Integrating imaging and clinical data for enhanced Lung cancer risk prediction: A weighted average ensemble approach

Mayur Munshi[1], Gareth Price[2], Andrew Reilly[1], Gordon Cowell[3]

Department of Clinical Physics and Bio-engineering[1]
NHS Greater Glasgow and Clyde[1,3] , University of Manchester[2]

PURPOSE

To improve lung cancer risk prediction among screening trial participants by integrating imaging and clinical data using a weighted average ensemble approach.

To compare ensemble performance with individual image-based and clinical risk models.

To benchmark model discrimination against published lung cancer risk prediction studies.

METHODS AND MATERIALS

Dataset
- Sequential LDCT scans from the first two screening years and clinical data were obtained from U.S. NLST participants, with balanced cancer and non-cancer groups based on retrospective follow-up confirmation.

Inclusions
- ➢ Participants with at least two years screening trial scans
- ➢ Training cohort (n =1,700)
- ➢ Testing cohort (n =408)
- ➢ Participants in test cohort with confirmed lung cancer between year 3 to 5 of follow-up

Exclusion
- ➢ Participants diagnosed with lung cancer inf first two years

- No manual lung nodule annotation was performed during training or testing.

Model design
- Image-based: Hybrid modified Resnet50-BiLSTM.
- Clinical risk-based: Voting classifier utilising fourteen independent clinical features.
- Weighted average ensemble : Combined prediction of image-based and clinical risk-based models.

Evaluation
- Internal evaluation with independent NLST test cohort for area under the receiver's operating curve(AUC), precision, recall, F1 score, and accuracy.
- Results from this study were compared against published lung cancer risk prediction benchmarks.

.

RESULTS

- **Model Performance (Current Study)**

- Image-based model: AUC 0.80 (95% CI: 0.76–0.85).
- Clinical risk-based model: AUC 0.84 (95% CI: 0.79–0.87).
- Weighted average ensemble model: AUC 0.91 (95% CI: 0.88–0.93).
- Ensemble model demonstrated the highest overall predictive performance.

Image-based model comparisons

- Table 1 compares current study's models with other similar model performances reported by of Gao et al. (2019) using NLST test dataset. .
- Current study's image-based model performed better than DLSTM model on recall, and F1 score showcasing improved detection of positive cases and better balance between sensitivity and precision.

- The weighted average ensemble model that combined predictive abilities of image-based and clinical risk-based models demonstrated consistent performance improvements over single modality models as shown in Tables 1 and 2.

- Other published results by Mikhael et al. (2023) demonstrated their 'Sybil' model developed with 3D CNN architecture achieved an AUC of 0.80 for year-three prediction with nodule annotation, that dropped to AUC of 0.72 without nodule annotation, which is a direct comparison with current study's image-based model that achieved an AUC of 0.80.

- The weighted average ensemble outperformed all image-based models on all key performance metrics with better balance of sensitivity and precision.

Clinical risk-based model comparisons

- Table 2 shows comparison of current study with other two published studies ,Zhou et al. (2021) and a systematic review by Feng et al. (2024).
- Feng et al. (2024) reported results for six different clinical risk-based models widely used in Europe for eligibility assessment for lung screening trials.

- Table 2 shows results of the highest and lowest performing models reported by Feng et al.(2024).

- The weighted average ensemble approached demonstrated second highest performance when compared with the published clinical risk-based models.

Table 1. Evaluation of image-based and weighted average ensemble Models from current study with the published experimental results from Gao et al. (2019) on NLST Data.

Model /Study (Researchers)	Model type	AUC	Precision	Recall	F1 score	Accuracy
Ori_CNN	3D CNN	74.18	83.24	38.07	52.18	71.94
MC-CNN	3D CNN	77.96	78.62	47.91	59.39	73.26
CNN-LSTM	2D CNN	80.84	78.68	59.92	67.85	77.05
CNN-tLSTM	2D CNN-LSTM	80.80	79.81	58.65	67.47	77.37
DLSTM	3D CNN-LSTM	82.55	83.38	61.61	70.85	78.96
Image based (Current study)	3D CNN-BiLSTM	80.0	78.0	72.0	75.0	70.0
Weighted Average Ensemble (Current study)	CNN-BLSTM +VC	91.0	82.0	89.0	85.0	85.0

Table 2. Evaluation of clinical risk-based and weighted average ensemble model with published clinical risk-based models.

