HORIZONS OF PREDICTION

Can we leverage the power of Machine Learning to predict complex psychobehavioral phenomena?

By: LT E.S. Vorm, PhD
Deputy Director, Laboratory for Autonomous Systems Research

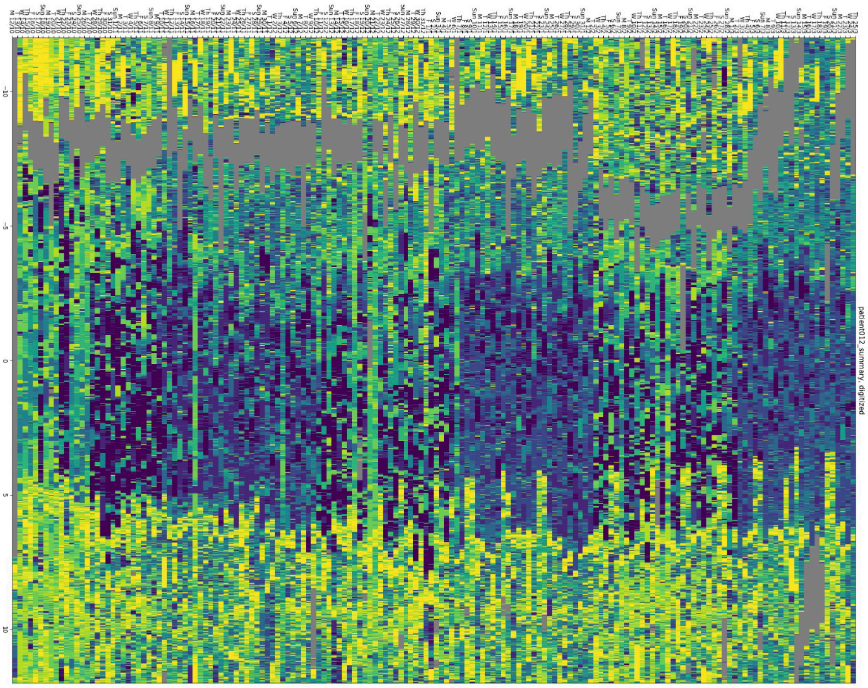
Popular depictions of military research and development are abundant. Unfortunately, however, visions of military scientists in white lab coats crafting futuristic weapons in batcave-like laboratories have an unintended side effect: the expectation that scientific breakthroughs happen frequently and immediately. The reality, unfortunately, is that most research is more mundane and time consuming. The building blocks of technologies like radar-deflecting stealth panels or laser-guided munitions come in the form of small but significant scientific baby steps, documented in technical reports, conference proceedings, and scientific journals. Vannevar Bush, first director of the Office of Scientific Research and Development and the person who ran the military's research and development during WWII, once famously said that basic research "creates the fund from which the practical applications of knowledge must be drawn" (1, Ch. 3). The US Government's strategy to invest in basic research represents its understanding of the fundamental relationship between scientific knowledge and its offspring practical applications. This article highlights one example of basic research (6.1 on the RDT&E funding spectrum) in the realm of military medicine. To set the stage and introduce the topic, let's begin with a brief scenario.

A Marine helicopter is performing routine training when suddenly one of the engines experiences a catastrophic failure. Despite the heroic efforts of the pilots, the helicopter crashes. Four of the six crew members on board, including both pilots, die in the crash. The remaining two crew members are both injured and are taken to the hospital where they both spend several weeks recovering from their injuries.

A year later, one of the surviving crew members has returned to her job and is experiencing no serious symptoms of prolonged stress after the accident. She has moved on from the traumatic event, and shows signs of continuing growth. She has strengthened her relationships with her loved ones and family, and has renewed her interests in hobbies such as playing ultimate frisbee and long-distance bicycling. Because of the positive interactions she experienced with her caregivers while in the hospital, she has begun taking college courses and is preparing to apply to a nursing program. She has a renewed sense of appreciation for life, and seeks to be mindful and thankful in her daily life, activities, and relationships.

The other crew member, however, has not fared as well. He has been experiencing serious adverse symptoms of prolonged stress since shortly after being released from the hospital. He experiences frequent nightmares, and is easily startled and reacts uncontrollably to being surprised to the point of becoming enraged. He angers easily, and is increasingly irritable. Depression, anxiety, and a prolonged sense of dread make it increasingly difficult for him to function in social circumstances. Because of frequent outbursts and unstable behavior at work, he was removed from flight status and his security clearance was frozen pending a medical review. Eventually his symptoms become so severe that he is admitted to an inpatient facility to manage his ever-growing inability to cope with daily life after he tells co-workers that it would have been better if he had died in the crash and that he wishes he was dead on a regular basis. He is eventually deemed no longer fit for active-duty service and is ultimately medically separated from the Marine Corps.

