant portion of background (hence, they may contain edges associated with the mask). Representative patches generated by this routine are shown in Fig. 3, including examples of image that are discarded according to our established criteria. The figure shows that we obtain about 30 usable patches from a typical soma image.

By applying this routine to our images, we obtained the patch dataset on which we trained and tested our CNN. In total, this dataset includes: 11390 patches (extracted from 208 soma images) for the classification problem HC vs SCZ in CHIR; 7065 patches (extracted from 132 soma images) for the classification problem HC vs SCZ in DMSO; 10735 patches (extracted from 175 soma images) for the classification problem DMSO vs CHIR in HC; 7720 patches (extracted from 165 soma images) for the classification problem DMSO vs CHIR in SCZ.

We remark that, while our CNN is designed to classify image patches, our end goal is to classify individual somas. To classify the soma images, we proceeded by concatenating the probability vectors of all image patches contained in any given soma and then computed the component-wise mean over the resulting probability vector. The argument max gives the network's predicted class for the soma image.

To train our CNN, we split our image patches into a training set and a validation set, by assigning 70% of the image patches for training and 30% for validation. Additionally, we augmented our training set by applying to each image patch rotations of 90, 180, and 270 degrees along with horizontal and vertical flips.

Our ViT network was similarly trained by assigning 70% of the images for training and 30% for validation.

During the training we used the Adam optimizer and one-cycle learning scheduler [25]. We trained our networks using a PC equipped with Intel Core i5-8250U CPU @ 1.6GHz, 4 Core(s), 8 Logical Processor(s) and 8GB of RAM, and a NVIDIA GeoForce GTX 1050.

3 Results

We applied our supervised learning framework, trained on image-based features of hiPSCs-derived neurons, to solve two classification tasks at the single cell level. The first task is the discrimination of disease (SCZ) vs. healthy control (HC); the second task is the discrimination of GSK3 inhibition (GSK3i) vs. treatment control (Vehicle).

In our tables below, we report the classification results obtained from our SVM, CNN and ViT methods. To also indicate the stability of the classification results, each accuracy result in our tables is an average over 10 independent runs of the same numerical experiments, that is, for the SVM, we run 10 experiments using different random training/tests splits; for the CNN and ViT networks, we trained the network using 10 different random initializations. In the tables, we also report the standard deviation associated with each accuracy measure.

3.1 Classification using single image channels

Tables 1 and 2 report the classification accuracy results obtained using our SVM, CNN and ViT classifiers trained on features taken from a *single* fluorescent image channels. As discussed in the Materials and Method section, there are four fluorescent channels: βIII tubulin, FGF12, MAP2 and Nav1.2. We consider a classification accuracy above 0.5 to be an indication of the predictive ability of the classifier. For instance, in Table 1, corresponding to the image channel FGF1.2 and the GSK3i cells, the SVM method is reported with accuracy 0.743 ± 0.046. This indicates that the probability of correctly assigning a given cell treated with GSK3i to either the class SCZ or HC is on average 74.3% (with standard deviation 4.6%). On the other hand, on the same classification problem, the CNN method is reported with accuracy 0.509 ± 0.045. This indicates that the probability of correctly assigning a given cell treated with GSK3i to either the class SCZ or HC using the CNN classifier is on average 50.9% (with standard deviation 4.5%), i.e., essentially the same as random chance.

Table 1 reports the classification performance of the classification problem SCZ vs HC applied to either Vehicle cells or GSK3i-treated cells. For the Vehicle cells (left section of the table), the ViT method reports the best classification performance corresponding to the βIII tubulin and FGF12 channels, while the SVM method reports the best performance for the MAP2 and Nav1.2 channels, even though in the latter case the performance is not significantly different than ViT. For the GSK3i cells (right section of the table), the SVM method reports the best classification performance for all four image channels. The CNN has the worst performance overall and in many cases (especially with Vehicle cell) the reported accuracy is around 0.5, meaning that

