



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 2, February 2025

Advanced Machine Learning Models for Early Detection of Severe Community-Acquired Pneumonia using Radiomics and Patient Data

Dr C K Gomathy, Dr V Geetha, B.Satya Meghana, R.Haritha

Department of CSE, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Enathur, Tamil Nadu, India

ABSTRACT: Severe community-acquired pneumonia (CAP) presents substantial clinical challenges, often resulting in significant morbidity and mortality. Early detection is vital for effective management and intervention. This article investigates the application of bioinformatics and computational biology techniques to develop advanced machine learning models for the early identification of severe CAP. We propose a novel framework that combines radiomic analysis of chest imaging data with extensive clinical and genomic patient information. By employing sophisticated feature extraction methods, we capture intricate patterns within radiological images that correlate with severe pneumonia, while integrating demographic, clinical, and laboratory data to enhance predictive performance. Our model is rigorously validated through cross-validation techniques, demonstrating improved sensitivity and specificity compared to conventional diagnostic approaches. This research underscores the transformative potential of integrating bioinformatics and computational biology in the diagnostic process for severe CAP, aiming to facilitate timely interventions and ultimately improve patient outcomes through earlier recognition and tailored treatment strategies.

KEYWORDS: BioInformatics, Radiomic Analysis, genomic patient information, cross-validation techniques.

I. INTRODUCTION

Severe Community-Acquired Pneumonia (SCAP) is the most dangerous form of CAP, often leading to ICU admission and associated with significant morbidity and mortality. Pneumonia severity assessment tools, such as the CURB-65 and the Pneumonia Severity Index (PSI), are widely used in clinical settings. However, each of these tools has its limitations. The current standard for identifying SCAP is derived from the 2007 guidelines of the Infectious Diseases Society of America (IDSA) and the American Thoracic Society (ATS). Prompt recognition and aggressive management of SCAP is critical but remains a challenge for clinicians.

Radiomics, first introduced by Lambin et al., proposes that medical images hold untapped data beyond what is visible to the human eye. This technique extracts large volumes of quantitative features from images to complement qualitative assessments by physicians. Radiomic data has been effectively used to train models for specific clinical predictions, such as diagnosing breast tumours from chest X-rays and non-invasively predicting the malignancy of lung nodules, often with high accuracy.

Compared to deep learning, machine learning-based radiomic feature extraction is more data-efficient, requiring fewer samples and offering interpretability. Radiomics-based machine learning models have shown promising results across a variety of predictive tasks. For instance, Prinzi et al. used radiomics to extract features from X-ray images and applied SVM and random forest classifiers for prognostic prediction, achieving a test set AUC of 0.819.

The goal of this study is to apply multiple machine learning algorithms for radiomics modelling and evaluate their performance alongside clinical metrics to enable rapid and precise identification of SCAP cases. Additionally, we aim to uncover image features in chest CT scans that are not visually recognizable by humans and to explore the relationship between these features and clinical data.

II. MATERIALS AND METHODS

Study Population

Clinical and imaging data were retrospectively collected from patients diagnosed with CAP at the North China University of Science and Technology Affiliated Hospital. Inclusion criteria were: (1) CAP diagnosis; (2) CT performed upon admission; and (3) lung lesions present on the CT scan. Exclusion criteria were: (1) absence of clinical data; (2) no CT performed during admission. A total of 174 patients were analysed: 110 in the non-SCAP group and 64 in the SCAP group. Diagnoses were made in accordance with the Guidelines for the Diagnosis and Treatment of Community-Acquired Pneumonia in Chinese Adults (2016 edition) and the IDSA/ATS guidelines.

CT Image Segmentation

Medical image segmentation can be manual, semi-automatic, or automatic. Manual segmentation requires radiologists to outline regions of interest (ROI), but is time-intensive and can result in loss of lesion contour accuracy. To address these challenges, we employed a combination of automatic segmentation, assisted by manual techniques, to accurately define ROIs.

The NNU-Net framework, based on U-Net architecture, was utilized for automatic segmentation. This adaptive deep learning model is tailored to medical imaging tasks, automatically determining network structures and training parameters based on dataset characteristics such as resolution and size. NNU-Net also automates key processes like data augmentation, learning rate adjustments, and early stopping.

III. FEATURE EXTRACTION

Radiomics Feature Extraction

To quantify grayscale features from each ROI, we used pyradiomics (v3.0), an open-source tool compliant with the Biomarker Standardization Initiative guidelines. Radiomics features were grouped into seven categories: (i) shape-based features, (ii) first-order statistics, (iii) grey-level dependence matrix, (iv) grey-level size zone matrix, (v) neighbouring grey-tone difference matrix, (vi) grey-level run-length matrix, and (vii) grey-level co-occurrence matrix.

