medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Psychiatric map diagnosis in mental disorders

Dongdong Li[1,2], Wenbin Liu[1], Henry Han[3]

[1]Institute of Computational Science and Technology, Guangzhou University, Guangzhou, Guangdong, 510006 China

[2]School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, 510641 China

[3]Department of Computer Science, School of Engineering and Computer Science, Baylor University, Waco, TX 76798 USA

Corresponding email: wbliu6910@gzhu.edu.cn, Henry_Han@baylor.edu

Abstract

The misdiagnosis between schizophrenia (SCZ) and bipolar disorder (BPD) has been a challenge in psychiatry. It is also a long-time unsolved mislabeled learning problem in machine learning and AI fields. In this study, we propose a psychiatric map (pMAP) diagnosis, which is built upon a novel feature self-organizing map (*fSOM*) algorithm, to tackle it. The psychiatric map summarizes the latent essential characteristics of each observation on a two-dimensional *fSOM* plane. It solves the misdiagnosis problem by providing high-accuracy psychiatric detection via automatically mislabeled observation identification. Furthermore, pMAP provides powerful and informative visualization for each observation in unveiling hidden psychiatric subtype discovery. This study also presents new insight into the pathology of psychiatric disorders by constructing the devolution path of psychiatric states via relative entropy analysis that discloses latent internal transfer and devolution road maps between different subtypes of the control, BPD, and SCZ groups. To the best of our knowledge, it is the first study to solve mislabel learning for high-dimensional data in machine learning and will inspire more future work in this field.

Keywords: mislabeled learning, bipolar disorder, schizophrenia, misdiagnosis, pMAP, fSOM

1 Introduction

Bipolar disorder (BPD) and schizophrenia (SCZ) are two severe worldwide psychiatric disorders [1]. They both are highly heritable, complex neuropsychiatric diseases that share some similar clinical symptoms [2][3]. Schizophrenia (SCZ) is a chronic and severe mental health disorder characterized by hallucinations, delusions, and disorganized thinking [4][5]. Bipolar disorder (BPD) is a chronic mental illness that causes dramatic shifts in a person's mood, energy, and ability to think clearly [6][7]. People with BPD have high and low spirits, known as mania and depression. According to the National Alliance on Mental Illness, it is reported that about 1% and 2.9% of Americans are diagnosed with schizophrenia and bipolar disorder respectively each year.

Schizophrenia (SCZ) and bipolar disorder (BPD) diagnosis remain a challenge in psychiatry though recent progress has been made from different perspectives [8]. Neuroimage studies show that SCZ patients can

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

exhibit a more characteristic pattern in brain imaging than normal people [9][10]. Similarly, brain abnormalities are found in people with bipolar disorder [11]. SCZ was reported to be related to neuronal-level changes causing cortical excitation-inhibition imbalance [12][13]. However, it can be difficult to conduct an accurate SCZ/BPD diagnosis only from the neuro-imaging perspective [14][15][16][17].

Molecular-based SCZ and BPD investigations from a translational approach have been surging recently with the development of omics [18][19][20][21]. It makes it possible to understand the molecular pathogenesis of SCZ and BPD, explore their genetic root, and provide a more robust clinical diagnosis via examining molecular signatures. For example, Sahu *et al* found that there were both unique and overlapping molecular signatures between SCZ and BPD from a systems biology perspective [22]. Li *et al* suggested the role of DNA methylation in the pathogenesis of BPD and SCZ [23]. Ellis *et al* reported that the BPD and SCZ transcriptomes were not significantly correlated [24].

Liu *et al* showed that SCZ and BPD shared common pathways and BPD could be a subtype of SCZ via manifold learning and pathway analysis. They also pointed out that the inevitable misdiagnosis issue between them from a translational bioinformatics perspective because their diagnoses are mainly based on clinical symptoms rather than real genetic differences. The label information generally contains a large portion of incorrect information because of the widely existed misdiagnosis between BPD and SCZ, which echoes their high misdiagnosis rates in clinical practice [25]. Therefore, it is equivalent to a mislabeled machine learning problem that is not well solved in AI and data science [26][27][28].

It is challenging to solve a mislabeled machine learning (ML) problem in AI, especially for high-dimensional data. First, such a problem is rarely investigated in classic ML domains, in which data is generally assumed to have correct labels. Second, dropping possible mislabeled data in the learning procedure may only work well for those with a large number of samples. However, data generally used in SCZ and BPD detection is high dimensional omics data that has relatively a few samples (e.g., $\sim O(10^2)$) but a large number of variables (e.g., $\sim O(10^5)$) [31][32]. It is infeasible to drop data because of the very limited number of observations. Third, high-dimensional data itself may put hurdles in the problem solving because it can be hard to seek meaningful variables in BPD and SCZ diagnosis. Therefore, it can be almost unlikely for any learning machine to achieve good performance under such a situation.

However, solving the problem will bring unpreceded impacts on psychiatry disorder detection and shed light on mislabel learning. It would provide more accurate BPD and SCZ diagnosis from a genetic level via an explainable AI resolution [31][32][33]. Furthermore, it would solve the challenging mislabeled learning problem for small-sized data (e.g., high-dimensional data) by enriching existing ML. Since small-sized data is challenging existing health, biomedical, and AI fields, the solution to their related mislabeled learning problem will be more valuable for the sake of more resilient and robust AI health [34].

In this study, we propose a novel psychiatric map (pMAP) diagnosis for BPD and SCZ detection. The psychiatric map can be viewed as a special characteristic map on a two-dimensional plane for each observation. It summarizes the essential characteristics of each observation by conducting novel feature self-organization maps (*fSOM)* learning. Unlike traditional self-organization map (SOM) [35], *fSOM* seeks the prototype of each observation in the subspace spanned by 'condensed features' on the *fSOM* plane rather than clustering them [29]. Technically, it is also a nonlinear dimension reduction to map high-dimensional data to its low-dimensional embedding. The proposed pMAP provides powerful visualization on the *fSOM* plane that captures the essential characteristics of each psychiatric observation.

Figure 1 illustrates the psychiatric maps of control, BPD, and SCZ under *fSOM* using SNP data, where the

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

same type of observations demonstrates similar or same patterns [1][3]. It shows those psychiatric maps of controls demonstrate clear differences with those of the BPD and SCZ samples. The proposed pMAP unveils latent subtypes via *fSOM* learning, in which those discovered two control, two BPD, and three SCZ subtypes are illustrated.

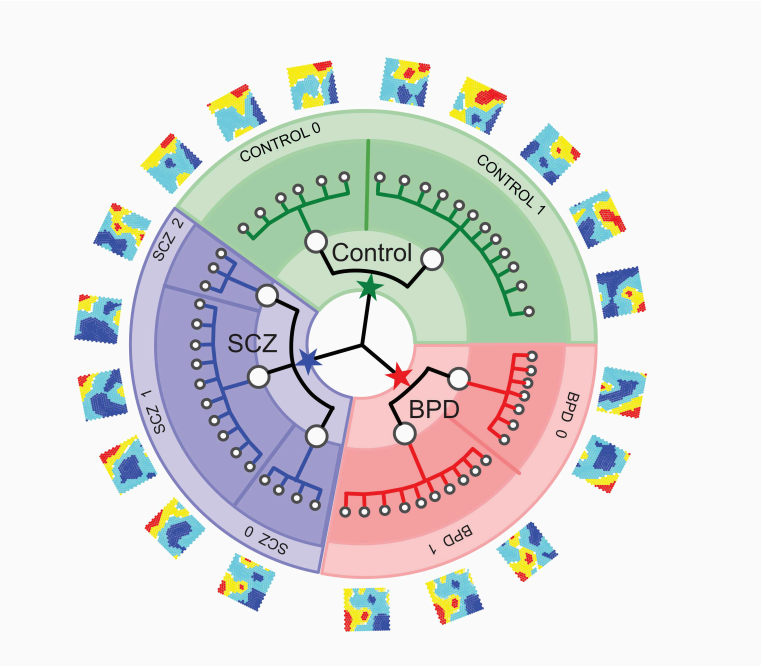

Fig 1. The psychiatric maps of control, BPD, and SCZ samples obtained under *fSOM* learning by using SNP data. A psychiatric map is a characteristic map of each observation such as a BPD/SCZ/control sample on the *fSOM* plane. The control psychiatric maps demonstrate clear differences from those of the BPD and SCZ. It also provides an effective latent subtype discovery in psychiatric analysis. The proposed pMAP diagnosis unveils latent two control, two BPD, and three SCZ subtypes: *control_0/1*, *BPD_0/1*, and *SCZ_0/1/2*.

The proposed approach has the following contributions. First, it effectively solves the long-term misdiagnosis issue between BPD and SCZ in psychiatry. It automatically detects mislabeled observations by generating psychiatric maps and relabels data with the help of density-based clustering (e.g., DBSCAN) [30]. It provides a novel path to solving mislabeled learning problems for high-dimensional data. It makes it possible to handle mislabeled learning problems for small-sized datasets. The small-sized datasets along with mislabeled information across data would fail almost all deep learning models because of the data scarcity issue [36]. Therefore, the proposed technology is especially valuable not only for psychiatry but also for most biomedical fields.

Second, the proposed pMAP presents novel visualization for high-dimensional SNP psychiatry data besides unveiling the latent subtypes for psychiatry samples. It transforms each high-dimensional observation into its psychiatric map, a characteristic 2D image that captures the sample prototype described by the key SNPs. To the best of our knowledge, it is the first work in mislabeled learning, psychiatry disorder, and extremely high-dimensional SNP data visualization, which generally can have hundreds of million features. The built-in powerful visualization in the pMAP diagnosis makes it an explainable AI technique with transparency and trustworthiness. The pMAP-based visualization can be exploited to provide knowledge-discovery-based visualization in other biomedical fields that use high-dimensional data. The knowledgeable pMAPs will assist doctors to understand the latent disease statuses/subtypes,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

achieve high-accuracy diagnoses, and enhance clinical decision-making.

Third, this work proposes a novel customized entropy analysis to explain the results of pMAP diagnosis in BPD and SCZ detection. We find that both BPD and SCZ samples tend to have lower entropies than the control samples. It suggests that molecular patterns of the samples with psychiatric disorders should contain a 'less random' information pattern than the ordinary samples from SNP data analysis. More importantly, we construct the devolution paths and internal transfers between different psychiatric states via relative entropy analysis to shed light on the pathology of psychiatric disorders.

This paper is structured as follows. Section 2 introduces pMAP generation and the principle of pMAP diagnosis besides presenting *fSOM* learning. Section 3 covers data and data preprocessing and section 4 conducts psychiatric map analysis from different perspectives, in which the novel devolution paths of psychiatric states via relative entropy analysis are unrolled for BPD and SCZ discovery. Section 5 presents the detailed results of pMAP diagnosis and compares it with the state-of-the-art ML and deep learning methods. Finally, we discuss the potential weakness and possible enhancements of our methods before concluding this study.

2 The principle of psychiatric map (pMAP) diagnosis

It can be technically hard to attain a satisfactory BPD and SCZ diagnosis by only applying existing ML or deep learning models to high dimensional omics data (e.g., SNP) with mislabeled information because of the widely existed misdiagnosis between BPD and SCZ in the clinical practice. The special mislabeled data handling techniques are urgent to solve this problem besides the effective high dimensional data feature selection.

