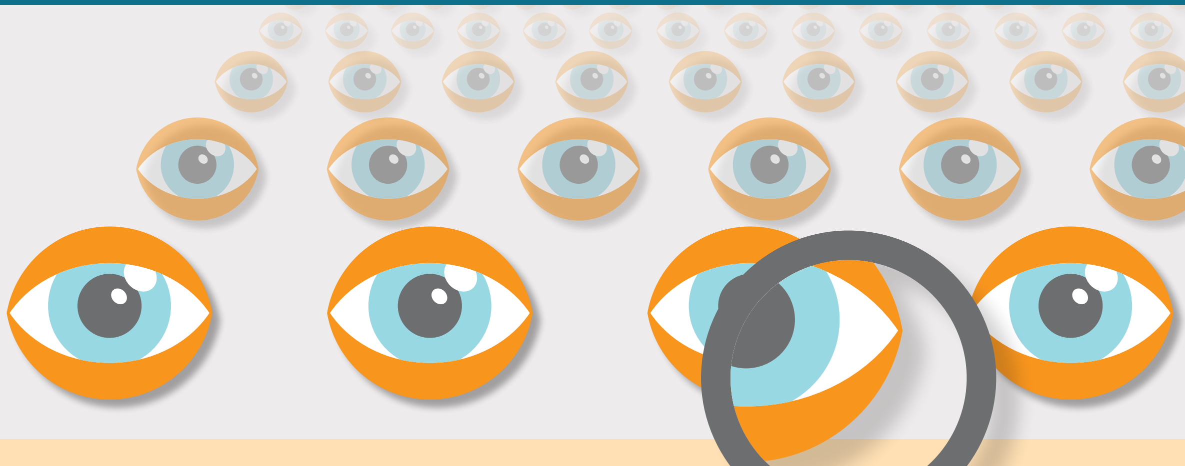
A Case Study on the Use of Unstructured Data in Healthcare Analytics

Analysis of Images for Diabetic Retinopathy

WPC HEALTHCARE

wisdom.performance.credibility



A Case Study on the Use of Unstructured Data in Healthcare Analytics:

Analysis of Images for Diabetic Retinopathy

Precision medicine is being advanced today by a combination of advanced analytics and physician expertise. The new frontier in this field is the analysis of unstructured data.

We know that most medical information resides in unstructured form – the most common forms are clinical notes and images. However, most of the analytics in healthcare today focuses on structured data, typically from hospital EMR and claims/reimbursement related systems.

In this paper, the authors explore the new frontiers in the analysis of unstructured data by discussing a specific case where images were used effectively in vision care, and discuss the potential uses of images across a wide range of situations and use cases.

Importantly, the case study is an illustration of how technology and advanced analytics complement a skilled physician's intuition and expertise, without necessarily supplanting it.

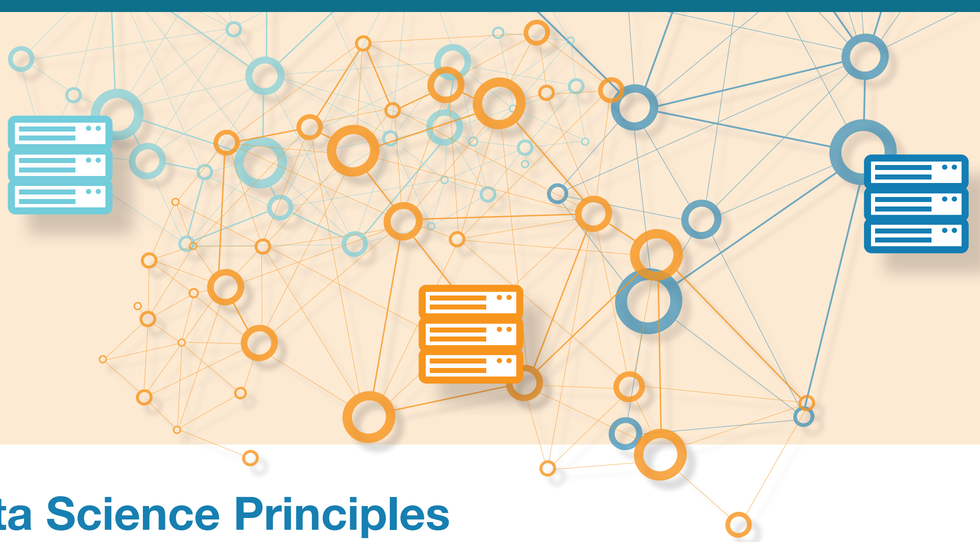
The Project

EyePACS, a provider of picture archive communication systems offered the original data set for analysis. WPC Healthcare, a Nashville-based advanced analytics company enabled the evaluation of diabetic retinopathy through an analysis of retinal images that gave early warning of disease progression. By applying the industry standard, a five-part scoring system, each individual patient was evaluated for degree of risk, and the results were then compared with the physician's original diagnosis.

