Biomarkers of brain diseases

Pascal Helson[1,2†] and Arvind Kumar[1,2†]

[1]School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology Stockholm, Lindstedsvagen 5, Stockholm, 10040, Sweden.
[2]Science For Life Laboratory, Tomtebodavagen 23, Solna, 17165, Sweden.

Contributing authors: pascal.hellson@gmail.com; arvkumar@kth.se;
[†]These authors contributed equally to this work.

Abstract

Despite the diversity of brain data acquired and advanced AI-based algorithms to analyze them, brain features are rarely used in clinics for diagnosis and prognosis. Here we argue that the field continues to rely on cohort comparisons to seek biomarkers, despite the well-established degeneracy of brain features. Using a thought experiment, we show that more data and more powerful algorithms will not be sufficient to identify biomarkers of brain diseases. We argue that instead of comparing patient versus healthy controls using single data type, we should use multimodal (e.g. brain activity, neurotransmitters, neuromodulators, brain imaging) and longitudinal brain data to guide the grouping before defining multidimensional biomarkers for brain diseases.

Introduction

Behavior is the most obvious readout of the brain function. Therefore, the diagnosis of a brain disease entails a number of behavioral assessments (e.g. the Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) for Parkinson's disease [1] and the Montreal Cognitive Assessment (MoCA) for Dementia [2]). Once the behavior-based diagnosis is narrowed down, the process to establish physiological causes or biomarkers of the behavior symptoms can start. Biomarkers are essential for deciding the treatment plan and prognosis of the disease.

More recently, there is also an impetus to define biomarkers for the diagnosis of brain diseases, especially for early diagnosis. Typically, the biomarkers are searched at the molecular level (e.g. gene expression and chemical imbalances) or morphological level (e.g. loss of neural or glial cells) or brain activity (fMRI/MEG/EEG/ECOG) and network structure (sMRI, PET, connectome) levels. However, despite years of research, such brain-based biomarkers have not superseded the behavior-based markers of brain diseases. It is counter intuitive that the brain does not play a crucial role in the diagnosis of its own disorders. It is possible that the current biomarker data is not detailed enough and analysis is not sophisticated enough.

Advances in AI/ML methods have brought fresh hope to define better biomarkers of brain diseases. At the same time, numerous online repositories now provide different measures of the brain from healthy controls and patients with a specific diagnosis [e.g. OpenNeuro (a large bank of public datasets including different brain diseases, https://openneuro.org/), Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA, https://enigma.ini.usc.edu/), the Alzheimer's Disease Neuroimaging Initiative (ADNI, https://adni.loni.usc.edu/), the Parkinson's Progression Markers Initiative (PPMI, https://www.ppmi-info.org/), the Autism Brain Imaging Data Exchange (ABIDE, https://fcon_1000.projects.nitrc.org/indi/abide/), and SchizConnect (a portal for data on schizophrenia, https://schizconnect.org/)] (also see [3, 4]). However, so far AI methods have rarely reached a classification accuracy useful for clinical practice and more crucial, AI models have severe reliability issues [5]. In this comment, we provide possible explanations of why our current approach to search for biomarkers is inherently flawed.

The case of brain activity based biomarkers

The brain is a multiscale system which can be described at many levels spanning from molecules to cells, to circuits, networks and networks of networks [6–8]. It stands to reason that "a person's mental activities are entirely due to the behavior of nerve cells, glial cells, and the atoms, ions, and molecules that make them up and influence them." [9]. This suggests that correlates of animal behavior (function/dysfunction) should be observed in the form of the activity of nerve cells, glial cells and the neurotransmitters/neuromodulators. Among these, neural activity seems to play the most prominent role. Eventually, behavior is expressed through muscle activity which are directly controlled by the spiking activity of neurons. Therefore, changes in the neural activity are essential to manifest any behavior. Consistent with this, every behavior can be decoded from the neural activity of one or more brain regions. Thus, one may infer that all dysfunctions of the brain could be decoded from the brain activity which can form a link between structural/chemical

changes and behavior. But then why brain activity is only used as a biomarker in a handful cases such as epilepsy focus detection? We answer this question using an AI model and a thought experiment.

AI model and thought experiment

Neural correlates have been observed in several lab experiments involving simple and repetitive tasks. However, the existence of neural correlates of behavior hides an important piece of information: the same behavior can be elicited by many brain activity patterns and the same neural activity pattern in a given brain region could result in different behaviors given the context and the specific individual [10]. This is called degeneracy which suggests that many different sets of parameters may lead to the same function [11].

