Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-3231

Reviewer A

The purpose of this work was to 'explore in-depth the relationship between AI-detected TNN and survival outcomes'.

It is well written, and the description of the analysis is detailed and clear.

Management of PN still poses several types of problems, therefore every new tool to aid the radiologist/surgeon is of great importance.

The work is of great scientific value and hopefully the model will be validated in the nearest future.

I appreciate the Tables and Figures. Especially Table 2 with the features of the detected nodules and Figure 5 in the Supplementary materials presenting a case report.

Please verify/change:

Page 7, line 249 - speculation to spiculation

We appreciate the reminder of this spelling mistake. Amendments have been made (see Page 8, line 274).

Reviewer B

Please describe how you dealt with data imbalance.

We appreciate these critical questions. Dealing with imbalanced data is a very challenging topic and we believe that the survival tree in our study is the most appropriate scenario to answer this question.

In our study, a tree-based model was built to confirm the value of TNN as a prognostic factor. We've noticed that the patient number in stage III cohort is slightly imbalanced (213 stage IIIA and 50 stage IIIB patients), which may generate a local minimum for any loss function we choose. So instead of building a classification tree model, which may use staging as a label, we used recursive partitioning approach, and turn the staging into a parameter, to avoid the influence by imbalanced data. Such approach starts with all data as a single cluster and then splits it into sub-clusters that have the smallest within cluster distances in some metric, which mimics a way of unsupervised learning or clustering algorithm. We choose log-rank test result as the metric and validated that TNN is a good prognostic factor in stage III NSCLC. Other approaches such as oversampling were not utilized in this study.

Describe challenges faced in dealing with this type of data

Imbalanced data is one of the major problems in machine learning, especially in supervised learning such as classification model training. The challenge of using imbalanced data is that the more imbalanced the data is, the more difficult it is for the model to beat the performance of blind guess, no matter what loss function was chosen. As a retrospective study, modeling with existing data is particularly vulnerable to data imbalance, especially when excessive subgroup analysis was done. So in this study, we performed traditional univariate-multivariate analysis in the beginning, and only used the survival tree as a validation method to confirm prior findings.

Describe future follow up studies needed

As for the follow up studies, we are investigating the biological nature of TNN. A prospective study is planned. We will focus on pathology findings of multiple nodules in stage III patients, and the correlation between TNN and tumor associated biomarkers such as CEA, CA12-5, and CYFRA21-1 etc.

Reviewer C

The authors analyzed the patients who underwent surgical resection for stage I-III NSCLC using convolutional neural networks in deep learning-based artificial intelligence algorithms to detect and classify pulmonary nodules. AI-detected total number of nodules was significantly associated with survival in patients with surgically resected stage III NSCLC.

The paper is well written and well-illustrated. There is a significant interest in using AI analysis to inform, guide, and improve clinical practice. My comment and question regarding the manuscript are listed below:

The authors concluded that TNN was an independent prognostic factor in patients with stage III NSCLC. I wonder whether high TNN represented pathologically pulmonary metastasis when the patients with stage III NSCLC underwent complete resection. The authors should demonstrate the ratio of pulmonary metastasis in patients with each stage.

We appreciate this valuable comment. Nodule evaluation is challenging and critical to diagnose multiple pulmonary nodules in the clinical practice. In this study, preoperative chest high-resolution CT scans, combined with PET/CT scans in some patients, were conducted and carefully reviewed by both radiologists and thoracic surgeons to identify pulmonary nodules with high malignancy probability. Along with the primary lesion, other suspicious nodules, no matter whether they were considered as multiple primary lung cancer or intra-pulmonary metastases, were resected together during the surgical intervention. Either the Martini-Melamed criteria or comprehensive histologic assessment was applied to differentiate multiple primary tumors from metastases. Pathological T stage was determined accordingly, and was upgraded once intrapulmonary metastasis was confirmed. In this cohort, approximately 0% (0/1626), 4% (10/237), and 11% (28/263) of stage I, II, and III NSCLC patients were diagnosed with intrapulmonary metastases, respectively.

Although the pathological stage in this study was the best we could have in a clinical scenario, whether those unresected small solid nodules detected by AI represented metastatic lesions or not remained unknown. We believed that most of these AI-detected nodules were benign lesions. However, like what we have demonstrated in Supplementary Figure 5, few unresected nodules did grow larger unexpectedly during the postoperative follow-up period, suggesting their potential malignant properties. Thus, we have the reason to suspect that higher TNN might represent higher probability of extra pulmonary malignant lesions, which could be a biological explanation of their inferior survival outcome. However, it is hard to distinguish these potential malignancies even with PET/CT scans, not to mention differentiating whether they were multiple primary tumors or intrapulmonary metastatic lesions, owing to the partial-volume effect of small pulmonary nodules. Future nodule-based prospective study is required to verify this inference.