What factors determine who will recover from trauma and who will experience prolonged psychological stress? Are these factors learned? Can they be trained? Or are these factors genetic? Questions such as these have interested human beings for as long as we have recorded human history. People as far back as the ancient Greeks noticed and commented on how some soldiers are able to experience traumatic adversity with resilience and carry on their lives after war without issue, while others seem to be permanently scarred and altered by their experiences. Despite many decades of modern-day research in genetics and the psychology of stress, however, little is known or understood about the complex phenomenon known as post-traumatic stress disorder (PTSD).



Our algorithm learns four separate scores from the data gathered with our wearable device. Each score is associated with a depressive characteristic: diurnal activity, anxiety, psychomotor retardation, and circadian rhythm (sleep). The image on the left is all the data from a single patient across a period of approximately six months. Patients wear the device while in treatment. Their activities and physiological data are recorded and used to create our predictive algorithm.

Traumatic events such as combat exposure, near-death experiences and sexual assault affect every person in some way. As the Psychiatrist Viktor Frankl once put it, "an abnormal reaction to an abnormal situation is normal behavior" (2, p. 20). Contrary to popular notions of PTSD, however, while many people develop short term effects (e.g., sleep loss, mild anxiety symptoms) following a trauma, most people recover within a short period of time, and relatively few people develop long-term PTSD. The trick to preventing serious PTSD is to identify early those who are more vulnerable and susceptible before chaotic expressions of PTSD are allowed to fully develop. But as we saw in the vignette earlier, predicting who is more or less likely to develop long term psychological effects of trauma is very difficult. As a result, organizations such as the US Department of Defense incur extremely high costs in terms of manning and medical treatments related to PTSD in service members.

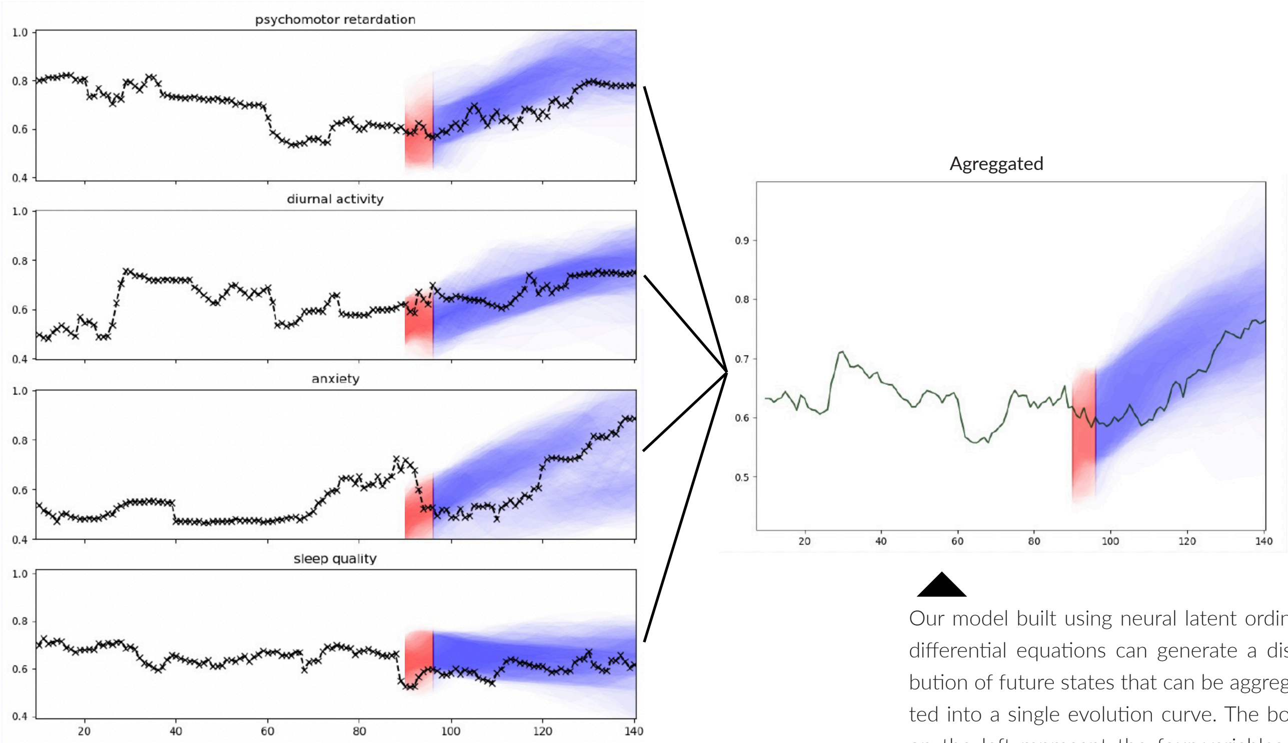
Estimates of the number of service members experiencing PTSD are alarming—between 13.5% and 30% of recently deploying troops have tested positive for PTSD and required significant medical treatment (3, 4). This number totals over 500,000 troops over the past 13 years of conflicts in the Middle East (5). The total costs of this is in the multiple billions of dollars, but even greater are the costs in terms of retention of qualified military personnel and medical readiness.