It is desirable to extract the prototype of each observation to distinguish possibly mislabeled samples from an unsupervised self-organizing approach under an effective SNP feature selection. The prototype contains the essential characteristics of each sample that can be a good discriminator to correct the mislabeled samples. To achieve this goal, we propose a novel feature self-organizing map (*fSOM)* learning to obtain the prototype of each observation. Unlike traditional SOM which dynamically looks for similarities between input samples, the proposed *fSOM* seeks to condense thousands of SNP features on a low-dimensional *fSOM* plane to derive all the prototypes of input data, which is called a psychiatric map **(**pMAP) in this study.

2.1 Psychiatric map (pMAP) diagnosis

The proposed pMAP diagnosis consists of three major steps: 1) pMAP generation with *fSOM* learning; 2) DBSCAN (Density-based spatial clustering of applications with noise) clustering for pMAPs; 3) Relabeled sample learning. We describe the detailed steps as follows.

pMAP generation. The pMAP diagnosis is built upon the SNP dataset after preprocessing which can be found in the Data and preprocessing section. The pMAP diagnosis first seeks a pMAP for each observation that is an SNP sample in our context. The psychiatric map (pMAP) is the corresponding prototype of each original observation in the input space $\mathbb{R}^n$, in which n represents the number of SNPs that can reach $O(10^6)$ before feature selection.

It captures the essential characteristics of each sample and generates its prototype in the low-dimensional

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

fSOM plane through proposed *fSOM* learning. Mathematically it is a reference vector in $\mathbb{R}^k, k \ll n$, representing the characteristic map of the sample on the $\sqrt{k} \times \sqrt{k}$ *fSOM* plane. Since the pMAPs are the prototypes of the original samples, those with similar pMAPs should belong to the same group, but the mislabeled samples will demonstrate different pMAP patterns as the group it 'belongs to'. For example, a BPD sample that is mislabeled as an SCZ type will have a dissimilar pMAP as the true SCZ. On the other hand, it would have similar or even the same pMAP patterns as the true BPD.

DBSCAN clustering for pMAPs. The pMAP diagnosis employs a DBSCAN (density-based spatial clustering of applications with noise) to cluster the pMAPs of input samples in the low-dimensional space that is represented as an *fSOM* plane geometrically. The clustering results are used to correct possible mislabeled samples. DBSCAN to conduct good clustering for its robustness and advantages compared to the widely used clustering methods such as K-means. K-means can be a good candidate for pMAP clustering because we already know the prior three groups of samples: control, SCZ, and BPD. However, there are two reasons for us to select the density-based clustering method DBSCAN rather than the popular K-means. The first is that K-means would require input data is convex. But we cannot guarantee the pMAPs generated from *fSOM* learning will satisfy it. Furthermore, we find that pMAPs can have concaved shapes on the *fSOM* plane. The second is that K-means would limit the possible new subgroup detection somewhat, i.e., the pMAP generation stage brings the prototypes of input samples that are not only limited to the original three types. Instead, the pMAPs generated from *fSOM* learning unveil new knowledge in the self-organizing learning process: the control, SCZ, and BPD groups all have their different subtypes. Thus, we do need a clustering algorithm that can automatically identify the number of clusters for input pMAPs for the sake of deep knowledge discovery. After the DBSCAN clustering, pMAPs will be grouped into different clusters for the following relabel sample learning.

Relabeled sample learning. We relabel the original samples and their pMAPs according to the DBSCAN clustering result. Although we have found the subtypes for each psychiatric group, we still label those from the same group (e.g., BPD) as one type rather than dividing them into different types for the sake of learning and peer comparisons. After the relabeling procedure, an ML model is used to conduct a BPD and SCZ diagnosis for the relabeled psychiatric maps. Theoretically, any ML model can be employed, but we prefer the ML model to satisfy the following characteristics for high-performance psychiatric diagnosis.

First, it should have good reproducibility so that the BPD and SCZ diagnosis would not change from run to run. Such good reproducibility has important clinical meaning to decrease false positive rates. In other words, we should avoid those ensemble learning methods such as random forests or deep learning models because their results may lack good reproducibility, though their learning performance can be good.

Second, it should have a built-in advantage to handle high-dimensional data, especially high-dimensional data with a small number of samples. In other words, its complexity should not increase much for high-dimensional data. The deep learning models generally need a large amount of data in the training stage to build sophisticated prediction functions, but they are not good for high-dimensional data with only a small number of samples in context [39].

Therefore, we employ multi-class support vector machines (multi-class SVM) because it satisfies the above two well [40]. It has a very good reproducibility because it is equivalent to solving a deterministic nonlinear programming problem with the least randomness involved. Furthermore, it is good at handling high-dimensional data because the kernel matrix calculated from high-dimensional data can be a relatively small and efficient one, which can speed up the whole SVM learning by avoiding possible computing overhead

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[40].

Entropy analysis for relabeled psychiatric samples. In the post-analysis of the pMAP diagnosis, we define a novel data entropy and relative entropy (KL divergence) to quantify each psychiatric map (pMAP) and investigate how entropies vary among different psychiatric subtype groups [41]. Traditionally, it is almost impossible to compute the entropy and relative entropy for a group of psychiatric samples, which is represented as a data matrix mathematically. With the help of the pMAPs, we have the following novel data entropy and KL divergence techniques for a group of psychiatric subtype samples. The details about entropy and KL-divergence calculation techniques can be found in the following section. We also examine the devolution paths of psychiatric states via relative entropy analysis.

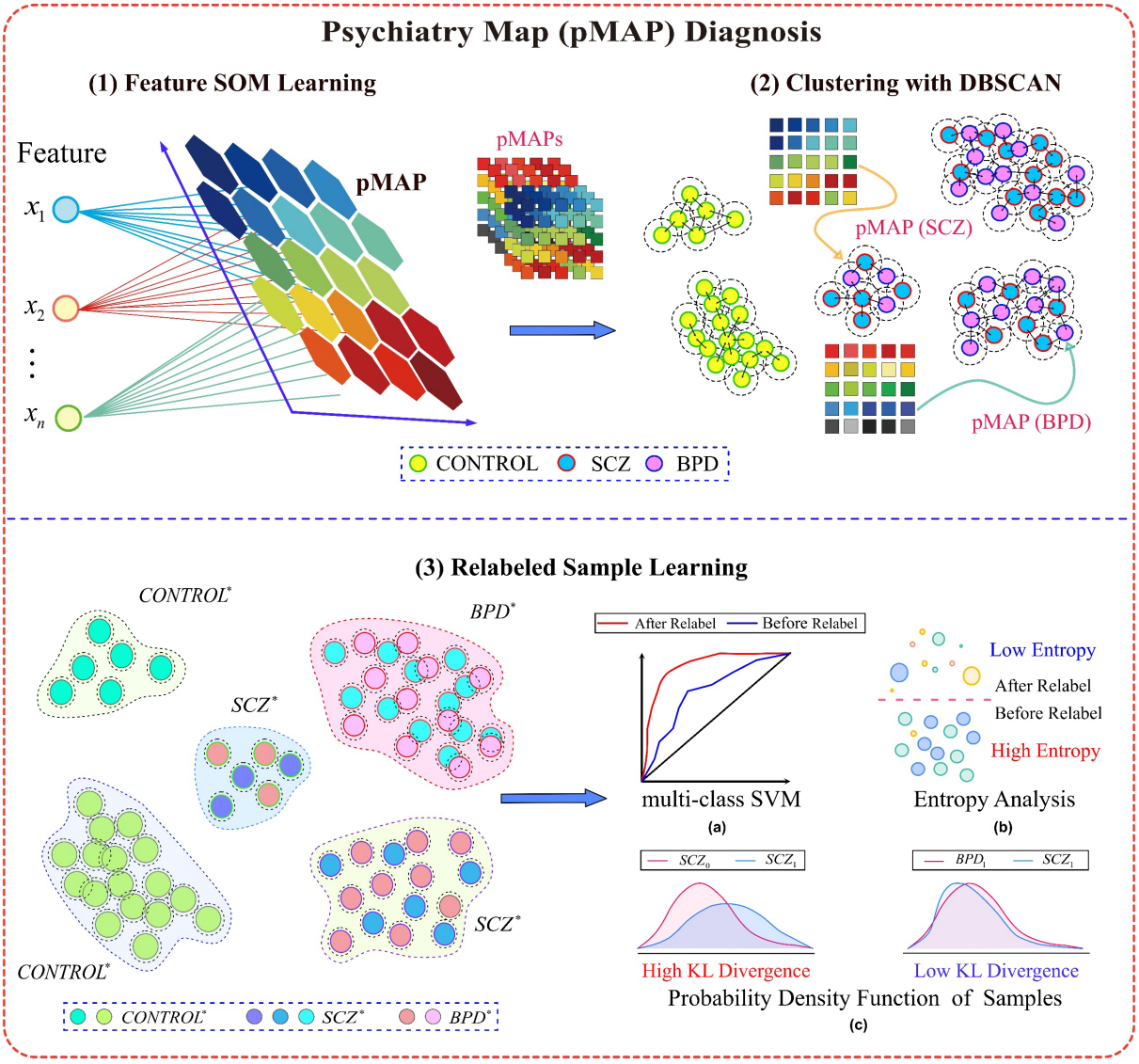

Fig 2. The flowchart of the proposed psychiatric map (pMAP) diagnosis. It employs feature self-organizing (*fSOM*) to generate a psychiatric map (pMAP) for each observation that can be a schizophrenia (SCZ), a bipolar disorder (BPD), or a control sample. The pMAPs are clustered under DBSCAN and further relabeled accordingly. Finally, the relabeled observations are inputted into the learning machine (e.g.,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

multi-class SVM) for diagnosis. In the post-analysis of the pMAP diagnosis, entropy analysis unveils the distributions of relabeled psychiatric sample entropies and their K-L divergence patterns.

Figure 2 illustrates the flowchart of the proposed pMAP diagnosis that consists of three major steps: 1) psychiatric map (pMAP) generation using the proposed feature self-organizing (*fSOM*) learning. 2) DBSCAN clustering for pMAPs. 3) Relabeled sample learning and entropy analysis. In this step, the psychiatric samples are further relabeled according to the DBSCAN clustering results for their pMAPs to correct the possible mislabeled samples before conducting psychiatry diagnosis with multi-class SVM. The following Algorithm 1 summarizes the proposed psychiatric maps (pMAP) diagnosis. The complexity of the proposed algorithm complexity is $O(nmk + mlogm) + O(\theta)$, where $O(\theta)$ is the complexity of the machine learning method used in psychiatric map diagnosis.