Reviewer D

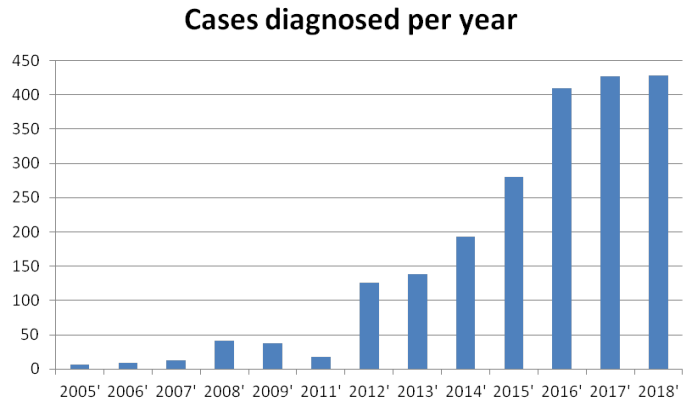
In this manuscript titled 'Total nodule number is an independent prognostic factor in resected stage III non-small cell lung cancer'. The study was conducted to detect the total lung nodules or subsolid by AI and higher total number is independent prognostic factor in Stage III. The manuscript was well written. However, I have some foundational questions.

Major

1.The authors should explain the differences between AI systems in previous reports and their deep-learning AI in PN detection in Introduction.

We have cited a few more references comparing the AI systems used for pulmonary nodule (PN) detection. In this study, the first deep learning-based AI algorithm for PN detection approved by FDA was utilized to guarantee the detection performance. Compared with AI algorithms reported in proof-of-concept studies, its robustness and generalizability were validated in multiple medical centers and proved to be valuable in enhancing imaging report standardization and improving clinical work flow (see Page 3, line 68).

2.Materials and Methods: Their follow-up time was less the 3years. However, their data was 2005 to 2018. The authors should provide why the duration in this study was relatively short.



Out data was collected from October 2005 to December 2018, and the median follow-up time of this cohort was 33 months (IQR 21-48). One possible explanation for the relatively short follow-up time is that most of the cases involved in the cohort were diagnosed after 2012 (see Histogram: Cases diagnosed per year). Moreover, some cases may loss to follow-up in a short period of time, which also contributed to the relatively short duration.

3.Results: Stage IIIA is T1T2a-bN2/T3N1/T4N0-1, while IIB is T1T2a-bN3/T3N2/T4N2. As for TNN, the pathological T difference is defined by tumor location as same lobe (T3) or ipsilateral other lobes (T4). I consider that higher TNN had a possibility of T4 or M1a. They provide the detail data in Figure 2 in each stage (I/IIA/IIB/IIIA/IIIB). In addition, in clinical stage, we sometimes cannot differ the pulmonary metastases from 2nd lung cancer. The authors should explain the M1a from 2nd primary after follow-up intervals.

Indeed, we agreed that higher TNN might have a possibility of T4 or M1a, which also correlated well with the phenomena that stage III NSCLC patients with higher TNN harbored inferior survival outcome. However, it was neither possible to distinguish the minor malignant nodules from the substantial amount of AI-detected pulmonary nodules, nor able to remove all the nodules during surgical intervention. Thus, the pathological stage in this study was the best we could have in the clinical practice. Moreover, according to the comment, we explained the criteria we used to differentiate multiple primary lung cancers and metastases in the Method section (see

Page 4, line 104).

4.The author commented that 'TNN may be a visualized representation of tumor burden in stage III NSCLC. Really? They should provide the total tumor burden by 3D-reconstrauction data or total area of solid component in 2D.