Construction of Radiomic Feature Set

To avoid overfitting and ensure model generalizability, we reduced feature dimensionality through multiple selection methods. Pearson's correlation coefficient was first used to eliminate highly redundant features. A threshold of 0.7 was set, with values above this indicating strong correlations.

Further feature selection was performed using model-based methods, including Random Forest, XGBoost, and Lasso models. Features were ranked based on their importance scores, and the top 15 were chosen for the radiomic feature set.

Clinical Feature Set Construction

Clinical data from patients were collected, including age, gender, smoking and drinking status, CURB-65 scores, and various lab test results (e.g., blood urea nitrogen, PCT, WBC, NEU, LYM, TP, ALB, SCR, ALT, AST, fibrinogen, D-dimer). We conducted correlation analyses between clinical and imaging features using Pearson's coefficient and visualized the results with heatmaps. Logistic regression (univariate and multivariate) was used to identify clinical features significantly associated with SCAP.

Model Establishment and Performance Comparison

Given the dataset size and feature dimensionality, we employed eight mainstream machine learning algorithms: AdaBoost, Logistic Regression, Random Forest, SVM (Radial Kernel), XGBoost, KNN, LightGBM, and Naive Bayes. Models were built using the radiomic feature set, clinical feature set, and their combination. A 10-fold cross-validation approach was used for model validation, ensuring robust performance evaluation.

Interpretability Analysis

Machine learning models often operate as "black boxes," leaving their decision-making process opaque to medical professionals. Interpretable analysis helps uncover the inner workings of these models, offering insights into important classification features.

Statistical Analysis

Univariate analysis was used to describe patient characteristics, and multivariate logistic regression identified independent risk factors for SCAP. Statistical significance was set at $p < 0.05$. Machine learning models were developed using Python 3.9.6 and the "sklearn" package, and ROC curves were plotted to calculate AUC values for model evaluation.

IV. CONCLUSION

This study highlights the transformative potential of advanced machine learning models for the early identification of severe community-acquired pneumonia (CAP), utilizing both radiomic signatures and comprehensive patient data. By integrating deep learning methodologies with traditional machine learning frameworks, we have developed a robust diagnostic tool that effectively elucidates subtle radiological indicators of severe pneumonia while incorporating critical clinical parameters, such as laboratory results and patient demographics.

Future research should focus on expanding the diversity of datasets, refining algorithmic techniques, and validating these models in prospective clinical trials. As we advance in the fields of bioinformatics and computational biology, our work contributes to the ongoing evolution toward precision medicine in the management of severe CAP, emphasizing the need for data-driven strategies in improving clinical decision-making and patient care.

REFERENCES

1. Dr.V.Geetha and Dr.C K Gomathy, Anomaly Detection System in Credit Card Transaction Dataset, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212564> Vol3028, Issue 01 2024
2. Dr.V.Geetha and Dr.C K Gomathy, Crime data analysis and prediction using machine learning, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212566> Vol3028, Issue 01 2024
3. Dr.C K Gomathy and Dr.V.Geetha House price prediction using machine learning, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212559> Vol3028, Issue 01 2024
4. Dr.V.Geetha and Dr.C K Gomathy, Identification of birds species using deep learning, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212968> Vol3028, Issue 01 2024
5. Dr.V.Geetha and Dr.C K Gomathy, Missing child recognition system using deep learning, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212567> Vol3028, Issue 01 2024
6. Dr.V.Geetha and Dr.C K Gomathy, Price forecasting of agricultural commodities, AIP Conference Proceedings,) <https://doi.org/10.1063/5.0212568> Vol3028, Issue 01 2024
7. Dr.V.Geetha and Dr.C K Gomathy, The customer churn prediction using machine learning, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212569> Vol 3028, Issue 01 2024
8. Dr.C K Gomathy and Dr.V.Geetha, Fall detection for elderly people using machine learning, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212561> Vol3028, Issue 01 2024
9. Dr.C K Gomathy and Dr.V.Geetha, Fall Navigation and obstacle detection for blind, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212560> Vol3028, Issue 01 2024
10. Dr.V.Geetha and Dr.C K Gomathy, Securing medical image based on improved ElGamal encryption technique, AIP Conference Proceedings,) <https://doi.org/10.1063/5.0212570> Vol3028, Issue 01 2024
11. Dr.C K Gomathy and Dr.V.Geetha, Software error estimation using machine learning algorithms, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212562> Vol3028, Issue 01 2024
12. Dr.V.Geetha and Dr.C K Gomathy, Web scraping using robotic process automation, AIP Conference Proceedings,) <https://doi.org/10.1063/5.0212571> Vol3028, Issue 01 2024
13. Dr.C K Gomathy and Dr.V.Geetha, Crypto sharing DAAP, AIP Conference Proceedings,<https://doi.org/10.1063/5.0212563> Vol3028, Issue 01 2024
14. Dr.V.Geetha and Dr.C K Gomathy, Company employee profile using QR code, AIP Conference Proceedings,) <https://doi.org/10.1063/5.0212572> Vol3028, Issue 01 2024
15. Dr.V.Geetha and Dr.C K Gomathy, Unified platform for advertising with predictive analysis, AIP Conference Proceedings,) <https://doi.org/10.1063/5.0212573> Vol3028, Issue 01 2024
16. Gomathy, C.K., Geetha, V., Lakshman, G., Bharadwaj, K. (2024). A Blockchain Model to Uplift Solvency by Creating Credit Proof. In: Mandal, J.K., Jana, B., Lu, T.C., De, D. (eds) Proceedings of International Conference on Network Security and Blockchain Technology. ICNSBT 2023. Lecture Notes in Networks and Systems, vol 738. Springer, Singapore. https://doi.org/10.1007/978-981-99-4433-0_39

