A Place Beyond Numbers: The Mathematics of Alzheimer's

Alzheimer's has been the subject of many excellent works of memoirs and journalism (some favorites are Diane Keaton's "Then Again"), John Bayley's "Elegy for Iris," Thomas DeBaggio's "Losing My Mind: An Intimate Look at Life with Alzheimer's," and David Shenk's "The Forgetting: Alzheimer's: Portrait of an Epidemic," it is a disease that distinctively impedes our ability to comprehend it through orthodox methods of investigation. In Shenk's book, Alzheimer's is essentially a radical slowing of death— "what is usually a quick flicker we see in super slow motion." In DeBaggio's first-person story, I can feel the agonizing beginnings of that gradual loss of self-awareness; in Bayley's and Keaton's, I can see the outward signs of late-stage Alzheimer's. Nevertheless, when I ponder about individuals who are prone to Alzheimer's, I'm left with contemplative queries that can only be addressed in fiction: How does it feel to lose oneself whilst still being alive? Is there a core component of selfhood that endures to the end? Those inflicted in the middle or later stages are unable to communicate it to us due to their aphasia thus it becomes an unexplained phenomenon.

More and more fiction writers are striving to enter that esoteric zone as baby boomers approach their seventies and Alzheimer's grows more prevalent. The release of Matthew Thomas's visionary and challenging "We Are Not Ourselves," the best Alzheimer's novel to date, seems like a good opportunity to re-evaluate the burgeoning genre and determine what its authors can and cannot tell us about the fate of the self as it succumbs to a disease that attacks the very center of selfhood.

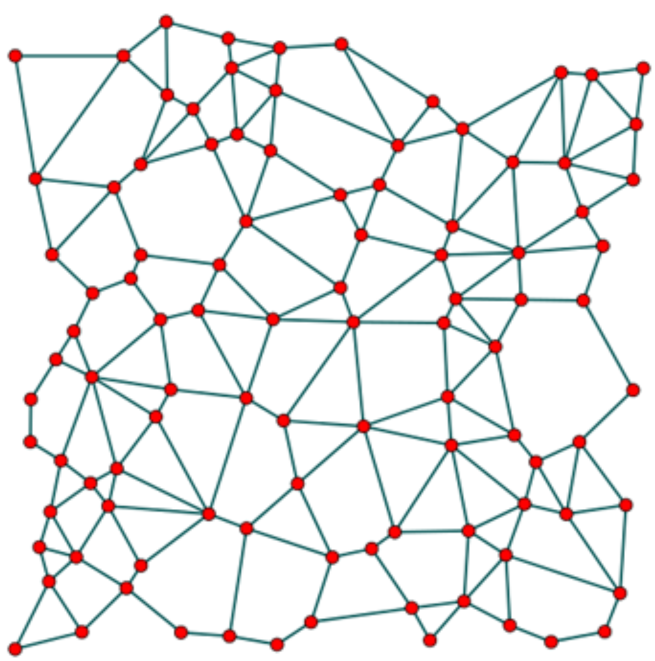
Alzheimer's is a debilitating condition that gradually develops but steadily worsens with time. Early diagnosis offers the best possibility of decreasing the disease's course and enhancing the patient's quality of life. Regrettably, there isn't a quick test to identify Alzheimer's. Doctors use a variety of tests and try to rule out other disorders with comparable symptoms to determine whether a patient has the disease. An automated technique that assists with the diagnosis has been created by a multidisciplinary team at the University of Cambridge to lessen the workload.

The team's approach, which was created under the direction of the mathematicians Carola-Bibiane Schönlieb and Angelica Aviles-Rivero, attempts to detect the illness before any noticeable symptoms have surfaced. Their machine-learning framework makes use of the numerous types of data that may be accessible to a patient, offering innovative approaches for analysing this substantial amount of information while requiring minimal human involvement. This modus operandi beats current machine learning algorithms for the early diagnosis of Alzheimer's disease, according to tests conducted on real patient data.