Algorithm 1: Psychiatric map (pMAP) diagnosis

Input:

Data: $X \in \mathbb{R}^{n\times m}$ with m observations across n features, $n>>m$
Machine learning model: θ (default: multi-class SVM)
The size of SOM plane $\sqrt{k} \times \sqrt{k}$ *(default:* $k = 20 \times 20$ *)*
Epoch of *fSOM* (default 1000)
$minpts$ (the minimum number of points in a core-point neighborhood: default 10)
eps (the minimum distance between two points in clustering: default 0.05)

Output:

Psychiatric maps (pMAPs): X_{pmap}
Predicted labels of test data in X

1. $X_{pmap} \leftarrow fSOM(X, k, epoch)$ // *Compute the psychiatric map (pMAP) for each observation*
2. $ClusteringIndex \leftarrow DBSCAN(X_{pmap}, minpts, eps)$ // *DBSCAN clustering for psychiatric maps (pMAPs)*
3. $(X_{pmap},\ label) \leftarrow Relabel(X_{pmap}, ClusteringIndex)$ //*Relabeling pMAPs according to clustering results*
4. $\theta \leftarrow fit(\theta, X_{pmap}^{train}, label)$ // *Train machine learning model θ using relabeled training data*
5. $\widehat{y_t} \leftarrow \theta.predict(X_{pmap}^{test})$ // *Prediction for test data*
6. **Return** $\widehat{y_t}$

2.2 Feature self-organizing map (*fSOM*) learning

Feature self-organizing map (*fSOM*) learning is an unsupervised learning model rooted in neurobiology for high-dimensional data. Unlike traditional SOM, *fSOM* conducts self-organizing for features rather than observations. It produces a set of condensed features on the SOM plane rather than the clustering of original observations. More importantly, *fSOM* can produce the prototype of each observation in a low-dimensional space spanned by condensed features. It also contributes to the meaningful knowledge-based representation of each obversion.

fSOM consists of an input high-dimensional dataset, $X \in \mathbb{R}^{n\times m}$ with n features across m observations, an SOM plane P, a two-dimensional lattice consisting of $\sqrt{k} \times \sqrt{k}$ *(e.g.,* k=400*)* neurons, and an unsupervised learning algorithm: θ: $fSOM = (X, P, \theta)$. The *fSOM* plane P is a set of neurons placed on a two-dimensional lattice. Each neuron $i, i = 1, 2 \cdots k$ on the *fSOM* plane has a reference vector $w_i \in \mathbb{R}^m$. *fSOM* maps the original n $(n \gg k)$ features as k characteristic features through self-organizing learning on the *fSOM* plane.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

As a nonlinear dimension reduction method, *fSOM* maps high-dimensional data $X \in \mathbb{R}^{n\times m}$ with n features and m observations to a low-dimensional dataset $W \in \mathbb{R}^{k\times m}$, which is also called a reference vector matrix, by maintaining and capturing the most representative features. The i^{th} column of W: W_i is called the prototype of the original i^{th} sample of X: X_i that captures the essential characteristics of the original observation in a low-dimensional feature space. Thus, *fSOM* maps input data to its prototype in a condensed feature space, i.e., $fSOM: X \to W \in \mathbb{R}^{k\times m}, k < n$. The reference vector matrix $W = [w_1, w_2 \cdots w_k]^t$ after the SOM finishes learning, the reference vector matrix W is the extracted prototype data for the input dataset X in the low-dimensional space.

The *fSOM* learning procedures. The *fSOM* learning process follows the general self-organizing learning procedure that consists of loops of three stages: competition, cooperation, and adjusting. In the competition stage, *fSOM* queries all neurons $i, i = 1,2\cdots k$ on the *fSOM* plane to find a candidate j^* with the closest reference vector W_{j^*} to a feature $x \in X$ according to a specific distance measure (e.g., Euclidean): $d(x - W_{j^*}) = \min_i\{d(x, W_i)\}$. The candidate j^*, the winning neuron in the competition, is identified as the best match unit (BMU) for the feature x.

In the cooperation stage, the reference vectors of the winning neuron j^* in the topological neighborhood of the BMU are updated as:

$$W_i(t+1) = W_i(t) + \alpha(t)h_{j^*i}(t)(x - W_i(t)) \quad (1)$$

to make them more and more 'similar' to the feature x. The $\alpha(t) \in (0,1)$ is the learning rate at time phrase t. The h_{j^*i} is the neighborhood function generally chosen as a Gaussian kernel function $h_{j^*i}(t) = e^{\|r_{j^*} - r_i\|^2/2\sigma^2(t)}$. The Gaussian function is more biologically meaningful and leads to the fast convergence. The r_{j^*} and r_i are the topological locations of the best match unit (BMU) and i^{th} neuron on the SOM plane, and the parameter σ denotes the radius of the topological neighborhood.

In the adjusting stage, the learning rate is adjusted, and the neighborhood radius decreases with respect to time exponentially to localize the reference vector matrix iterations. That is, $\sigma(t) = \sigma(0)e^{\frac{-t}{\tau_1}}$, $\alpha(t) = \alpha(0)e^{\frac{-t}{\tau_2}}$, $t = 0,1,2,\cdots$. The $\alpha(0)$ and $\sigma(0)$ are the initial learning rate and neighborhood size respectively and τ_1 and τ_2 are pre-selected time constants.

The three stages are repeated until the reference vector matrix W converges. The SOM learning is equivalent to a gradient-based optimization algorithm. If the objective function is defined as the local energy function at the time for a neighborhood centered in the winning neuron j:

$$e(t) = \sum_{i=1}^{k} h_{ji}(t)\|x - W_i(t)\|^2 \quad (2)$$

where the distance measure is any specified distance metric. Thus, the iteration scheme in the cooperation stage is equivalent to finding the optimal reference vector in a neuron to minimize the specified local energy function specified by the following gradient optimization problem:

$$W_i(t) = W_i(t) - \alpha(t)I\frac{\partial e(t)}{\partial W_i(t)} \quad (3)$$

where I is an identity matrix and $\alpha(t)I$ is the learning rate matrix. Similarly, the global energy function $e(t)$ can be defined as the sum of all local energy for all entries in the training data set. The computing complexity of one epoch of training to minimize the global energy function is $O(nmk)$.

The described learning algorithm in *fSOM* is called sequence learning. There is an acceleration learning

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

algorithm called batch learning to speed up sequence learning. In batch learning, total data samples are involved in each update of the reference vector matrix W. Compared with the sequential learning, the batch learning algorithm converges fast and is free from learning rate adjustments but is more likely to trap in a local minimum. Batch learning is employed in our implementation for the sake of its powerful convergence speed.

2.3 Density-based spatial clustering of applications with noise (DBSCAN)

DBSCAN is a density-based clustering algorithm that handles arbitrary-shaped clusters with noise [41]. DBSCAN classifies points as core, reachable, and outliers (noise). A core point simply refers to a point whose neighborhood has enough points under a radius ε. A reachable point is a point that can be reached by one or a sequence of core points and an outlier is an unreachable point, i.e., noise. In our context, it will be a transaction with exceptional trading behaviors that are potential to be trading markers. The core points form clusters because of their high densities, the reachable points form the edge of clusters, and the outliers stand out as noise in clustering.

The primary idea of DBSCAN can be described briefly as follows. Given a point to be clustered, DBSCAN retrieves its ε-neighborhood. If the neighborhood size is ≥ the minimum number of points (*minpts*) required to form a 'dense region', i.e., a region with an enough number of close points, the neighbor will be initialized as a cluster and the point is marked as a core point. Otherwise, the point is marked as an outlier. If the point is a reachable point for a cluster, its ε-neighborhood will be marked as a part of that cluster. All points in the ε-neighborhood will be added to the cluster until the density condition is satisfied. This procedure continues until all clusters and outliers are identified. The average running time complexity of DBSCAN is $O(nlogn)$ if a meaningful neighborhood radius ε is selected, though the worse time complexity is $O(n^2)$.

3 Data and preprocessing

We briefly introduce SNP data and nonnegative singular value approximation (nSVA), which is an effective feature selection algorithm proposed by Han for high-dimensional data [38].

The original SNP data (GSE71443) downloaded from the NCBI GEO database includes 74 control, 65 bipolar disorder (BPD), and 64 schizophrenia (SCZ) subjects [42][43]. To obtain the significant differentially expressed SNP loci, we first filter those SNPs with missing annotations, not on autosomes or sex chromosomes, or diverged from Hardy-Weinberg equilibrium (HWE) with $p\_value < 10^{-10}$ [42]. We still have a total of 627,693 SNPs left after the initial filtering. We further screen statistically significant SNPs using ANOVA with $FDR < 0.01$ and remove those SNPs in the linkage disequilibrium (LD) using $R^2 > 0.25$ [42]. Finally, we have a total of 5,843 SNPs across 74 control, 65 BPD, and 64 SCZ samples.

It is desirable to seek the most important features from the preprocessed SNP data. We conduct an effective SNP feature selection by using nonnegative singular value approximation (nSVA) [38]. Unlike the traditional model-driven methods that generally assume SNP data subject to a specific probability distribution, nSVA feature selection is a purely data-driven feature selection method. It ranks the importance of each SNP by taking advantage of the nonnegativity of input SNP data. Our previous work shows that it can identify meaningful feature selection for high-accuracy downstream analysis such as classification and pathway analysis. Therefore, we conduct nSVA feature selection from the data with 5843 SNPs to obtain the datasets with the top-ranked 10%, 20%, 30%, …, 90%, and 100% features. The datasets will be employed in the proposed psychiatric map diagnosis.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

4 Psychiatric map analysis

The pMAPs generated from the *fSOM* learning step unveil the latent subtypes of control, BPD, and SCZ. Unlike the original assumptions that samples are partitioned as control, BPD, and SCZ groups, the pMAPs discover that there are two subtypes for control, two subtypes for BPD, and three subtypes for SCZ. It suggests that the pMAP generation procedure provides a meaningful knowledge discovery process to disclose intrinsic data characteristics, which will be essentially important for mislabeled data.

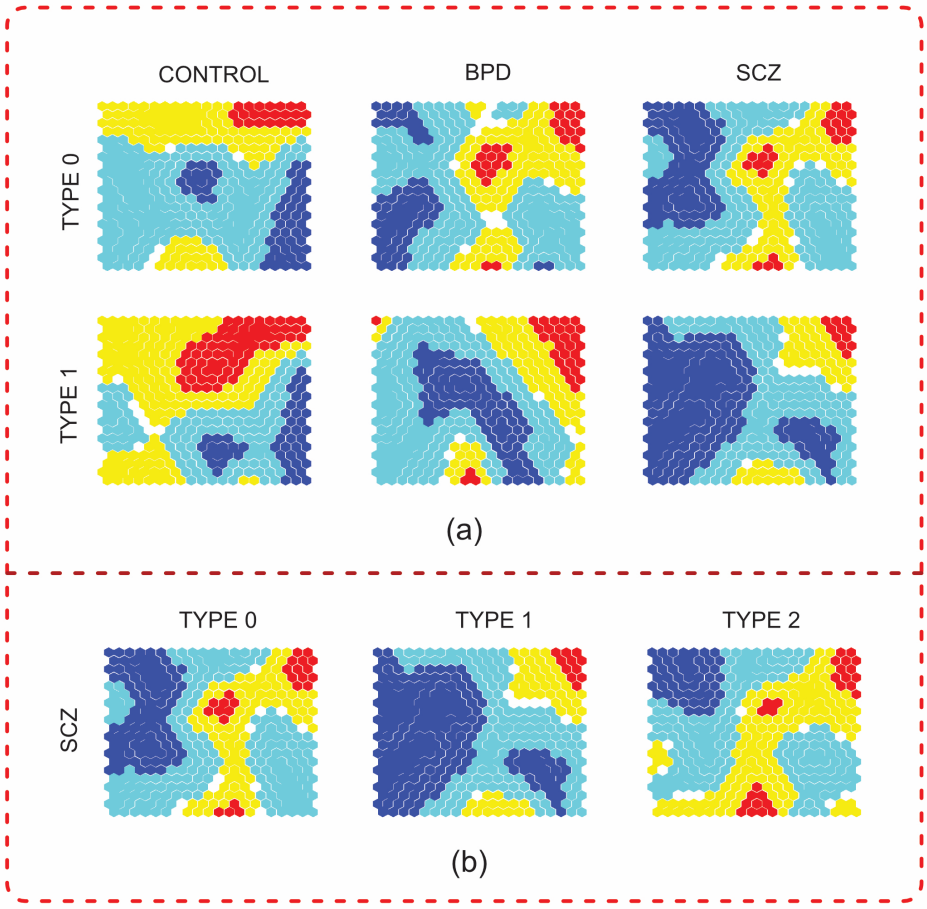

Fig 3. The pMAPs of control, BPD, and SCZ generated from *fSOM* learning. (a) shows the type 0 and type 1 pMAPs of control, BPD, and SCZ. (b) illustrates the pMAPs of all the three discovered SCZ subtypes. The pMAPs of controls demonstrate obvious differences from those of the BPD and SCZ. The pMAPs of SCZ 0 ('type 0'), SCZ 2 ('type 2), and BPD 0 ('type 0') share very similar patterns indicating they are highly potentially mislabeled ones.

