Seismic Medicine

Accelerometers and Machine Learning Peek Inside Your Neck

The Problem:

Take a day in the life of my third grade teacher, Mrs. Bungee. School is almost over. She is way beyond the breaking point. All day she has been trying to calm down her students, maybe try to get them to stop talking when she’s talking, only to eventually have to concede defeat and elevate her voice to overcome the chatter. Perhaps she had yard duty today - tasked with the everlasting duty of shouting at the boys for chasing the girls cross the playground - or maybe the students were being particularly argumentative about a new teaching method she thought to try out.

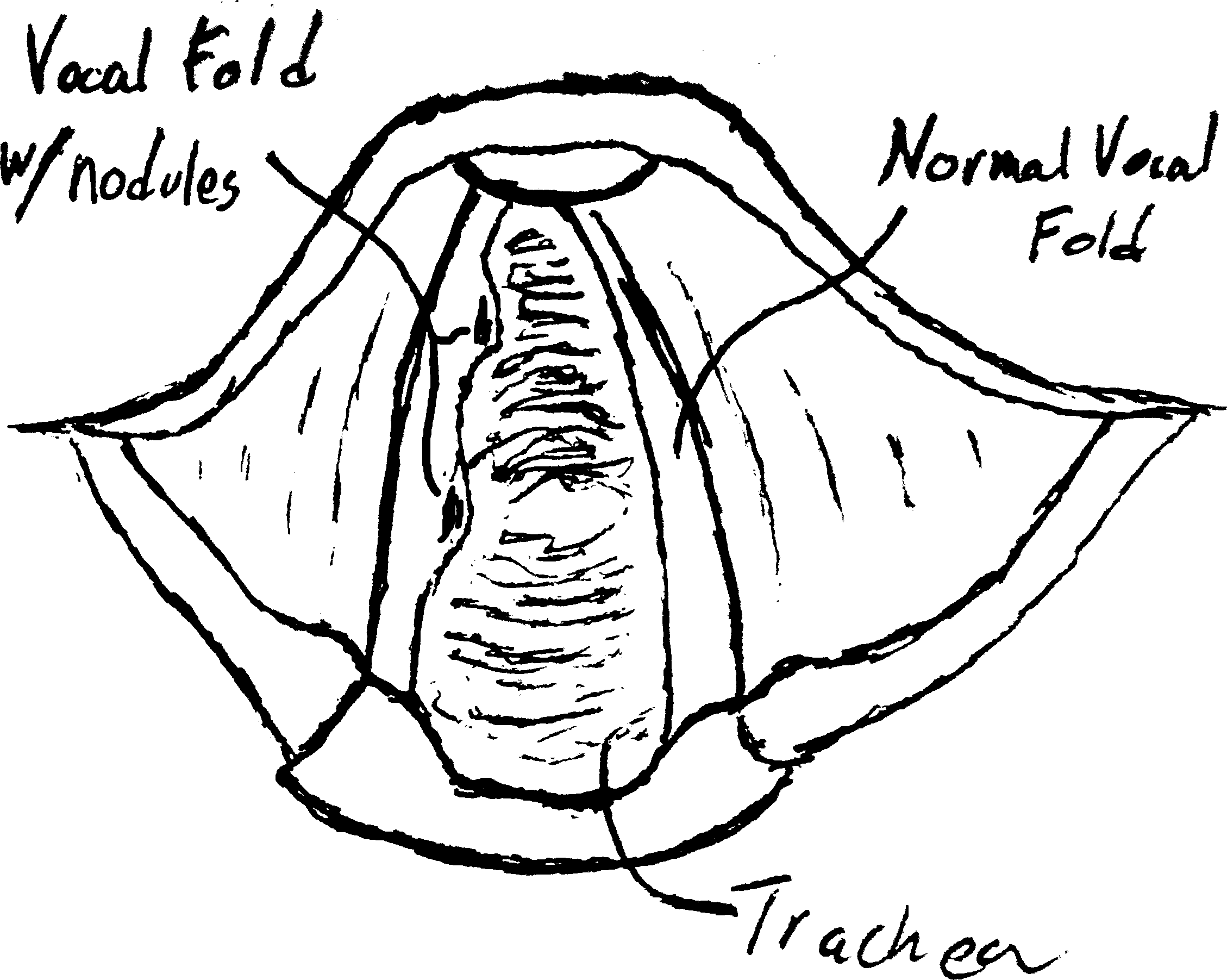
Either way, her voice has been pushed all day, and she’s starting to feel the same old symptoms: soreness, tiredness… it’s taking more and more effort to output at the same volume, and she’s left silently praying none of the children cause too big a stir, holding out hope she’ll get out of raising her voice again today. The days of her playing the piano and singing to the class might be coming to an end, if this keeps up much longer.