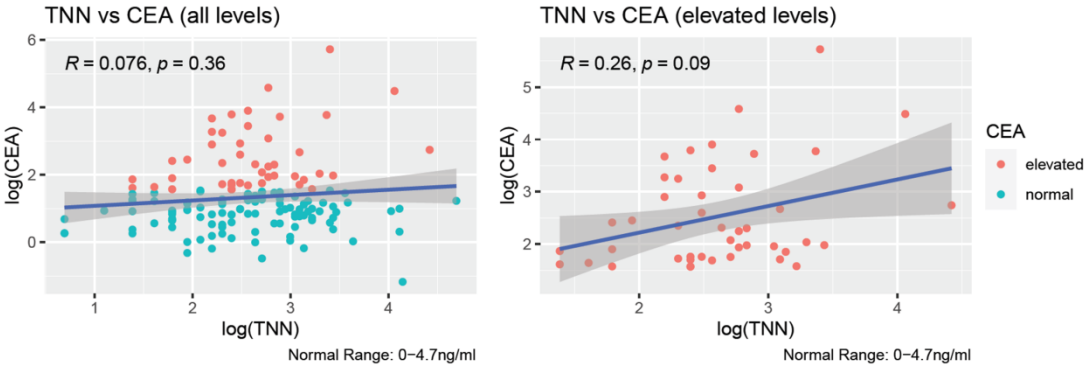
The wording "visualized representation" fails to convey the idea. We have amended the expression as "TNN may be a simplified representation of tumor burden in stage III NSCLC" (see Page 7, line 257). Also, we have shown a 3D-reconstruction demonstration in figure 1B as visualization. Since the emphasis of this article is the abstract aspect of the value of TNN, we would avoid the 3D-reconstruction demonstration of all cases.

5. CEA is well-known to be a surrogate marker for tumor volume (Okada M,Nishio W, Sakamoto T, et al. Prognostic significance of perioperative serum carcinoembryonic antigen in non-small cell lung cancer: Analysis of 1,000 consecutive resections for clinical stage I disease. Ann Thorac Surg 2004; 78: 216-21, etc). They should provide the data of correlation between CEA and TNN.

We have conducted a further study regarding the association between TNN and CEA, since CEA is a good indicator of tumor burden in CEA elevated patients.

We retrospectively reviewed the CEA level of all stage III patients. Of all 263 patients, 201 have pre-operative CEA result. A total of 3 test methods were used for CEA measurement, distinct by the threshold of elevation defined as 4.7 (151 patients), 5 (21 patients) and 10 (29 patients) ng/ml. We selected patients using the first method (normal range: 0-4.7ng/ml) for most of the analysis to ensure the data is comparable.

First, we investigated the correlation between TNN and CEA both as continuous variables. Since neither TNN nor CEA fits Gaussian distribution, Spearman's correlation was used. As shown in the figure below, no correlation was seen between TNN and CEA in all stage III patients (R = 0.076, p = 0.36). Even in CEA elevated patients, the correlation between these two markers is unconvincing (R = 0.26, p = 0.09).



We then consider TNN as a discrete variable and stratify patients accordingly. TNN-high group has significantly higher median CEA comparing to TNN-low group (3.33 (2.26, 5.81) vs. 2.58 (1.91, 4.31), p = 0.050, Wilcoxon rank sum test).

If we treat both TNN and CEA as discrete variables, a trend of positive correlation between TNN and CEA was seen (p = 0.117, Chi-squared test). Such trend became statistically significant when increasing the sample size by adding patients using 2 other methods (p = 0.026, Chi-squared test).

Method 1 population	Normal CEA	Elevated CEA

Lower TNN	31	7
Higher TNN	75	38
All methods population	**Normal CEA**	**Elevated CEA**
Lower TNN	43	8
Higher TNN	100	50

Due to limited time, we cannot review CEA data for all enrolled patients, but the result in stage III patients have shown some notable trend of the correlation between TNN and CEA level.

Although CEA may represent the tumor burden in CEA elevated NSCLC patients, its value in all NSCLCs is yet controversial (Grunnet 2012, Carcinoembryonic antigen (CEA) as tumor marker in lung cancer, LUNG CANCER). Since CEA suffered from less optimal sensitivity in NSCLC (Jian 2020, The combination of CTCs and CEA can help guide the management of patients with SPNs suspected of being lung cancer, BMC Cancer), TNN may be a good supplementary indicator of tumor burden due to its universality and ease of measurement.

Based on the above mentioned results, we are planning for a further study to investigate the association between TNN and all tumor markers. Hopefully it will be presented in a separate paper.

Minor

1. The authors should provide the abbreviation in abstract. Because they used many abbreviations.

   Abbreviations were added in the abstract (see Page 2, line 27 etc).

2. The authors should provide the method to obtain the cut-off value in TNN in abstaract

   Maximally selected log-rank statistics were mentioned in abstract (see Page 2, line 26).