4.1 Psychiatric subtype visualizations via pMAPs

The pMAPs provide a powerful visualization technique to discover different new subtypes for the original control, BPD, and SCZ samples. Figure 3 illustrates the pMAPs of the three types of samples as well as their new subtypes discovered in *fSOM* learning. It shows that control has two types of psychiatric maps named 'control' type 0 and type 1, both of which have blue regions, in which the reference vectors have small values, on the right boundary of the SOM plane. The pMAPs of the control samples demonstrate quite clear differences from those of the BPD and SCZ samples. It concurs with our previous results that general

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

machine learning can achieve 98% diagnostic accuracy between control and BPD/SCZ, which is approximately a linearly separable problem.

The pMAPs unveil subtypes with different patterns for those from the same group. For example, Fig 3 (a) illustrates the two different subtypes of pMAPs of the BPD samples, in which the red and blue regions lie on the right and left sides of the *fSOM* plane respectively. Similarly, Fig 3 (b) shows that the SCZ samples consist of three different hidden types of pMAPs, named SCZ 0, 1, and 2 respectively.

However, it is possible that some pMAPs from different types share similar patterns because of widely existing misdiagnosis between BPD and SCZ [25][26]. Under this situation, it is rational to doubt they represent mislabeled samples due to possible misdiagnosis. For example, the pMAPs of SCZ 0 and BPD 0 share very similar patterns, but they are labeled as different types. Since the pMAPs are obtained from unsupervised feature self-organizing map learning (*fSOM*), the pMAPs with similar patterns should stem from the same type rather than different ones. As the pMAPs are 'discovered latent patterns' from input SNP data, the label information of the samples should be consistent with their pMAPs. Therefore, those samples (e.g., SCZ 0 and BPD 0) with similar pMAPs but different labels are highly likely to be mislabeled ones. In other words, the pMAPs can unveil mislabeled samples via a unique visualization viewpoint and build a solid foundation for solving mislabeled learning problems.

4.2 Psychiatric subtype probability density function investigation

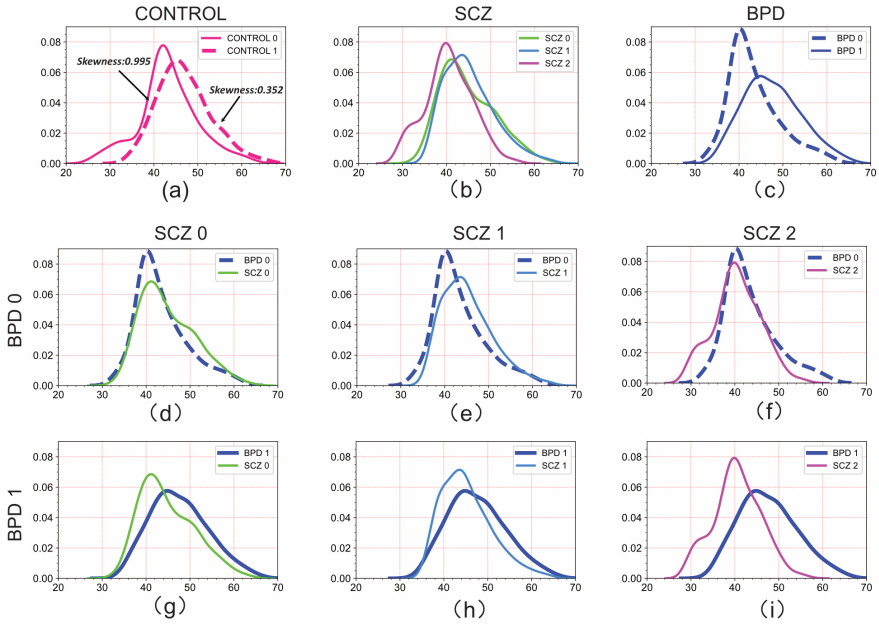

Fig 4. The *p.d.f.*s of the different subtypes of the control, BPD, and SCZ samples. The subplots (a) to (c) illustrate the *p.d.f.*s of the subtypes of the three groups and suggest quite good differences between the control and two psychiatric groups for their different shapes as well as different skewness and kurtosis values. The subplots (d) to (i) illustrate the pairwise comparisons of the *p.d.f.*s of the BPD and SCZ subtypes. It also strongly suggests the similarity between the subtypes BPD 0, SCZ 0, and SCZ 2, which may indicate the occurrence of mislabeled types.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The pMAPs also provide a powerful way for us to build the probability density function (*p.d.f*) for each type of sample, which will contribute to detecting mislabeled samples in a more rigorous way [44][45]. We apply Gaussian kernel density estimation to the reference vector matrix of each sample, which is reshaped as a 1-dimensional vector, to estimate the probability density functions of different types of samples [46][47]. Figure 4 compares the probability density functions of the control, BPD, and SCZ samples. It shows that the two control subtypes have clear differences with BPD and SCZ in their *p.d.f.*s.

The pMAPs provide a more rigorous way to detect different discovered psychiatric subtypes by constructing the probability density function (*p.d.f.)* for each subtype. Traditionally, it is almost unlikely to obtain the *p.d.f.s* of different psychiatric types for high-dimensional data due to the lack of rigorously statistical theory and techniques. We estimate the *p.d.f.* of each subtype by employing the reference vector matrix of each psychiatric sample on the *fSOM* plane, which is reshaped as a corresponding 1-dimensional vector, via Gaussian kernel density estimation techniques [46][47].

Figure 4 further compares probability density functions (*p.d.f.*s) of the subtypes of the control, BPD, and SCZ samples from different perspectives, where the horizontal direction represents the gene expression levels. The subfigures from (a) to (c) summarize the *p.d.f.*s of the subtypes in each group and indicate their obvious differences, where the *p.d.f.*s of the two control subtypes have very different skewness. For example, the skewness and kurtosis values of 'control 0' are 0.9946 and -0.3445, but those of 'control 1' are 0.3517 and -1.3085.

The subfigures from (d) to (i) pairwisely compare the *p.d.f.*s of the subtypes of the BPD and SCZ groups. It also strongly suggests the similarity between the subtypes BPD 0, SCZ 0, and SCZ 2, which may indicate the occurrence of mislabeled types. On the other hand, the ranges of the SNP expression levels of the different *p.d.f.*s of the subtypes fall in different intervals on the *fSOM* plane. For example, the range of the SNP expressions of *p.d.f.s* of the control 0 and control 1 fall in [20, 70] and [28,70] respectively. It illustrates the good sensitivity of *fSOM* learning in uncovering the latent data characteristics of each group in a low-dimensional space.

4.3 Mislabeled sample detection and relabeling via pMAP clustering

It is desirable to seek the ground truth labels and tackle the mislabeled issue by clustering the pMAPs, because those samples from the same psychiatric type will be more likely to be grouped in the same cluster for the similarities of their pMAPs patterns. As a widely used clustering algorithm with simplicity and good explainability, K-means could be a good candidate to accomplish it, especially because the general number of clusters is already known in our context [48]. However, K-means clustering only works well for convex data but may fail badly for non-convex data [49]. However, there is no guarantee that the pMAPs from the *fSOM* learning would be convex. Furthermore, K-means clustering is sensitive to noise that can be interpreted as some mislabeled samples. Besides, it may lead to possible sub-optimal clustering results because of the trapping in the local minimum easily [50].

Therefore, we may need a clustering algorithm that can handle both convex and nonconvex data as well as demonstrate robustness to noise. DBSCAN is a state-of-the-art clustering algorithm that works well for arbitrary-shaped data and is robust to noise. More importantly, DBSCAN can automatically find the number of clusters for input data, which is particularly useful to seek different psychiatric subtypes and unveil latent clustering structures. Thus, DBSCAN prepares itself as a good candidate for pMAP clustering to screen the possible mislabeled samples to look for the ground truth.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The results of DBSCAN clustering demonstrate that those originally mislabeled samples will be more highly likely to fall into the same cluster because of the proximity of their pMAPs. For example, a BPD 0 sample will be clustered into the same cluster as an SCZ 0 or SCZ 2 sample because their pMAPs share good similarities that suggests the possible mislabel happen to them. Technically, they are misdiagnosed patients in clinical practice because of their similar psychiatric symptoms according to the current BPD and SCZ categorization standards [51].

Figure 5 shows the DBSCAN clusters the pMAPs of the three groups as 5 subclusters, where the control samples are separated well from those of the BPD and SCZ samples. However, the BPD 0, SCZ 0, and SCZ 2 samples are clustered in the two close subclusters and BPD 1 and SCZ 1 samples are clustered as a relatively independent subcluster. Since the samples with different labels falling in the same cluster may indicate the possible mislabel issue, we relabel them to form new psychiatric groups BPD* and SCZ* separately to reflect the ground truth better, i.e., the cluster consisting of BPD 1 and SCZ 1 samples from the BPD* group and the clusters consisting of BPD 0, SCZ 0, and SCZ 2 generate the SCZ* group.