Here, we illustrate this fact using a simple AI model. We considered two RNNs (RNN1 and RNN2) connected to their respective readout layer (Model 1 and Model 2). Both RNNs were identical in their neurons, architecture, and initial synaptic weights. Both were trained using the same learning rule to perform the same task: predict the sum of the last 5 elements (output) of a time series (input) (Fig. 1A). Our input set was composed of 32 time series of time-length 20 and dimension 10 (dimension 3 in Fig. 1A), so 32 matrices of size 10×20. Each element of a time series was drawn independently from a normal distribution. Given an input matrix $I = (i_{kt})_{0\leq k\leq 9, 0\leq t\leq 19}$, the associated output O was the sum of all the elements of the final 5 time steps, $O = \sum_{t=14}^{19} \sum_{k=0}^{9} i_{kt}$ (Fig. 1A). Both RNN1 and RNN2 learned the task to the same proficiency (Fig. 1B). However, after training, they developed different weights due to the training's stochasticity. That is, the neural activity in both RNNs was 'healthy' as it enables the two RNNs to perform the task when connected to their respective readout.

Next, we changed the setup and connected the output of the first RNN to the second readout (Model 3). As expected the performance of RNN1 degraded significantly (Fig. 1B). Now based on the task performance, RNN1 appears 'dysfunctional'. We will also get a different dysfunctional response if we instead use the readout layer of the first RNN with the second RNN's activity.

We can extend this example of RNNs to the brain through a thought experiment: take the motor cortical activity (RNN1) from a healthy person when he/she is performing a reaching task – the readout here is the mapping from the motor cortex activity (RNN1) to the arm movement (output 1). Map this activity on the motor cortex of another person (replace RNN2 by RNN1). Will the second person also perform the reaching task in the same way as the first person? The answer is no. That is, healthy activity of a brain region
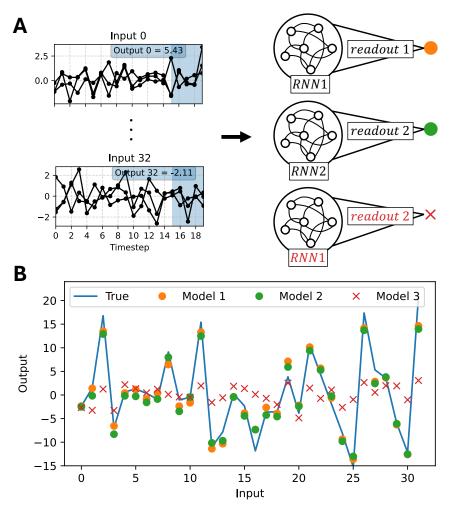
Fig. 1 AI model: learning the same task can be done via different RNN weight optimization. **(A)** Left: Example of two time series (inputs) along with their corresponding outputs. Time series are in dimension 3 for easier visualization. The blue area shows the points over which the sum is done to get the output displayed at the top. Right: Schematic of the models: Model 1 (RNN1 → readout 1), Model 2 (RNN2 → readout 2), and Model 3 (RNN1 → readout 2). Classical single-layer Elman RNN followed by single linear readout layer were used with the non-linearity $tanh$, hidden layers of size 50, learning rate 0.01 and 100 epochs. **(B)** Comparison of the outputs of the three models versus the true output.

is 'healthy' given the state of the rest of the brain and would make the other person non-functional or dysfunctional.

The AI model and thought experiment reveal two important issues of biomarker of a brain disease: degeneracy prevents from resorting to cohort comparison and considering single type of data leads to confounding.

Cohort comparison approach to identify biomarkers

When we search for brain activity based biomarkers of a brain disease, we define a certain descriptor of brain activity. Next, we obtain the distribution of this descriptor from a patient group and a matched healthy control group. If we can classify the patients from healthy controls based on the brain activity descriptor, we can argue that we have found a biomarker. This cohort comparison approach implicitly assumes that the brain activity of all healthy subjects (resp. all patients) is statistically identical. But, as we have illustrated with the RNN example and the thought experiment, even identical brain activity could manifest itself as healthy and dysfunctional behavior. That is comparing one cohort against another is not a useful way to identify biomarkers.

Indeed, cases for which comparing cohorts is useful are rare. For example, epileptiform activity in patients with epilepsy or persistent beta oscillations in the basal ganglia of patients with Parkinson's disease are very different from anything seen in healthy subjects. However, in most cases, cohort comparison has not revealed any novel biomarker that can be used in clinical setting [12, 13].

