

Results Analysis:

The study analyzed the performance of Convolutional Neural Networks (CNNs) and Random Forest models on the IQ-OTHNCCD lung cancer dataset and the LC25000 lung cancer dataset. The Random Forest model outperformed CNNs in terms of accuracy, achieving more accurate results. Data augmentation was applied to increase variability and improve model generalization. The Random Forest Classifier was employed as the classification model, aggregating multiple decision trees for robust predictions. Grid Search CV was used to optimize the model hyper parameters. The results showed that the Random Forest model outperformed CNNs in terms of accuracy, F1-score, recall, and precision. Future work could explore hybrid approaches that combine the strengths of both CNNs and Random Forests or investigate other advanced models to improve classification accuracy and generalizability.

- ⇒ গবেষণাটি IQ-OTHNCCD ফুসফুসের ক্যান্সার ডেটাসেট এবং LC25000 ফুসফুসের ক্যান্সার ডেটাসেটে কনভোলিউশনাল নিউরাল নেটওয়ার্ক (CNNs) এবং র্যান্ডম ফরেস্ট মডেলগুলির কর্মক্ষমতা বিশ্লেষণ করেছে। র্যান্ডম ফরেস্ট মডেল নির্ভুলতার পরিপ্রেক্ষিতে সিএনএন-কে ছাড়িয়ে গেছে, আরও সঠিক ফলাফল অর্জন করেছে। পরিবর্তনশীলতা বৃদ্ধি এবং মডেল সাধারণীকরণ উন্নত করতে ডেটা বৃদ্ধি প্রয়োগ করা হয়েছিল। দৃঢ় ভবিষ্যদ্বাণীর জন্য একাধিক সিদ্ধান্ত গাছ একত্রিত করে র্যান্ডম ফরেস্ট ক্লাসিফায়ারকে শ্রেণীবিভাগের মডেল হিসাবে নিযুক্ত করা হয়েছিল। মডেল হাইপার প্যারামিটার অপ্টিমাইজ করতে গ্রিড সার্চ সিভি ব্যবহার করা হয়েছিল। ফলাফলগুলি দেখায় যে র্যান্ডম ফরেস্ট মডেলটি নির্ভুলতা, F1-স্কোর, প্রত্যাহার এবং নির্ভুলতার পরিপ্রেক্ষিতে সিএনএনকে ছাড়িয়ে গেছে। ভবিষ্যত কাজ হাইব্রিড পদ্ধতির অন্বেষণ করতে পারে যা CNN এবং র্যান্ডম ফরেস্ট উভয়ের শক্তিকে একত্রিত করে বা শ্রেণীবিভাগের নির্ভুলতা এবং সাধারণীকরণের উন্নতির জন্য অন্যান্য উন্নত মডেলগুলি তদন্ত করে।

Discussion:

The study compares the performance of CNNs and Random Forest models in lung cancer imaging analysis on two datasets: the IQ-OTHNCCD and LC25000 Lung Datasets. Random Forest, when enhanced with optimization techniques like Genetic Algorithms (GA) and K-fold cross-validation, offers exceptional accuracy in classifying lung cancer subtypes. Its adaptability to complex datasets outperforms CNN in experiments, particularly on the LC25000 dataset. However, CNN's performance on the LC25000 dataset was lower, suggesting potential limitations in deep learning models for certain medical data types. Gene Transformer, a deep learning framework that integrates CNN and attention mechanisms, is particularly effective at handling high-dimensional data, such as gene expression or medical imaging datasets. It also addresses the interpretability gap in Random Forest models, allowing researchers to gain insights into specific genes or imaging features driving classification decisions. Future studies could integrate clinical and histopathological data with deep learning models for more comprehensive cancer analysis.

