Development and Validation of a Deep Learning Model for Preoperative Screening of Myasthenia Gravis in Patients with Thymoma based on CT Images

Ying Zhu
Sun Yat-sen University First Affiliated Hospital
Zhen-guo Liu
Sun Yat-sen University First Affiliated Hospital
Lei Yang
Sun Yat-sen University First Affiliated Hospital
Kefeng Wang
Sun Yat-sen University First Affiliated Hospital
Ming-Hui Wang
Sun Yat-sen University First Affiliated Hospital
Xiao-yu Yang
Sun Yat-sen University First Affiliated Hospital
Xi Wu
Sun Yat-sen University First Affiliated Hospital
Xi Tian
Advanced Institue, Infervision, Beijing
Rong-Guo Zhang
Advanced Institue, Infervision, Beijing
Jie-chao Ma
Sun Yat-Sun University, College town, Panyu district
Yu-Jie Yuan
Sun Yat-sen University First Affiliated Hospital
zunfu Ke (kezf8688@hotmail.com)
Sun Yat-sen University First Affiliated Hospital

Research article

Keywords: Thymoma, Myasthenia gravis, Artificial Intelligence (Computer Vision Systems), Deep Learning,Tomography, X Ray Computed

Posted Date: May 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-26909/v1

Abstract

Objectives: Thymoma-associated myasthenia gravis (TAMG) is the most common paraneoplastic syndromeof thymoma. The screening of TAMG before thymoma resection is required to avoid severe perioperative complications, especiallyrespiratory failure. Herein, we developed a 3D DenseNet deep learning (DL) model based on preoperative computed tomography (CT) to detect TAMGin thymoma patients.

Methods:A large cohort of 230 thymoma patientswere enrolled. 182 thymoma patients (81 with TAMG, 101 without TAMG) were used for training and model building. 48 cases from another hospital were used for external validation. A 3D-DenseNet-DL model and five machine learning models with radiomics features were performed to detectTAMG in thymoma patients. A comprehensive analysis by integrating 3D-DenseNet-DL model and general CT image features,named 3D-DenseNet-DL-based multi-model, was also performed to establish a more effective prediction model.

Results: By elaborately comparing the prediction efficacy,the 3D-DenseNet-DL effectively identified TAMG patients, with a mean area under ROC curve (AUC), accuracy, sensitivity and specificity of 0.734, 0.724, 0.787 and 0.672, respectively. The effectiveness of the 3D-DenseNet-DL-based multi-model was further improved as evidenced bythe following metrics: AUC 0.766, accuracy 0.790, sensitivity 0.739 and specificity 0.801. External verification results confirmed the feasibility of this DL-based multi-model with metrics: AUC 0.730, accuracy 0.732, sensitivity 0.700 and specificity 0.690,respectively.

Conclusions: Our 3D-DenseNet-DL model can effectively detect TAMG in patients with thymoma based on preoperative CT images. This model may serve as a non-invasive screening method or as a supplement to the conventional diagnostic criteria for identifyingTAMG.

Key points:

Thymoma-associated myasthenia gravis (TAMG) is a common paraneoplastic syndrome.

3D-DenseNet-DL model can effectively detect TAMG based on preoperative CT images.

This model may serve as a supplement for identifying TAMG.

Introduction

Thymoma is the most common neoplasm of the anterior mediastinum in adults, and often associated with various autoimmune paraneoplastic syndromes (PNS) [1]. Thymoma-associated myasthenia gravis (TAMG) is the most common PNS, accounting for 30–50% of all PNS [2]. TAMG is an autoimmune disease, involving antibodies against the postsynaptic nicotinic acetylcholine receptors (AChR) at neuromuscular junctions, resulting in variable weakness of voluntary muscle [3]. Patients with TAMG can experience severe cardiopulmonary complications [4, 5]. One of the most severe complications of TAMG is postoperative myasthenic crisis, which can rapidly worsen, leading to respiratory failure and even death [6]. Moreover, the incidence of postoperative myasthenic crisis was high (ranges from 11.5 to 18.2%), and patients with crisis have a high mortality rate [7–9]. According to NCCN clinical guidelines for thymomas, all patients suspected

of having thymomas (even those without symptoms) should be carefully evaluated for the presence of TAMG before surgical procedure in order to avoid respiratory failure during the operation [2, 10, 11]. However, muscle weakness, as a symptom, is common in many other diseases, attributing to frequently missed or delayed diagnosis of MG in patients experiencing mild weakness or in individuals with weakness restricted to only a few muscles [12]. In addition, although the current criteria for diagnosing MG (including immunological, electrophysiological, and pharmacological approaches) have improved [13], a simple, non-invasive method for preoperative screening of TAMG, especially for patients with inconspicuous or atypical symptoms, has important clinical significance.