Our example suggests that it is more useful to compare healthy state activity with disease-related activity of the same individual. To paraphrase the line about material models by Rosenblueth and Wiener [14], "a patient is a good model of a patient, preferably the same patient".

The problem of subsampling

In hotel rooms, we have two two-way switches to control the lights in the room – one at the entrance and the other is typically close to the bed. Can we determine whether the lights are on or off just by looking at the state of one of the two switches? This is not possible because the state of the light bulb is determined by the two switches together. That is, partial sampling is likely to give a random result. In fact, one of the first pseudo-random number generators (middle-square method) invented by von Neumann exactly exploited this fact about partial sampling.

In our RNN example, the activity and readout weights together allow correct task performance. Just based on the RNN's activity, we cannot predict whether the model will perform the task correctly or not. However, prediction can become possible by incorporating additional context. For instance, if we know that the only difference between two networks is their bias terms – analogous to neuromodulatory levels like dopamine in the brain – we can accurately distinguish them and predict their performance. That is, it is crucial to integrate multiple data types to understand neural function and dysfunction.

We illustrate this fact in Fig. 2A with a simple example showing how data clearly separable when considered in a two dimensional space becomes indistinguishable when sampled from only one dimension. In the brain, the situation is of course more complex and each disorder is affected by a large number of factors.
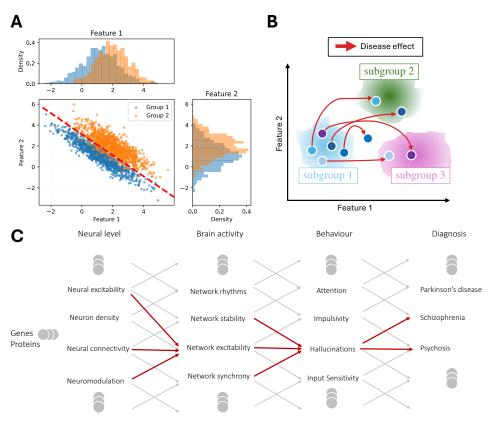

Fig. 2 How multimodal and longitudinal data can help identifying biomarkers of brain diseases. (A) Simple illustration of data easy to separate when considering two features but not when only one is analyzed. Right and top plots are respectively the marginal distribution of feature 1 and 2 for the two groups. **(B)** Assuming you can split your data on healthy controls in three subgroups. When people get a disease, the brain tries to compensate and find another working point – illustrated by the red arrows – which may be described by another subgroup's feature collection. If you look at individual changes you will observe some, but if you look at the global level, healthy controls have the same features' distributions than patients. **(C)** Example of different possible chains of changes occurring at different levels of the brain leading to the same behavior present in several brain diseases. Same symptoms (e.g. hallucination) could be shared by multiple diseases. The symptom can be manifested in different features of brain activity (red arrow) and those brain activity changes could come from multiple changes at the neuron/synapse level (red arrows). This makes it essential to study and identify correlates of a diseases at multiple scales of brain organization.

Beyond brain-activity based biomarkers

The necessity of multimodal data

The problem in the brain is far more complex than in our AI model and thought experiment. As stated earlier, the brain is a multi-scale system and its function or dysfunction is determined by multiple levels simultaneously (Fig. 2C). Sampling only one level or few levels is not likely to give a correct diagnosis. That is, measurements of one of a few of the brain's state variables is bound to fail.

Indeed, we have focused so far on brain activity but our arguments extend to other potential biomarkers of brain diseases. This is not a surprise that genetic, chemical and morphological features alone are usually not sufficient to classify patients from healthy controls. Counter-examples are rare but there exist for example one gene diseases like Huntington's disease (HD) and the spinal muscular atrophy (SMA) [15]. However, this does not mean that there is a cure: SMA has several – especially solving this specific gene mutation – but HD do not have any. On the other hand, several potential biomarkers failed in the past. Indeed, there are examples of 'healthy' subjects with less cortical density than patients with Dementia [16]. Similarly, it has been observed that 100s genes are implicated in Schizophrenia but each one may not explain even 1% of the risk of the disease [17].