Or, take my seventh grade acting techniques teacher, Mrs. Adam. She has a long history of vocal performance, but the countless hours in front of the microphone have finally gotten to her. For her, vocal hyper exertion has taken its toll in a permanent way, and she has been left forever raspy, with no hope of any future voice based gigs. While she enjoys her new job as an acting teacher, it isn’t hard to see that she wishes she had caught her vocal condition early, in hope that she might have been able to keep the damage from developing to the point it is at now.

These problems are in no way unique to my two teachers, or even teachers in general. It is estimated that over 30% of adults have experienced vocal conditions stemming from overexertion, with about 7% experiencing some vocal troubles at any given point in time (Mehta et al. 2015). These rates only grow when you look professions such as teachers, singers, and nurses. And while remedies do exist, it can be hard to get properly diagnosed, as a diagnosis requires specialists and complicated apparatus that are always inconvenient to access for the standard working adult, and its especially hard to get people to seek help when they themselves might not know that there’s help out there, or that they need it in the first place.

So What’s Being Done?

Fortunately, the medical industry has in fact been working on this problem, and there exist treatment programs that can take a patient’s voice, once in a state of constant tiredness, to being indistinguishable from any normal person’s (Ghassemi 2015). The recovery process itself isn’t complicated: after being diagnosed, a patient will see a vocal therapist for a few appointments and practice various vocal exercises to strengthen the vocal cords and rehabilitate a person’s speaking ability.

Unfortunately, the diagnosis process for these conditions has historically been overly complicated, with various needed trips to specialists and long appointments involving complicated apparatus to measure all sorts of possibly relevant data, sometimes even such invasive procedures as forcing telescopic tubes down a patient’s throat to visually examine the vocal cords for the tell tale “vocal fold nodules”, small lumps (as pictured) on the sides of a persons vocal cords (Ghassemi et al. 2013). However, while these lumps work as smoking guns for critical vocal damage, they are only present in the most extreme cases, causing diagnosis to not be as a simple peek.

In the end, the current diagnosis process presents a few major problems: first and foremost, the invasive procedures and large difficulty in accessing a specialist who is able to diagnose an individual present a huge barrier for the layperson to get the medical diagnosis and treatment they need. The second problem is subtler, up until a patient finds himself or herself in the doctor’s chair, being forced to talk ‘normally’ with all that equipment on their face: it turns out that it is simply impossible to get an accurate measure of a person’s day-to-day vocal state from a single, or even multiple, short, out of the norm appointments (Ghassemi et al. 2013).

In short: doctors are faced with the impossible: to accurately and efficiently inspect vocal cords over a long period of time, from days to weeks, in such a way that the patient is able to act and talk completely normally (and privately) while being investigated constantly.