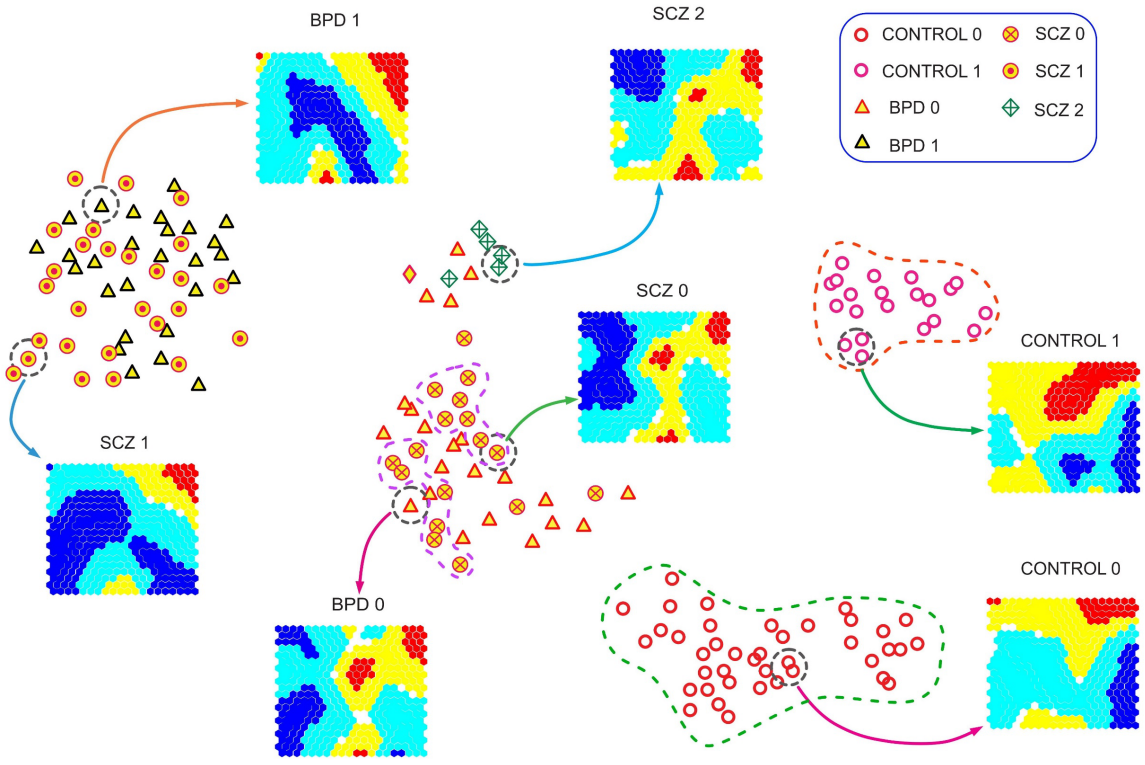

Fig 5. The structure of the pMAP clustering. The pMAP clustering result consists of 5 subclusters under DBSCAN: the control group partitioned as the control 0 and 1 subclusters is clearly separated from the mixed BPD and SCZ groups consisting of three subclusters. Different colors indicate different scales of the values in the pMAP, i.e., red and yellow symbolize the largest numerical values, and dark blue indicates the corresponding numerical values close to zero. The BPD 0, SCZ 0, and SCZ 2 samples, which are clustered in the two subclusters for their similar pMAPs, form a new psychiatric group BPD*. So are the clustered BPD 1 and SCZ 1 samples that generate another psychiatric group SCZ*.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

4.4 The devolution paths of psychiatric states via relative entropy analysis

Devolution path and intrinsic transfer. To further demonstrate the possible pathological devolution path, we conduct a novel relative entropy analysis for the pMAPs after the relabeling procedure. The devolution path refers to the generic devolution process from a normal psychiatric state to dysregulated psychiatric states such as bipolar disorders or schizophrenia states. Similarly, we call the change between two subtypes of dysregulated psychiatric states an internal transfer.

The devolution path can be inferred by calculating the relative entropy, i.e., the K-L divergence of the different psychiatric states. Unlike the traditional symmetric distances, the non-symmetry of the K-L divergence provides a good measure to evaluate the devolution distance between the two psychiatric states. It is almost theoretically impossible to achieve it using the original SNP data because the probability distribution of high-dimensional SNP data is unknown. However, we can define the K-L divergence by using the pMAPs of each psychiatric group as follows.

Given two datasets $X = \{x_i\}_{i=1}^{n}, in\ \mathcal{R}^{n\times p}$ and $T = \{t_i\}_{i=1}^{m}, \mathcal{R}^{m\times p}$ representing two different psychotic groups, the KL divergence between them is defined as,

$$KL(p \parallel q) = \sum_{i=1}^{n} p_i log \frac{p_i}{q_i} \tag{4}$$

where $p_j = s_j / \sum_{j=1}^{p} s_j, s_j$ is the j^{th} singular value of X and $q_j = u_j / \sum_{j=1}^{p} u_j, u_j$ is the j^{th} singular value of T.

We use the pMAPs of each psychotic group (e.g., control 1) to represent their original one to calculate their K-L divergence. Figure 6 illustrates the K-L divergences from the two control subgroups to the other two dysregulated psychiatric states. The results seem to echo our previous relabeling result as well as provide more insights for possible devolution paths. Figure 6 (a) shows that the control 1 subgroup is closer to BPD 1 in terms of KL-divergence than control 0. It suggests that the control 1 state tends to go devolution to the BPD 1 state more likely compared to the control 0 state. That both control 0 and control 1 have almost the same K-L divergence to the SCZ 0 suggests they have the same likelihood to devote to the 'SCZ 0' state. Similarly, the 'control 1' state seems to devote to the SCZ 2 with a higher likelihood than the 'control 0' state. Figure 6 (b) illustrates that the. The 'BPD 1' state is least likely to conduct an internal transfer to the SCZ 2 compared to other dysregulated psychiatric states.

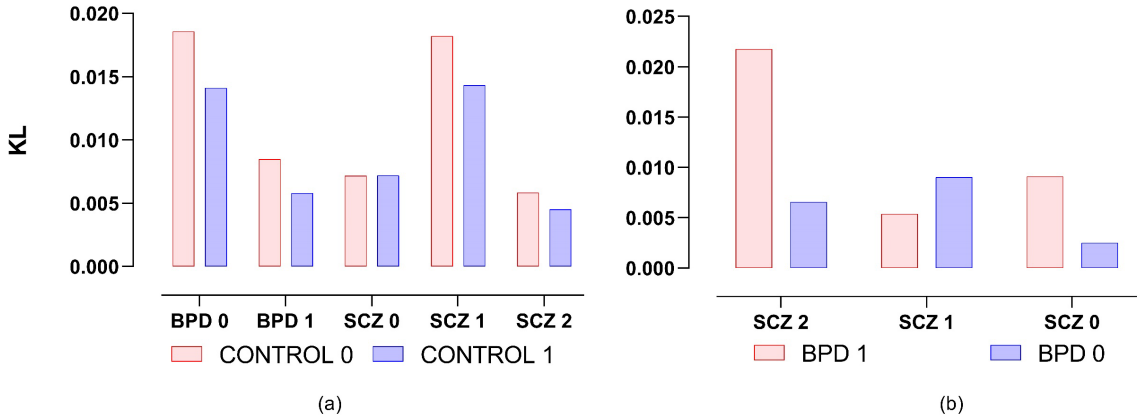

Fig 6. The K-L divergence analysis between the different psychiatric groups based on pMAPs. The subfigure (a) shows the K-L divergences of the controls 0 and 1 with respect to BPD 0, 1, and SCZ 0,1,2 respectively. The subfigure (b) describes the K-L divergence relationships between BPD 0, 1, and SCZ 0,1,2. It suggests the possible devolution paths and internal transfers between different psychiatric states.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We find that the control group demonstrates interesting devolution paths with respect to the BPD* and SCZ* groups according to their pMAP pattern analysis. The devolution path refers to the generic devolution process from a normal psychiatric state to a dysregulated psychiatric state such as bipolar disorders or schizophrenia states. We discover that the pMAP patterns of the control, BPD*, and SCZ* demonstrate a more and more complicated tendency even if the control & BPD* and BPD* & SCZ* somewhat show some level of similarity.

Integrating with the previous relabeling, we have the following interesting devolution path information. The control groups have a shorter devolution path genetically to the BPD* group for their relatively less K-L divergence values. On the other hand, a longer devolution path can exist from the control to the SCZ* group because they have a larger K-L divergence.

Figure 7 illustrates the possible devolution paths of the control groups to the BPD* and SCZ* respectively according to their K-L divergences as well as internal transfers between the disease states. Classic psychiatric studies seem to support the devolution path because a subject can start from a normal psychiatric state to a more dysregulated, complicated, or unstable state in a gradual or abrupt manner. Moreover, previous studies also support the finding because it was reported that bipolar disorder can be a transition state between the normal and schizophrenia states [52][53][54][55]. Figure 7 shows the possible devolution paths and internal transfers between different psychiatric states, where the KL divergence value is marked for each path or internal transfer.

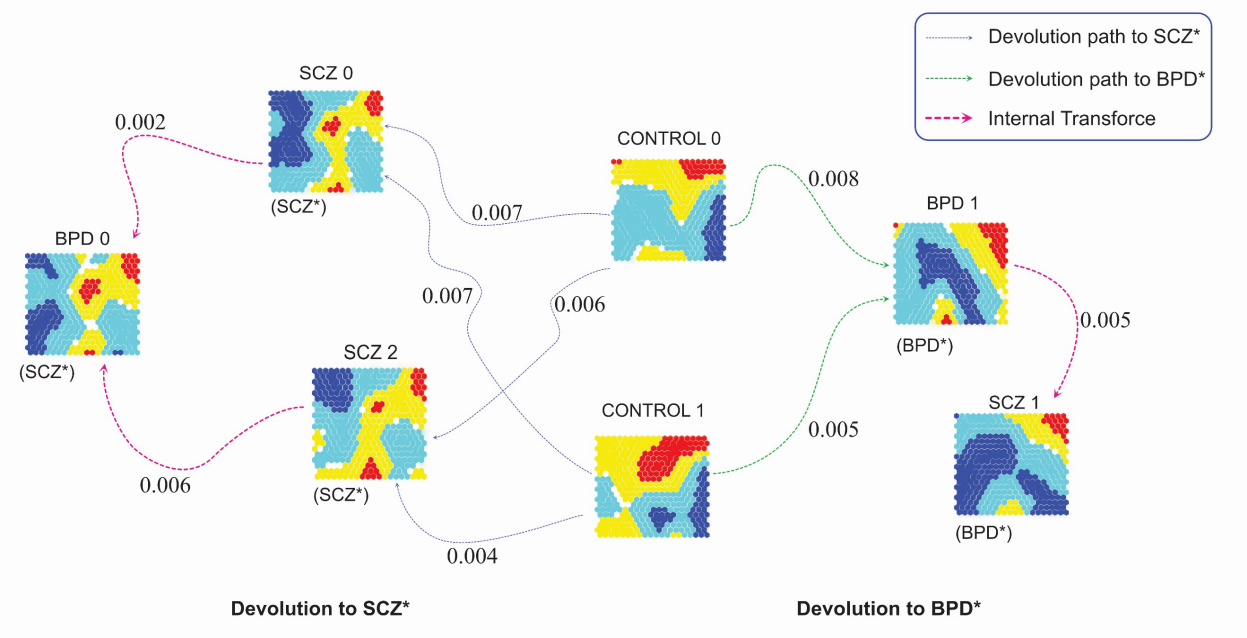

Fig 7. The possible devolution paths of the control groups to the BPD* and SCZ* according to their pMAP patterns. The pMAPs of the BPD* and SCZ* groups have more complicated patterns than those of controls. It indicates that the BPD* is an intermediate psychiatric state between the SCZ* and control.

5 Comparisons of psychiatric map diagnosis with peer methods

We conduct control, BPD, and SCZ diagnosis with the relabeled data to validate the correctness and

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

effectiveness of the relabeling. The correct relabeling should lead to good improvements in diagnostic accuracy. We mainly employ multi-class support vector machines (SVM) to conduct the psychiatric map diagnosis for SVM's good reproducibility, transparency, and interpretability in this study [40]. Although deep neural networks (DNN) and ensemble learning methods such as random forests (RF) and extremely randomized trees (Extree), can achieve decent performance also, they especially lack good reproducibility for their built-in randomness, which is essential for clinical psychiatric diagnosis [56][57][58][59]. We further employ nonnegative singular value approximation (nSVA) for SNP feature selection for its proven effectiveness and efficiency for high-dimensional data [60][61].

Since the traditional classification measures are neither efficient nor interpretable in assessing different machine learning models' performance, we extend the proposed diagnostic index (d-index) measure under binary classification to provide a more explainable and sensitive learning performance evaluation. This is because the traditional classification measure assessment may only reflect one aspect of classification performance. As a result, it is inconvenient to compare many classification measures for different machine learning model performances on different datasets in a more explainable approach. The d-index definition of the binary classification can be found in the following section and more d-index information can be found in [61].

