

Medical image analysis on the frontline of clinical research

by Steve Taylor BSc, MSc Co-founder Zegami, Co-Lead Centre of Computational Biology, MRC WIMM

Introduction

Expenditure on big data analytics generally in healthcare is estimated to see a compound annual growth rate of over 19% between now and 2027. By then, the sector could be worth over \$95 billion.

There are a number of factors fuelling this growth including the increased adoption of data-driven decision making, the emergence of cloud-based business intelligence solutions for the industry, an increased focus on reducing costs and the emergence of mobile-based healthcare business intelligence solutions.

Specifically, high throughput image analysis (working with large image datasets or frames from video) is a rapidly growing application and its success is largely due to significant advancements in machine learning as well as the ubiquity of affordable imaging platforms.

The amount of clinical data alone that has been produced between 2013 and today has increased by over 1,400% (2) and data visualisation tools are critical to managing this, reducing costs in the medical sector and ultimately improving care and treatments.

We predict that the medical profession will be one of the biggest adopters of image analysis and data visualisation tools, and we are already working on several projects in this sector.

They include:

- Gene therapy: Collaborating with MRC Weatherall Institute for Molecular Medicines (MRC WIMM) Centre for Computational Biology to help clean its data and assist with the training of its machine learning models, specifically around developing a better understanding of which proteins in genes bind, and where they do this in the genome.
- Oesophageal cancer: After winning Cancer Research UK's Early Cancer Detection Sandpit Challenge in 2019, we are currently working with Professor Barbara Braden (Oxford University), Dr Xiohang Gao (Middlesex University) and Dr Wei Pan (Herriot Watt University) on a range of cancer projects. One such project is linked to Oesophageal cancers, most of which are detected by endoscopy when they have reached an advanced stage and treatment is less effective and patient prognosis is poor.
- Cardiovascular diseases: Working with Oxford University researchers to help improve its diagnosis of cardiovascular diseases such as coronary heart and aortic diseases. It will do this by applying data visualisation techniques to echocardiogram images and videos, making it easier to categorise them and identify trends and patterns.

CASE STUDY #1

Accelerating research into cardiovascular disease

Accelerating research into cardiovascular disease

Heart and circulatory diseases cause more than a quarter of all deaths in the UK, and some 7.4 million people are living with cardiovascular conditions. Researchers at the Oxford Cardiovascular Clinical Research Facility are using Zegami to accelerate research into cardiovascular disease.

The team is applying data visualisation techniques developed by Zegami to hundreds of thousands of images of the heart, making it easier to categorise them and identify new trends and patterns.

Oxford Research Echocardiography Core Laboratory supports and manages some of the largest cardiovascular trials and studies in the world, involving multiple different hospitals and countries, and millions of images. Head of the Laboratory, Cristiana Monteiro, believes that data visualisation tools address many of the problems they face when trying to navigate and make sense of these large complex imaging datasets.

Specifically, cardiovascular imaging yields large data sets, both for image analysis as well as incorporation with other clinical data. Most medical images today are stored in digital formats as pixels or voxels describing a small area or volume. Millions of pixels or voxels form one image, and data visualisation techniques can help analyse this more effectively.

Medical researchers are studying the cardiovascular health of thousands of different patients at the same time, the insights generated through data visualisation could support the development of machine learning algorithms that identify and predict heart disease in patients.

CCRF	61 out of 638 items (9.56%)	T Filters Applied: 1 Q. Reset Zoom Level					Exit collection 🕀
٩	View Types						
т			mm	Mr.		and the star for the first of the star	
<u>100</u>	ttii Scatter Table	The second secon	Li Mili Mili Mili Mili Mili Mili Mili Mi		Net Man John Street Course Day		<u>ه</u> د
	Select column for colour	<u>ę`</u>					
	unit v Calour control	* *** *** ** *** ***	-		***	****	ether we there
	o 100% unit IEmptyl Parameters						, Č
	Include empty values Column						
	ID V Sot Ordered by frequency (count) V Target number of bins 1						Ê
۵	Row V						S
e G							A
٠				**	ES TPH504		

Looking for outliers in echocardigrams to predict disease.

