NetraMark

Next-Gen AI for Disease Definition, Patient Stratification,
and Placebo Effect

AUTHORS

Mike Tsay, Director of Bioinformatics, NetraMark Corporation

Dr. Joseph Geraci, CEO of NetraMark Corporation and Assistant Professor of Molecular Medicine, Queen's University

Abhishek Agrawal, Associate Director, Strategy & Operations, Solid Tumor Franchise, Novartis

Challenges in the Current Drug Industry

The Cost of Getting a Drug to Market

The average cost of developing and releasing a drug to market rose to a landmark \$2.56 billion USD, according to the 2016 Tufts Center for the Study of Drug Development. Earlier reports from the same group suggested the fully capitalized cost of a new drug had increased almost 245% from \$1.04 billion in the previous decade and an astounding 619% from \$413 million in the 1980s/early 1990s.¹

Figure 1. Decline of drugs brought to market per billion USD. Scannell et al. 2012.

These increases are a major concern for pharmaceutical companies and government agencies, as well as patients seeking newer and more effective treatments. Indeed, the Tufts report confirms what other studies have shown regarding not only increasing costs, but also the steady decline of the number of drugs per billion dollars expended, and this despite advances in biological knowledge, sophisticated machinery, and artificial intelligence (AI).²

In part, the rise in overall costs can be attributed to the landscape of increasingly successful drugs and how newer drugs require more effort and resources to compete. Scannell also attributes the change in costs to: spikes in regulatory caution, which tend to occur shortly after post-market withdrawal of drugs; and the industry's tendency to resolve developmental pipeline issues with brute force and ever-increasing injections of money.²

In a traditional lab-based paradigm, these approaches

have proven effective yet expensive, and financially unsustainable. As of 2018, financial resources attributed to 22% of prematurely terminated Phase 3 clinical trials, leaving another 78% failing for clinical or regulatory factors.³

Such failures, often addressed with additional financial and physical resources, mostly stem from decisions made early in the development process. Patient inclusion and exclusion criteria, for example, ideally represent the intended population, yet many times, unanticipated criteria or poorly understood disease pathology lead to the wrong cohort being selected. Depending on the drug in question, the marketing to and recruitment of study patients can be significant, particularly when subpopulations are poorly defined.

If future costs to bring drugs to market are to decrease significantly, the industry will need to adopt technologies, such as next-generation AI platforms, and reconsider the paradigm that centers on lead compound identification. Similarly, a better understanding of disease states and the proper identification of patient subpopulations will significantly alter the drug development landscape.

Given that downstream clinical studies and approvals depend on such an understanding, efforts to reduce costs and increase efficiency could potentially transform everything from study design to appropriate patient recruitment and trial execution, and ultimately, clinical outcomes.

CHAPTER 1

The State of Artificial Intelligence in Drug Development

How First-Generation AI is Trying to Break Into the Market

The current strategy for drug development begins with the identification of a biologically relevant compound that is then characterized for efficacy using assays and animal-based models. This is followed by the systematic and iterative physicochemical modification of the lead to provide a library of compounds with high therapeutic potential but few or no detrimental effects.⁴ The success of such a strategy is seen in the rational design of

drugs such as Dorzolamide, Saquinavir, Zanamivir and Oseltamivir, among others.⁵

Automation of traditional drug discovery techniques such as 3D structure prediction, ligand design, and docking is a natural evolution of these largely intuitive methods. The incorporation of machine learning (ML) into high-throughput screening and automation has elevated these strategies even further with companies such as Atomwise, Cyclica, Insilico Medicine, and Numerate at the forefront of this niche.

A smaller segment of the AI community, however, has focused their ML efforts on improving our understanding of disease. For example, BenevolentAI and Berg LLC extrapolate disease information by using deep-learning methods on data from journal articles, patents, clinical trials, and even cell biology.⁶ Meanwhile, companies such as GNS Healthcare have focused their attention on the identification of patient subpopulations.⁷ Insights into disease and patient population are truly valuable when the variables describing them are transparent and explainable. Deep-learning methods are handicapped by their inability to explain their insights while simpler machine learning methods, such as Bayesian networks used by GNS, may not be sophisticated enough to integrate many variables at once nor provide interpretability, making it difficult or impossible to decipher heterogeneous diseases. Machine Learning and Artificial Intelligence must evolve if they are to make inroads into truly solving complex diseases.