In addition, it would greatly surprise us if a doctor prescribed medicine only using a single biomarker. A major part of medical training of physicians is about differential diagnosis, i.e. after looking at multiple symptoms and clinical indicators, they can identify the most likely causes of the symptoms. For instance, a neurologist observes a subject with a slight tremor in one hand and a shuffling gait. She then assesses these behavioral and motor symptoms with the MDS-UPDRS [1]. Based on this exam, a clinical diagnosis of Parkinson's disease is made. The core biomarker and cause is the progressive loss of dopamine-producing neurons in the substantia nigra compacta region of the brain. Confirming the dopamine deficit (chemical biomarker) with PET scans directly informs the treatment plan. Consecutive PET scans provide a rate of neuronal loss hence helping the neurologist to provide a prognosis for disease progression.

But this essential part of medicine is completely ignored when it comes to the use of AI to develop/identify novel diagnostic biomarker – we assume that AI will solve the problem of degeneracy and redundancy by itself [18].

The web is full of repositories of various kind of data collected from healthy controls and patients with specific diagnosis. In most cases, these repositories used to contain one specific measurement e.g. levels of protein α-synuclein or brain images or brain activity.

They now diversified and multimodal as well as longitudinal data are available (e.g. PPMI and ENIGMA).

Based on our arguments and thought experiment, we would claim that a data repository without data about multiple levels of the brain (from genetics to brain activity and behavior) from the same subject is not useful [19].

What is a brain disease: Third person versus first person perspective

The idea of comparing one cohort of healthy controls with a cohort of subjects with a certain disease tacitly assumes that the disease condition pushes the patient out of the healthy distribution. This is a third person perspective on the what should be called a disease.

But there is a first person perspective. An individual compares him/herself against their previous state (longitudinal analysis). If there is a big deviation, then they go for a medical consultation. However, this would mean that every individual tracks their brain activity data and statistically significant deviations with respect to the past data should be taken as signs of a brain disease.

As we have highlighted with our AI model and thought experiment, we need to abandon the cohort comparison approach (third-person perspective) and develop protocols that involve individual specific longitudinal measurements of multimodal (proteins, neurotransmitters, neuromodulator, neural activity etc.) brain data. Such high dimensional longitudinal data should enable us to subgroup subjects (not patients) [12] – see below the "subgroup hypothesis" – to gain more understanding on the disease and the possible interactions between the multiple modalities at hand.

A way forward

The subgrouping hypothesis

Despite each brain possessing a unique learning history and a capacity to compensate for slowly accumulating pathologies [11], they have evolved to solve similar complex tasks in dynamic environments. This shared functional burden – the necessity to solve a multitude of survival-critical problems like energy efficiency, perception, action, and learning – powerfully constrains their evolutionary and developmental possibilities. This principle aligns with the concept of multi-task constraint in systems theory [20, 21]: the more tasks a system solves, the more constrained it becomes. Just as an artificial neural network trained on multiple tasks must find a shared representation that narrows its viable parameter space, the brain is tuned toward a limited set of robust architectural solutions [20].

Consequently, the features of a functional brain must lie on rather low-dimensional subspaces (manifolds). These subspaces may consist of several distinct "islands" of solutions (see Fig. 2B), but they are vastly fewer in number than the possibilities for a single task (e.g. our AI model).

The "solution island" idea suggests that we should be able to find ways to subgroup subjects within this high dimensional space spanned by multimodal and longitudinal data. We refer to this as the subgrouping hypothesis (see Fig. 2B). In particular, within a given subgroup, we may observe similar changes due to a disease that we could only observe through longitudinal data. Hence, we must stop looking for a single answer for everyone. Instead, we need to group subjects (not patients) based on their common biological/brain features [12] – the "islands". This is how we will finally run studies that yield clear, statistically powerful biomarkers.

Interpretable AI and model-driven data analysis

Interpretable AI is an essential tool to address this huge data analysis problem. By interpretable AI we mean that it is not used only to classify but we know what it uses to classify. One way to reach such interpretability is to guide the features extraction with biologically plausible models [13] – we call this model-driven data analysis. Mechanistic models of the brain at different scales suggest a number of measurements that can reveal an underlying dysfunction e.g. excitation-inhibition balance [22], critical slowing [23] etc. Moreover, sensitivity analysis can help identify variables in which small change can lead to a larger effect in the model output and vice versa. A combination of model-driven feature engineering can complement AI-based methods to find patterns across massive, diverse datasets that humans simply cannot see [24]. With more data available than ever – from genetics to brain scans to digital health monitoring – investing in a model-assisted AI-driven strategy is the most promising path forward for brain health [5, 25]. Important breakthrough in the discovery of biomarkers indeed require a collaborative effort between theoreticians, data analysts, experimentalists and clinicians.