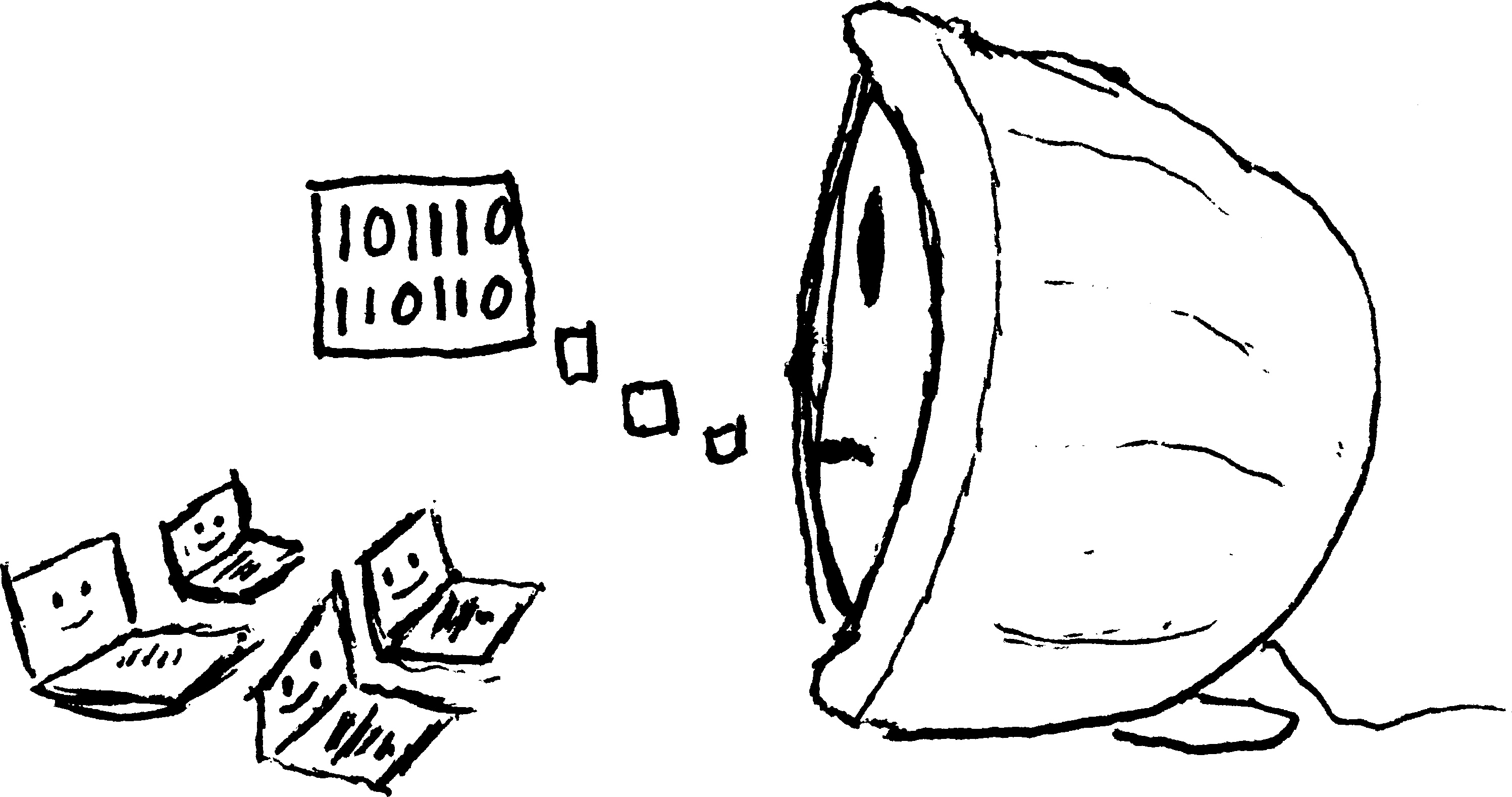
Achieving the Impossible:

