ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB411

QUANTIFYING DNA DAMAGE IN COMET ASSAY IMAGES USING NEURAL NETWORKS

S. Dhinsey, T. Greenshaw, C. P. Welsch[1]
Department of Physics, University of Liverpool, U.K.
J. L. Parsons, Cancer Research Centre,
Department of Molecular and Clinical Cancer Medicine, University of Liverpool, U.K.
[1]also at The Cockcroft Institute, Daresbury, Warrington, U.K.

Abstract

Proton therapy for cancer treatment is a rapidly growing field and increasing evidence suggests it induces more complex DNA damage than photon therapy. Accurate comparison between the two treatments requires quantification of the DNA damage the cause, which can be assessed using the Comet Assay. The program outlined here is based on neural network architecture and aims to speed up analysis of Comet Assay images and provide accurate, quantifiable assessment of the DNA damage levels apparent in individual cells. The Comet Assay is an established technique in which DNA fragments are spread out under the influence of an electric field, producing a comet-like object. The elongation and intensity of the comet tail (consisting of DNA fragments) indicate the level of damage incurred. Many methods to measure this damage exist, using a variety of algorithms. However, these can be time consuming, so often only a small fraction of the comets available in an image are analysed. The automatic analysis presented in this contribution aims to improve this. To supplement the training and testing of the network, a Monte Carlo model will also be presented to create simulated comet assay images.

INTRODUCTION

Cancer treatments rely on causing lethal damage to the DNA in tumour cells in order to kill them and halt tumour growth. Proton therapy has advantages over photon therapy in this respect due to the higher linear energy transfer of heavy charged particles, in particular at the end of the proton track in the Bragg peak distal end [1,2]. This results in more complex DNA damage which is more difficult for the cell to repair, and is directly linked to killing tumour cells [3]. A deeper understanding of the biological outcome of proton therapy is therefore needed [4]. Currently, an average relative biological effectiveness (RBE) of 1.1 is used as a dose conversion from photons to protons, but research suggests this value needs tailoring for the specific energies used and the specific biology of the tumour [5].

The comet assay is a technique widely used to determine the degree of radiation-induced DNA damage. The methods used to analyse comet assay images vary considerably. Here, progress is described towards the implementation of artificial intelligence through the Mask-RCNN architecture to allow robust, efficient and versatile analysis of comet assay images.

THE COMET ASSAY

The Single Cell Gel Electrophoresis (or Comet) assay involves first embedding irradiated cells into a gel-like substance called agarose which has a mesh structure with many pores. Lysis is performed with an high salt and detergent solution, breaking down the cellular and nuclear membranes so that the DNA is left within the hole in the agarose originally occupied by the cell. The agarose is then flooded with a neutral or alkaline solution, to identify double or single strand breaks, respectively. Electrophoresis is performed by applying an electric field across the agarose; the DNA migrates along the field direction due to the negatively charged backbone of the DNA strands. This diffusion of DNA fragments is what causes the cells to form comet-like structures, as shown in Fig.1, and gives the assay its name. Shorter fragments are able to move more easily through the agarose and therefore migrate further than longer fragments. The DNA is then stained with a fluorescent marker and imaged. A detailed description of the assay used to produce the images analysed in this paper can be found in [6].

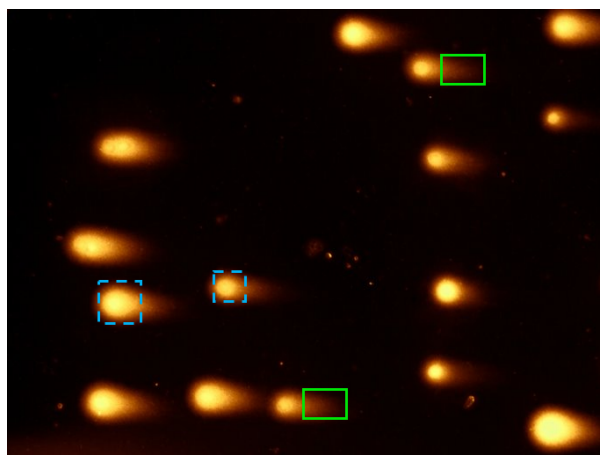

Figure 1: An example comet assay image showing the DNA damage of cells following irradiation. Comet head examples are highlighted in dashed cyan whilst example tail regions are shown in solid green.