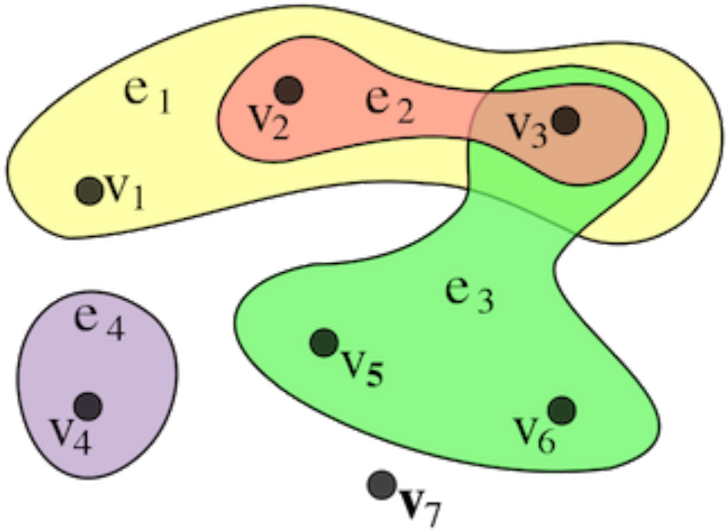
The diagnosis of Alzheimer's disease requires a wide range of patient data. A patient's brain can be seen in medical imagery such as MRIs and PET scans. Non-imaging data, such as genetic data and identifying details like the patient's age, also include important information. Analysing the pertinent data to see if it has hidden patterns that could aid in the task at hand—in this case, identifying a disease—is one strategy frequently utilized in machine learning. Avilies-Rivero and her co-workers intended to use all the various forms of information that are available about specific patients, whereas this stratagem often concentrates on only one sort of data. How can they combine this wealth of heterogeneous data, consisting of an array of data modalities, as well as obtain a meaningful understanding from it?

Firstly, analysing each source of data separately. In substantial sets of images, significant features can be automatically detected using existing image analysis techniques. This may be due to the existence of specific patterns or geometric structures in the image, which, in the case of these MRI images, may represent the state of particular brain regions in Alzheimer's patients. The feature space, a high-dimensional mathematical space where each axis of the space corresponds to a certain image characteristic, is defined by these features. Images that share a certain feature are positioned close to one another in the direction of that axis in the feature space since the images are represented as points in it.

By creating a hypergraph, the researchers can use this sort of depiction of the MRI scans to capture the correlations between various elements in these images. A typical graph is essentially a network made up of nodes, or points, connected by edges. The advantage of a hypergraph is that it may depict more intricate connections between the nodes. Whereas a hypergraph's hyperedges can connect an unlimited number of nodes, a graph's edges can only connect exactly two nodes at a time. In order to find any structure in the MRI data, the researchers connect the points that are close to one another in the feature space with a hyperedge. Additionally, they can create a separate hypergraph just for the structure of the PET imaging data in the same way.



A complex-looking graph, where each edge relates exactly two nodes – represented by a line joining those two nodes.



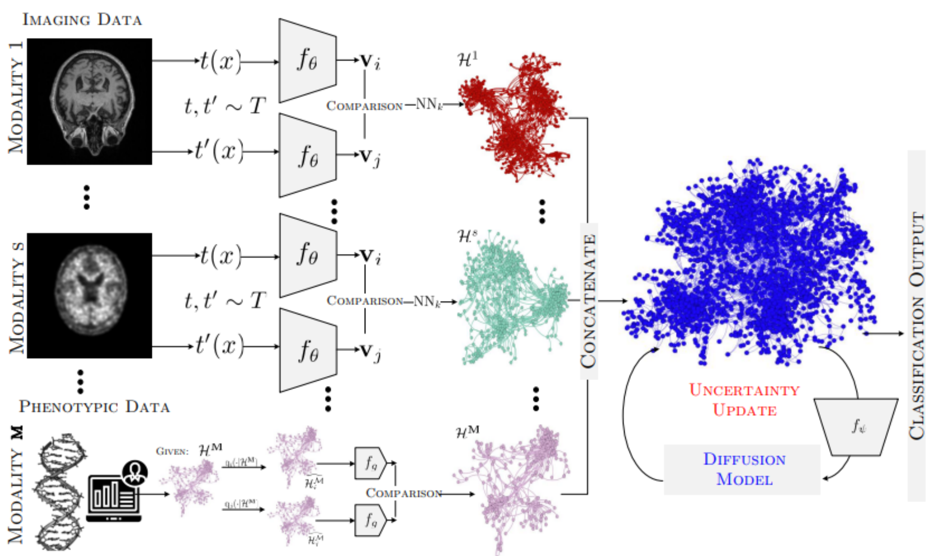
Example of a hypergraph where the hyperedges can relate any number of nodes. Now the hyperedges (e_1, e_2, e_3 , and e_4) are represented by coloured groupings; and can connect any number of nodes. (Image by Kilom691 – CC BY-SA 3.0)

Image analysis approaches fail when it comes to non-image data, such as genetic information, but Aviles Rivero and her colleagues were able to utilize the proper mathematical measurements of similarity to create hypergraphs for each of these various forms of data as well. Lastly, by connecting all the various samples for each patient in a hyperedge, all the hypergraphs for the various modalities of data—the imaging and non-imaging data—are brought together. The multimodal hypergraph that is produced allows for the capture of the highly intricate linkages between all the various forms of data. According to Aviles Rivero, "the hyperedges allow us to create higher-order connections in the data, to go beyond pairwise data interactions."

