http://dx.doi.org/10.35630/2023/XIIsc4hlth/110

PROVIDING NOVEL DIAGNOSTIC TOOLS FOR PRIMARY PEDIATRIC AND ADULT CNS MALIGNANCIES WITH MACHINE LEARNING BASED WORKFLOWS

Vasilii Khammad[1,2], Natalia Kharchenko[1], Jose Javier Otero[2], Gadzhimurad Zapirov[1], Yolanda Cabello Izquierdo[2], Francisco Garagorry[2]

[1] Peoples' Friendship University of Russia, Moscow, Russia;
[2] James Cancer Hospital and Solove Research Institute, Columbus, Ohio, USA

INTRODUCTION

Primary Central Nervous System malignancies consist of various groups of tumors with a wide range of morphologies, molecular alterations, and clinical behaviors. Under the current World Health Organization (WHO) classification of CNS tumors, integration of histomorphology and molecular data is critical for diagnosis and directly impacts prognosis and treatment-related approaches. However, in Russia, US and worldwide scarce resources are available to perform all the required tests routinely.

Aim:

Our long-term goal was to improve and standardize testing and diagnoses for brain tumor patients worldwide by validating new diagnostic workflows with implementation of digital imaging, immunohistochemical tests, open-source computing platforms and machine learning algorithms. We focused on improving diagnostic capabilities for primary CNS tumors with a particular emphasis on resource-poor diagnostic scenarios.

METHODS

Our working group has reviewed all cases available at the Ohio State University's digital pathology archive from 2013 to present. Fully digital workflows in neuropathology at James Cancer Hospital enabled us to quickly obtain and review data on over 500 brain biopsies. Retrospective analysis was used as a ground truth upon which to validate our predictive model generated in activity. The cases were re-reviewed by board-certified neuropathologists and features annotated for insertion to the models generated in R Studio. We used modern, innovative approaches to augment our diagnostic workflow, including correlation of radiographic and neuropathological data to evaluate tumor heterogeneity and implementing deep learning models to augment histopathological interpretation in whole slide imaging.

RESULTS

We completed the large meta-analysis study by incorporating data from all primary brain cancers classified by the WHO Classification of CNS tumors. The data from this meta-analysis was successfully used to generate a population model of all primary CNS tumors (pediatric and adult cases). For each model, we chose machine learning algorithms that do not require strong data assumptions. Models trained with decision tree algorithms (randomForest, XGBoost and C5.0) showed high overall accuracy in predicting diagnoses of our models (93%, 74% and 87% respectively) even in the absence of IHC and molecular data. Importantly, despite the obvious value in further clinical decisions, the inclusion of all molecular data (BRAFV600E, FISH codeletion 1p19q, EGFR amplification, MGMT status and etc.) did little to contribute to distinguishing between different primary CNS malignancies in our study. All our generated models allowed to establish clinical diagnosis for primary brain lesions using just basic clinical data (age, gender, tumor location), ki67 levels and distinct features of histological architecture with high accuracy.

CONCLUSIONS

Molecular data managed to slightly improve the accuracy of our prediction models only by approximately 5–6% compared to data on ki67 alone (37%), suggesting that the proliferation markers should be considered as a more reliable tool for accurate staging and morphological evaluation of CNS tumors. Each of the described above algorithms can be implemented into clinical practice and serve as a reliable second opinion in resource-strained settings.

Keywords:

primary CNS malignancies, neuropathology, ki67 level, machine learning algorithms, diagnostic workflows