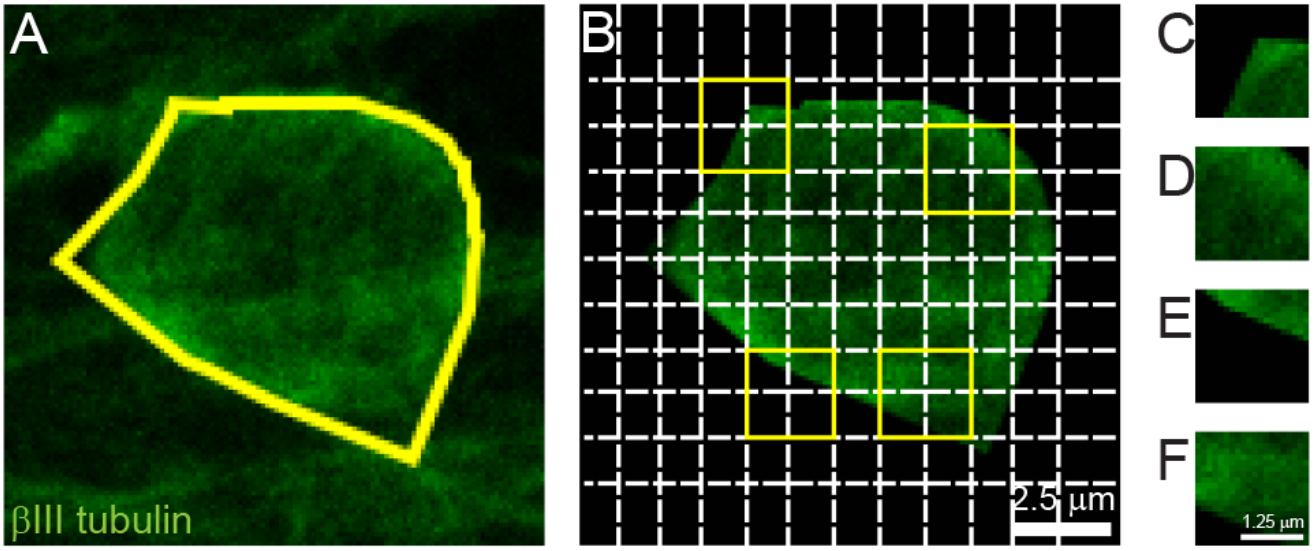

Fig. 3: Automated patch extraction. A: Soma image in the βIII tubulin fluorescent channel. B: Patches of size 16×16 pixels are extracted sequentially by rows allowing a 8 pixel slide. Patches including less than 10% of 0-intensity pixels are retained (panels D and F) otherwise they are discarded (panels C and E). The image contains a total of 32 usable patches.

Image channel	Control cells (Vehicle)			GSK3i cells		
	SVM	CNN	ViT	SVM	CNN	ViT
βIII tubulin	0.576 ± 0.060	0.530 ± 0.029	$\mathbf{0.669} \pm 0.084$	$\mathbf{0.687} \pm 0.052$	0.605 ± 0.042	0.598 ± 0.039
FGF12	0.614 ± 0.066	0.524 ± 0.039	$\mathbf{0.695} \pm 0.070$	$\mathbf{0.743} \pm 0.046$	0.509 ± 0.045	0.593 ± 0.051
MAP2	$\mathbf{0.697} \pm 0.068$	0.564 ± 0.049	0.634 ± 0.077	$\mathbf{0.691} \pm 0.047$	0.576 ± 0.081	0.586 ± 0.045
Nav1.2	$\mathbf{0.702} \pm 0.070$	0.511 ± 0.025	0.653 ± 0.058	$\mathbf{0.651} \pm 0.050$	0.513 ± 0.028	0.581 ± 0.038

Table 1: Individual soma classification performance SCZ vs HC using features from a single image channel. Bold font indicates best performance by row (within the same classification experiment)

Image channel	HC cells			SCZ cells		
	SVM	CNN	ViT	SVM	CNN	ViT
βIII tubulin	0.638 ± 0.041	0.562 ± 0.098	$\mathbf{0.642} \pm 0.017$	$\mathbf{0.677} \pm 0.044$	0.487 ± 0.097	0.604 ± 0.026
FGF12	$\mathbf{0.671} \pm 0.037$	0.522 ± 0.125	0.627 ± 0.014	0.558 ± 0.049	0.512 ± 0.091	$\mathbf{0.603} \pm 0.030$
MAP2	0.613 ± 0.030	0.527 ± 0.090	$\mathbf{0.632} \pm 0.014$	$\mathbf{0.621} \pm 0.051$	0.572 ± 0.056	0.611 ± 0.033
Nav1.2	0.618 ± 0.017	0.486 ± 0.079	$\mathbf{0.624} \pm 0.010$	0.607 ± 0.040	0.585 ± 0.023	$\mathbf{0.633} \pm 0.046$

Table 2: Individual soma classification performance GSKi vs Vehicle using features from a single image channel.

classification is unreliable. We explain the difference in performance among the 3 classification methods with the observation that: (i) due to our selection of handcrafted features, the SVM approach is overall very effective at capturing microstructural alterations occurring especially in the GSKi cells; (ii) the CNN approach is not effective at learning discriminating features using single image channels; (iii) the ViT is able to learn discriminating features using single image channels but not as effectively as the SVM for the GSKi cells.