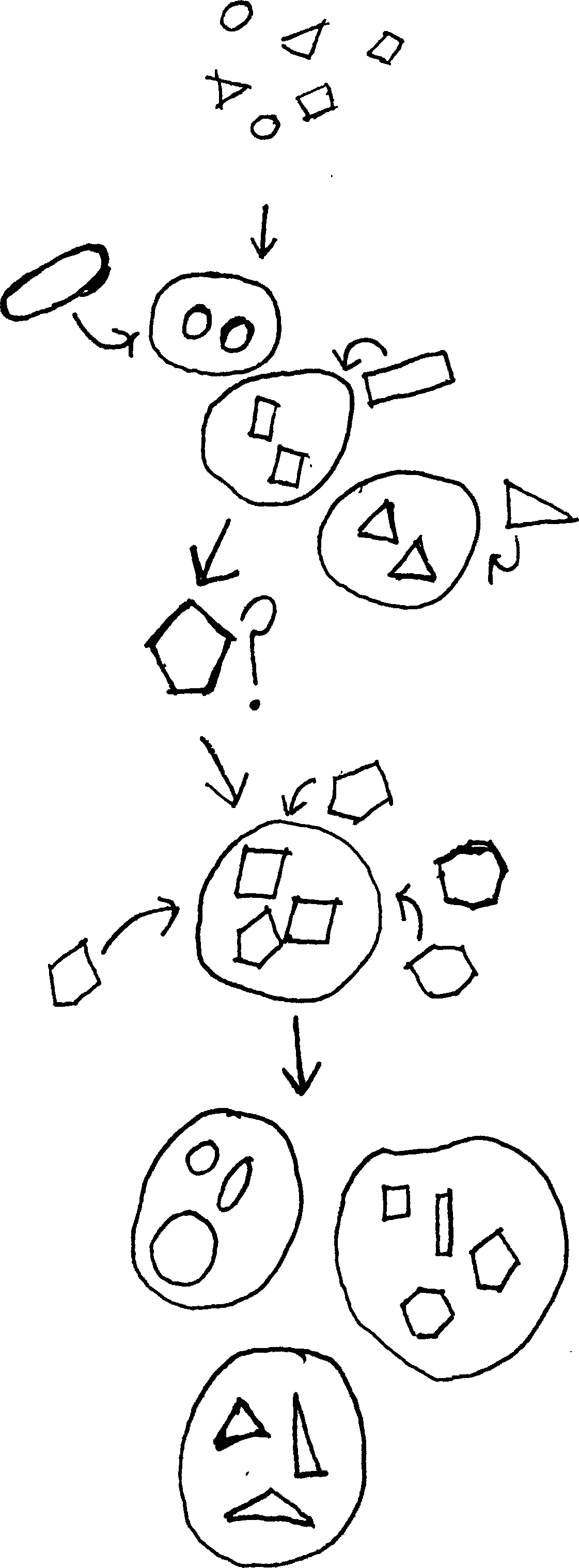
Luckily, where classical medicine falls, modern technology steps up to take its place. Through leveraging the constant presence of technology in day-to-day lives, researchers have been able to develop techniques that rely not on the measurements from many different apparatus over a short period of time, but rather an investigation of at long-term measurements taken by a single instrument: a simple neck mounted accelerometer, with data collection facilitated by something nearly everyone has access to: a smartphone app (Ghassemi et al. 2013).

Via the accelerometer, researchers are able to measure metrics about the vibrations of a person’s vocal cords such as vocal frequency, the total distance traveled by the vocal cords in a given time frame, and various characteristics of the actual shape of a person’s vocal pressure waves, such as weather there is more of an emphasis at the beginning of the pulse or the end, or maybe it is evenly balanced throughout (Ghassemi et al. 2013).

Researchers recorded over 15,345 five minute windows of accelerometer data, collected from 12 persons evaluated positively for vocal disorders, and their 12 corresponding normal counterparts, controlled for age and occupation (Mehta et al. 2015). As one can imagine, with this amount of data, there is no way to simply look at the results and figure out a correlation in one’s head, or even with the standard techniques of statistics of the previous centuries. This is where the Machine Learning comes in.

Machine Learning: Kindergarten for Computers?