A very puzzling, but interesting characteristic of MG is that most of patients have histopathological abnormalities in their thymus, such as hyperplastic thymus and thymoma [14]. Thymomas are now stratified into six entities (types A, AB, B1, B2, B3, and TC (carcinoma)) on the basis of the morphology of epithelial cells and the lymphocyte-to epithelial cell ratio [15]. TAMG is the specific subtype of MG, which is closely related to the different pathological subtypes of thymoma: TAMG is more common in type B (B_1, B_2 and B_3) than type A and AB thymomas and absent in TC [14, 16]. In thymoma, the correlation with grading and staging of thymoma has been widely analyzed based on imaging features or quantitative texture analysis [17, 18].In recent years, deep learning (DL) and radiomics in the medical imaging field have been studied intensively to explore the potential of utilizing various medical images as diagnostic, predictive, or prognostic information of human diseases, including the possibility of identifying tumor pathological subtypes, tumor phenotypes and the gene-protein signatures [19, 20]. Therefore, with rapid advancement in machine learning algorithms, it is possible to determine the status of TAMG in patients with different pathological subtypes of thymoma using imaging data from preoperative routine CT scan of thymoma.

Here, we designed this study to explore the effectiveness of 3D-DenseNet-DL model and five radiomics as predictive methods for detecting TAMG based on preoperative chest CT image. The final optimal model, named as 3D-DenseNet-DL based multi-model, integrating with general CT image features was ultimately established to detect MG in thymoma patients.

Materials And Methods

Patients

For this study, 182 patients diagnosed with thymoma who had undergone thymectomy at the First Affiliated Hospital of Sun Yat-sen University from Jan 1st, 2011 to Jun 31st, 2018 were included for analysis and model building (Table 1). Another 48 thymoma patients admitted to the Sun Yat-sen Memorial Hospital of Sun Yat-sen University from Jan 1st, 2017 to Mar 31st, 2019 were used as the external validation cohort (Table 1). All cases had undergone enhanced preoperative CT examination and had been clearly staged based on pathological examination and clinical manifestation. All patients in our study were evaluated by neurologists to determine the status of myasthenia gravis (MG) syndrome or other autoimmune diseases before operation. This project was approved by the Ethics Committee and Institutional Review Board of Sun Yat-sen University. Informed consent was waived due to the retrospective nature of this study.

Table 1
Baseline characteristic of the 230 patients with thymoma from two medical centers.

Variables	Training set (n = 182)				External validation set (n = 48)			
	Number	without MG (n = 101)	with MG (n = 81)	P-value*	Number	without MG (n = 34)	with MG (n = 14)	P-value*
Sex				0.500				0.230
Male	115(63.2%)	66	49		27(56.3%)	21	6	
Female	67(36.8%)	35	32		21(43.7%)	13	8	
Age (year, mean ± SD)	NA	51.5 ± 13.1	47.5 ± 12.1	0.035†	NA	51.6 ± 13.8	50.4 ± 15.1	0.791†
WHO histologic classification				< 0.001				0.227
A	22(12.1%)	19	3		8(16.7%)	7	1	
AB	37(20.3%)	22	15		13(27.1%)	11	2	
B1	21(11.5%)	15	6		5(10.4%)	4	1	
B2	72(39.6%)	26	46		19(39.6%)	10	9	
B3	21(11.5%)	10	11		3(6.25%)	2	1	
C	9(4.95%)	9	0		0	0	0	
Masaoka staging				0.006				0.151
I	84(46.2%)	43	41		41(85.4%)	28	13	
IIA	40(22.0%)	24	16		1(2.08%)	0	1	
IIB	16(8.79%)	5	11		0	0	0	
IIIA	20(10.9%)	13	7		5(10.4%)	5	0	
IIIB	13(7.14%)	7	6		1(2.08%)	1	0	
IV	9(4.94%)	9	0		0	0	0	
Smoking history				0.215				0.835
No	161(88.5%)	92	69		35(72.9%)	24	11	

Note: *Chi-square test or Fisher's exact test; † Student's t test; # Some patients' data were missing; NA-Not Applicable.

Variables	Training set (n = 182)				External validation set (n = 48)			
	Number	without MG (n = 101)	with MG (n = 81)	P-value*	Number	without MG (n = 34)	with MG (n = 14)	P-value*
Yes	21(11.5%)	9	12		13(27.1%)	10	3	
Surgical approach#				< 0.001				0.051
Thymoma resection	31(17.0%)	30	1		4(8.33%)	4	0	
Thymectomy	30(16.5%)	29	1		30(62.5%)	23	7	
Extended thymectomy	111(61.0%)	34	77		14(29.2%)	7	7	
Postoperative pneumonia#				0.003				-
No	137(75.7%)	83	54		-	-	-	
Yes	39(21.5%)	13	26		-	-	-	

Note: *Chi-square test or Fisher's exact test; † Student's t test; # Some patients' data were missing; NA-Not Applicable.