We also observe that the SVM approach trained with image features from the FGF12 channel on GSK3i cells is the most effective at discriminating HC vs SCZ, yielding accuracy 74.3%. Remarkably, using features from the same fluorescent channel, the SVM classification accuracy drops to 61.4% on vehicle cells. Similarly using features from the βIII tubulin channel, the SVM classification performance drops from 68.7% on GSK3i cells to 57.6% on Vehicle cells. By comparison, there is no drop in the SVM classification performance for the MAP2 and Nav1.2 image channels. This observation points out to the role of the GSK3 signaling pathway, suggesting that βIII tubulin and FGF12 might be differently regulated in SCZ cells as compared with HC cells (cf. additional comments in Discussion and Conclusion.

Table 2 reports the classification performance of the classification problem GSK3i vs Vehicle applied to either HC cells or SCZ cells. For the HC cells (left section of the table), the SVM method in combination with the FGF12 image channel reports the best classification accuracy at 67.1%; the SVM classification performance for the other image chan-

nels is comparable (between 61.3% and 63.8%) and the ViT classification performance is also comparable (between 62.4% and 64.2%). For the SCZ cells (right section of the table), the SVM method in combination with the βIII tubulin image channel reports the best classification accuracy at 67.7%; the SVM classification performance for the other image channels is slightly lower (between 55.8% and 62.1%) and is comparable with the ViT classification performance (between 60.4% and 63.3%). Similar to Table 1, the CNN has the worst performance overall and in many cases the reported accuracy is around or close to 0.5, meaning that classification is unreliable.

3.2 Classification using concatenated image channels

As discussed in the Materials and Method section, to improve classification accuracy, we trained our classifiers on concatenated image features derived from all four fluorescent image channels. Results reported in Tables 3 show the outcomes of two classification experiments, Vehicle vs SCZ3i and HC vs SCZ, obtained using our 3 classification methods.

We first examine the Vehicle vs GSK3i experiment (left side of the table). The best classification is achieved by the ViT classifier on the SCZ cells with accuracy 72.1% even though the the SVM approach, at 69.6%, has a comparable performance. On the HC cells, the best classification performance is achieved by the SVM classifier at 67.4% with the ViT classifier yielding a comparable classification performance at 64.0%. In both cases, the CNN classifier has a slightly worse performance. We remark that, as compared with the classification results obtained using features from a single image channel in Table 2, the classification accuracy is now higher for all classifiers; this is especially true for the CNN approach that is now able to discriminate Vehicle vs SCZ3i with accuracy significantly above 50%. This observation confirms that the concatenation of image features improves the performance of the classifiers.

We next examine the SCZ vs HC experiment (right side of the table). The best classification is achieved by the SVM approach on the Vehicle cells and by the ViT approach on the GSK3i cells; in both cases the classification accuracy is 80.8%. However, the performance of all 3 classifiers is comparable on both Vehicle and GSK3i cells. Similar to our observation above, we remark that, as compared with the classification results obtained using features from a single image channel in Table 1, the classification accuracy is now higher for all classifiers, with the largest improvement observed for the CNN approach. The last observation suggests that the concatenation of image features allows the CNN classifier to better exploit the complementary information associated with the different image channels.

4 Information Sharing Statement

The imaging dataset presented in this study was previously released [26] and is found in the Mendeley Data repository: `http://dx.doi.org/10.17632/jgcsykf8xv.1`

Our source codes, as well as our pre-trained filters and imaging data, are available on GitHub at: `github.com/nfularczyk/SCZ_project/tree/main`

5 Discussion and Conclusion

We have introduced a new computational framework for the analysis of fluorescent images of hiPSCs of SCZ patients that is aimed at predicting the presence of disease or GSK3 inhibition using supervised cell classification methods, namely SVMs, CNN and ViT. All our classifiers are trained on image-based features extracted from multichannel fluorescent images of the soma regions of hiPSC derived neurons. However, while our SVM employs a class of kernel filters derived from GLD statistical matrices to capture local image patterns, the CNN and a ViT methods are fully data-based. ViT, in particular, is designed to consider not only the local feature of an image (as our CNN does) but also their relationship within the image.

