

# A DEEP LEARNING MODEL FOR AUTOMATED QUANTIFICATION OF DNA REPAIR FOCI IN SOMATIC MAMMALIAN CELLS

*S. Shadmehri<sup>1,\*</sup>, T. Bezhanyan<sup>1,2</sup>, M. Bondarev<sup>2</sup>,  
O. I. Streltsova<sup>1,2</sup>, M. I. Zuev<sup>1</sup>, A. Chigasova<sup>3,4,5</sup>, A. Osipov<sup>3</sup>,  
N. Vorobyeva<sup>3,4</sup>, A. N. Osipov<sup>3,4</sup>*

<sup>1</sup> Joint Institute for Nuclear Research, Dubna, Russia

<sup>2</sup> Dubna State University, Dubna, Russia

<sup>3</sup> Semenov Federal Research Center for Chemical Physics, RAS, Moscow, Russia

<sup>4</sup> State Research Center — Burnasyan Federal Medical Biophysical Center,  
Federal Medical Biological Agency, Moscow, Russia

<sup>5</sup> Emanuel Institute for Biochemical Physics, RAS, Moscow, Russia

Double-strand breaks (DSBs) are the most lethal DNA damages induced by ionizing radiations. Here, we present a DSB quantification method based on the analysis of DNA repair foci (pATM and  $\gamma$ H2AX) in somatic mammalian cells as the most reliable biomarkers of DSBs. Human dermal fibroblasts and human mesenchymal stem cells were irradiated with 1, 2, and 5 Gy of X rays and their fluorescent immunocytochemistry microscopy images were captured at 24 h after the irradiation. The captured images were annotated in CVAT platform and the exported data were obtained in Yolo (You-Look-Only-Once) format. The deep learning approach consisted of two stages; a computer vision algorithm and a neural network were used to extract the cells from each image thus providing our features (cell images) and corresponding labels, then these data were split as the required train/test/validation set for the deep learning model based on a Yolo algorithm to count the number of the two mentioned foci per cell. The results of our model were compared with the biologists counting by DARFI software. The developed algorithm and training models were based on the ML/DL/HPC ecosystem of the HybriLIT Heterogeneous Computing Platform.

Двунитевые разрывы (ДР), вызванные действием ионизирующего излучения, являются наиболее опасными для жизнеспособности клеток повреждениями ДНК. В данной работе представлен метод количественной оценки ДР ДНК, основанный на анализе фокусов репарации ДНК (pATM и  $\gamma$ H2AX) в соматических клетках млекопитающих как наиболее надежных биомаркеров ДР ДНК. Дермальные фибробlastы человека и мезенхимальные стволовые клетки человека облучали рентгеновским излучением в дозах 1, 2 и 5 Гр и через 24 ч после облучения получали их флуоресцентные иммуноцитохимические микрофотографии. Полученные изображения были аннотированы на платформе CVAT

---

\* E-mail: shadmehri@jinr.ru

и экспорттировались в формате Yolo (You-Look-Only-Once). Для автоматизации подсчета количества фокусов был разработан алгоритм на базе моделей глубокого обучения, который состоит из двух этапов. Первый этап — это детекция клеток на изображении, для чего использовались методы компьютерного зрения и предобученная нейронная сеть. Второй этап — детекция фокусов на изображении каждой клетки, предварительно вырезанной из начального изображения, что осуществлялось с использованием обученной нейросетевой модели Yolo. Результаты разработанного алгоритма сравнивались с подсчетами биологов и с результатами, полученными с помощью программного пакета DARFI. Аннотирование данных, обучение моделей и разработка алгоритмов проводились на ресурсах экосистемы ML/DL/HPC гетерогенной платформы HybriLIT.

PACS: 84.35.+i; 87.53.-j