Peer review file

Article information: http://dx.doi.org/10.21037/jtd-20-1956

Reviewer A

This study is to establish an artificial neural network (ANN) model to predict pathologic nodal involvement in clinical stage I-II ESCC patients. 27 variables were collected and 9 variables were identified as candidate variables for an ANN model. The training set was used to construct the ANN prediction models on 350 patients, and the test set was used on 173 patients to verify and evaluate the accuracy of the ANN models. The ANN model presented good performance for predicting pathologic lymph node metastasis.

Comment 1: First of all, the self-learning ability of the ANN is quite innovative compared with traditional statistical techniques. However, it should be established and based on the reliable data.

Clinical N stage (cN0 and cN+) was included in the model. But, clinical N staging was not identified as one of 4 important predictors in this ANN model. Because only 48/216 (22%) of patients with pN+ were predicted as cN+ preoperatively. This clinical N staging is far from reliable one establishing the model. Clinical N stage was established only using a combination of EUS-FNA and PET-CT. How about CT scan and EUS?

Reply 1: Because only patients with clinical staging <T3N1M0 were enrolled (patients diagnosed with N2 or N3 preoperatively were excluded). Thus, both sensitivity and accuracy in the detection of nodal metastasis may have been relatively

low. The ANN model constructed in this study can be used as a supplement to the current clinical staging system to reduce the false negative rate of preoperative lymph node assessment (if the patient is diagnosed as clinical stage I-II, we can use the model to screen out patients with suspicious N+). Thank you for your comment. We have declared it in the “limitation” part of the manuscript and the revised part have been marked with red (see Page 12, line 258-267).

CT has been widely used for preoperative evaluation, but it is known to be nonsensitive for the identification of transmural spread and the detection of metastases to lymph nodes[1]. PET examinations have been shown to have a significantly higher specificity and accuracy compared with the conventional CT examinations for lymph nodal metastasis diagnosis, and particularly the specificity is reported to be higher[2]. The accuracy of EUS for estimating the depth of penetration of the primary tumor has been validated, but EUS has been shown to be inaccurate in the evaluation of nodal status[1]. Even when combined with CT, EUS has a reported sensitivity for detecting involved lymph nodes of only 11% to 54% and a specificity of 90% to 95%[3]. Tissue acquisition by EUS-FNA remains the optimal way to assess a (non-peritumoural) node for malignant involvement.

Comment 2: Of the 27 variables, a total of 9 variables with P<0.20 were identified as candidate variables (including age, smoking history, dysphagia, CEA, CA125, clinical T stage, clinical N stage, tumor differentiation, and tumor length). CA125 was included but did not showed P<0.20. Tumor location showed P<0.20 but was not included. Why was CA125 included and why was tumor location not

included?

Reply 2: Thanks to the reviewer for pointing out the flaws in our manuscript. We added the tumor location to the ANN model, and deleted CA125; then reconstructed the ANN model. We re-uploaded Figure 2 and Figure 3; the revised part has been marked with red in the manuscript (see Page 8, line 182-183).

Comment 3: According to the prediction outcomes of the ANN model, 523 patients were divided into with low risk lesions and with high risk lesions. Was high risk means predicted as pN+? The definition should be described clearly.

Reply 3: Patients with high risk lesions were defined as patients more likely to diagnosed with pN+ by the prediction of the ANN model. However, considering that postoperative variables that would seriously affect survival were not incorporated in this model and the short follow-up time of enrolled patients, we decided to delete this controversial part according to Reviewer B's advice.

Comment 4: Symptoms were dysphagia and obstruction. Did all of patients with cT1 tumor complain dysphagia or obstruction? Was no patient with no symptom?

Reply 4: Obstruction was defined as patient feels discomfort or pain when swallowing hard foods. Dysphagia was defined difficulty in swallowing liquids or soft foods. All the patients enrolled in this study had symptoms when swallowing foods before surgery including patients with cT1 lesions. This single-center study is limited by its retrospective nature and the small number of patients. We have declared it in the "limitation" part of the manuscript and the revised part have been marked with red.

Comment 5: Line 287: “The eighth edition of the AJCC cancer clinical staging system is based solely on T staging and N staging and does not take into account tumor differentiation, tumor size, histology, or symptoms. In this study, the above 4 indicators were identified as important predictors influencing model decision.” Tumor size and histology were not included in this study.

Reply 5：Thanks to the reviewers for pointing out the inaccurate statements in our manuscript. We have modified the relevant content and revised part has been marked with red (see Page 12, line 258-267).

Reviewer B

The authors of the study aimed to construct and validate an artificial neural network model to improve the accuracy of clinical N staging in patients with ESCC. This is a very important topic because an accurate preoperative clinical staging is critical in these patients. We are all aware of the limitations of the current staging model, and thereby developing strategies to improve preoperative staging is commendable. Congratulations to the authors for an interesting and well written manuscript. I have minor concerns about the study:

Comment 1: I am surprised that some well-studied variables related to nodal involvement in T1 tumors were not included in the ANN (e.g. depth of submucosal invasion in T1b tumors, lymphovascular invasion, etc). The authors should explain and comment why such variables were not included in the model.

Reply 1：Thanks for reviewer’s advice. We divided T1a into m1, m2 and m3; T1b

were divided into sm1, sm2 and sm3. We redefine the variable “clinical T stage” as “tumor invasion depth” (see Table 1 and Table 2). We added data about lymphovascular invasion of patients. After incorporating the new variables, the new ANN model constructed presented better C-index, SEN, SPE, NPV, PPV and DA. The new ANN model now incorporated 10 variables (including tumor invasion depth, tumor length, tumor differentiation, LVI, dysphagia, CEA, clinical N stage, age, smoking history and tumor location). 5 variables (tumor invasion depth, tumor length, tumor differentiation, LVI, dysphagia) were identified as important predictors (normalised importance＞80%). (see Page 8, line 181-192)

Comment 2: The purpose of the study was to build an ANN model able to predict nodal involvement. The survival analysis should be removed from the study because multiple postoperative variables that significantly affect survival are not included in this model (postoperative morbidity and mortality and use of adjuvant therapy among others). In addition, very few patients had 3 years of follow-up.

Reply 2: Thank you for your comment. Considering that postoperative variables that would seriously affect survival were not incorporated in this model and the short follow-up time of enrolled patients, we decided to delete the survival analysis as advised.

Comment 3: Although the authors divided the cohort in a training set and a test set, external validity is a potential limitation of a model created and validated with the same cohort of patients. This should be addressed in the limitations of the study.

Reply 3: We have declared it in the “limitation” part of the manuscript and the

revised part have been marked with red (see Page 12, line 258-267). Thank you for your advice.