5.1 Diagnostic index (d-index)

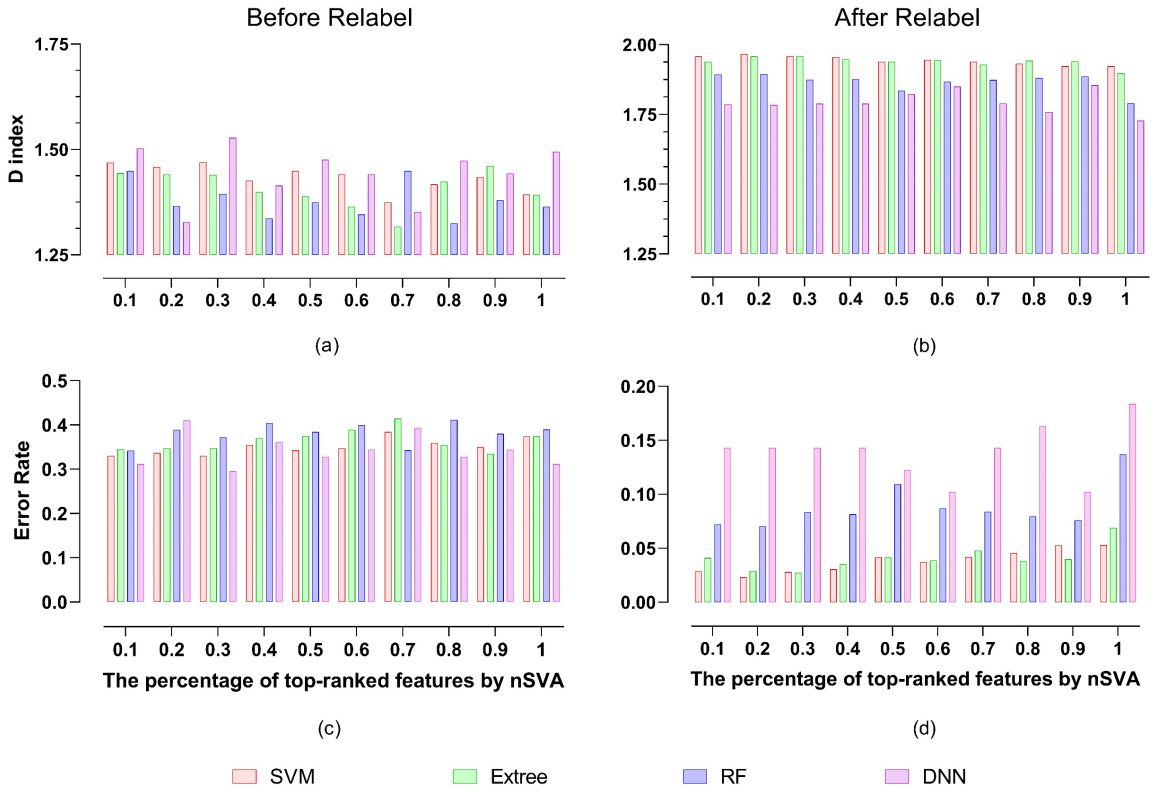

Fig 8. The comparisons of control, BPD, and SCZ diagnoses before and after relabeling under nSVA feature selection under four machine learning models.

Given a prediction function $\hat{f}(x): x \to \{-1,1\}$ constructed from training data $X_r = \{x_i, y_i\}_i^m$ under a machine learning model Θ, where each sample $x_i \in \mathcal{R}^p$ and its label $y_i \in \{-1,1\}$, $i = 1,2,\cdots m$, d-index evaluates the performance of $\hat{f}(x)$ in predicting the class of test data $X_s = \{x_j', y_j'\}_j^l$. It is defined as

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$d = log_2(1+a) + log_2(1+\frac{s+p}{2}) \quad (5)$$

where a, s and p represent the corresponding accuracy, sensitivity, and specificity in diagnosing test data respectively. The larger the d-index value, the better the predictability of $\hat{f}(x)$, *i.e.,* the better learning performance achieved by the machine learning model Θ. The maximum value of d-index is 2 where classification has the perfect results. The minimum value of the d-index is $2log_2(\frac{3}{2})$ if there is no underfitting [61].

Figure 8 compares the d-index values and misclassification rates before and after relabeling in detecting control, BPD, and SCZ, under four machine learning models: SVM, RF, ET, and DNN, where training and test datasets have 80% and 20% samples of the total samples respectively. We employ nSVA to select p% (p=10, 20, …100) features to observe how the relabeling results impact those datasets with different percentages of SNP features. The d-index values of the diagnoses after relabeling are much higher than those before relabeling for all datasets across all four models under nSVA feature selection. It strongly suggests the correctness and effectiveness of the relabeling. Furthermore, all the models have low misclassification rates after relabeling and SVM had the lowest ones, which indicates the strong reproducibility of the proposed psychiatric map diagnosis.

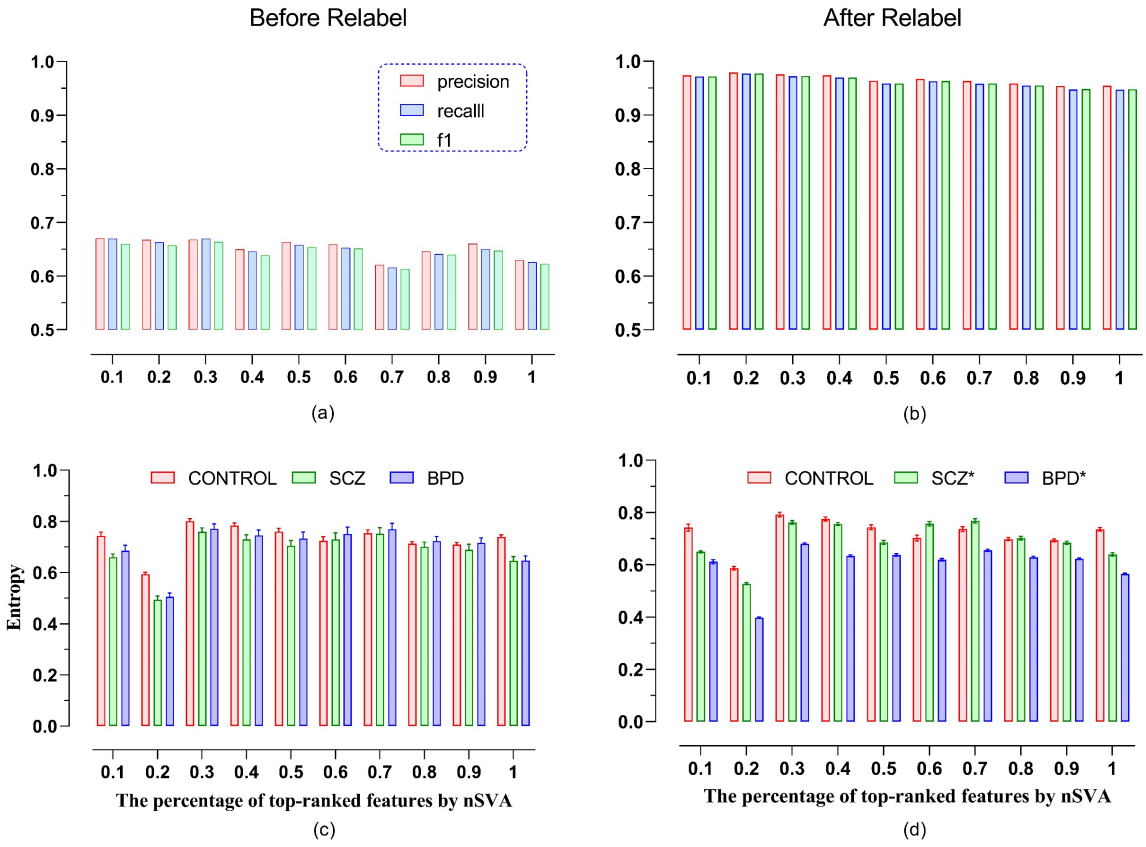

Fig 9. Classification metric and entropy comparisons before and after relabeling under different percentages of top-ranked features by nSVA. The subfigures (a) and (b) compares the values of recall precision, F1-score before and after relabeling before and after relabeling under multi-class SVM. The subfigures (c) and (d) illustrate the differences of entropies of control, SCZ, and BPD.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 9 compares the precision, recall, and F1 values under multi-class SVM as well as entropy values before and after relabeling under different percentages of features selected by nSVA. The precision, recall, and F1 values are consistent with the d-index values well. Interestingly, we have found that the BPD and SCZ groups have relatively smaller entropy values than the control group. However, the entropies of the relabeled subgroups demonstrate regular patterns: the relabeled BPD* subgroup regularly has the smallest entropies, and SCZ* has the second-smallest entropies under different SNP feature sets selected under nonnegative singular value approximation (nSVA). It not only validates the correctness of relabeling but also suggests that the psychiatric disorder samples may have more special SNP patterns than those of the normal ones.

5.2 Comparison with state-of-the-art machine learning methods

We compare our results with state-of-the-art deep learning methods that include one-shot learning [62][63], convolutional neural networks (CNN) [64][65], residual neural networks (ResNet) [66][67], long short-term memory (LSTM) [68][69], Transformer [70][71][72], and generative adversarial networks (GAN) [73][74], as well as support vector machines (SVM) [40].

We briefly describe the deep learning models for the convenience of description. CNN is characterized by a partially connected layer structure and different layers have different functionalities such as convolutional, and max/average pooling. It demonstrates powerful learning capabilities, especially for image data besides decent feature extraction. ResNet is an enhanced CNN using residual learning techniques to tackle the challenges of gradient disappearance and explosion in CNN learning. It demonstrates advantages over CNN in handling big and more complicated data. LSTM, which is widely employed in time-series data analysis, overcomes the weakness of general recurrent neural networks (RNN) in handling long-time information dependence by employing LSTM cells that consist of three different gates. GAN employs two different neural networks: a generator and discriminator (e.g., CNN and LSTM) to contest with each other to accomplish learning. GAN stops at the point when the discriminator was completed 'confused' by the learning results from the generator. In addition, Transformer is a special feedforward neural network taking advantage of the self-attention mechanism in topology and learning. It improves the parallelism of the model and decreases its reliance on long-term memory. One-short learning aims to handle the data scarcity issue in deep learning, i.e., input data itself is small enough to satisfy the training demand for the number of observations in training. It creates models that can accurately identify test samples with a limited quantity of training data. More details about the models can be found in the literature [62][63].

Figure 10 compares the proposed pMAP diagnosis with the state-of-the-art deep learning models as well as SVM which is a representative of the classic learning method, under different levels of feature selection by nSVA. It is obvious that the pMAP diagnosis demonstrates its superiority to the rest of the methods no matter in learning effectiveness or stability. It suggests that all deep learning models show quite poor performance on the original data. For example, LSTM obviously fails the whole learning process by encountering overfitting.