CASE STUDY #2

Speeding up cancer diagnosis 11

Speeding up cancer diagnosis

Much has been said and written about delays in cancer diagnosis and treatment in the UK. However, with a growing and aging population, they are under great and growing pressure. Machine learning and data visualisation has a huge role to play in terms of helping doctors make diagnoses earlier.

After winning Cancer Research UK's Early Cancer Detection Sandpit Challenge in 2019, we are currently collaborating with Professor Barbara Braden (Oxford University), Dr Xiohang Gao (Middlesex University) and Dr Wei Pan (Herriot Watt University) on a cancer project.

The oesophageal cancer, which is detected by endoscopy when it has reached an advanced stage and treatment is less effective and patient prognosis is poor. Currently, the miss rate for early oesophageal cancer during endoscopy is 25% - the cancer lesions are very hard to spot and can be life threatening if overlooked. Detecting early cancer on the other hand, offers a significantly higher chance of cure, as the tumour can be easily removed during an endoscopic examination.

We are speeding up the training of machine learning models using anonymised video footage from endoscopy. Our aim is to develop a real-time computer system which can run during the procedure (using an endoscope with inbuilt algorithm in it), highlighting areas of concern by overlaying markers to guide biopsy taking. Early detection is paramount, because the cancer is curable as long as it is defined to the mucosa whilst the 5 year survival is less than 20% in more advanced stages.

This not only helps speed up the process of detection of cancerous lesions faster, with more accuracy, and less discomfort for the patients, so reduces the chances of missing early signs of cancer, saving both time and lives.



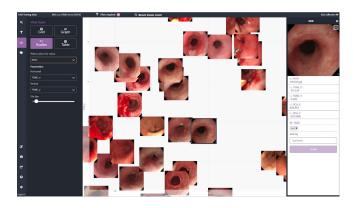


Image similarity comparison of oesophageal cancer video frames to find cancer lesions to train the ML model

CASE STUDY #3

Creating personalised medicines with gene therapy

Creating personalised medicines with gene therapy

The MRC Weatherall Institute of Molecular Medicine (MRC WIMM) at the University of Oxford was founded in 1989 to foster research in molecular and cell biology, with the aim of improving human health. Through its excellent basic and applied research, it has become a leading centre for translational medicine. Its research has resulted in improved understanding, diagnosis and treatment of a wide range of human diseases.

Our partnership with the MRC WIMM Centre for Computational Biology is designed to help clean its data and assist with the training of its machine learning models. This is specifically focussed on developing a better understanding of which proteins in genes bind, and where they do this in the genome. All of this is key to the development of personalized gene therapy and medicine.

We're supporting scientists in this renowned institute by speeding up the process of training machine learning tools to find the locations of important proteins that bind to the human genome that turn genes on and off. When these genes are incorrectly activated/deactivated they can cause disease, so accurately finding their control mechanisms is vital. Existing software often gets this wrong but using machine learning the Taylor and Hughes group have trained computers to 'see' these signals to get more accurate catalogues of their positions.

The MRC Weatherall Institute of Molecular Medicine at Oxford University have already achieved significant gains through the partnership. Jim Hughes, Professor of Gene Regulation states: "Our ability to sequence and reconstruct entire genomes has transformed our approach to research and medicine. We can now investigate, on the scale of the whole genome, how and in what situations, parts of that blueprint are used. This is enhancing our understanding of how our genome works in health and disease, but it also means we are generating huge amounts of data." This an exciting opportunity for the Institutes also presents a problem in how best to ensure these data are clean, and accurate so that they can train machine learning methods effectively and efficiently to create new insights.

They use Zegami to visualize, sort, filter and label vast amounts of biological data for use in training machine learning models, solving a key challenge in this field. Zegami also allows the easy publication of all the data for scientists to understand how the models were created to help address the machine learning black box problem.