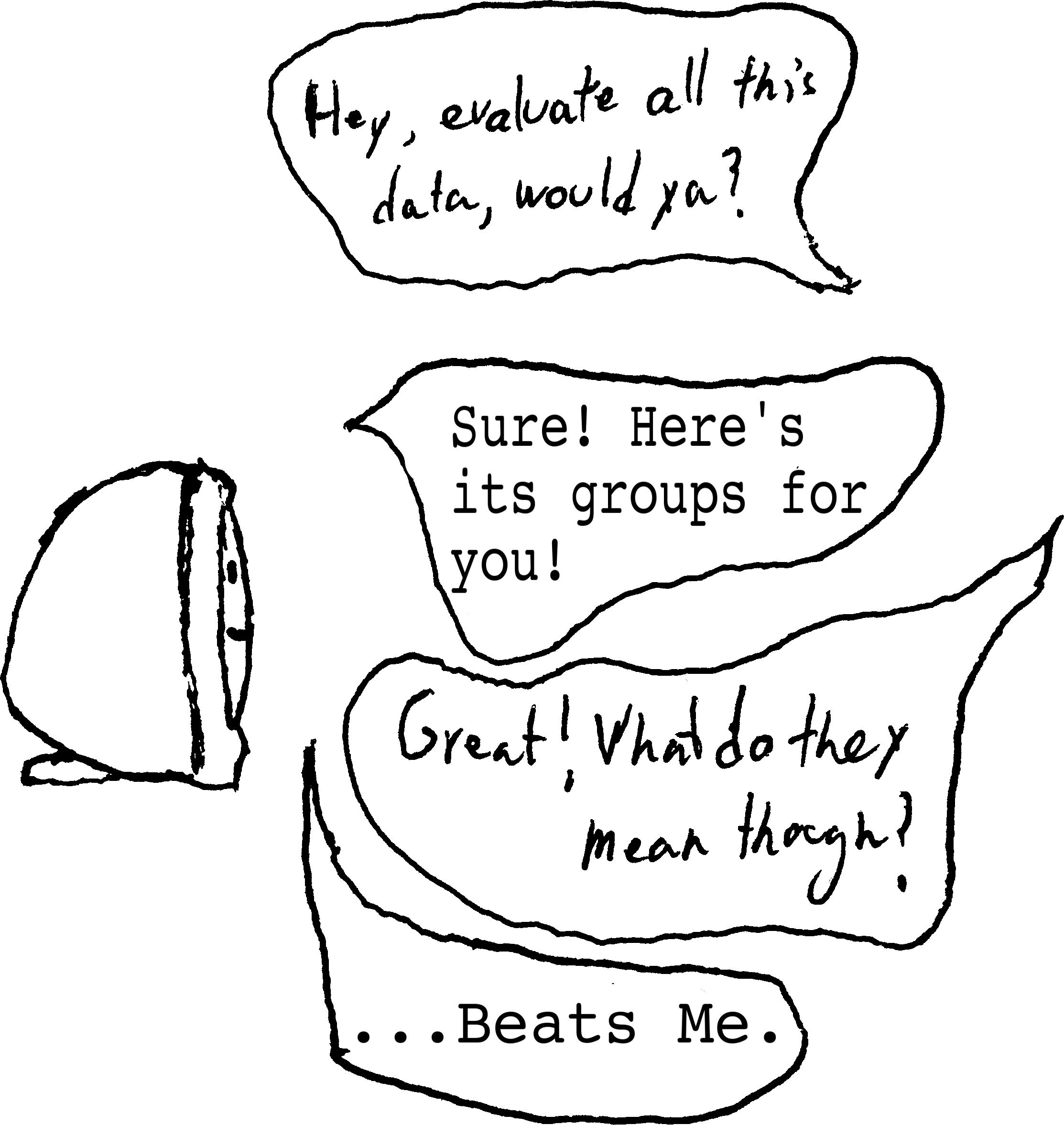
The phrase Machine Learning gets thrown around so much nowadays, it almost ranks up there with such buzzwords as “The Cloud” or “Across all your Devices”. But for Machine Learning, the actual definition isn’t nearly as obvious. To those that don’t know, it might sound almost like we’re teaching computers just as one might teach a kindergartener how to read or write. Interestingly enough, for some machine learning applications, such as the one used by one of the groups working on this project, that’s actually exactly what’s happening. This is the classification algorithm, known as “K-Means”.



K, Now What Does K-Means Mean?

Going back to the kindergarten analogy, how might a toddler to group blocks? Lets say you tell a small child to group the blocks you give her into 3 categories. Then, you give her exclusively white triangles, squares, and circles. In almost all cases, she’ll group circles together, triangles together, and so on. Not too hard. But then what if you throw in a pentagon? Is that closer to a circle or square? The toddler would be forced to make a decision based off of how closely she feels the pentagon resembles the square versus the circle. Lets say after some deliberation she says its closer to the square. Then, you give her another pentagon. Instantly, she groups it with the square again. Why’s that? She has just changed her algorithm to reflect the growing stream of data she’s receiving of course! What’s more, now that she has pentagons in with the squares, she’s probably more likely to put hexagons in there too, after all, her algorithm has already established that the group is valid for squares and pentagons, another side isn’t too big a jump from there. Eventually, the toddler will have a fully functioning classification algorithm able to sort arbitrary shapes into ellipsoids, triangles, and polygons, *with no idea of what any of those things are.*

This basic technique is easily adapted to the computer in a technique called *K-Means*. *K*, because the number of groups (three above: ellipses, triangles, and squares) is variable, and is typically referred to with the letter ‘K’, and *Means* because the cluster is really just the average, or ‘Mean’ of all the data points within it (in the above example you might say the mean of the triangle cluster’s side counts is three, that for the polygons is around six, and for the ellipses is either zero or some large number).

