Title: XAI-AttBiLSTM: an explainable two-stage feature selection framework with Attention-based BiLSTM for gynecological cancer risk module identification

Authur names: Mst. Mostary Khatun, Francis Rudra D Cruze · Mst. Amina Khatun, Md. Faruk Hosen, Muhammad Shahin Uddin , Monir Morshed1

Abastact: Gynecological cancers (GC) rank among the most common cancers affecting women and are leading causes of cancerrelated morbidity and mortality worldwide. Despite notable clinical progress, effective early diagnostic biomarkers and targeted therapeutic options are still lacking. In this study, we propose a comprehensive machine learning and bioinformatics framework to identify potential biomarkers and therapeutic compounds for GC. Initially, four gene expression microarray datasets were used to extract a preliminary set of differentially expressed genes (DEGs). We employed a two-step feature selection approach: first, Boruta was applied, followed by LIME as the final feature selection technique. The final gene subset was then fed into our Attention-based Bidirectional Long Short-Term Memory (AttBiLSTM) model, which demonstrated better predictive performance compared to other classifiers. Thirty common genes, resulting from the intersection of DEGs and the gene subset derived from machine learning, were identified. Gene Ontology (GO) and KEGG pathway enrichment analysis revealed enrichment in biological processes associated with GC. The protein-protein interactions (PPI) network identified 10 hub genes: TK1, CRMP1, KIF5A, TUBB2A, METTL20, TSC2, RHEB, A2M, UBC, and RBM48. The transcription factor (TF) regulatory network added the TF-gene and TF-miRNA interactions and identified five core regulators that had regulatory network associations in the PPI. Molecular docking also identified that five of the candidate targets had strong binding affinity with four GC drugs. These findings indicate that the identified hub genes show potential as biomarkers for early detection and prognosis, while the associated compounds may serve as promising candidates for the therapeutic development of GC

Keywords: Gynecological cancers · AttBiLSTM · Boruta · LIME · Biomarkers · PPI network · Hub genes · Gene regulatory network · Molecular docking

https://link.springer.com/article/10.1007/s13721-026-00739-x