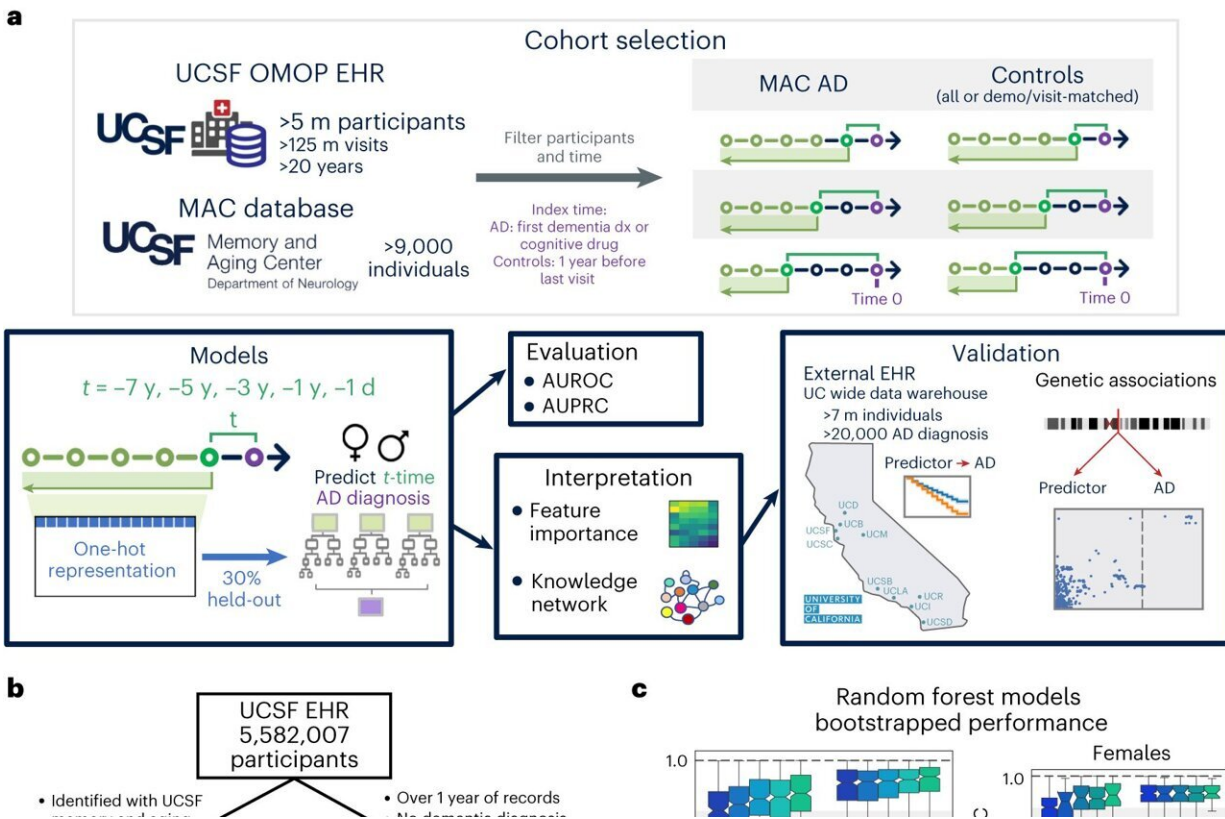# How AI can help spot early risk factors for Alzheimer's disease