Overall, our approach demonstrated a superb ability to extract predictive features from the imaging data. Given a fluorescent image of the soma region of a hiPSC derived neuron, our SVM and ViT classifiers (trained on image-based features extracted from the βIII tubulin, FGF12, MAP2 and Nav1.2 image channels) can predict if the neuron belongs to a SCZ patient with accuracy 80.8% on vehicle and GSK3i cells respectively. The ViT classifier predicts the perturbation by GSK3 inhibition with accuracy 72.1% on SCZ cells and the SVM classifiers achieves 67.4% accuracy on H cells.

Our results using single-channel image features exhibit lower prediction accuracy but are informative of the sensitivity of selects molecular soma constituents to disease and perturbation. Specifically, given the image of a hiPSC derived neuron with GSK3 inhibition, our SVM approach, trained on image-based features extracted from the βIII tubulin or FGF12 image channels, predicts SCZ with accuracy 68.7% and 74.3%; by contrast, the same method is unable to predict disease on Vehicle cells. Since our SVM is trained on image features associated with local image patterns, this result suggests that micro-structures of the soma cytoskeleton associated with βIII tubulin and FGF12 might be differently altered in SCZ vs HC cells upon perturbation by GSK3 inhibition. This is observation can be seen in the context of the existing literature about protein kinase inhibitors that differentiate between SCZ and HC samples. For instance, studies in postmortem tissue from schizophrenia patients have shown dysregulation of protein kinase B (AKT) [8] and mTOR [18], which are both upstream components of the GSK3

Control (Vehicle) vs GSK3i				HC vs SCZ			
HC		SCZ		Control (vehicle)		GSK3i	
CNN	0.618 ± 0.039	CNN	0.622 ± 0.052	CNN	0.797 ± 0.013	CNN	0.793 ± 0.014
ViT	0.640 ± 0.026	ViT	$\mathbf{0.721} \pm 0.049$	ViT	0.773 ± 0.038	ViT	$\mathbf{0.808} \pm 0.047$
SVM	$\mathbf{0.674} \pm 0.046$	SVM	0.696 ± 0.059	SVM	$\mathbf{0.808} \pm 0.056$	SVM	0.803 ± 0.024

Table 3: Individual soma classification performances GSKi vs Vehicle using features from all four image channels (βIII tubulin, FGF12, MAP2 and Nav1.2).

pathway. In [26], differences in the expression of βIII tubulin were observed in neurites of SCZ at baseline (DMSO) and in response to GSK3 inhibition. Changes in βIII tubulin distribution in neurons of SCZ patients were reported more generally in the literature, although some discrepancy exists [3]. For instance, βIII tubulin was found to be decreased in postmortem in the cortex in SCZ patients [17, 19]; studies in olfactory neurons showed a change in the pattern of distribution of βIII tubulin in the soma of neurons of SCZ patients [4]; plasma-levels of βIII tubulin were observed to be increased in SCZ patients treated with clozapine (which may indicate more severe or treatment resistant SCZ) compared to HC, while patients treated with other antipsychotics or who were in first-episode of psychosis (no treatment) showed no change in βIII tubulin [23].

Recent work using iPS cells has indicated that there may be sex-specific differences in differentiation, gene expression and protein expression [1, 29], a factor that, due to the gender distribution of patients included in this study, could have influenced our analysis and limit the impact of our conclusions. This is potentially a significant limitation in this work that will need to be further assessed in future studies. It is worth noting though that none of the genes coding for the cellular markers examined in this paper are on sex chromosomes nor were found to contribute to SCZ diagnosis by sex interaction in a previous study in neurons derived from iPS cells [29].

We finally observe that several studies in the literature have already identified a number of phenotypes as potentially associated with neural cells from SCZ patient-derived hiPSCs many results. While the analysis found in these studies is usually carried out in the form of hypothesis testing over samples of populations, in this work, we have pursued a machine learning approach to predict disease status (SCZ vs HC) or perturbation (GSK3i vs Vehicle) by interrogating imaging data at a single cell level. As discussed in a recent perspective by Li and Tong [15], hypothesis testing and supervised classification are different in nature where "the former concerns an unobservable population-level property of a feature, while the latter pertains to an observable label of an instance". The two methods can successfully complement each other, with hypothesis testing informing the construction of a classifier and, in turn, the classifier motivating the formulation of scientific questions to be investigated by hypothesis testing. Specifically, our application of supervised classification pertains the prediction of disease or GSK3 inhibition based on single-cell image-based features pointing out to the role of a class of cytoskeleton alterations. By formulating the study of structural alterations in fluorescent images of hiPSCs of SCZ patients as supervised classification problem, our approach provides a framework to establish with statistical rigour a quantitative relationship between image-based features associated with alterations of the cytoskeleton and SCZ or GSK3 inhibition. The same method can be implemented in combination with other image-based markers to reveal additional molecular or cellular components that could be relevant to explain critical mechanisms underlying SCZ and possibly other psychiatric disorders.