There are mainly two reasons to interpret the poor performance from the deep learning models. The first is the dataset is too small to take advantage of the powerful learning capabilities of the deep learning models. The second, which can be more important, is the problem itself is a mislabeled learning problem, but almost all deep learning models assume training data are correctly labeled and there are no techniques available to handle this in deep learning. It also can explain why one-short learning encounters underfitting.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

On the other hand, why the pMAP diagnosis leads all the other methods lies in that it is a specifically designed algorithm for mislabeled learning. It exploits *fSOM* learning to gain the pMAP for each SNP observation and density clustering to seek the similarities between the pMAPs. It takes advantage of the DBSCAN clustering results of the prototypes of the original observations to relabel data to decrease or even eliminate mislabeling information at the most level. Finally, the kernel method SVM is employed to conduct psychiatry prediction by exploiting its reproducibility and efficiency in learning. Therefore, the pMAP diagnosis is more effective, efficient, and robust in handling the mislabeling psychiatry learning problem than its possible peer methods.

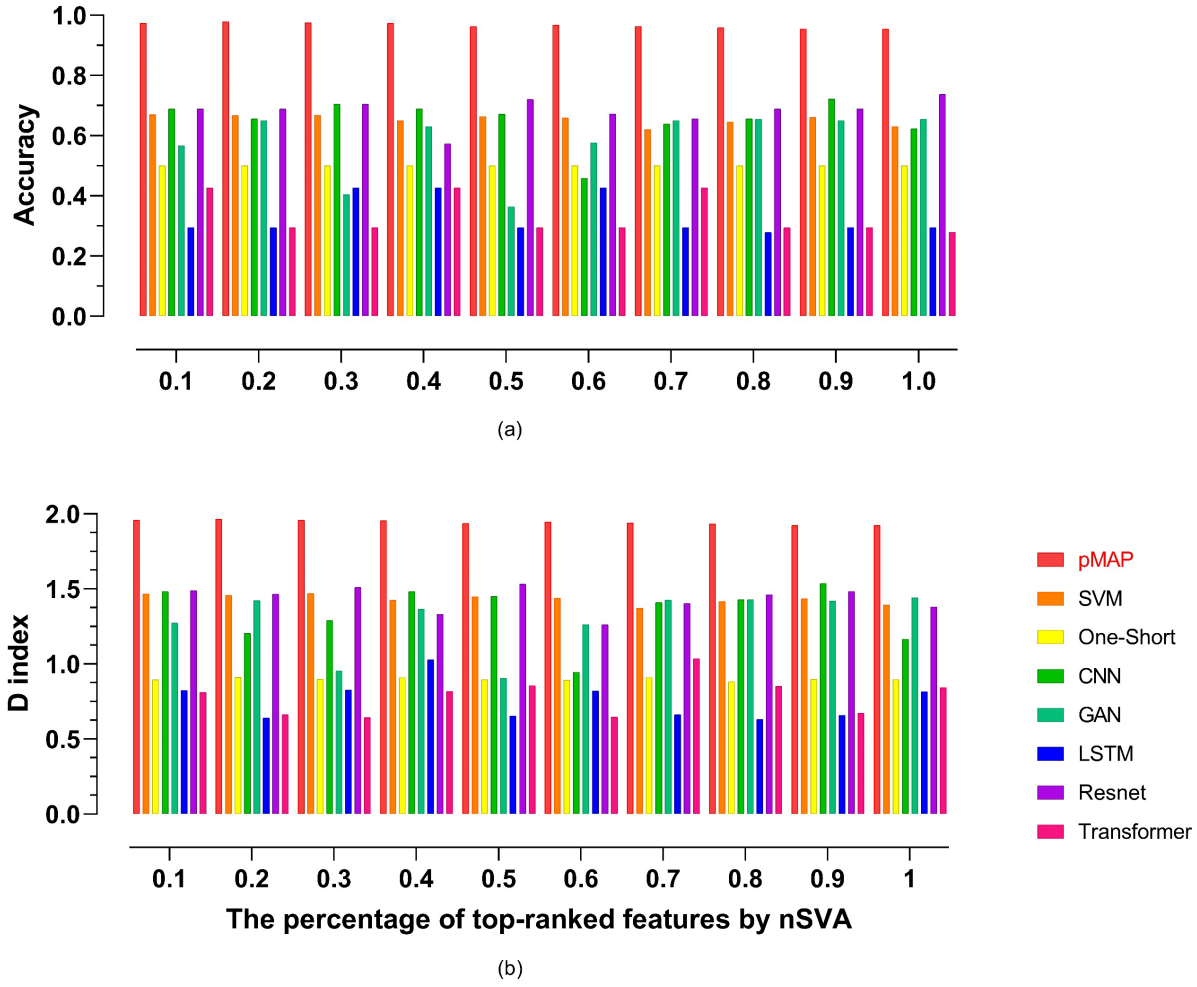

Fig 10 The comparisons of the proposed pMAP diagnosis with its peer methods: SVM, one-short learning, CNN, GAN, LSTM, ResNet, and Transformer. The pMAP diagnosis demonstrates stably leading performance compared to its peers under different levels of feature selections. Almost all deep learning models show poor performance for the original data. Both one-short learning and LSTM encounter underfitting.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

6 Discussion and conclusion

We point out that misdiagnosis between BPD and SCZ can be unavoidable due to the existing psychiatric standards in psychiatry. The existing behavior-based definition and categorization for BPD and SCZ do not include genetic analysis that should provide a more accurate classification. On the other hand, detecting BPD and SCZ using SNP or other bioinformatics data via ML is to handle a mislabeled learning problem for high dimensional data, because the label information provided from the psychiatry practice can count mislabeled information. To the best of our knowledge, mislabeled learning is a rarely investigated but essentially important and challenging problem in modern AI and data science. With the surge of big data and AI, more and more mislabeled learning problems need serious investigations according to their 'data background'. Simply assuming the label information is correct would cause ML to encounter mediocre or poor performance and produce a serious misdiagnosis that is happening in many AI-driven disease diagnoses such as mental disorder detection, especially for those data with a limited number of observations.

The proposed psychiatric map diagnosis employs the feature self-organizing map (*fSOM*) learning to tackle the high-dimensional mislabel learning problem successfully. It generates the low-dimensional prototype: pMAP for each observation that synthesizes the essential characteristics of each observation. The pMAPs discover and unveil new knowledge for input data, i.e., it identifies different unknown hidden subtypes for each group. For example, it finds that there are two subtypes in the control group, 2 subtypes in the BPD, and 3 subtypes in the SCZ group. The pattern similarities between some subtypes from different groups further validate the existence of the mislabeled samples. We employ a novel relabel technique to correct label information according to DBSCAN clustering. Finally, we conduct reproducible SVM prediction based on the relabeled data and achieved leading performance in comparison with peer deep learning methods. To the best of our knowledge, it is the first time to handle misdiagnosis between BPD and SCZ via inventing novel ML methods to overcome the challenge of mislabeled learning from high-dimensional data.

Furthermore, the devolution path of psychiatric states via relative entropy analysis provides insights into existing pathological psychiatry. The novel devolution path analysis unveils latent internal transfer and devolution road maps between different subtypes of the control, BPD, and SCZ groups, which have been rarely investigated in the existing psychiatry studies and bioinformatics research. It will inspire more future studies on this topic via similar devolution path analysis ways. However, it still needs more data and experiments or genetic findings to validate its effectiveness furthermore.

The proposed pMAP diagnosis is an explainable machine learning method for high-dimensional data mislabel learning, because of its transparent structure, good learning efficiency, and trustworthy learning results. Compared to the existing deep learning models with thousands of parameters, the pMAP diagnosis has only a few parameters that contribute to a more interpretable learning architecture and learning performance evaluation.

Although this method is designed to tackle mislabeled learning for high-dimensional data, it can be easily applied or extended to more general mislabeled learning because the essential components of the pMAP diagnosis such as *fSOM*, DBSCAN, multiclass SVM can essentially apply to any kind of data. However, it is noted that *fSOM* may need a huge computing demand in generating pMAPs especially because the dimensionality of input data is high. We have spent about 2 weeks completing *fSOM* learning on a 20x20 *fSOM* plane to generate 203 pMAPs on an Intel Xeon E5-2620 machine under OS Ubuntu 20.04 LST (Focal Fossa) with RAM 128Gb and CPU speed 2.1Ghz. We are seeking to implement *fSOM* via an FPGA approach

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

to tackle the high computing demands [75][76].

In addition, there are quite a few aspects to be improved in the proposed pMAP diagnosis. For example, it is desirable to design more customized kernels in multiclass SVM rather than rely on some standard nonlinear kernels (e.g., Gaussian kernel), which can be especially important for other types of data. Furthermore, we are interested in exploring a hierarchical *fSOM* to decrease its complexity for the sake of its good adaptability, especially for big data besides enhancing the current batch processing [77].

Data availability

The data of this work is publicly available at: https://github.com/AidenLee1994/Psychiatric-map-diagnosis-in-mental-disorders/data

Code availability

The code of this work is publicly available at: https://github.com/AidenLee1994/Psychiatric-map-diagnosis-in-mental-disorders

Acknowledgment

This work was supported in part by the National Science Foundation of China under Grant 61572367, Grant 61573017, and McCollum endowed chair startup fund.

Declaration of competing interests

The authors declare they have no conflicts of interest. All authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

Authors' contributions

Han designed the methods and algorithms. Li, Liu, and Han analyzed the data. Liu and Han wrote the manuscript. Han and Liu finalized the paper.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Reference

[1] McGuinness, A. J., et al. "A systematic review of gut microbiota composition in observational studies of major depressive disorder, bipolar disorder and schizophrenia." *Molecular psychiatry* 27.4 (2022): 1920-1935.

[2] Smail, Marissa A., et al. "Similarities and dissimilarities between psychiatric cluster disorders." *Molecular psychiatry* 26.9 (2021): 4853-4863.

[3] Karege, Félicien, et al. "Genetic overlap between schizophrenia and bipolar disorder: a study with AKT1 gene variants and clinical phenotypes." *Schizophrenia research* 135.1-3 (2012): 8-14.

[4] Vita, Antonio, et al. "Effectiveness, core elements, and moderators of response of cognitive remediation for schizophrenia: a systematic review and meta-analysis of randomized clinical trials." *JAMA psychiatry* 78.8 (2021): 848-858.

[5] Legge, Sophie E., et al. "Genetic architecture of schizophrenia: a review of major advancements." *Psychological medicine* (2021): 1-10.

[6] Goodwin, Frederick K., and S. Nassir Ghaemi. "Bipolar disorder." *Dialogues in Clinical Neuroscience* (2022).

[7] Ghaemi, S. Nassir, James Y. Ko, and Fred er ick K. Goodwin. "“Cade's disease” and beyond: misdiagnosis, antidepressant use, and a proposed definition for bipolar spectrum disorder." *The Canadian Journal of Psychiatry* 47.2 (2002): 125-134.

[8] Dubovsky, Steven L., et al. "Psychotic depression: Diagnosis, differential diagnosis, and treatment." *Psychotherapy and Psychosomatics* 90.3 (2021): 160-177.

[9] Birur, Badari, et al. "Brain structure, function, and neurochemistry in schizophrenia and bipolar disorder—a systematic review of the magnetic resonance neuroimaging literature." *NPJ schizophrenia* 3.1 (2017): 1-15.

[10] Oh, Jihoon, et al. "Identifying schizophrenia using structural MRI with a deep learning algorithm." *Frontiers in psychiatry* 11 (2020): 16.

[11] Martino, Matteo, et al. "Contrasting variability patterns in the default mode and sensorimotor networks balance in bipolar depression and mania." *Proceedings of the National Academy of Sciences* 113.17 (2016): 4824-4829.

