doi 10.18699/CRISPR-2025-77

Artificial intelligence for predicting chromosomal rearrangements in cancer genomes

P.E. Karitskaia*, Y.V. Vyatkin
Novosibirsk State University, Novosibirsk, Russia
** p.karitskaya@g.nsu.ru*

Key words: machine learning; deep learning; structural variations; chromothripsis; cancer diagnostics

Chromothripsis is a catastrophic genomic event involving massive chromosomal rearrangements, often linked to poor prognosis in cancer [1]. Detecting such events is clinically valuable, but current methods are either manual or data-specific [2, 3]. We aimed to develop predictive models for chromothripsis using classical machine learning and deep learning. Data from ChromothripsisDB yielded 19713 events after preprocessing [4].

Classical methods used aggregated features of structural variations. Logistic regression, SVM, random forest, and XGBoost were tested. SMOTE and class weighting addressed class imbalance (11.4% positive cases). XGBoost performed best (ROC AUC 0.93, PR AUC 0.78).

Deep learning employed a BiLSTM with attention, trained on sequences of 200 SV events, each described by 10 features [5]. Weighted sampling and combined Focal + F1 loss addressed imbalance. 10-fold stratified cross-validation showed BiLSTM outperformed classical models (ROC AUC 0.95, PR AUC 0.94).

These results confirm that deep learning models are more effective at capturing complex spatial and contextual patterns in structural variation data compared to aggregated feature-based methods. The proposed approach demonstrates strong potential for developing accurate and automated diagnostic tools for genome instability in cancer.

References

1. Stephens P.J., Greenman C.D., Fu B. et al. Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell*. 2011;144:27-40
2. Govind S.K., Zia A., Hennings-Yeomans P.H. et al. ShatterProof: operational detection and quantification of chromothripsis. *BMC Bioinformatics*. 2014;15:78. doi 10.1186/1471-2105-15-78
3. Yu J., Chen N., Zheng Z. et al. Chromothripsis detection with multiple myeloma patients based on deep graph learning. *Bioinformatics*. 2023;39(7):btad422. doi 10.1093/bioinformatics/btad422
4. Cortés-Ciriano I., Lee J.J.-K., Xi R. et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet*. 2020;52:331-341
5. Schuster M., Paliwal K.K. Bidirectional recurrent neural networks. *IEEE Trans Signal Process*. 1997;45:2673-2681