6 acknowledgements

NF and DL acknowledge financial support by NSF-DMS 1720487 and 1720452; they also acknowledge the Hewlett Packard Enterprise Data Science Institute at the University of Houston for providing computational resources. FL was supported by 1R01MH124351 and JDR was supported by a National Institutes of Health training grant (T32ES007254).

References

1. Montserrat C Anguera, Ruslan Sadreyev, Zhaoqing Zhang, Attila Szanto, Bernhard Payer, Steven D Sheridan, Showming Kwok, Stephen J Haggarty, Mriganka Sur, Jason Alvarez, et al. Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell stem cell*, 11(1):75–90, 2012.
2. Shabeesh Balan, Manabu Toyoshima, and Takeo Yoshikawa. Contribution of induced pluripotent stem cell technologies to the understanding of cellular phenotypes in schizophrenia. *Neurobiology of disease*, 131:104162, 2019.
3. Deborah Elaine Bauer, Vahram Haroutunian, Robert E McCullumsmith, and James H Meador-Woodruff. Expression of four housekeeping proteins in elderly patients with schizophrenia. *Journal of neural transmission*, 116(4):487–491, 2009.
4. G Benítez-King, M Valdés-Tovar, C Trueta, T Galvan-Arrieta, J Argueta, S Alarcon, A Lora-Castellanos, and H Solis-Chagoyan. The microtubular cytoskeleton of olfactory neurons derived from patients with schizophrenia or with bipolar disorder: Implications for biomarker characterization, neuronal physiology and pharmacological screening. *Molecular and Cellular Neuroscience*, 73:84–95, 2016.