[12] Wada, Masataka, et al. "Dopaminergic dysfunction and excitatory/inhibitory imbalance in treatment-resistant schizophrenia and novel neuromodulatory treatment." *Molecular Psychiatry* (2022): 1-18.

[13] Reale, Marcella, Erica Costantini, and Nigel H. Greig. "Cytokine imbalance in schizophrenia. From research to clinic: potential implications for treatment." *Frontiers in Psychiatry* (2021): 82.

[14] Swathi, N., and S. X. Prabha. "Survey on Structural Neuro imaging for the Identification of brain Abnormalities in Schizophrenia." *Current Medical Imaging* (2022).

[15] Moreau, Clara A., et al. "Dissecting autism and schizophrenia through neuroimaging genomics." *Brain* 144.7 (2021): 1943-1957.

[16] Haukvik, Unn K., et al. "In vivo hippocampal subfield volumes in bipolar disorder—A mega-analysis from The Enhancing Neuro Imaging Genetics through Meta-Analysis Bipolar Disorder Working Group." *Human brain mapping* 43.1 (2022): 385-398.

[17] Ching, Christopher RK, et al. "What we learn about bipolar disorder from large-scale neuroimaging: findings and future directions from the ENIGMA Bipolar Disorder Working Group." *Human brain mapping* 43.1 (2022): 56-82.

[18] Dong, Xianjun, Chunyu Liu, and Mikhail Dozmorov. "Review of multi-omics data resources and integrative analysis for human brain disorders." *Briefings in functional genomics* 20.4 (2021): 223-234.

[19] Guan, Fanglin, et al. "Integrative omics of schizophrenia: from genetic determinants to clinical classification and risk prediction." *Molecular Psychiatry* 27.1 (2022): 113-126.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[20] Zhang, Chu-Yi, et al. "An alternative splicing hypothesis for neuropathology of schizophrenia: evidence from studies on historical candidate genes and multi-omics data." *Molecular Psychiatry* 27.1 (2022): 95-112.

[21] Du, Yang, et al. "Exosome transplantation from patients with schizophrenia causes schizophrenia-relevant behaviors in mice: an integrative multi-omics data analysis." *Schizophrenia bulletin* 47.5 (2021): 1288-1299.

[22] Sahu, Ankur, et al. "Integrative network analysis identifies differential regulation of neuroimmune system in Schizophrenia and Bipolar disorder." Brain, Behavior, & Immunity-Health 2 (2020): 100023.

[23] Li, Shufen, et al. "Altered DNA methylation of the Alu y subfamily in schizophrenia and bipolar disorder." *Epigenomics* 11.6 (2019): 581-586.

[24] Ellis, Shannon E., et al. "Transcriptome analysis of cortical tissue reveals shared sets of downregulated genes in autism and schizophrenia." *Translational psychiatry* 6.5 (2016): e817-e817.

[25] Liu, Wenbin, Dongdong Li, and Henry Han. "Manifold learning analysis for allele-skewed DNA modification SNPs for psychiatric disorders." *IEEE Access* 8 (2020): 33023-33038.

[26] Witt, Stephanie H., et al. "Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia." *Translational psychiatry* 7.6 (2017): e1155-e1155.

[27] Ayano, Getinet, et al. "Misdiagnosis, detection rate, and associated factors of severe psychiatric disorders in specialized psychiatry centers in Ethiopia." Annals of general psychiatry 20.1 (2021): 1-10.

[28] Semmo, N. "Schizophrenic Patient with Drug Induced Liver Injury Initially Misdiagnosed as Wilson's Disease with Final Diagnosis of Celiac Disease." *Ann Clin Med Case Rep* 6.18 (2021): 1-5.

[29] Kohonen, Teuvo, et al. "Self-organization of a massive document collection." *IEEE transactions on neural networks* 11.3 (2000): 574-585.

[30] Schubert, Erich, et al. "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN." *ACM Transactions on Database Systems (TODS)* 42.3 (2017): 1-21.

[31] Zeng, Z., et al. "Common SNPs and haplotypes in DGKH are associated with bipolar disorder and schizophrenia in the Chinese Han population." *Molecular psychiatry* 16.5 (2011): 473-475.

[32] Geschwind, Daniel H., and Jonathan Flint. "Genetics and genomics of psychiatric disease." *Science* 349.6255 (2015): 1489-1494.

[33] Sadeghi, Delaram, et al. "An overview of artificial intelligence techniques for diagnosis of Schizophrenia based on magnetic resonance imaging modalities: Methods, challenges, and future works." *Computers in Biology and Medicine* (2022): 105554.

[34] Yuan, Weiwei, Guangjie Han, and Donghai Guan. "Learning from mislabeled training data through ambiguous learning for in-home health monitoring." *IEEE Journal on Selected Areas in Communications* 39.2 (2020): 549-561.

[35] Kohonen, Teuvo. "Self-organizing neural projections." Neural networks 19.6-7 (2006): 723-733.

[36] Bansal, Ms Aayushi, Dr Rewa Sharma, and Dr Mamta Kathuria. "A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications." *ACM Computing Surveys (CSUR)* (2021).

[37] Chen, Xing, et al. "A novel relationship for schizophrenia, bipolar and major depressive disorder Part 5: a hint from chromosome 5 high density association screen." *American journal of translational research* 9.5 (2017): 2473.

[38] Han, Henry. "A novel feature selection for RNA-seq analysis." *Computational biology and chemistry* 71 (2017): 245-257.

[39] Han, Henry, et al. "Enhance Explainability of Manifold Learning." *Neurocomputing* 500:877-895 2022.

[40] Wu, Ting-Fan, Chih-Jen Lin, and Ruby Weng. "Probability estimates for multi-class classification by pairwise coupling." *Advances in Neural Information Processing Systems* 16 (2003).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[41] Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." kdd. Vol. 96. No. 34. 1996. Goldberger, Jacob, Shiri Gordon, and Hayit Greenspan. "An Efficient Image Similarity Measure Based on Approximations of KL-Divergence Between Two Gaussian Mixtures." *ICCV*. Vol. 3. 2003.

[42] Gagliano, Sarah A., et al. "Allele-skewed DNA modification in the brain: relevance to a schizophrenia GWAS." *The American Journal of Human Genetics* 98.5 (2016): 956-962.

[43] Meaburn, Emma L., Leonard C. Schalkwyk, and Jonathan Mill. "Allele-specific methylation in the human genome: implications for genetic studies of complex disease." *Epigenetics* 5.7 (2010): 578-582.

[44] Haworth, Daniel Connell. "Progress in probability density function methods for turbulent reacting flows." *Progress in Energy and combustion Science* 36.2 (2010): 168-259.

[45] Wang, Aiping, and Hong Wang. "Survey on stochastic distribution systems: A full probability density function control theory with potential applications." Optimal Control Applications and Methods 42.6 (2021): 1812-1839. Chen, Yen-Chi. "A tutorial on kernel density estimation and recent advances." *Biostatistics & Epidemiology* 1.1 (2017): 161-187.

[46] Chen, Yen-Chi. "A tutorial on kernel density estimation and recent advances." *Biostatistics & Epidemiology* 1.1 (2017): 161-187.

[47] Wang, Zhipeng, and David W. Scott. "Nonparametric density estimation for high-dimensional data—Algorithms and applications." *Wiley Interdisciplinary Reviews: Computational Statistics* 11.4 (2019): e1461.

[48] Arthur, David, and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. *Stanford*, 2006.

[49] Bhargav, Sushant, and Mahesh Pawar. "A review of clustering methods forming non-convex clusters with missing and noisy data." *IJCSE* 4 (2016): 39-44.

[50] Russo, Stefania, and Stefano Bonassi. "Prospects and Pitfalls of Machine Learning in Nutritional Epidemiology." *Nutrients* 14.9 (2022): 1705.

[51] Gratten, Jacob, et al. "Large-scale genomics unveils the genetic architecture of psychiatric disorders." *Nature neuroscience* 17.6 (2014): 782-790.

[52] Fogarty, F., et al. "Mania." *Acta Psychiatrica Scandinavica* 89 (1994): 16-23.

[53] Maier, Wolfgang, Astrid Zobel, and Michael Wagner. "Schizophrenia and bipolar disorder: differences and overlaps." *Current Opinion in Psychiatry* 19.2 (2006): 165-170.

[54] Perlis, Roy H., et al. "Transition to mania during treatment of bipolar depression." *Neuropsychopharmacology* 35.13 (2010): 2545-2552.

[55] Santa Cruz, Elisa C., et al. "Association between trace elements in serum from bipolar disorder and schizophrenia patients considering treatment effects." *Journal of Trace Elements in Medicine and Biology* 59 (2020): 126467.

[56] Montavon, Grégoire, Wojciech Samek, and Klaus-Robert Müller. "Methods for interpreting and understanding deep neural networks." *Digital Signal Processing* 73 (2018): 1-15.

[57] Song, Hwanjun, et al. "Learning from noisy labels with deep neural networks: A survey." *IEEE Transactions on Neural Networks and Learning Systems* (2022).

[58] Chen, Xi, and Hemant Ishwaran. "Random forests for genomic data analysis." *Genomics* 99.6 (2012): 323-329.

[59] Geurts, Pierre, Damien Ernst, and Louis Wehenkel. "Extremely randomized trees." *Machine learning* 63.1 (2006): 3-42.

[60] Han, Henry. "Interpretable Machine Learning Assessment." Available at SSRN: https://ssrn.com/abstract=4146556

[61] Han, Henry, and Ke Men. "How does normalization impact RNA-seq disease diagnosis?." *Journal of biomedical informatics* 85 (2018): 80-92.

[62] Vinyals et al. "Matching networks for one shot learning." *Advances in neural information processing systems* 29 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278675; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[63] Fei-Fei, Li, Robert Fergus, and Pietro Perona. "One-shot learning of object categories." *IEEE transactions on pattern analysis and machine intelligence* 28.4 (2006): 594-611.

[64] Gu, Jiuxiang, et al. "Recent advances in convolutional neural networks." *Pattern Recognition* 77 (2018): 354-377.

[65] O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." *arXiv preprint arXiv*:1511.08458 (2015).

[66] Xie, Saining, et al. "Aggregated residual transformations for deep neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. 2017.

[67] Zagoruyko, Sergey, and Nikos Komodakis. "Wide residual networks." *arXiv preprint* arXiv:1605.07146 (2016).

[68] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

[69] Van Houdt, Greg, Carlos Mosquera, and Gonzalo Nápoles. "A review on the long short-term memory model." *Artificial Intelligence Review* 53.8 (2020): 5929-5955.

[70] Han, Kai, et al. "Transformer in transformer." *Advances in Neural Information Processing Systems* 34 (2021).

[71] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv*:1409.0473 (2014).

[72] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv*:1810.04805 (2018).

[73] Creswell, Antonia, et al. "Generative adversarial networks: An overview." *IEEE Signal Processing Magazine* 35.1 (2018): 53-65.

[74] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).

[75] Tisan, Alin, and Marcian Cirstea. "SOM neural network design–A new Simulink library based approach targeting FPGA implementation." *Mathematics and Computers in Simulation* 91 (2013): 134-149.

[76] Kurdthongmee, W. "A novel hardware-oriented Kohonen SOM image compression algorithm and its FPGA implementation." *Journal of Systems Architecture* 54.10 (2008): 983-994.

[77] Ontrup, Jörg, and Helge Ritter. "Large-scale data exploration with the hierarchically growing hyperbolic SOM." *Neural networks* 19.6-7 (2006): 751-761.