Generic AI and Machine Learning is not up to the task

ML should aid in generating better hypothesis associated with disease therefore enabling proper patient cohort selection helping improve R&D productivity. To seriously investigate complex heterogeneous disease systems, an approach must solve three fundamental handicaps of current AI and ML systems:

1 Current ML packages and algorithms are largely designed for simple tasks.

Open-source algorithms arising from academic research, for example, are designed to solve sets of well-defined parameterized tasks, such as image recognition and recommender systems. This is seen in the recent surge in neural network adoption for diagnostic imaging, occurring three times as often as for all other uses combined.⁴

2 Almost all ML techniques depend on large amounts of high-quality, clean data.

The described trend to neural network frameworks and variants has only exacerbated this need. The reality, however, is that most available data is low-resolution and noisy, aspects that can bias interpretation and limit predictive power.

3 Many AI systems, used to learn about disease complexity, cannot clearly identify variables that drive the understanding of their models.

Such analyses can result in forced and faulty models. Although such limitations may be acceptable for more generic projects, they complicate adoption of AI in the highly regulated pharmaceutical industry via the obfuscation of mechanisms in “black boxes”, which hamper the understanding of disease and patients. Whereas some companies use more “explainable” ML to reveal important variables, these algorithms lack the sophistication required to reveal subtleties within complex disease states. More advanced ML techniques, conversely, try to tackle the complexity, but end up sacrificing explainability.

CHAPTER 2

Next-Gen AI: NetraAI Shows a Different Perspective of Defining Disease States and Patient Subpopulations

Defining Disease – Understanding Heterogeneity

Current generation machine learning is capable of learning about classes that are known, exemplified by the model in Figure 2 involving non-small cell lung cancer patients from two public datasets (GSE10245 and GSE18842). Using Random Forest and t-SNE (Figure 2) or UMAP (not shown), patients can be correctly classified into two meta-categories: squamous cell carcinoma and adenocarcinoma. Based on the literature, however, the heterogeneity associated with these two subpopulations is likely much more complex than our usual clinical definitions.⁸⁻¹⁰ Is

it possible to discover subtypes of disease using datasets labelled according to our usual clinical understanding?

Figure 2. Classification of adenocarcinoma (blue) vs. squamous cell carcinoma (red) from balanced data (GSE10245 and GSE18842) using t-Distributed Stochastic Neighbor Embedding (t-SNE).

How NetraAI Sees Disease and Patient Populations

NetraAI is NetraMark's answer to the aforementioned issues with modern machine learning. It incorporates novel algorithms with other ML algorithms derived from publicly accessible literature. In some ways, NetraAI can be viewed as a platform that fuses the benefits of other machine learning algorithms but overcomes their detriments with special elements from NetraMark's DeepCrush algorithm, and our augmented intelligence interface, NetraPlay, described below. This not only allows NetraAI to utilize next-gen algorithms, but also enables scientists to interact and understand data with exceptional precision.

Our understanding of disease pathology and its treatment is a perpetually moving target. Each year, new insights, research methods, and technologies help us refine our understanding of drug design, and target selection. In recent years, AI has emerged as the technology most likely to consistently produce new targets, drug candidates and methodologies.⁶ Can we harness the true power of AI to combine rich domain knowledge with complete sets of clinical, patient, and genetic datasets to produce high-quality insights and definitions?

Over the past decade, NetraMark has developed and refined NetraAI, a computer engine capable of

discovering high-resolution, previously unrecognized disease subtypes and patient subpopulations. Using the same datasets described above, NetraAI distinguishes three clear subtypes of squamous cell carcinoma and four clear subtypes of adenocarcinoma. It also identifies similarity between individuals and clusters based on variables that drive separation.

Figure 3. Classification of adenocarcinoma (blue) vs. squamous cell carcinoma (red) from balanced data (GSE10245 and GSE18842) using NetraAI shows 3+ distinct subpopulations in each disease.

Algorithm Fundamentals

NetraAI's core algorithms are based on the concept that accessible variables may be unable to explain the clinical characteristics of all patients. Instead, the system identifies subpopulations that are both explainable and substantially correlated with available variables. A secondary concept that drove the engineering of these algorithms is that there are multiple ways in which to view a patient population. Thus, the system extracts different perspectives of a patient population and can merge several to form supremely robust models.

These algorithms are designed to learn models using smaller data while still seeking the requested goal through a supervised algorithm, all while simultaneously discovering unanticipated substructures. The algorithm knows what it cannot explain and so it provides very precise and honest insights into what is driving the disease state. This offers the user insights into disease heterogeneity, and extracts models leading to personalized treatment protocols.