One of the most common aims in PTSD research, therefore, has been to create models of patient trajectories that could serve as an early warning for patients who are likely to need more help recovery from a traumatic event. Patient models could indicate whether patients are improving or whether they are declining and moving towards full PTSD in ways that current medical practices do not afford. Clinicians could use these models to inform their treatment decisions, which would then improve patient outcomes. And today there is reason to be optimistic that model-based approaches to predicting PTSD might be possible thanks to recent advancements in artificial intelligence.

AI is everywhere today—in our cities, in our homes, and in some cases even in our bodies (6, 7). Due to the recent explosion of computer processing speed and power, machine learning (ML) approaches to data science have yielded exciting new opportunities to learn more about ourselves and the world around us than ever before. ML's advantage comes from its ability to perceive relationships between variables at extremely high dimensions. This means that ML can learn patterns and associations between variables at levels far beyond human comprehension, and in a fraction of the time it would take for humans to perform those same calculations. ML has been used to successfully predict astonishing things with remarkable accuracy, from discovering new drugs (8) to predicting who might live or die in the intensive care ward of a hospital (9). Most recently, my team and I began to wonder whether ML could predict something as complex as PTSD, and whether the future Department of Defense could use predictive analytics to provide early warnings for personnel who are most at risk at developing severe PTSD. Inspired by this question and armed with some very bright minds and powerful computers, we set out in early 2020 to answer this basic question.

Through a cooperative agreement between the U.S. Naval Research Laboratory and the Office of Naval Research Global, teams of scientists and mathematicians from France and the U.S. teamed up to explore how machine learning could be used to predict PTSD. In order to utilize machine learning, we first must have enough data to train the system. To get this data, our first step was to utilize an existing bespoke wearable device to facilitate data collection. The wearable that we chose had been developed in partnership with the Office of Naval Research Global, and has already been approved for human participants experimentation through the l'agence nationale de sécurité du médicament et des produits de santé (ANSM), which is the French equivalent of the U.S. Food and Drug Administration. This wearable device combines a photoplethysmography (PPG) meter, with an actimeter and electrodermal activity (EDA) meter. Each of these represents the state-of-the-art in wearable technologies and facilitate things like detailed sleep analysis, evaluations of stress, overall physical activity, blood pressure, blood flow, and oxygen saturation. The device is meant to be worn 24 hours a day, and continuously collects data for months at a time.

With this device we collected six months' worth of patient data through a cooperation with local treatment centers in



Our model built using neural latent ordinary differential equations can generate a distribution of future states that can be aggregated into a single evolution curve. The boxes on the left represent the four variables we are interested in, collected from our wearable device. The box on the right is the agreggated score. To the right of the red line is projected data based on the data to the left of the red line. In this example, the model was trained on the first 90 days of data, and then asked to predict future states. The data is robust and reliable out to several weeks in the future. Using a system such as this, clinicians could possibly detect the onset of severe conditions or relapse early enough to intervene--something that is impossible to do today.

France, as well as participation with the French Army. We collected data from 200 patients who had been diagnosed with major depressive disorder, 200 patients who had been diagnosed with PTSD, and approximately 2,600 healthy patients.