While the example above was only in one dimension, namely number of sides, a computer would be able to evaluate data objects consisting of a large number of dimensions, going to perhaps ten or more (dimensions simply meaning different aspects of a data point being recorded: take for example not only looking at number of sides above, but also color, weight, size, and material).

However, this increased detail comes with a trade off: while in the example of the preschooler is it easy to look back at the groupings and identify why any given object was placed in a given group, as the number of groups gets higher, and the data gets more abstract, just as the child did not know what any of the names of the groupings she made were, a computer *cannot determine what makes a group different from an other in any meaningful way*. This is one of the major caveats of K-Means, because while the computer may classify ad infinitum, it will never be able to say what those classifications mean. This stems form K-Means being an *unsupervised* learning approach, meaning as data points are given to the algorithm, just as when blocks were given to the child, they *cannot* be named.

This may seem strange: why can’t we just add some ‘is afflicted’ data point to the data points as we pass them in? The simple answer is: when we later go back to try and classify new data with the means we have established, we will have no ‘is afflicted’ data to go off of. The state of affliction is indeed what the algorithm is meant to determine!

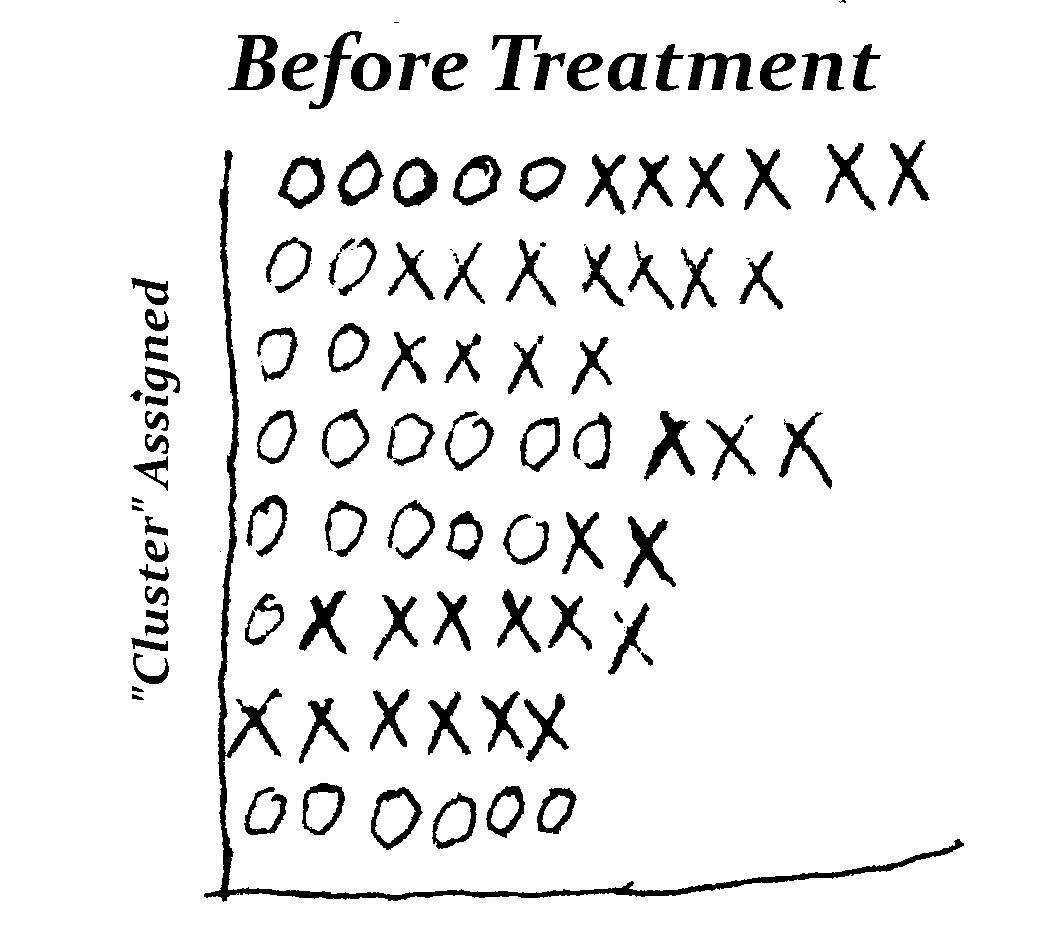
Eventually, after having the computer look at a large set of sample, called *training,* data, from both healthy and afflicted individuals, researchers may begin to pass in arbitrary collected data points from new patients for categorization, in order to produce a diagnosis. Based on whether these new points are assigned to groups containing more training data from the healthy or afflicted individuals, the computer is able to produce an estimate of weather or not this new data comes from an afflicted individual.