NetraAI uses advanced mathematics, statistics and machine learning to combine high-resolution data separation on low quantities of data to reveal mechanisms

Mathematical Background

Critical to the ability of NetraAI to view patient populations from a multiscale perspective is its unique mathematical basis. This allows it to extract models wherever it can extract relationships between patients, such that novel genetically defined subclasses become apparent.

This multiscale perspective comes from the fusion of random matrix theory, contraction mappings and reinforcement learning, which allows us to merge insights from several different ML algorithms. The particular algorithm that utilizes these mathematical principles is called DeepCrush. There are no deep neural networks involved in DeepCrush alone, but the machine does perform a type of compression on the data, hence the ‘Crush’.

The NetraAI technology has been used to analyze psychiatric clinical scales, gene expression, copy number variation, methylation, miRNA, RNA-Seq, and proteomics data with excellent results.

NetraMaps

A major challenge with ML is explainability, which is the capacity to describe which and how much any variable, and combination of variables, contributes to a

solution. This is typically a non-trivial task considering the ample combinatorial complexity involved in learning from datasets consisting of a large number of variables. DeepCrush together with NetraMaps addresses this challenge by offering users unique ways to engage and interact with high-dimensional models, assuring that anyone can extract information about known and unanticipated patient subpopulations. Users can interrogate models simply by using a looping function to explore which variables draw a group of patients together. Alternatively, the same tool can help elucidate which variables are driving two or more subgroups. This process provides true Augmented Intelligence: the fusion of human and computer intelligence.

The NetraHealthAtlas

Current biological research is based on small, understandable experiments that iteratively assemble knowledge toward a final summation. Where these projects are designed to answer specific questions, however, breakthroughs in disease understanding and treatment may require a more global view.

The NetraHealthAtlas was created to address this specific challenge. It is a complete database infrastructure of public and private knowledge that permits access to all aspects of the clinical, biological and algorithmic landscape to establish a foundation for previously unknown disease descriptors. The database is driven by a set of computational biology and machine learning software, including DeepCrush. Central to this idea is a collection of novel disease definitions, derived from our current broader understanding of disease, and a list of compounds which may potentially be used to treat them.

Figure 4. NetraHealthAtlas and NetraAI Engine derive disease definitions, patient subpopulations and placebo response which have direct impact on drug development and clinical trials.

The NetraHealthAtlas is intended to create a curated, high-quality repository leading to significant discoveries and savings for established and emerging drug discovery companies.

Although many public and private datasets exist, most have been created to serve a very specific and often narrow scope. Thus, it is difficult to combine large datasets to identify high-level associations; there is a large amount of data, but with little overlap and thus, low quality.

The NetraHealthAtlas aims to build a unified landscape, using data of variable quality to comprehensively describe disease states, identify associations on a global level, produce entry points and leads for drug discovery, and identify treatment options for the public and private sector.

CASE STUDY

Validation of NetraHealthAtlas in Parkinson's Disease, Alzheimer's Disease, and Aging

The NetraHealthAtlas is tasked to understand the inherent patient population heterogeneity of disease states. The goal is for the machine to replace our current bulky definitions of disease with a set of basic biological elements through which a set of patients cohere. The NetraHealthAtlas enables strategies to precisely target disease states, due to the fact that basic biological elements, such as gene expression, methylation, miRNA, are able to reveal the mechanisms at play. Here we describe some recent results derived from testing our system for Parkinson's Disease, Alzheimer's Disease, and aging.

Parkinson's Disease

Private data (confidential organization) consisting of gene expression (~20000 genes) from cell lines derived from the substantia nigra of Parkinson's patients vs. controls was processed by the NetraHealthAtlas which divided patients into several different subpopulations. Each subpopulation was defined by a set of 16 genes and can be accessed through a set of NetraMaps, considered the subconscious of the algorithms.

The machine reported several molecules that could be repositioned and here we indicate two molecules, simvastatin and nilotinib. Of interest is that, not only did the algorithm correctly identify these molecules, but the NetraHealthAtlas also produced a Parkinson's subtype definition which may best respond to these potential treatments. This presents a compelling proposition for developing a digital therapeutics offering wherein multiple drug combinations using the NetraPlatform could be developed for managing a disease condition.