The most commonly reported measures of DNA damage following a comet assay are tail DNA and tail length [7]. Tail length refers to the length of the region known as the comet tail. This is the area formed of DNA fragments that have migrated away from the initial cell nucleus or "head", Fig.1, where all the DNA was situated prior to electrophoresis. Tail

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB411

DNA is the sum of the intensity of the tail pixels as a proportion of the whole comet's intensity, usually expressed as a percentage. The methods used to calculate these measures vary and are often not described in publications.

IMAGE ANALYSIS

Since the comet assay's first implementation in 1984 by Ostling and Johanson [8], the methods for quantifying DNA damage have evolved significantly. Initially, a numbered ranking system of 1-5 was used [9] to categorise the degree of damage seen in assay images, which of course introduces high variability and is dependent on an individual's judgement, thus introducing potential bias. The ranking also fails to differentiate between images that have been grouped into the same category but display different degrees of damage. The use of computational analysis removes some of the human error and bias of the ranking method, but some software still requires a lot of user input. An example of such software is Komet 7 [10], which requires the user to select where in an image comet bodies appear and then also define the corresponding comet head regions. Bias may result from selecting only favourable comets and, as manual selection of the images is time consuming, often only a small proportion of the comets in an image are analysed. The program described here aims to minimise human intervention and create a reproducible and standardised means of performing quantitative analysis of comet assay images in a time-efficient manner, whilst maximising the data available.

MASK RCNN

The first aspect investigated was automating comet identification from the image background. As image quality, intensity and resolution vary significantly between experiments, a comet identification process based on shapes and structures rather than intensity range was explored, using artificial intelligence. The Mask-RCNN (Mask Region based Convolutional Neural Network) architecture is comprised of a regional convolutional neural network and fully convolutional network that performs instance segmentation [11]. It works by combining bounding box identification with pixel-wise classification in order to produce segmented masks of the detected object. The learning process is based on the assay images and their corresponding annotations. The annotations are a series of x/y coordinates describing the vertices of polygons that encompass the comet bodies, made using the VGG Image Annotator software [12]. In order to overcome dataset size limitations and the computing resources required, Microsoft's COCO model was used to implement transfer learning [13]. COCO is an extensive model, trained on over 200,000 images split into 80 object categories. Transfer learning enables the re-use of an existing model that has already learned a set of features; training of the final few layers of the model is all that is required to learn features of the new data.

Inference is applied to test the model on new images and produce segmentation masks of identified comets, Fig. 2. Parameters within the training configuration have been set to produce masks of confidence 90% and above, giving a reliable and accurate outcome for each mask. These are produced by taking the highest probability regions of interest the network has detected as belonging to the segmentation class. Measurements of tail DNA and tail length are performed on the detected comet masks to quantify the DNA damage presented in that cell. Currently, the radius of the comet head is defined to be the distance from the central head pixel (assumed to be the brightest region of the comet) to the left-most comet pixel. All pixels within the circle defined using this centre and radius are assigned as head pixels and the remaining are tail pixels. Measurements of comet area, tail DNA and tail length are then exported as a csv file for each comet found in each analysed image alongside corresponding plots, allowing for any irregularities in the model to be identified. However, due to a lack of data there have been limits on testing the model and its full capabilities remain currently unrealised.

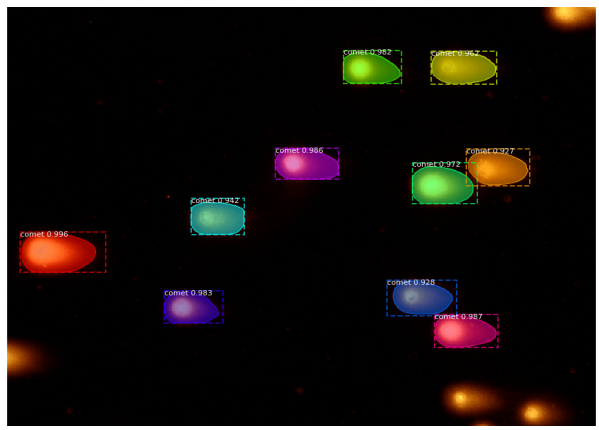

Figure 2: A comet assay image overlaid with the bounding boxes and segmented masks the model has defined.