February 21 2024



Overview of participant selection and RF model performance. a, From the UCSF EHRs and the UCSF Memory and Aging Center (MAC) database, participant and clinical information was extracted, filtered and prepared for time points before the index time. All clinical features extracted were one-hot encoded and trained on random forest (RF) models to predict future risk of AD diagnosis. Models were evaluated on a 30% held-out evaluation set to compute AUROC/AUPRC and interpreted based on feature importances and using a heterogeneous knowledge network (SPOKE). Top features were then further

validated in external databases. b, Filtering a consistent set of individuals with AD and controls from the UCSF EHR for model training and testing. Filtered participant cohorts are shown in Table 1 and split with 30% held-out set for testing. c, Bootstrapped performance of RF models on the held-out evaluation set (n = 300 bootstrapped iterations of 1,000 participants, prevalence of AD on held-out set = 0.003). Bootstrapped AUROC performance for models trained and tested on female strata and male strata are also shown. The box shows quartiles (25th, 50th and 75th percentiles), whiskers extend to 1.5 times the interquartile range, and the remaining points are outliers. Credit: *Nature Aging* (2024). DOI: 10.1038/s43587-024-00573-8

UC San Francisco scientists have found a way to predict Alzheimer's disease up to seven years before symptoms appear by analyzing patient records with machine learning.

The conditions that most influenced prediction of Alzheimer's were [high cholesterol](link), and for women, the bone-weakening disease osteoporosis.

The work demonstrates the promise of using [artificial intelligence](link) (AI) to spot patterns in [clinical data](link) that can then be used to scour large genetic databases to determine what is driving that risk. The researchers hope that one day it will hasten the diagnosis and treatment of Alzheimer's and other complex diseases.

"This is a first step towards using AI on routine clinical data, not only to identify risk as early as possible, but also to understand the biology behind it," said the study's lead author, Alice Tang, an MD/Ph.D. student in the Sirota Lab at UCSF. "The power of this AI approach comes from identifying risk based on combinations of diseases."

The findings [appear](link) in *Nature Aging*.

# Clinical data and the power of prediction

Scientists have long sought to discover the biological drivers and early predictors of Alzheimer's disease, a progressive and ultimately fatal form of dementia that destroys memory. Alzheimer's affects some 6.7 million Americans, nearly two-thirds of whom are women. The risk of getting the disease increases with age, and women tend to live longer than men, but that does not fully explain why more women than men have it.

The researchers used UCSF's clinical database of more than 5 million patients to look for co-occurring conditions in patients who had been diagnosed with Alzheimer's at UCSF's Memory and Aging Center in comparison to individuals without AD and found they could identify with 72% predictive power who would develop the disease up to seven years prior.

Several factors, including hypertension, high cholesterol and vitamin D deficiency, were predictive in both men and women. Erectile dysfunction and an enlarged prostate were also predictive for men. But for women, osteoporosis was a particularly important predictor.

This does not mean that everyone with the bone disease, which is common among older women, will get Alzheimer's.

"It is the combination of diseases that allows our model to predict AD onset," said Tang, "Our finding that osteoporosis is one predictive factor for females highlights the biological interplay between bone health and dementia risk."

## A precision medicine approach

To understand the biology underlying the model's predictive power, the researchers turned to public molecular databases and a specialized tool developed at UCSF called SPOKE (Scalable Precision Medicine Oriented Knowledge Engine), which was developed in the lab of Sergio Baranzini, Ph.D., a professor of neurology and a member of the UCSF Weill Institute for Neurosciences.

SPOKE is essentially a database of databases that researchers can use to identify patterns and potential molecular targets for therapy. It picked up the well-known association between Alzheimer's and high cholesterol, through a variant form of the apolipoprotein E gene, APOE4. But, when combined with genetic databases, it also identified a link between osteoporosis and Alzheimer's in women, through a variant in a lesser-known gene, called MS4A6A.

Ultimately, the researchers hope the approach can be used with other hard-to-diagnose diseases like lupus and endometriosis.

"This is a great example of how we can leverage patient data with [machine learning](link) to predict which patients are more likely to develop Alzheimer's, and also to understand the reasons why that is so," said the study's senior author, Marina Sirota, Ph.D., associate professor at the Bakar Computational Health Sciences Institute at UCSF.

  **More information:** Alice S. Tang et al, Leveraging electronic health records and knowledge networks for Alzheimer's disease prediction and sex-specific biological insights, *Nature Aging* (2024). [DOI: 10.1038/s43587-024-00573-8](link)

Provided by University of California, San Francisco