ViewSel	::WIMM Day - H3K27ac (ENCODE) - MACS2	peak called peak analysi	s 1854 out of 1854 item	s (100%) 🕇 Filters App	lied: 🔘 🔍 Reset Zoo	m Level	Exit	collection 🔂
٩	Search		المحاقيقا أأفق والمعاليين	And a station of the station of the				فاستند أيل
T	Available Filters Find filters	posti kost		And	termine for the first termine for the	Data Para Para Para Para Para Para Para P	20060 N Lone Lone	para press pr
ш	ш id	and the second			-	and the second second		1. A
	A chromosome	Maldan 217	Polenie Graf prone gra	2004WD Got printen 1791	in hate hate	near Interest Proper Proper	w Passe Passe	200wp 0.7
		and Mult	يتعالى اللاطر فالعيمان	. 11	and a state of the state			In cash of
			August and a second and a secon	205/00/10	200400 N	Econto m	200000 N prose prove prove	200MB N
	A Image Name	he are					non have have have	Partie La
	凹 H3K4me1 Score	and the second						
	■ Noise Score	pi224a 25=	Score of A	house hot de-	These for the form	Jooned Crit	2010400 04.7 pelitien (p.mbe) 1 497-	200482-0-7
	III ATAC Score		Landin . An and a	وبالتعالى بلغ أستحد والمخالف وال		<u>a de la composition</u>		and a subscription of the
	H3K4me3 Score	Calles at Minks						
		11400 Priess Pri	Press (1)	· Prove prove pr	20000 P(10000 P(10000 P(10000 P(10000	Printe Printe Printe	Parte Parte Parte	POORE N HORSE
	Ш H3K27ac Score				and the second		and the second second	1
	A Genes(Ref Genes)	(*****	2008EE 0/7 177800 177800 12/2000 1 6. 1737-	2004020 007 101 108 102010 17-	Accesso out 120000 The Tables	SCOMED OV1 SIZE P	200800 0.4 He P2004 05.9 15 17-	COOMED DUT
		N		Lade attles I at .	to sales be as use	A at adult is a		4.11
۵	₩ PCA2	and the second se		Anger	A REAL PROPERTY AND A REAL		and allow all a second second	Angen
~	⊯ tSNE1	2400 - QUEZ 3400	200000 0 galana galana galana galana 270	THE PARTY PA	and transformer processo proceso processo processo processo processo processo proces	40000 N 40000 00530	- the trans the trans to the trans to	200480 N galovas
Ċ	쁘 tSNE2			1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 -	and the state		a de catteries	and the
0		- finance	200460 OVI prime prime	20440 COJ STRAND COJ	SCOME CO.	ECOMO DAT DOMAS CONTRA	200mmD 644 presse granue presse	Zoowati out gateriae
			1	in the	المر بالعلم بال	4	and delay	A del
٠		Harris with	Appendia and Ampendia Appendia	And a second sec	And the second s	and the second s	A DALA MANAGEMENT	The second se
zegami		n <i>q</i> asani banan	100400-M 425M0 623M0	STOWED IN FARMER FARMER FARMER	200460 N 470318 472848 473034	SCONED IN SLOWLE BLOOM BLOOM	Buanto Pranto August August Paul	200WED N JAO NO

Looking for characteristic peak shape to find where proteins bind to the genome

About the author



Steve Taylor

Founder and CSO of Zegami. Steve's personal mission is to disrupt the analysis of complex data to improve the world. He wants to put powerful but easy to use analysis methods into the hands of everyone - not just computer experts - by using innovative visualisation and interaction technologies. As well as CSO at Zegami, he heads the Analysis, Visualisation and Informatics research group at Oxford working on understanding big data in genomics and health care.

About Zegami

Zegami combines advanced analysis tools with a unique visualisation interface and allows you to rapidly categorise, label and clean large image datasets, invaluable for many applications including training machine learning models.

Zegami was launched out of Oxford University in 2016, and is being used by academics and commercial leaders across the world to gain a competitive edge from their data.



Our Clients









Sign up <u>here</u> for your free trial of Zegami.



<u> </u>zegami

Oxford Sciences Innovation - Grassroots 46 Woodstock Road Oxford OX2 6HT United Kingdom

Email : info@zegami.com

Website: https://zegami.com