MONTE CARLO IMAGES

To address the lack of data required to train an independent, extensive model from scratch, a Monte Carlo (MC) model is in development to produce simulated comet assay images. This will provide more images for the development, training and testing of the mask segmentation model. The MC model performs DNA breaks and drifts the broken fragments according to their length. To reduce the computing power and time required to produce the MC images, the DNA is simulated as being comprised of a total ~7.6×10^5 segments instead of the actual number of base pairs in the human genome which is ~3.2×10^9 [14]. This number of segments dictates the number of break sites and the smallest possible fragment size. Parameters such as the probability of a break, p_{break} and the drift extent of fragments that form the tail have been tuned to give images that reasonably match those obtained experimentally, Fig. 3. The average number of breaks per cell is found by multiplying p_{break} by the number of units comprising a strand. Further analysis of

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB411

the simulated images is required to determine if they can be used alongside/in place of experimental images for training comet detection models.

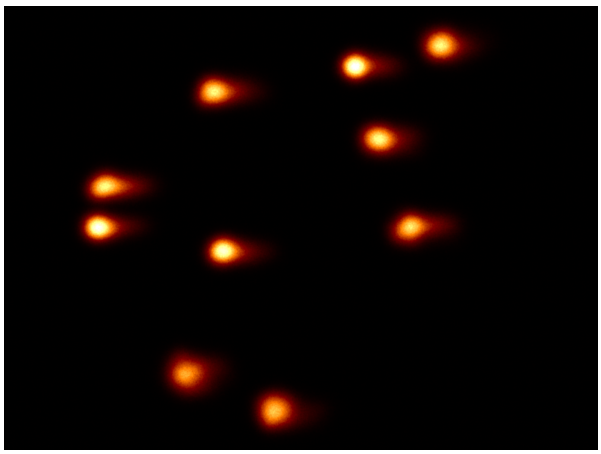

Figure 3: An example Monte Carlo image produced from the simulation of breaking and drifting DNA strands. The strand break probability for this image was set at 0.65.

Initial analysis of MC images of varying p_{break} parameter has been performed to test if the broken fragments drift away from the comet head as expected. Figures 4 and 5 show an overall monotonic relationship between p_{break} and asymmetry/tail DNA. To determine the asymmetry, each comet is split into its left and right pixels, the centre determined from the head centre. The assumption of this measure is that cells with little/no damage will appear spherical and thus, the head centre is also the comet centre. However damaged cells will have lost their spherical structure and gain an elongated comet tail. This creates an imbalance between left and right pixels in favour of the tail pixels on the right. As expected, as p_{break} is increased the asymmetry also increases. At $p_{break} = 0.8$ we see an unexpected decline in asymmetry. This is due to the images produced from this level of damage being exceptionally low in intensity and long in tail length. These type of comets are not seen experimentally as cells are not irradiated for long enough to undergo such high levels of damage and still survive. Similarly, Fig. 5 shows a decrease in tail DNA at $p_{break} = 0.8$. This again is due to the highly damaged comets that are not identified by the neural network, which detected fewer comets per image for the highest p_{break} value.

CONCLUSIONS

The analysis of comet assay images is very varied and a lot of methods introduce a high degree of human bias. The methodology outlined here aims to remove human bias and maximise the measurement output from assay images by incorporating neural network architecture that performs instance segmentation on images of damaged cells. A lack of data has been overcome by developing a Monte Carlo model of the comet assay process to create images similar to those taken experimentally. Initial analysis indicates that the MC

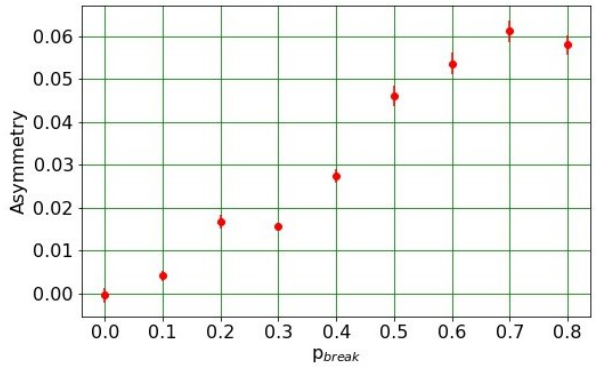

Figure 4: A plot of how comet asymmetry varies with p_{break} value.