- ⇒ গবেষণাটি দুটি ডেটাসেটে ফুসফুসের ক্যান্সার ইমেজিং বিশ্লেষণে সিএনএন এবং র্যান্ডম ফরেস্ট মডেলের কর্মক্ষমতা তুলনা করে: IQ-OTHNCCD এবং LC25000 ফুসফুসের ডেটাসেট। র্যান্ডম ফরেস্ট, যখন জেনেটিক অ্যালগরিদম (GA) এবং কে-ফোল্ড ক্রস-ভ্যালিডেশনের মতো অপ্টিমাইজেশান কৌশলগুলির সাথে উন্নত করা হয়, তখন ফুসফুসের ক্যান্সারের

সাবটাইপগুলিকে শ্রেণীবদ্ধ করার ক্ষেত্রে ব্যতিক্রমী নির্ভুলতা প্রদান করে। জটিল ডেটাসেটের সাথে এর অভিযোজনযোগ্যতা পরীক্ষায় CNN কে ছাড়িয়ে যায়, বিশেষ করে LC25000 ডেটাসেটে। যাইহোক, LC25000 ডেটাসেটে CNN-এর কর্মক্ষমতা কম ছিল, যা কিছু মেডিকেল ডেটা প্রকারের জন্য গভীর শিক্ষার মডেলগুলিতে সম্ভাব্য সীমাবদ্ধতার পরামর্শ দেয়। জিন ট্রান্সফরমার, একটি গভীর শিক্ষার কাঠামো যা সিএনএন এবং মনোযোগের প্রক্রিয়াকে একীভূত করে, বিশেষ করে জিন এক্সপ্রেশন বা মেডিকেল ইমেজিং ডেটাসেটের মতো উচ্চ-মাত্রিক ডেটা পরিচালনার ক্ষেত্রে কার্যকর। এটি র্যান্ডম ফরেস্ট মডেলের ব্যাখ্যাযোগ্যতার ব্যবধানকেও সম্বোধন করে, যা গবেষকদের নির্দিষ্ট জিন বা ইমেজিং বৈশিষ্ট্যগুলির মধ্যে অন্তর্দৃষ্টি লাভ করতে দেয় যা শ্রেণিবিন্যাসের সিদ্ধান্তগুলি চালায়। ভবিষ্যতের অধ্যয়নগুলি আরও ব্যাপক ক্যান্সার বিশ্লেষণের জন্য গভীর শিক্ষার মডেলগুলির সাথে ক্লিনিকাল এবং হিস্টোপ্যাথোলজিকাল ডেটা একীভূত করতে পারে।

Conclusion:

The study uses Convolutional Neural Networks (CNNs) and Random Forest models to classify lung cancer subtypes using two datasets: the IQ-OTHNCCD lung cancer dataset and the LC25000 Lung Dataset. The Random Forest model consistently achieved perfect performance, with 100% accuracy, F1-score, precision, and recall. However, CNN's performance on the LC25000 dataset was moderate, indicating potential areas for improvement. The study suggests that traditional machine learning models can be powerful tools in medical imaging analysis.

- ⇒ গবেষণাটি দুটি ডেটাসেট ব্যবহার করে ফুসফুসের ক্যান্সারের উপপ্রকার শ্রেণিবদ্ধ করতে কনভোল্যুশনাল নিউরাল নেটওয়ার্ক (CNNs) এবং র্যান্ডম ফরেস্ট মডেল ব্যবহার করে: IQ-OTHNCCD ফুসফুসের ক্যান্সার ডেটাসেট এবং LC25000 ফুসফুসের ডেটাসেট। র্যান্ডম ফরেস্ট মডেলটি ধারাবাহিকভাবে 100% নির্ভুলতা, F1-স্কোর, নির্ভুলতা এবং স্মরণ সহ নিখুঁত কর্মক্ষমতা অর্জন করেছে। যাইহোক, LC25000 ডেটাসেটে CNN-এর কর্মক্ষমতা মাঝারি ছিল, যা উন্নতির সম্ভাব্য ক্ষেত্রগুলি নির্দেশ করে। গবেষণাটি পরামর্শ দেয় যে ঐতিহ্যগত মেশিন লার্নিং মডেলগুলি মেডিকেল ইমেজিং বিশ্লেষণে শক্তিশালী সরঞ্জাম হতে পারে।