These patterns evolve between constructive and destructive interference as the delay-induced phase varies from 0 to π. The soliton signature emerges as a broad background signal superimposed on the Ramsey pattern and involves higher light intensities (optical fields accumulated in the pulse-shaped soliton) that, consequently, produce higher-order sidebands.

Nonlinear optical behavior can also be observed through the formation of chaotic light fields, trains of solitons, or individual solitons, depending on how the input light wavelength and its power are tuned. These phenomena are driven by the silicon nitride nonlinear optical response (*5*, *8*). As the laser wavelength and power were varied, Yang *et al.* could probe the transitions between nonlinear optical regimes because of the strong sensitivity of PINEM to the light field intensity and its distribution around the ring. The data of Yang *et al.* could be explained through a combination of well-established theoretical models for both the nonlinear optical response of microcavities (*9*) and the electron–photon interaction (*2*, *3*). Simulations based on these models were in good quantitative agreement with the authors' experiments.

The interaction of electrons with enhanced light fields enabled by the microring geometry creates opportunities to obtain previously inaccessible information on light propagation inside integrated optical circuits. This includes the degree and spatial distribution of coherence associated with nonlinearly generated light pulses such as solitons. Furthermore, electron–soliton interaction enables a disruptive approach to shaping the electron probability density in space and time. Also, solitons circulating the microring cavity can interact with a continuous electron beam at a high repetition rate approaching the terahertz, which is inaccessible with current ultrafast optics technology. The resulting electron modulations could be synchronized with the optical excitation, presenting new ways to perform electron microscopy and probe ultrafast dynamics in material systems with sub-fs/nm spatiotemporal resolution. ■

REFERENCES AND NOTES

1. B. Barwick *et al.*, *Nature* **462**, 902 (2009).
2. F. J. García de Abajo *et al.*, *Nano Lett.* **10**, 1859 (2010).
3. A. Feist *et al.*, *Nature* **521**, 200 (2015).
4. Y. Yang *et al.*, *Science* **383**, 168 (2024).
5. M. Liebtrau *et al.*, *Light Sci. Appl.* **10**, 82 (2021).
6. Y. Liu *et al.*, *Science* **376**, 1309 (2022).
7. J.-W. Henke *et al.*, *Nature* **600**, 653 (2021).
8. T. Herr *et al.*, *Nat. Photonics* **8**, 145 (2014).
9. L. A. Lugiato, R. Lefever, *Phys. Rev. Lett.* **58**, 2209 (1987).

ACKNOWLEDGMENTS

The authors receive support from European Research Council (ERC) 789104/eNANO; ERC 01019932/QEWS; and European Commission FET-Proactive 101017720/EBEAM.

10.1126/science.adn1876

GRAPHIC: F. PETZSCHNER. ADAPTED BY K. HOLOSKI/*SCIENCE*

MEDICINE

Practical challenges for precision medicine

The prediction of individual treatment responses with machine learning faces hurdles

By **Frederike H. Petzschner**

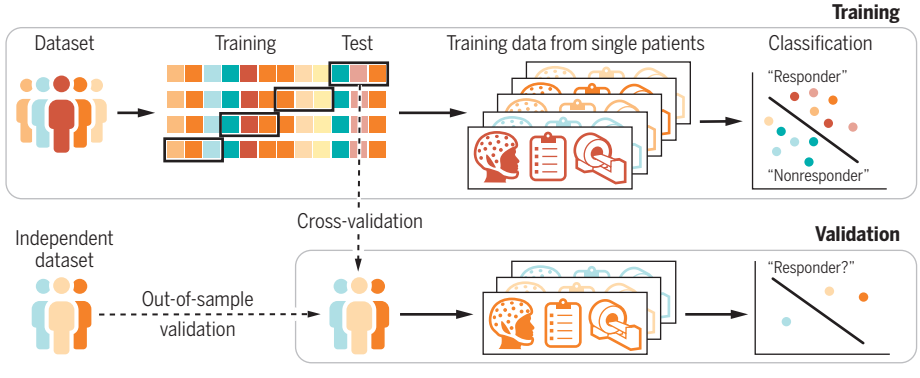
Precision medicine promises treatments tailored to individual patient profiles. Machine learning models have been heralded as the tools to accelerate precision medicine by sifting through large amounts of complex data to pinpoint the genetic, sociodemographic, or biological markers that predict the right treatment for the right person at the right time. However, the initial enthusiasm for these advanced predictive tools is now facing a sobering reality check. On page 164 of this issue, Chekroud *et al.* (*1*) show that machine learning models that predict treatment response to antipsychotic medication among individuals with schizophrenia in one clinical trial failed to generalize to data from new, unseen clinical trials. The findings not only highlight the necessity for more stringent methodological standards for machine learning approaches but also require reexamination of the practical challenges that precision medicine is facing.