17. CK.Gomathy, MangantiDhanush, SikharamSaiPushkar, V.Geetha ,Helmet Detection and Number Plate Recognition using YOLOv3 in Real-Time 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2023) DVD Part Number: CFP23K58-DVD; ISBN: 979-8-3503-4362-5,DOI:10.1109/ICIMIA60377.2023.10425838, 979-8-3503-4363-2/23/\$31.00 ©2023 IEEE
- 18.Dr.V.Geetha and Dr.C K Gomathy, Cloud Network Management System, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.69 ISSN: 1308-5581 Vol 14, Issue 05 2022
- 19.Dr.C K Gomathy and Dr.V.Geetha,Fake Job Forecast Using Data Mining Techniques, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.70 ISSN: 1308-5581 Vol 14, Issue 05 2022
- 20.Dr.V.Geetha and Dr.C K Gomathy,Cyber Attack Detection System, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.71 ISSN: 1308-5581 Vol 14, Issue 05 2022
- 21.Dr.V.Geetha and Dr.C K Gomathy,Attendance Monitoring System UsingOpencv, International Journal of Early Childhood Special Education (INT-JECSE) DOI:DOI:10.9756/INTJECSE/V14I5.68ISSN: 1308-5581 Vol 14, Issue 05 2022
22. Dr.C K Gomathy and Dr.V.Geetha,The Vehicle Service Management System, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.66ISSN: 1308-5581 Vol 14, Issue 05 2022
- 23.Dr.C K Gomathy and Dr.V.Geetha, Multi-Source Medical Data Integration AndMining For Healthcare Services, International Journal of Early Childhood Special Education (INT-JECSE) DOI: DOI:10.9756/INTJECSE/V14I5.67ISSN: 1308-5581 Vol 14, Issue 05 2022
- 24.Dr.V.Geetha and Dr.C K Gomathy, An Efficient Way To Predict The Disease UsingMachine Learning, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.98ISSN: 1308-5581 Vol 14, Issue 05 2022
- 25.Dr.C K GomathyandDr.V.Geetha, Music Classification Management System, International Journal of Early Childhood Special Education (INT-JECSE) DOI: DOI:10.9756/INTJECSE/V14I5.72ISSN: 1308-5581 Vol 14, Issue 05 2022
26. Dr. C.K. Gomathy ,Dr.V.Geetha ,G.S.V.P.Praneetha , M.Sahithisucharitha. (2022). Medicine Identification Using OpenCv. Journal of Pharmaceutical Negative Results, 3718–3723.<https://doi.org/10.47750/pnr.2022.13.S09.457>
27. Dr.V.Geetha ,Dr. C.K. Gomathy , KommuruKeerthi , NallamsettyPavithra. (2022). Diagnostic Approach To Anemia In Adults Using Machine Learning. Journal of Pharmaceutical Negative Results, 3713–3717.<https://doi.org/10.47750/pnr.2022.13.S09.456>
28. Dr. C. K. Gomathy, " A Cloud Monitoring Framework Perform in Web Services, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 3, Issue 5, pp.71-76, May-June-2018.
29. Dr. C. K. Gomathy, " Supply Chain - Impact of Importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 3, Issue 6, pp.01-04, July-August-2018.
30. Dr.C.K.Gomathy, Dr.V.Geetha, PeddireddyAbhiram, "The Innovative Application for News Management System," International Journal of Computer Trends and Technology, vol. 68, no. 7, pp. 56-62, 2020. Crossref, <https://doi.org/10.14445/22312803/IJCTT-V68I7P109>
31. Dr. C. K.Gomathy, " A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 3, Issue 1, pp.1568-1573, January-February-2018.
32. Gomathy, C. K., et al. "A Location Based Value Prediction for Quality of Web Service." International Journal of Advanced Engineering Research and Science, vol. 3, no. 4, Apr. 2016.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details