5. Eduard Bentea, Erica AK Depasquale, Sinead M O'Donovan, Courtney R Sullivan, Micah Simmons, James H Meador-Woodruff, Ying Zhou, Chongchong Xu, Bing Bai, Junmin Peng, et al. Kinase network dysregulation in a human induced pluripotent stem cell model of disc1 schizophrenia. *Molecular omics*, 15(3):173–188, 2019.
6. Kristen Brennand, Jeffrey N Savas, Yongsung Kim, Ngoc Tran, Anthony Simone, Kazue Hashimoto-Torii, Kristin Grant Beaumont, Hyung Joon Kim, Aaron Topol, Ian Ladran, et al. Phenotypic differences in hipsc npcs derived from patients with schizophrenia. *Molecular psychiatry*, 20(3):361–368, 2015.
7. Kristen J Brennand, Anthony Simone, Jessica Jou, Chelsea Gelboin-Burkhart, Ngoc Tran, Sarah Sangar, Yan Li, Yangling Mu, Gong Chen, Diana Yu, et al. Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, 473(7346):221–225, 2011.
8. Radhika Chadha, Khaled Alganem, Robert E Mccullumsmith, and James H Meador-Woodruff. mtor kinase activity disrupts a phosphorylation signaling network in schizophrenia brain. *Molecular Psychiatry*, pages 1–12, 2021.
9. Pyradiomics Community. pyradiomics.readthedocs.io/en/latest/index, 2016.
10. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
11. Christa W Habela, Hongjun Song, and Guo-li Ming. Modeling synaptogenesis in schizophrenia and autism using human ipsc derived neurons. *Molecular and Cellular Neuroscience*, 73:52–62, 2016.
12. Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC–3, 1973.
13. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
14. Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
15. Jingyi Jessica Li and Xin Tong. Statistical hypothesis testing versus machine learning binary classification: Distinctions and guidelines. *Patterns*, 1(7):100115, 2020.
16. Wenyi Lin, Kyle Hasenstab, Guilherme Moura Cunha, and Armin Schwartzman. Comparison of handcrafted features and convolutional neural networks for liver mr image adequacy assessment. *Scientific Reports*, 10(1):1–11, 2020.
17. Francesca Marchisella, Eleanor T Coffey, and Patrik Hollos. Microtubule and microtubule associated protein anomalies in psychiatric disease. *Cytoskeleton*, 73(10):596–611, 2016.
18. Jennifer L McGuire, Erica A Depasquale, Adam J Funk, Sinead M O'Donnovan, Kathryn Hasselfeld, Shruti Marwaha, John H Hammond, Vahram Hartounian, James H Meador-Woodruff, Jarek Meller, et al. Abnormalities of signal transduction networks in chronic schizophrenia. *NPJ schizophrenia*, 3(1):1–10, 2017.
19. Mark S Moehle, Richard F Luduena, Vahram Haroutunian, James H Meador-Woodruff, and Robert E McCullumsmith. Regional differences in expression of β-tubulin isoforms in schizophrenia. *Schizophrenia research*, 135(1-3):181–186, 2012.
20. Behnam Neyshabur, Zhiyuan Li, Srinadh Bhojanapalli, Yann LeCun, and Nathan Srebro. The role of over-parametrization in generalization of neural networks. In *International Conference on Learning Representations*, 2018.
21. Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
22. In-Hyun Park, Natasha Arora, Hongguang Huo, Nimet Maherali, Tim Ahfeldt, Akiko Shimamura, M William Lensch, Chad Cowan, Konrad Hochedlinger, and George Q Daley. Disease-specific induced pluripotent stem cells. *cell*, 134(5):877–886, 2008.
23. Daniela Rodrigues-Amorim, Tania Rivera-Baltanás, María del Carmen Vallejo-Curto, Cynthia Rodriguez-Jamardo, Elena de Las Heras, Carolina Barreiro-Villar, María Blanco-Formoso, Patricia Fernández-Palleiro, María Álvarez-Ariza, Marta López, et al. Plasma β-iii tubulin, neurofilament light chain and glial fibrillary acidic protein are associated with neurodegeneration and progression in schizophrenia. *Scientific reports*, 10(1):1–10, 2020.
24. Kazem Safari, Saurabh Prasad, and Demetrio Labate. A multiscale deep learning approach for high-resolution hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 2020.
25. Leslie N. Smith and Nicholay Topin. Super-convergence: very fast training of neural networks using large learning rates. In Tien Pham, editor, *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006, pages 369 – 386. International Society for Optics and Photonics, SPIE, 2019.
26. Laura Stertz, Jessica Di Re, Guangsheng Pei, Gabriel R Fries, Emily Mendez, Shenglan Li, Laura Smith-Callahan, Henriette Raventos, Jerricho Tipo, Rohan Cherukuru, et al. Convergent genomic and pharmacological evidence of pi3k/gsk3 signaling alterations in neurons from schizophrenia patients. *Neuropsychopharmacology*, 46(3):673–682, 2021.
27. C. Sun and W.G Wee. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 23(3):341–352, 1983.
28. G. Thibault, J. Angulo, and F. Meyer. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE transactions of biomedical imaging*, 61(3):630–7, 2014.
29. Jari Tiihonen, Marja Koskuvi, Markus Storvik, Ida Hyötyläinen, Yanyan Gao, Katja A Puttonen, Raisa Giniatullina, Ekaterina Poguzhelskaya, Ilkka Ojansuu, Olli Vaurio, et al. Sex-specific transcriptional and proteomic signatures in schizophrenia. *Nature communications*, 10(1):1–11, 2019.
30. Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.
31. V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.
32. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
33. Zhexing Wen, Ha Nam Nguyen, Ziyuan Guo, Matthew A Lalli, Xinyuan Wang, Yijing Su, Nam-Shik Kim, Ki-Jun Yoon, Jaehoon Shin, Ce Zhang, et al. Synaptic dysregulation in a human ips cell model of mental disorders. *Nature*, 515(7527):414–418, 2014.
34. Ross Wightman. Pytorch image models. `https://github.com/rwightman/pytorch-image-models`, 2019.
35. Diana Xuan Yu, Francesco Paolo Di Giorgio, Jun Yao, Maria Carolina Marchetto, Kristen Brennand, Rebecca Wright, Arianna Mei, Lauren Mchenry, David Lisuk, Jaeson Michael Grasmick, et al. Modeling hippocampal neurogenesis using human pluripotent stem cells. *Stem cell reports*, 2(3):295–310, 2014.
36. Alex Zwanenburg, Martin Vallières, Mahmoud A Abdalah, Hugo JWL Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof J Beukinga, Ronald Boellaard, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, 2020.