Baiming Zou[1], James Xenakis[2], Meisheng Xiao1, Apoena Ribeiro[1], Kimon Divaris[1], Di Wu[1], Fei Zou[1]

[1]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2]Harvard University, Cambridge, MA, USA

A Deep Learning Feature Importance Test Framework for Integrating Informative High-dimensional Biomarkers to Improve Disease Outcome Prediction

Many human diseases result from a complex interplay of behavioral, clinical, and molecular factors. Integrating low-dimensional behavioral and clinical features with high-dimensional molecular profiles can significantly improve disease outcome prediction and diagnosis. However, while some biomarkers are crucial, many lack informative value. To enhance prediction accuracy and understand disease mechanisms, it is essential to integrate relevant features and identify key biomarkers, separating meaningful data from noise and modeling complex associations. To address these challenges, we introduce the high-dimensional feature importance test (HdFIT) framework for machine learning models. HdFIT includes a feature screening step for dimension reduction and leverages machine learning to model complex associations between biomarkers and disease outcomes. It robustly evaluates each feature's impact. Extensive Monte Carlo experiments and a real microbiome study demonstrate HdFIT's efficacy, especially when integrated with advanced models like deep neural networks (DNN), termed HdFIT-DNN. Our framework shows significant improvements in identifying crucial features and enhancing prediction accuracy, even in high-dimensional settings. This approach offers a promising avenue for advancing precision medicine, understanding complex disease mechanisms, and ultimately improving patient outcomes through more accurate diagnostics and targeted therapies.

Keywords

Complex association, Dimension reduction, Interpretable and scalable predictive modeling, Non-parametric feature selection, Stable deep neural network

Biography

Dr. Baiming Zou is an Associate Professor in the Department of Biostatistics at the University of North Carolina at Chapel Hill. With extensive experience in statistical methodologies and their applications in biomedical and public health research, Dr. Zou has contributed significantly to advancing quantitative approaches that support evidence-based decision-making in health sciences. His work spans biostatistical modeling, clinical trial analysis, and the development of innovative techniques to address complex data challenges in epidemiology and medical research.

doi.org/10.51219/URForum.2025.Baiming-Zou ISBN: 978-1-917892-34-6