Diabetic retinopathy is a condition suffered by diabetics that causes progressive damage to the retina. There is a classification system used dividing the retina into quadrants and counting hemorrhages, exudates, neovascularization, and other defects. Counting hemorrhages is an objective measurement but classification by visual analysis is subject to human error.

This gives rise to a lot of variance. The premise here is that a data science approach, driven by technology, could reduce/eliminate this variance by enabling the analysis of a large number of images to more accurately detect patterns that can inform a physician's diagnosis.

Most ophthalmologists see about 5,000 patients per year (or 10,000 eyes). In this case, the analysis included a half a million images to train a computer to recognize the disease stages by "looking at" a more expansive view of the image, and discounting errors. So, 50 years of clinical experience compressed into a single 24-hour period. The resulting benefits from an enhanced ability to analyze a vast number of images using software greatly improve the quality of diagnosis. These benefits can be extended to other use cases in the field of radiology in general.



The Data Science Principles and Approach to the Problem

Although human beings learn over time, gathering knowledge and applying it, a computer retains everything that is learned and continues to evolve at a staggering rate.

From a data science perspective, the approach was to treat this as a multi-class classification problem. In this case there were 5 classes, 4 disease states and no disease state. The computer was trained to identify disease states based on the images. The nature of the problem required the team to create a neural network algorithm. There was a small set of images that came to us originally (35k images). Neural network algorithms need large datasets to expand the opportunity to find issues, hence the data sets were expanded synthetically to train the computer to see an ``infinite" number of things. The value of using this approach was to enable to the physician to benefit from computerized backup with the ability to see more based on a better process and greater history.

Data and Computing Challenges

All healthcare data is "dirty" from an analytic standpoint.

When dealing with images, there are often defects on the images resulting from poor image quality that may erroneously appear to indicate disease. The software learns to discriminate between a defect in the image (poor quality, incorrect position such as left and right eye), and actual disease. This provides more context to the provider to arrive at a final diagnosis.

The volume of the total dataset required an appropriate computing infrastructure. The Graphics Processing Unit (GPU) infrastructure enabled the processing of the images in 24 hours to train the model. On a regular CPU, it would have taken 2 weeks or more. The goal was to have enough data to minimize the outliers and train the computer to read the image correctly.

Data is everywhere, images are no different than 0s and 1s. But many data scientists don't touch imaging because it's layered and has a spatial aspect, it's not flat and not labeled. A computer reads this kind of data differently and can retain the memory of what is being viewed.

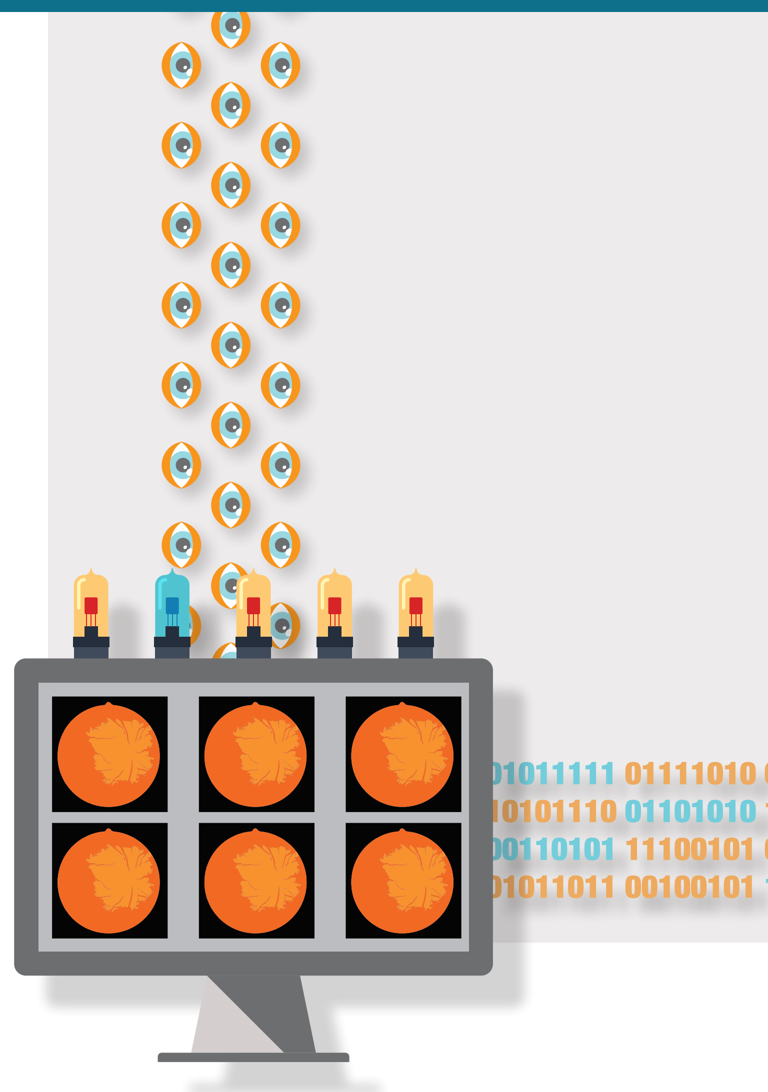
How We Executed the Project and What We Learned from the Process Conclusions