She then states: "We hope to develop automatic tools that are reliable enough for diagnosing patients in a genuine clinical context." Yet in order to accomplish this, these technologies must be equipped to deal with real-life data.

For instance, although two medical scans of the same individual, such as two MRIs, should have the same information about their brain, they won't always be identical. Two MRIs acquired by separate machines may have tiny differences, slightly different orientations, or minor distortions as a result of handling or digitally saving the picture.

It's brilliant how the scientists construct their mathematical models of the patient data since they are inherently robust to these real-world data issues. The framework automatically creates a rendition of the image data as points in the analogous feature space as previously mentioned. Next, it repeats this procedure with a modified data set in which the images have undergone minor alterations. A machine learning technique called contrastive self-supervised learning is used to compare the generated representations. This shows how to precisely map the images to the feature space so that, even with such minute changes in the images, the final representation is as similar as possible. In spite of any minuscule variations in real-life data, they may be confident that the hypergraph they create from this representation accurately reflects the meaning contained in the image data.



A visual overview of the machine learning framework. (Image from the paper *Multi-Modal Hypergraph Diffusion Network with Dual Prior for Alzheimer Classification*)

The meaning captured inside an image as a whole is not significantly altered by a small modification to the pixel data or the direction of the image. To ensure that the representation of the non-imaging data is also robust, a distinct adaptation to the method is required because even little changes to genetic data might notably alter the data's significance. Instead, Aviles-Rivero and her associates investigate if slight modifications to the corresponding hypergraphs themselves affect the overall structure and meaning of the data.

According to Aviles-Rivero, "Our framework offers several benefits, one of which is the building of a robust hypergraph." "These technological benefits improve the performance for diagnosing Alzheimer's disease."

To date, the algorithm has effectively extracted meaning from different available data sets without any supervision. Yet, a patient's diagnosis does require some human input. Instead of explicitly teaching a computer how to perform a task (as in a standard computer program), machine learning algorithms allow the computer to learn by doing the task itself numerous times.

One choice, known as supervised learning, is a set of data for patients who have already undergone an evaluation by a physician for Alzheimer's disease. Based on this training data, the algorithm can try to diagnose the patients, and then compare its findings to the accurate diagnosis provided by the doctor. (You may see it as the computer being in school and comparing its responses to those provided by a teacher.) In this manner, the algorithm discovers patterns in the data that point to the presence of the disease in a patient.

Although supervised learning produces good results, it requires sizable training data sets that have been labeled with the desired output. "Labelling training data is always time-consuming, but is even trickier in the medical domain because of the clinical expertise that is required," says Aviles-Rivero. "[The problem] is how to develop [machine learning] tools that rely less on labelling."Instead, semi-supervised learning has been the subject of ingenious techniques created by Aviles-Rivero and her associates. Compared to the greater quantity of unlabelled training data, only a very small number of human-labeled training data are provided in semi-supervised learning.

The algorithm then takes as much structure as it can from the labeled and unlabeled data, creating in this instance a robust multi-modal hypergraph for the patient data. The algorithm then propagates the labels across the entire data set using this structure. “Only 15% of the total data set is labeled when they begin,” according to Aviles-Rivero. Next, "they propagate these labels to the entire hypergraph by taking advantage of the links in the data." Imagine doing this by giving each piece of the labelled data a colour that corresponds to their specific diagnosis (for example, green for "healthy," blue for "mild cognitive impairment," and red for "Alzheimer's disease"), and then having these colours diffuse through the connections in the hypergraph to the unlabelled data, much like a drop of colored dye spreading through paper. The researchers employ a brand-new mathematical diffusion model that is motivated by a certain hypergraph property. Although this issue is not new, they offer an alternative approach to the hypergraph diffusion problem.

The two phases that have just been described - creating the hypergraph from the data and then dispersing the labels through the hypergraph - are looped through in their semi-supervised technique, with each step improving the one before it. Avilés Rivero says: "This is a hybrid model. It takes advantage of the mathematical modelling and the [machine] learning techniques,". "It takes the best of both worlds."

Comparing this new method to existing machine learning approaches is more effective at diagnosing Alzheimer's disease. It was put to the test using actual data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set by Aviles-Rivero and her colleagues. They discovered that their technique outperformed all others by including all the various types of data and capturing higher-order linkages within this multi-modal data.

Alzheimer's is a horrific disease, and both the sufferer and their loved ones find the diagnosing procedure stressful. The goal of Aviles-Rivero and her team's research is not to take the position of the specialists who care for these patients. Instead, the framework for machine learning that has been built intends to assist clinicians, freeing up important time to perform tasks that only humans can execute. It’s evident that mathematics is linked to every aspect of life hence acting as a problem-solving tool for medical diseases-indicative of the case above.