Unfortunately, in reality, the seeming simplicity disappears, because as much as we might like to be able to tell the algorithms to create two clusters (or in technical parlance: set *K* equal to two), and just hope one is afflicted and one is normal, a situation like that simply doesn’t happened with the high degree of complexity of big data. Rather than two big clusters, one of afflicted and one of normal, the data better is better represented by many different clusters, often with the same clusters containing data points from both unaffiliated and afflicted patients. This is where the statistics of Machine Learning begin to kick in.

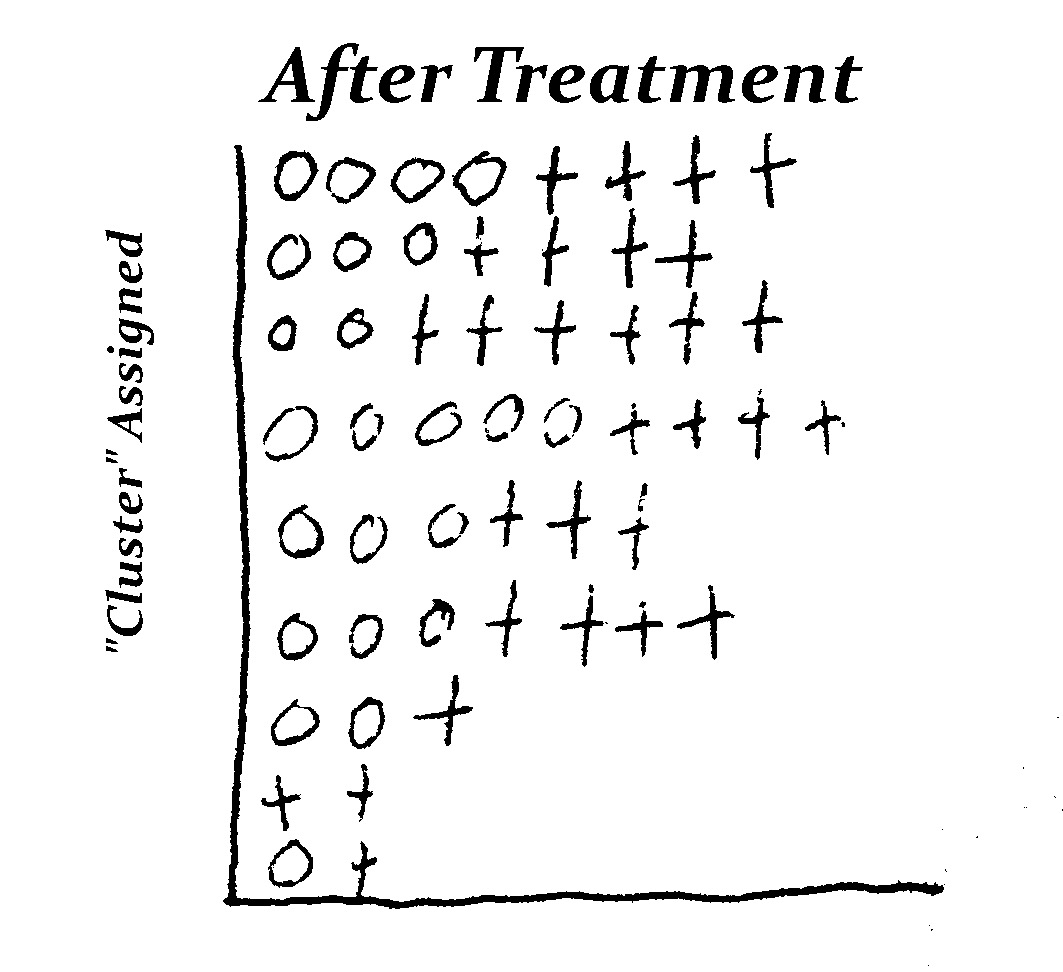
Statistics: Why Math Knows You Better Than You Do