What predicts whether a patient will benefit from a particular treatment? The answer may lie in their genetics, biology, sociodemographic background, social environment, past experiences, or a myriad of other potential factors. Machine learning techniques have the capacity to analyze large datasets and identify the most effective combination of features that accurately predicts a variable of interest. They thus offer a promising avenue for discovering relevant features or biomarkers that predict individual treatment responses. Typically, this involves training the model on a dataset for which the outcome, such as the response to a given treatment, is already known. This is known as supervised learning. One common pitfall of this method is overfitting. Overfitting occurs when a model is too flexible relative to the data it is trained on, which limits its generalizability. A sign of overfitting is when the model accurately predicts outcomes on the data it was trained on but performs poorly on new, unseen data. To address the issue of overfitting, it is essential to validate models on unseen data. Cross-validation is a widely used technique for this purpose. It involves repeatedly dividing the data into subsets, training the model on one subset, and then evaluating its prediction accuracy on the remaining "held-out" data (see the figure).

However, cross-validation is not infallible. Chekroud *et al.* revealed that models trained to predict responses to antipsychotic medication in schizophrenia within a specific clinical trial using cross-validation failed to predict treatment responses in other independent

Individual treatment prediction using machine learning

Supervised machine learning for individual treatment prediction is based on the development of classifiers. To prevent overfitting, model validation is crucial, typically achieved through cross-validation or out-of-sample validation. Out-of-sample validation requires a completely independent dataset and is more resource-intensive, but this approach is less susceptible to overfitting and can provide more generalizable results.



clinical trials. One reason that cross-validation can inadvertently result in overfitting the held-out data is that the modeler, through iterative model adjustments, may eventually use all the available data. The issue is likely more widespread than typically acknowledged. For example, a comprehensive review of 116 studies across various psychiatric diagnoses found signs of overfitting specifically in studies with small sample sizes (<50 participants) (*2*). Small sample sizes also cause large variance in cross-validation results, and although these issues are well known in statistics and machine learning, many studies still do not follow best practices to improve the outcomes of cross-validation (*3*).

A reliable way to assure the generalizability of machine learning models lies in validating their predictive accuracy on a truly independent, untouched validation sample, known as out-of-sample validation. Often, this approach is not used in clinical studies owing to the challenges associated with acquiring larger datasets and the need for stringent rules governing data acquisition and usage. However, the study by Chekroud *et al.* adds to a growing body of evidence that underscores the necessity of these more robust validation standards to avoid overly optimistic results from machine learning models that fail to generalize to wider clinical contexts.

Even with models that are properly validated and supported by large sample sizes, attempts to predict the clinical outcome or treatment response for individual patients can be unreliable. In the study by Chekroud *et al.*, even when data from multiple clinical trials were pooled to train the model, its predictions still failed to generalize to a new independent trial. The reasons for this are complex and multifaceted. A primary factor is the inherent heterogeneity in data from clinical populations. This issue is particularly prominent in psychiatric disorders, which are typically defined by sets of symptoms (syndromes). Patients with the same diagnostic label may exhibit vastly different symptom profiles that warrant different treatments. Moreover, identical symptoms in different individuals might have distinct biological underpinnings and thus require different therapeutic strategies (*4*). Basing machine learning models purely on diagnostic labels without taking this type of heterogeneity into account can lead to inaccuracies when predicting effective treatment strategies.

A promising approach to address this challenge is to stratify patients into more precisely defined categories, for example, based on underlying symptom causes. This can be achieved, in part, through the use of theory-driven computational models that aim to describe underlying disease mechanisms, a method gaining traction in the field of computational psychiatry. These models are increasingly being used alongside data-driven machine learning techniques, forming powerful tools to tackle the issue of heterogeneity in patient populations (*5, 6*).