Alzheimer's Disease

ALZHEIMER'S DISEASE

- 💰 Average Development cost per Drug is \$5.6 Billion dollars.¹¹
- 🕒 Time from discovery to approval is 13 years.¹²
- ❗ In the decade leading up to 2012, the failure rate of potential drugs that failed before reaching the market is 99.6%.¹³

AD is widely accepted to be a heterogeneous disease likely consisting of many subtypes. The recent flood of clinical trial failures in Alzheimer's drug development is indicative of some basic shortcoming of the current understanding of AD. We primed the NetraHealthAtlas with a gene expression dataset consisting of controls vs AD patients (GEO dataset GSE84422). NetraAI found several interesting novel molecules that could potentially be repurposed for AD treatment along with the definition of the subtypes that they could treat. Two candidates the NetraHealthAtlas determined to have therapeutic use were cetuximab and vorinostat along with a third molecule which is a statin: lovastatin. These findings are known potential candidates and support that several mechanisms may be at play in Alzheimer's. These insights present an opportunity to understand and confirm valuable hypotheses, some that are pertinent for the recent amyloid based failures. Preliminary explorations of certain Alzheimer's maps have revealed that certain subpopulations are driven by phosphodiesterase in varying degrees but

also that the protein, presenelin, can have major differences between subpopulations. This reveals heterogeneity in the population with respect to the mechanisms at play behind the accumulation of plaques and tangles. It is also understood that this type of heterogeneity plays a major role in oncology and NetraAI's ability to augment personalization efforts is vital to this area of medicine.

Aging

Using a private data set consisting of gene expression from fibroblasts derived from skin samples from a population ranging in age from 1 to 96 years old, the NetraHealthAtlas was asked to distinguish between people under 50 and over 50 in an attempt to find subtypes from this initial binary labelling. The NetraMaps produced reveal that the algorithm was able to extract distinct subpopulations of people with ages beyond 75. The variables driving this insight further indicated putative targets for anti-aging. The technology's findings were confirmed when the machine proposed several interesting compounds including Rosiglitazone and Bezafibrate, both of which have been the subject of antiaging in the scientific literature.^{14,15}

CASE STUDY

Classifying patients and identifying trends using the 'Data Microscope', NetraPlay

A growing body of evidence shows that the variations in the taxonomic and functional composition of the human gut microbiome can contribute not only to patients' risk of disease development but also to individualized manifestations of disease and responses to particular therapies.

Applying NetraAI to a privately processed dataset including microbiome samples from 55 healthy controls and 153 patients with a common gastrointestinal disease (GD), we were able to clearly distinguish between several classes based on a concept called purity.

Figure 5. When 153 gastrointestinal disease patients and 55 control patients are analyzed using NetraAI, patients with gastrointestinal disease are classified into 4 distinct subpopulations. At a global viewpoint, the control group forms a tight cluster. NetraAI is designed to “zoom in” to the landscape to see the true heterogeneity of the global “control group”.

Purity is how the algorithm is rewarded for making good choices – if the machine finds a subpopulation that mostly consists of the same type of patient or sample, it is rewarded. The parameters of this reward can be varied to find different types of models. Viewed macroscopically, NetraAI identified four different subtypes of the gastrointestinal disorder, whereas all control patients grouped to the top left corner of the chart.

Drilling microscopically into the control group, however, we discovered an entirely new “map” describing this cluster. Although most control patients clustered, a small spread of diseased patients was evident.

Closer examination of this subset showed that many were patients who displayed symptoms of the disease yet still registered as healthy controls. This enables us to generate several investigable hypotheses as well as design stronger patient controls. Patients who are “near” the controls may more easily recover from disease onset than those with more severe gastrointestinal disease. Further investigation of this group may reveal treatments that correlate well with specific patient clusters, maximizing efficacy while reducing the risks of adverse side effects or inappropriate treatment.

NetraPlay is the process that we have engineered to allow a user to zoom into subpopulations, explore subpopulations, and to generate insights. The goal in developing this technology was to provide a true augmented intelligence experience for users. The graphic in Figure 5 was derived from NetraPlay, as this technology allows us to zoom into a NetraMap.

The data from this project consisted of gene abundance data coupled with metabolomic mass spectrometry data from the gut microbiome of these patients. NetraMark is focused on developing insights from microbiome data as our particular technologies are capable of extracting deep insights from these kinds of non-traditional data. We believe that this integration can have an impact on the personalized treatment of many disease states including neurodegeneration, psychiatric, and oncological.