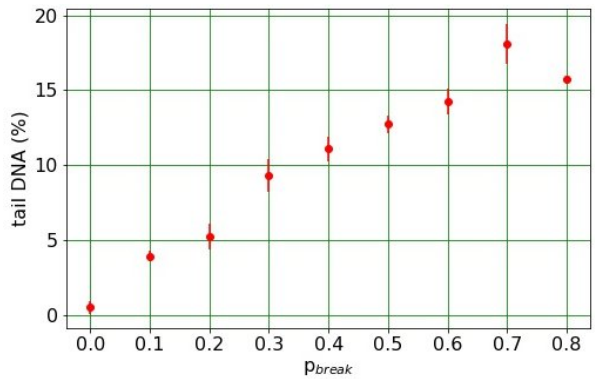

Figure 5: A plot of how tail DNA varies with p_{break} value.

model is performing as expected in terms of breaking and drifting DNA fragments. The next steps will be to investigate if neural networks trained from MC data are able to perform instance segmentation accurately enough on experimental data, allowing more thorough models to be trained and tested from a larger data set. The MC produced images will also act as a benchmark to calibrate the measurements of tail DNA and tail length.

ACKNOWLEDGEMENTS

This work was supported by the STFC Liverpool Centre for Doctoral Training on Data Intensive Science (LIV.DAT) under grant agreement ST/P006752/1.

REFERENCES

[1] W. D. Newhauser and R. Zhang, "The Physics of Proton Therapy", *Phys. Med. Biol.*, vol. 60, pp. R155-R209, 2015. doi:10.1088/0031-9155/60/8/R155

[2] E. T. Vitti and J. L. Parsons, "The Radiobiological Effects of Proton Beam Therapy: Impact on DNA Damage and Repair", *Cancers (Basel)*, vol. 11, no. 7, p. 946, Jul. 2019. doi:10.3390/cancers11070946

[3] R. J. Carter *et al.*, "Complex DNA Damage Induced by High Linear Energy Transfer Alpha-Particles and Protons Triggers a Specific Cellular DNA Damage Response", *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 100, no. 3, pp. 776-784, Nov. 2017. doi:10.1016/j.ijrobp.2017.11.012

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB411

[4] S. Girdhani, R. Sachs, and L. Hlatky, "Biological Effects of Proton Radiation: What We Know and Don't Know", *Radiat. Res.*, vol. 179, no. 3, pp. 257-272, Feb. 2013. doi:10.1667/RR2839.1

[5] J. Ödén, P. M. DeLuca Jr., and C. G. Orton, "The use of a constant RBE=1.1 for proton radiotherapy is no longer appropriate", *Med. Phys.*, vol. 45, no. 2, pp. 502-505, Feb. 2018. doi:10.1002/mp.12646

[6] C. M. Nickson and J. L. Parsons, "Monitoring regulation of DNA repair activities of cultured cells in-gel using the comet assay", *Front Genet.*, vol. 5, p. 232, Jul. 2014. doi:10.3389/fgene.2014.00232

[7] T. S. Kumaravel *et al.*, "Comet Assay measurements: a perspective". *Cell Biol. Toxicol.*, vol. 25, no. 1, pp. 53-64, Feb. 2009. doi:10.1007/s10565-007-9043-9

[8] O. Ostling and K.J. Johanson, "Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells", *Biochem. Biophys. Res. Commun.*, vol. 123, no. 1, pp. 291-298, Aug. 1984. doi:10.1016/0006-291x(84)90411-x

[9] A. R. Collins, "The comet assay for DNA damage and repair", *Mol. Biotechnol.*, vol. 26, p. 249, Mar. 2004. doi:10.1385/MB:26:3:249

[10] Komet 7, https://andor.oxinst.com/products/komet-software/komet-7

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN", Jan. 2018. arXiv:1703.06870v3

[12] VGG Software, https://www.robots.ox.ac.uk/~vgg/software/via/.

[13] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context", Feb. 2015. arXiv:1405.0312v3

[14] T. A. Brown, *Genomes*, Oxford, UK: Wiley-Liss, 2002.