CT imaging Characteristics and Scan Protocol

Enhanced chest CT images were acquired within one week prior to operation. Imaging features were carefully evaluated through PACS reading workstation by two experienced radiologists specializing in chest CT imaging that were blinded to the MG statuses of the patients. CT Imaging characteristics that were evaluated included (Table 2): maximum diameter (3-D Maximum diameter); degree of enhancement (increment of enhanced CT value, HU); enhancement (homogeneous or heterogeneous); necrosis/cystic component (divided into 0%-25%,26%-50%, 51%-75%, 75%-100% according to its volume percentage); shape (round or oval, lobulated, irregular); contours (smooth or irregular); presence of calcification, adjacent organ invasion, effusion(pleural/pericardial), and lymphadenopathy. All preoperative enhanced chest CT images were obtained with a 64-row multidetector CT scanner (Aquilion 64; Toshiba Medical, Tokyo, Japan). Scan parameters: x-ray tube voltage of 120 kVp; maximum of 500 mA with automatic tube current modulation.

Axial thin-section CT images of the whole lung were reconstructed with a section thickness and spacing of 1.0 mm. Iopromide at 80–100 ml/per patient (300 mg I/m1, Schering Pharmaceutical Ltd) was injected at 3–4 ml/s flow rate and applied to contrast enhanced scanning protocol.

Table 2
Image characteristics of patients with thymoma.

Variables	Number	Preoperative MG		P-value
		No (n = 101)	Yes (n = 81)	
Maximum diameter †	NA	6.13 ± 2.93	4.91 ± 2.27	0.065
Dgree of enhancement (HU) †	NA	32.56 ± 22.17	30.86 ± 20.06	0.972
Enhancement				0.074
Homogeneous	81(44.5%)	39	42	
Heterogeneous	101(55.5%)	62	39	
Necrosis/cystic component				0.029
0%-25%	71(39.0%)	36	35	
26%-50%	78(42.9%)	39	39	
51%-75%	16(8.79%)	12	4	
75%-100%	17(9.34%)	14	3	
Shape				0.027
Round or oval	91(50.0%)	50	41	
Lobulated	37(20.3%)	27	10	
Irregular	54(29.7%)	24	30	
Contours				0.030
Smooth	163(89.6%)	86	77	
Irregular	19(10.4%)	15	4	
Calcification				0.827
No	147(80.8%)	81	66	
Yes	35(19.2%)	20	15	
Adjacent organ invasion				< 0.001
No	157(86.3%)	79	78	
Yes	25(13.7%)	22	3	
Effusion (Pleural/Pericardial)				0.028
No	169(92.9%)	90	79	
Yes	13(7.14%)	11	2	

Note: † Data are mean ± standard deviation; NA-Not Applicable.

Variables	Number	Preoperative MG		P-value
		No (n = 101)	Yes (n = 81)	
Lymphadenopathy				0.030
No	166(91.2%)	88	78	
Yes	16(8.79%)	13	3	
Note: † Data are mean ± standard deviation; NA-Not Applicable.				

Machine Learning

1. Datasets

Thymoma on CT images were segmented manually using the annotation tool “ITK-SNAP” (www.itksnap.org) [21]. “ITK-SNAP”, as a free software, is widely used for medical image annotation and labeling. In this work, ITK-SNAP was applied for thymoma lesion segmentation. The output from ITK-SNAP are NIFTI files containing mask information of the thymoma for each sequence of CT images. We then used the mask information to extract the area of thymoma, namely the regions of interests (ROI) (Figure S1). For feature extraction in radiomics analysis, the segmented thymoma was used directly. For deep learning modeling, a further preprocessing step was designed to prepare the segmented data for the convolutional neural network.

2. Radiomics analysis

2.1 Radiomic analysis procedure

Radiomics analysis involved several steps: feature extraction, feature selection and machine learning. First, feature extraction was performed to convert raw images to structural data with radiomics information that could be processed by machine learning algorithms. Then, several methods were applied to further select high-quality features based on variance or regression. Finally, the data with selected features are used as inputs for several mainstream machine learning algorithms to train and test the model.

2.2 Radiomic features

The radiomic features were extracted using open source PyRadiomics software (http://pyradiomics.readthedocs.io) [22]. The categories of features include: shape descriptors (2D and 3D), First Order Statistics, Gray Level Matrices (GLM) based: Gray Level Cooccurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM) and Gray Level Dependence Matrix (GMDM). These features were extracted not only from original images, but also from derived images filtered using Laplacians of Gaussians (LoG), Wavelet Decompositions, Square, Square Root, Logarithm and Exponential filters. In total, 1390 radiomic features were extracted, covering the popular features used in research.

