Predicting DNA Methylation Patterns and Early Cancer Diagnosis Using Deep Learning Models

深層学習モデルを用いたDNAメチル化パターンの予測とがんの早期診断

Abstract

DNA methylation is an important epigenetic modification that controls gene expression and is involved in various biological processes, such as cancer development. Abnormal DNA methylation patterns often appear in the early stages of cancer, making them useful for early diagnosis. This research aims to use deep learning models to analyze DNA methylation data to predict early onset of different cancer types, exploring its application in early cancer diagnosis. By designing and optimizing deep learning architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer models, this study aims to accurately predict DNA methylation patterns linked with early cancer onset. The research will also focus on identifying specific DNA methylation sites as reliable biomarkers, enhancing model interpretability, and developing benchmark datasets for systematic model evaluation. The outcomes are expected to improve the accuracy and utility of deep learning models in early cancer detection, significantly impacting clinical practices and contributing to the advancement of precision medicine.

概要：

DNAメチル化は、遺伝子のスケジュールを制御する重要なエピジェネティック修飾であり、がんの発生を含むさまざまな生物学的プロセスにおいて重要な役割を果たしています。その異常は通常、癌の初期段階で見られ、がんの早期発見のマーカーとなっています。本研究の目的は、ディープラーニングを用いてDNAメチル化データを分析し、さまざまな種類のがんの早期発症を予測することで、早期がん発見への応用を考察することです。畳み込みニューラルネットワーク（CNN）、リカレントニューラルネットワーク（RNN）、トランスフォーマーモデルなどのディープラーニングアーキテクチャの設計と改善によって、本研究では、がんの早期発症に関連するDNAメチル化を正確に予測することを目指しています。また、信頼できるバイオマーカーとして、特定のDNAメチル化部位を鑑別し、モデルの解釈可能性を高め、モデルの系統的評価のためのベンチマークデータセットを開発します。最後に、研究の成果については、早期がん検出のためのディープラーニングモデルの精度と有用性を向上させ、臨床診療に大きな影響を与え、精密医療の発展を促進することが期待されています。

Research Background

DNA methylation is a key epigenetic modification that controls gene expression and is crucial in various biological processes, including cancer development. Abnormal DNA methylation patterns often appear in early cancer stages, making them potential markers for early diagnosis. Deep learning models are promising for analyzing complex biological data due to their ability to capture intricate patterns and relationships within the data. This research proposal aims to investigate the use of deep learning models to analyze DNA methylation data for predicting the early onset of different cancer types, thereby exploring its application in early cancer diagnosis.

Research Objectives

The primary objectives of this research are to design and optimize deep learning models capable of accurately predicting DNA methylation patterns associated with early cancer onset. I plan to assess the performance of various deep learning architectures, such as CNNs, RNNs, and Transformer models, in predicting early cancer diagnosis. Another objective is to identify specific DNA methylation sites that are highly predictive of early cancer stages, aiding in the understanding of cancer epigenetics. Additionally, my research seeks to develop methods that improve the interpretability of the deep learning models, ensuring that the predictions can be understood and trusted by clinicians and researchers. Lastly, I intend to create benchmark datasets and systematically compare the performance of various deep learning models in this application.

Previous Research

MethylNet: An Automated and Modular Deep Learning Approach for DNA Methylation Analysis: This study presents MethylNet, a framework designed to encode DNA methylation features using deep learning techniques. MethylNet shows superior performance in clustering DNA methylation patterns and has been successfully applied to predict various attributes, such as age estimation, cellular proportion estimation, and disease classification, making it a versatile tool for DNA methylation-based analysis [1].

Integrated Deep Learning Model for Predicting DNA Methylation and Tumor Types from Histopathology in Central Nervous System Tumors: This paper introduces DEPLOY, a model that predicts DNA methylation beta values from histopathology images and classifies CNS tumors. DEPLOY integrates three components: direct classification from histopathology images, indirect prediction of DNA methylation beta values, and classification from patient demographics. The study demonstrates DEPLOY's potential in improving the diagnostic accuracy of CNS tumors using inferred methylation data [2].

DNA Methylation Markers for Pan-Cancer Prediction by Deep Learning: This research identifies DNA methylation markers capable of diagnosing multiple cancer types simultaneously. Using deep learning models, the study achieved high sensitivity and specificity in identifying cancer markers, which can be applied in liquid biopsies for early cancer detection. The model demonstrated strong performance across various cancer types, showcasing the potential of DNA methylation markers in pan-cancer diagnostics [3].

These studies form a solid base for applying deep learning models to predict DNA methylation patterns and improve early cancer diagnosis. Integrating multimodal data with advanced machine learning can greatly boost the accuracy and robustness of these predictive models.

Research Methods

For data collection and preprocessing, I will use publicly available DNA methylation datasets from resources such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Preprocessing steps like normalization, feature selection, and dimensionality reduction will ensure the data is ready for deep learning models. In terms of model development and training, I will implement and fine-tune various deep learning architectures, including CNNs, RNNs, and Transformers. Hyperparameter tuning will be conducted using techniques such as grid search and Bayesian optimization to achieve the best performance. I will also explore transfer learning approaches by pre-training models on large-scale epigenetic data and fine-tuning them on cancer-specific datasets.

To evaluate the models, I will define evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) to comprehensively assess model performance. Benchmarking will be done systematically across different models using standardized datasets and evaluation protocols. For interpretability and visualization, we will use techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to interpret model predictions and identify key methylation sites. Additionally, I will develop visualization tools to illustrate the relationships between methylation patterns and cancer types, aiding in the interpretability of the results.

Expected Outcome

The expected outcomes of this research include the development of deep learning models that can accurately predict DNA methylation patterns associated with early cancer diagnosis. This may lead to identifying specific DNA methylation sites as reliable biomarkers for early cancer detection. Furthermore, I aim to improve methods for interpreting model predictions, making them more accessible and useful for clinical applications. I also plan to create benchmark datasets and evaluation protocols that can be used for future research in this area. Finally, I plan to publish our research findings in reputable journals or conferences, contributing to the field of cancer epigenetics and AI in healthcare.

Significance

This research aims to improve the early diagnosis of cancer through the prediction of DNA methylation patterns using deep learning models. By enhancing the accuracy and interpretability of these models, I hope to develop robust tools for early cancer detection, which could significantly impact clinical practices and patient outcomes. The findings of this research may also provide a foundation for future studies in cancer epigenetics and the application of AI in biomedical research, ultimately contributing to the advancement of precision medicine.

References

[1] Levy, J.J., Titus, A.J., Petersen, C.L. *et al,* MethylNet: an automated and modular deep learning approach for DNA methylation analysis. *BMC Bioinformatics* 21, 108 (2020). doi: 10.1186/s12859-020-3443-8

[2] E. D. Shulman, D.-T. Hoang, R. Turakulov, *et al*, "Abstract 886: Integrated deep learning model for predicting DNA methylation and tumor types from histopathology in central nervous system tumors," *Cancer Res.*, vol. 84, no. 6_Supplement, p. 886, Mar. 2024. doi: 10.1158/1538-7445.AM2024-886

[3] Z. Lin, S. Yu, Z. Kuang, *et al*, "Multimodality Helps Unimodality: Cross-Modal Few-Shot Learning with Multimodal Models," in *2023 IEEE/CVF CVPR*, Jun. 2023, doi: 10.1109/CVPR52729.2023.01852.