Several conclusions and business benefits can be derived from this unique project:

1. Software developed from this use case could be extended and applied to imaging software (along with pre-authorization software for example). So in a real-time context, the provider could actually use this to validate and ensure that their impression is accurate. It thus creates a real opportunity to enhance quality so that treatment matches diagnosis.
2. In rural settings, physicians can read images with software and obtain “second opinions” without requiring patients to go anywhere else.
3. This can be extended to high-volume radiology use cases such mammography that typically requires a dedicated radiologic resource. Most radiologists read everything (broken bones to cancer scans) with a computer backup providing expertise that a single radiologist can’t gather in a lifetime, physician training is enhanced and mistakes minimized. Of course, a physician review would still be a necessity.

The feedback from, Daxx Dunn, O.D., a provider with years in the field provides further insight:

“This concept could provide a more consistent means for doctors to confidently classify various stages of diabetic retinopathy. It could also be a valuable tool for students as well as established clinicians in giving them an instant, objective second opinion. Beyond the benefit to the doctors, there is the potential added benefit to the patients (time, money, and peace of mind) and cost saving benefits to the healthcare system (less money spent on facilities, tests, multiple visits for the same diagnoses, etc.)”



The use of unstructured data is a new frontier in healthcare data analytics. The analysis of images, in particular, requires specialized skills and advanced software. Typically, the computing infrastructure has to be more robust than in standard environments. Since images form a significant part of patient medical history, it is important to explore this frontier and expand it to other use cases such as radiology departments.

The ability to rely on a computer to process large volumes of images and approve the results is a game changer in precision medicine. The resulting ability to scale would be tremendous and the accuracy would certainly be improved as well. Finally, the cost savings could also be significant.

Ultimately, the benefits will be in the form of early detection and treatment which will result in improved quality of care and lower costs of care for patient populations.

Authors' Bios

Paddy Padmanabhan

Paddy Padmanabhan is an experienced and accomplished business leader & entrepreneur with extensive experience in Technology and Analytics in the Healthcare sector. Paddy is CEO of Damo Consulting Inc, (www.damoconsulting.net) a management consulting firm focused on Global Sourcing Advisory, Analytics Consulting, and Digital Content Solutions for healthcare enterprises and technology providers.

Prior to founding Damo Consulting, he was a part of Accenture's Healthcare practice. He has also been in two silicon valley start-ups focused on healthcare analytics. Paddy is a frequent writer and speaker on information technology in healthcare.

You can follow him @paddypadmanabha.

Damian Mingle

Damian Mingle is Chief Data Scientist for WPC Healthcare, a premier provider of cloud-based operational, financial, and clinical analytic solutions. In this role, Mingle manages a team of experts transforming data into meaningful strategic insights and offers hospitals systems, payers, and the HIT vendors descriptive analytics, exploratory data analysis, inferential analytics, predictive analytics, and prescriptive analytics.

Prior to WPC Mingle held positions with Hospital Corporation of America (HCA), Coventry Healthcare, Morgan Stanley. He is ranked in the top 1% globally as a data scientist through regular competitions to solve intricate data-related problems for organizations like the Mayo Clinic, Merck, Practice Fusion, Pfizer, and the California Healthcare Foundation.

©2014, WPC Healthcare

WPC HEALTHCARE wisdom.performance.credibility
1802 Williamson Court, Brentwood, TN 37027 (615) 913-8850 www.wpchealthcare.com