Acknowledgments

References

[1] Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., *et al.*: Movement

disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society **23**(15), 2129–2170 (2008)

[2] Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H.: The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society **53**(4), 695–699 (2005)

[3] Sandström, M., Abrams, M., Bjaalie, J.G., Hicks, M., Kennedy, D.N., Kumar, A., Poline, J.-B., Roy, P.K., Tiesinga, P., Wachtler, T., *et al.*: Recommendations for repositories and scientific gateways from a neuroscience perspective. Scientific Data **9**(1), 212 (2022)

[4] Eickhoff, S., Nichols, T.E., Van Horn, J.D., Turner, J.A.: Sharing the wealth: Neuroimaging data repositories. Neuroimage **124**(Pt B), 1065 (2016)

[5] Grote, T., Freiesleben, T., Berens, P.: Foundation models in healthcare require rethinking reliability. Nature Machine Intelligence **6**(12), 1421–1423 (2024)

[6] Churchland, P.S., Sejnowski, T.J.: Perspectives on cognitive neuroscience. Science **242**(4879), 741–745 (1988)

[7] Bell, A.J.: Levels and loops: the future of artificial intelligence and neuroscience. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **354**(1392), 2013–2020 (1999)

[8] Hemberger, M., Pammer, L., Laurent, G.: Comparative approaches to cortical microcircuits. Current opinion in neurobiology **41**, 24–30 (2016)

[9] Crick, F., Clark, J.: The astonishing hypothesis. Journal of Consciousness Studies **1**(1), 10–16 (1994)

[10] Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., Poeppel, D.: Neuroscience needs behavior: correcting a reductionist bias. Neuron **93**(3), 480–490 (2017)

[11] Albantakis, L., Bernard, C., Brenner, N., Marder, E., Narayanan, R.: The brain's best kept secret is its degenerate structure. Journal of Neuroscience **44**(40) (2024)

[12] Insel, T.R., Cuthbert, B.N.: Brain disorders? precisely. Science **348**(6234), 499–500 (2015)

[13] Woo, C.-W., Chang, L.J., Lindquist, M.A., Wager, T.D.: Building better biomarkers: brain models in translational neuroimaging. Nature neuroscience **20**(3), 365–377 (2017)

[14] Rosenblueth, A., Wiener, N.: The role of models in science. Philosophy of science **12**(4), 316–321 (1945)

[15] Rust, N.C.: Elusive Cures: Why Neuroscience Hasn't Solved Brain Disorders—and How We Can Change That. Princeton University Press, Princeton, NJ (2025)

[16] Stern, Y.: Cognitive reserve in ageing and alzheimer's disease. The Lancet Neurology **11**(11), 1006–1012 (2012)

[17] Pantelis, C., Papadimitriou, G.N., Papiol, S., Parkhomenko, E., Pato, M.T., Paunio, T., Pejovic-Milovancevic, M., Perkins, D.O., Pietiläinen, O., *et al.*: Biological insights from 108 schizophrenia-associated genetic loci. Nature **511**(7510), 421–427 (2014)

[18] Huang, G., Li, R., Bai, Q., Alty, J.: Multimodal learning of clinically accessible tests to aid diagnosis of neurodegenerative disorders: a scoping review. Health Information Science and Systems **11**(1), 32 (2023)

[19] Vieira, S., Bolton, T., Schöttner, M., Baecker, L., Marquand, A., Mechelli, A., Hagmann, P.: Multivariate brain-behaviour associations in psychiatric disorders. Translational psychiatry **14**(1), 231 (2024)

[20] Sterling, P., Laughlin, S.: Principles of Neural Design. MIT press, Cambridge, MA (2015)

[21] Baxter, J.: A model of inductive bias learning. Journal of artificial intelligence research **12**, 149–198 (2000)

[22] Gao, R., Peterson, E.J., Voytek, B.: Inferring synaptic excitation/inhibition balance from field potentials. Neuroimage **158**, 70–78 (2017)

[23] Scheffer, M., Carpenter, S.R., Lenton, T.M., Bascompte, J., Brock, W., Dakos,

V., Koppel, J., Leemput, I.A., Levin, S.A., Van Nes, E.H., *et al.*: Anticipating critical transitions. science **338**(6105), 344–348 (2012)

[24] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence **1**(5), 206–215 (2019)

[25] Yang, G.R., Wang, X.-J.: Artificial neural networks for neuroscientists: a primer. Neuron **107**(6), 1048–1070 (2020)