One challenge in modeling PTSD is that it has an enormous range of expressions, meaning that two people with identical diagnoses may look very different in terms of the symptoms they are experiencing. The range of symptoms and tendency of those symptoms to be subjectively derived via self-reported instruments (i.e., surveys filled out by patients and interpreted by clinicians) further complicates the use of sophisticated mathematical models such as those used by machine learning. Rather than attempting to model and predict PTSD directly, therefore, we determined that we would first attempt to model and predict another malady that frequently occurs in coordination with PTSD—major depressive disorder (MDD). It is estimated that between 50-70% of patients diagnosed with PTSD are also diagnosed or diagnosable with MDD (10). In addition to this significant overlap, major depressive disorder can be more readily identified and diagnosed through the use of physiological data, such as the types we were collecting. One additional benefit to attempting to predict MDD for this project was that patients in our study were also evaluated by licensed clinicians using the Montgomery-Asberg depression rating scale (MADRS). MADRS is a questionnaire that patients complete at various intervals while in treatment, and is used by clinicians to document and quantify the severity of symptoms. Each patient in our sample received six MADRS evaluations, one per month. With this MADRS data, we constructed patient trajectories, representing ground truth for how each patient fared during our data collection period. This ground truth served as the baseline against which our team could evaluate the accuracy of our machine learning algorithm.

With these trajectories constructed and our data collected, we trained our semi-supervised neural network and began to explore how it learned what features (e.g., heart rate variability, sleep disruption, daily step count) provide the best predictive power, and how accurately we could predict patient outcomes with the data we collected.

To do this we fed our neural network the first 60 days' worth of patient data, and then asked it to predict patient trajectories for the remaining portion of our six-month window. We then compared these machine-generated trajectories with the ground truth we constructed from MADRS data. The results were very robust.

As you can see from the diagrams Agree, the trajectories correlate very highly with MADRS data, which is a strong indication that this automated approach to detecting complex psychological suffering is entirely feasible. Most notable of our findings is that we were able to make accurate predictions of patient trajectories (e.g., who is likely to relapse, who is likely to recover) with only 60 days of data. This means that a clinician, armed with data from a wearable and an algorithm like ours, could potentially intervene weeks or months before symptoms become severe. This is of immense practical value if you consider that the current standard of practice for diagnosing and quantifying major depressive disorder (i.e., MADRS) can only describe how a patient is feeling at any given moment, but cannot accurately predict what a patient will do in the near or far-term future. Even the most skilled and seasoned clinician is likely unable to make accurate predictions about patient outcomes, especially when dealing with complex pathologies such as major depression or PTSD. As you can see in the figure below, a clinician treating this patient at Time 50 may believe this patient is getting significantly better, when in reality they will soon experience a significant relapse and a return to clinically significant symptoms shortly after this data point. It is worth repeating that our algorithm accurately predicted this trajectory, despite only having the first 60 days of data. Only through the added computational power of machine learning can clinicians hope to gain an edge in forecasting future patient states.

We have demonstrated the feasibility of an approach by utilizing technical and scientific expertise. Out of this small study may grow a technology that could one day greatly reduce the long-term effects of trauma, but there are many other studies necessary before the U.S. Department of Defense might be willing to embrace this approach (or one like it) at full scale. Such is the nature of basic 6.1 research. Through a small investment in time and resources, we have grown our knowledge and explored something new. The next step in the RDT&E journey (6.2) would be to expand this research and focus it further.

If you would like to read a more in-depth account of this research, you can read our article published in the International Journal Human-Intelligent Systems Integration:

Fompeyrine, D.A., Vorm, E.S., Ricka, N. et al. Enhancing human-machine teaming for medical prognosis through neural ordinary differential equations (NODEs). Hum.-Intell. Syst. Integr. (2021). https://doi.org/10.1007/s42454-021-00037-z

REFERENCES

1- Bush, V. Science The Endless Frontier. A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945. United States Government Printing Office, Washington: 1945.

2. Frankl, V. E. (1984). Man's search for meaning: An introduction to logotherapy. New York: Simon & Schuster.

3- Eber S, Barth S, Kang H, et al. The National Health Study for a New Generation of United States Veterans: methods for a large-scale study on the health of recent veterans. Mil Med. 2013;178:966–969.

4- Tanielian T, Jaycox LH, editors. Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery. Santa Monica, California: RAND Corporation; 2008.