Another form of heterogeneity may stem from systematic differences across studies, locations, or time points. As a result, predictions of machine learning models trained on data from a specific context—a population, country, setting, or time period—might rely on features that are associated but not causally related with a clinical outcome in a given study but are not predictive in other contexts. One way to address this heterogeneity is to pool data across multiple studies and sites.

Unreliable predictions may also be the result of outdated outcome measures. Many existing symptom scores are based on questionnaires that may no longer align with understanding of the disease and potentially lead to inaccurate assessments of treatment response. For example, the positive and negative syndrome scale (PANSS) used in the clinical trials from Chekroud *et al.* is gradually being supplanted by more contemporary assessment tools, specifically in the context of negative symptoms in schizophrenia (*7*). If a questionnaire fails to fully capture the true disease burden, it might not accurately detect genuine improvements resulting from treatment. This discrepancy can lead to misclassification of who has or has not benefited from the treatment, which hinders the accurate training of the machine learning model. Similar to the heterogeneity within diagnostic categories, outcome measures will become more accurate with increasing insight into the underlying disease mechanism.

The challenges of using machine learning to predict individual treatment response in medicine, specifically in the context of psychiatry, stem from a complex interplay of issues related to model validation standards, diagnostic heterogeneity, and the relevance of outcome measures used. Addressing these challenges is essential for impactful clinical research and to enable progression toward effective precision medicine. ■

REFERENCES AND NOTES

1. A. M. Chekroud *et al.*, *Science* **383**, 164 (2024).
2. T. Wolfers *et al.*, *Neurosci. Biobehav. Rev.* **57**, 328 (2015).
3. G. Varoquaux *et al.*, *Neuroimage* **145** (Pt B), 166 (2017).
4. K. E. Stephan *et al.*, *Lancet Psychiatry* **3**, 77 (2016).
5. M. P. Paulus, W. K. Thompson, *Psychopharmacology* **238**, 1231 (2021).
6. Q. J. M. Huys *et al.*, *Nat. Neurosci.* **19**, 404 (2016).
7. B. Kirkpatrick *et al.*, *Schizophr. Bull.* **32**, 214 (2006).

ACKNOWLEDGMENTS

F.H.P. was supported by the Brainstorm Program at the Robert J. and Nancy D. Carney Institute for Brain Science.

Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI, USA. Email: frederike_petzschner@brown.edu

10.1126/science.adm9218

MATERIALS SCIENCE

The bumpy road to friction control

The frictional properties of material interfaces can be rationally designed

By **Viacheslav Slesarenko**[1,2] *and* **Lars Pastewka**[1,2]

Friction controls daily life, often without being noticed. It allows walking without slipping, holds sandcastles together, and determines the perceived cleanliness of hair. Little resistance is desired when pedaling bikes, yet the expectation of pulling the brakes is to stop moving. Overall, machines use 20% of the world's energy production to overcome frictional resistance (*1*). Present-day strategies to tune friction, derived from more than a century of engineering insights, often involve the lubrication of interfaces with oils or greases. On page 200 of this issue, Aymard *et al.* (*2*) report an alternative strategy of rationally designing the frictional properties of interfaces. Their approach to friction control may lead to the development of surfaces that adapt to the environment in real time.

Aymard *et al.* show that small bumps of identical radii (*3*) constitute simple building blocks that can be combined into a frictional metainterface. By using many such bumps on a surface and adjusting their height distribution, the authors could prescribe a desired, even nonlinear, dependence of the frictional force that resists sliding motion on the external load that pushes the sliding interfaces together.

The effect of surface topography on friction has long been known. Charles-Augustin Coulomb, one of the founders of tribology (the science of friction), wrote in 1779 about the interlocking of asperities (*4*), the name given to "bumps" on rough surfaces. Surface topography determines the amount of actual contact that two bodies make. Thus, two bodies typically

[1]Department of Microsystems Engineering, University of Freiburg, 79110 Freiburg, Germany. [2]Cluster of Excellence *liv*MatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, 79110 Freiburg, Germany. Email: viacheslav.slesarenko@livmats.uni-freiburg.de; lars.pastewka@imtek.uni-freiburg.de

Practical challenges for precision medicine

Frederike H. Petzschner

Science **383** (6679), . DOI: 10.1126/science.adm9218