Let’s take a look back at the data from the Vocal Accelerometers. There, researchers took all the collected data from the accelerometers, those frequencies, tone pulse characteristics, and so on, and put all the data through the K-Means algorithm many times over, with K ranging from 2 to 38. From the returned clusters, researchers were able to take millions of different samples, and reduce them into K letters in what they describe as an “alphabet of shapes” (Ghassemi 2015). Once again, the K-Means can’t say what any of the shapes meant, only that a given shape looks a lot like the shapes in this other group of shapes.

Once all the clusters have been made – or the alphabet has been formed, so to speak – the real magic can happen. After getting the clusters from the computer, researchers are able to go back and label each of the data points as coming from a control or patient. Then, if there is a *statistically significant* (meaning ‘very hard to attribute simply to randomness’) tendency for a given cluster to contain data from either the patients or controls, researchers can develop an association between a given vocal pattern (‘letter’, to continue the alphabet analogy) and either normal voce or the voice of someone with vocal hyperexertion. Continuing the analogy, if almost all of the training data that got identified as an ‘A’ turned out to come from people with vocal hyperexertion, and a new subjects data comes in and is identified as ‘A’, researchers can conclude that it is likely that this new patient suffers from this condition.

In the most recent example, Ghassemi 2015, it was shown that the difference between the placement of those with vocal disorders versus those without is indeed statistically significant. In the ‘Before Treatment’ plot to the right, one can see how the clustering obtained via the K-Means does show a significant trend towards ‘X’s (those with vocal conditions) being grouped with other ‘X’s, and the same for ‘O’s (those without conditions). In fact, the chance of such a polarized clustering from random chance is less than 0.00001% (Ghassemi 2015). This shows that if a new subject’s data gets classified into say the first or fourth group form the bottom, there’s a good chance that they are an ‘O’, or healthy, but if a subject is sorted into say the second or third group form the bottom, then they probably have some vocal condition.

In practice, this technique works exceedingly well. In a recent study (Ghassemi 2015), given data from 24 patients, the computer was able to correctly diagnose 22 of them. It is not unreasonable to assume that with more time, and more training data, the percentage of correct diagnoses will only rise.

On the other hand, this *statistically significant* separation isn’t always present. However, this lack of a separation can give information just as valuable as if it were there. This can be seen in practice when researchers were able to confirm that vocal therapy techniques actually are effective at removing signs of vocal hyperexertion. This evidence for this comes from researchers training a new algorithm with data from both normal people and previously afflicted patients that had gone through vocal therapy. This time, the distinct grouping of data into clusters of like sources was gone, showing there to be no real difference between the voices of post therapy patients, and any normal person. In other words, the therapy works!

A Bright and Sound Filled Future?

So then what does that mean for the rest of us, such as my two teachers at their various stages of vocal struggle? In short, it’s actually looking good for them. As this technology progresses onward, it will become easier and easier for them and others like them to get properly diagnosed. Once diagnosed, these people will be well on their way to a better life, as this new data shows that if they then seek vocal therapy, they should expect to see a full recovery.

On the technological side, this use of computers in the medical industry has a lot of potential that is being realized every day. As wearable health and fitness tracking technology becomes more and more accessible, the realized uses for computers and machine learning will only rise. For this case in particular, as more data is analyzed, and perhaps different algorithms are tested, the reliability of these at-home diagnoses will only rise, giving everybody a chance at a healthier and more productive life.