5- Thompson M. Unlocking the secrets of PTSD. Time. 2015; 185:40–43

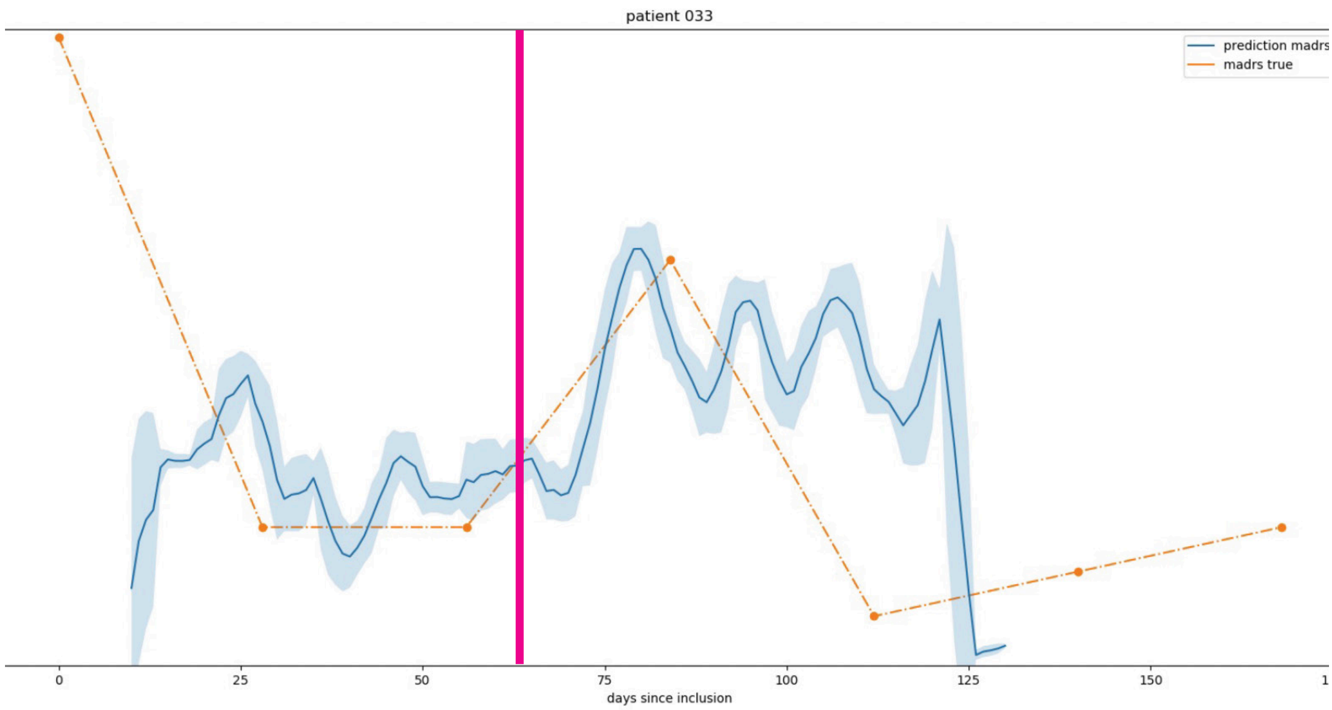
6- Reardon, S. AI-controlled brain implants for mood disorders tested in people. Nature 551, 549–550 (2017). https://doi.org/10.1038/nature.2017.23031

7- A. N. Balaji and L.-S. Peh, "AI-on-skin: Enabling On-body AI Inference for Wearable Artificial Skin Interfaces," Ext Abstr 2021 CHI Conference on Human Factors of Comput Systems, pp. 1–7, 2021, doi: 10.1145/3411763.3451689.

8- Vamathevan, J., Clark, D., Czodrowski, P. et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 18, 463–477 (2019). https://doi.org/10.1038/s41573-019-0024-5

9- A. Avati, K. Jung, S. Harman, L. Downing, A. Ng, and N. H. Shah, "Improving palliative care with deep learning," 2017, pp. 311–316, doi: 10.1109/bibm.2017.8217669 [Online]. Available: https://ai.stanford.edu/~avati/bibm17.pdf

10- Flory JD, Yehuda R. Comorbidity between post-traumatic stress disorder and major depressive disorder: alternative explanations and treatment considerations. Dialogues Clin Neurosci. 2015;17(2):141-150. doi:10.31887/DCNS.2015.17.2/jflory



An example of our algorithm's prediction of patient outcome. Blue line is the algorithm's prediction. Yellow dotted line is the actual MADRS score. Vertical magenta line indicates the 60-day mark. To the left of the magenta line the system was trained on actual data. To the right of the magenta line the system is projecting patient outcomes based on physiological data collected from a wearable device. Our algorithm correlates very closely with ground-truth MADRS, and is capable of operating far beyond human horizons of prediction.