CASE STUDY

Using Clinical Data to Identify the Causes of the Placebo Effect

Although poorly understood, the placebo effect is a well known confounder of many clinical trials, where thresholds to identify treatment efficacy can be raised to such a point that a possibly effective therapy is otherwise considered to have an insignificant or insufficient impact on disease progression or patient outcomes. Attempts to elucidate the placebo effect, to understand why some patients respond to non-intervention, using traditional ML methods have largely failed. In part, commodity algorithms are incapable of teasing out subtleties in data due to the supplemental complexity of unquantified or unconsidered factors such as patient actions and expectations. Analyzing a clinical dataset that incorporated such behavioral and attitudinal factors (CTSS or Clinical Trials Site Scale), NetraAI correctly distinguished placebo responders from non-responders while doing proof of principle work for a failed mood disorder drug for a top 20 pharmaceutical company. DeepCrush extracted a subset of elements from this scale, from hundreds

Figure 6. Using the CTSS Scale alone, NetraAI is able to correctly separate placebo non-responders (red) and super responders (blue). Unique to NetraAI, unexplainable patients are pulled into a No Call Region (lower left grouping). This effectively removes noise and ensures purity and confidence of the other clusters.

of clinical variables, which it used to predict placebo response.

Furthermore, whereas traditional ML would force patients into specific categories, the underlying heuristic sub-classification of patients in NetraAI enabled the identification of “No Call Patients”, those who could not be reliably classified. Effectively removing these outliers allowed NetraAI to define highly pure patient classes, and to become a reliable decision maker for clinical trial recruitment.

Applying NetraAI & NetraHealthAtlas to an even broader global dataset could permit insights into the fundamental nature of the placebo effect.

“ The CTSS together with NetraAI is unique in that they have shown the ability to identify placebo responders in multiple trials. The potential to use the CTSS to identify placebo responders has the ability to significantly reduce trial costs, reduce time approval and ultimately get new treatments to patients in need quickly at a lower cost than traditional approaches. ”

- Dr. Robert Morlock, Creator of the CTSS, Managing Director, YourCareChoice, Ann Arbor, MI

NetraMark is Uniquely Positioned to be Paradigm-Disrupting via Next-Generation Methods

A new paradigm for understanding disease states and patient subpopulations will have an immediate impact on all phases of drug development, from research to clinical trials. At NetraMark, we have become disease cartographers with our unique set of tools, featuring a compelling next-gen ML algorithm which enables us to reveal the inherent entropy in complex systems. We believe this can become transformative information for intelligent experimental design and particularly for any firm focused on therapeutics. NetraAI has the

capability of impacting R&D by reframing disease pathology, optimizing target discovery, defining patient subpopulations and understanding the placebo effect. By reframing a disorder as a series of disease pathologies and the patients as distinct subpopulations, this new approach aims to improve the success and efficiency of the pharmaceutical undertaking. An established infrastructure can lay the foundation for marketed products such as digital therapeutics for combination therapy and developing stronger campaigns based on disease pathology. Further to our vision, the ability to integrate and understand patients from the 'here-to-stay' onslaught of medical data through an augmented intelligence framework, will allow for the ideal fusion of human and machine intelligence.

Bibliography

1. Dimasi, Joseph A., Henry G. Grabowski, Ronald W. Hansen. Innovation in the pharmaceutical industry: New estimates of R and D costs. *Journal of Health Economics*, V 47. May 2016, 20-33.
2. Scannell et al. Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Reviews Drug Discovery*. V 11, March 2012 191-200.
3. Fogel. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communication*, V 11 2018, 156-164.
4. Mak, KK and M. Pichika. Artificial intelligence in drug development: present status and future prospects, *Drug Discovery Today*, March 2019.
5. Singh, Dev Bukhsh. Success, Limitation and Future of Computer Aided Drug Designing. *Translational Medicine* 2014, 4:4.
6. Fleming, Nic. Computer-calculated compounds. *Nature*. 557, S55-S57 (2018)
7. Ware, V. Artificial Intelligence, Real Solutions: How AI Can Accelerate Drug Discovery. (<https://www.technologynetworks.com/informatics/articles/artificial-intelligence-real-solutions-how-ai-can-accelerate-drug-discovery-311063>)
8. Motoi N, et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. *Am J Surg Pathol* 2008;32:810-27.
9. Russell PA, et al. Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. *J Thorac Oncol* 2011;6:1496-504.
10. Zandberg DP et al. Intratumor genetic heterogeneity in squamous cell carcinoma of the oral cavity. *Head Neck*. 2019 Mar 14
11. Scott et al. Economic analysis of opportunities to accelerate Alzheimer's disease research and development. *Ann NY Acad Sci*. 2014 Apr; 1313(1): 17-34
12. Cummings et al. Alzheimer's disease drug development pipeline: 2018. 2018 Volume 4, Pages 195-214
13. Cummings et al. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther*. 2014 Jul 3;6(4):37
14. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743759/>
15. <https://www.ncbi.nlm.nih.gov/pubmed/25703056>