Model /Study (Researchers)	Dataset	Purpose	AUC
Decision Tree (Zhou et al., 2021)	PLCO	Lung cancer risk prediction	0.927
LASSO logistic regression (Chandren et al., 2023)	Optum EHR	Lung cancer risk in three years	0.76
PLCOM2012, LLPv3, BATCH, LCRAT, LCDRAT, and UCLD (Feng et al., 2024)	ATBC, HUNT, CONSTANCES, and EPIC	Screening eligibility assessment in nine European countries	Lowest (LLPv3) 0.68 Highest (UCLD) 0.83
Clinical risk-based (Current study)	NLST clinical risk features	Lung cancer risk prediction in three years in future	0.84
Weighted Average Ensemble (Current study)	NLST LDCT and clinical risk features	Lung cancer risk prediction in three years in future without nodule annotation	0.91

Discussion

- The image-based model in current study was trained and tested with two years of sequential LDCT screening trial scans without cancer nodule annotation. It demonstrated strong performance compared with other image-based models designed for nodule detection. The current study instead focused on predicting lung cancer risk using longitudinal changes over a two-year screening period, highlighting its ability to capture disease progression rather than detect existing nodules.

- Among the best performing clinical risk-based models by Zhou et al. (2024 developed using PLCO dataset included only 2.2% lung cancer cases, resulting in substantial class imbalance. In contrast, the current study used a retrospectively balanced dataset across cancer and non-cancer groups in both training and testing cohorts to improve model generalisability and robustness.

- Published clinical risk-based models in table 2 were originally developed for lung screening eligibility using population-based datasets. In contrast, the current study focused on post-screening follow-up risk prediction, making direct comparison challenging. However, this approach complements existing clinical risk-based models by addressing post-screening decision-making and the high false-positive rates seen in screening programmes.

- A key limitation is the lack of external validation when compared with the Sybil model (Mikhael et al., 2023). External validation is required to confirm model robustness and generalisability across independent datasets.

Conclusion

- Weighted average ensemble model achieved the highest performance (AUC 0.91).

- Combining image-based and clinical risk-based features improves prediction accuracy.

- Image-based and hence weighted average ensemble models demonstrated strong performance despite no nodule annotation.

- This approach demonstrated superior lung cancer risk prediction based on longitudinal screening data.

Future work

- Perform external validation using independent datasets.

- Evaluate model performance in prospective screening cohorts.

- Explore integration into clinical pathways.

REFERENCES

- Chandran, U et al. (2023).'Machine Learning and Real-World data to predict cancer risk in routine care. Cancer Epidemiology, Biomarkers & Prevention, 32(3), pp. 337-343.

- Feng, X et al. (2024). 'Evaluation of risk prediction models to select lung cancer screening participants in Europe: a prospective cohort consortium analysis', The Lancet Digital Health, 6 (9), pp. e614-e624.

- Gao, R et al. (2019). Distanced LSTM: Time-Distanced Gates in Long Short-Term Memory Models for Lung Cancer Detection. In: Suk, HI., Liu, M., Yan, P., Lian, C. (eds) Machine Learning in Medical Imaging. MLMI 2019. *Lecture Notes in Computer Science ()*, vol 11861. Springer, Cham. https://doi.org/10.1007/978-3-030-32692-0_36.

- Zhou, J et al. (2021). 'Lung Cancer Prediction Using Curriculum Learning Based Deep Neural Networks', Proceedings – 2021 *IEEE International Conference on Digital Health*, Virtual, Chicago, Il, USA, 5-10 September, pp. 11-18. Doi: 10.1109/ICDH52753.2021.00013.