2.3 Radiomic feature extraction

Feature selection was conducted to select a subset of features from all extracted features for use in model building. The aims of this step were to reduce the dimensions of features, simplify the model and enhance generalization by reducing overfitting. A multi-level selection approach was adopted, which involved three algorithms in the order of: variance threshold method, k-best method, and the least absolute shrinkage selection operator (LASSO). Variance based method was adopted at first to select features with variance larger than a threshold (threshold = 0.1 in this study, data were normalized to a range of -1 to 1). Then, top k (k = 300 in this paper) features were further selected based on top ANOVA F-value between feature and the label. Finally, LASSO with five-fold cross-validation was adopted to automatically select the more effective features (Figure S2).

2.4 Radiomics model building

The performance of radiomics analysis was evaluated using five popular machine learning algorithms: Random Forest, XGBoost, Multilayer Perceptron, Logistic Regression and Support Vector Machine.

3. Deep learning

3.1 Data preprocessing

For deep learning, images with fixed dimension 160 × 160 × 64 (pixels) were used as input of the model. The images were constructed with equal width and length of 160 pixels and channels of 64 pixels. The size of the input image was determined by statistical analysis of the region of all the thymomas in this dataset.

3.2 3D-DenseNet

DenseNet [23] is a type of convolutional neural network (CNN). DenseNet composes of four dense blocks, as shown in the schematic diagram. Dense connections between layers within dense blocks are present in DenseNet. We chose DenseNet as the base model in this study due to its various advantages. First, DenseNet can be used to reduce over-fitting. Second, DenseNet is computationally efficient as it requires less than half of the parameters of ResNet. Although DenseNet was first designed for two-dimensional images, our study targeted 3D CT sequences. As most medical images are three-dimensional, we designed a 3D-DenseNet where the kernel of each convolutional and pooling layer was modified to 3D versions. In the proposed 3D DenseNet model, rectified linear unit (ReLu) was used as activation function in each layer, and softmax function was applied in the last layer of our network to obtain the probability for each sample (Fig. 1 and Table S1). Batch normalization was applied before activation layer. The loss function of our model was due to binary cross-entropy, which was optimized using Adam with mini-batch size of 16.

3.3 Training Process Optimization

Two kinds of data augmentation were applied during the training stage of deep learning (Figure S3) to avoid overfitting. First, random cropping was implemented by randomly placing the segmented thymoma image in the fixed cube with shape. Second, a fixed window center (WC) and window width (WW) of 300 were applied for input images with original CT values. A random change was applied for training data with WW value ranging from − 10 to 10 and WC value from − 5 to 5. Transfer learning was also applied to obtain benefit

such as acceleration of the training stage from the pretrained model, which boosted the training speed significantly compared with the other initialization methods (such as Xavier).

3.4 Evaluation Metrics of Machine Learning

Five machine learning models with radiomics features (RF, XGBoost, SVM, MLP, LR) and the 3D-DenseNet-DL model were evaluated in the training and validation cohort using stratified five-fold cross-validation, and the parameter alpha was chosen with Mean Square Error (MSE) at minimum value. The metrics Area Under ROC Curve (AUC), ACC (accuracy), Sensitivity (SN) and Specificity (SP) were used to compare the performance of these models.

The deep learning model was implemented using MXNet (version 1.2.0, Apache Software Foundation, Forest Hill, MD, USA) library [23], and the model was trained using four NVIDIA GeForce GTX 1080 GPUs (NVIDIA, Beijing, China).

Statistical Analyses

Statistical analyses were performed using SPSS 22.0 (IBM, USA). Variables were grouped based on the presence of MG. Categorical variables were compared using the Chi-squared test. Continuous variables were compared using the T-test or Mann Whitney U test for variable with abnormal distribution. Multivariate logistic regression analysis was used to explore independent predictors of MG. Variables included in this analysis included age, gender, enhancement heterogeneity, necrosis/cystic component rate, contours, shape, adjacent organ invasion, pleural/pericardial effusion and lymphadenopathy. $p < 0.05$ was considered as statistically significant. The area under the ROC curve (AUC), accuracy, sensitivity, and specificity were measured in order to evaluate the accuracy of models. An AUC estimate of more than 0.7 was regarded as outstanding.

Results

Clinical characteristics of patients

230 patients were included in this study, 182 were used for training and model building and an independent cohort of 48 cases was used for external validation. In the training set, 182 patients were divided into two groups: thymoma-associated myasthenia gravis (TAMG) group (n = 81) and thymoma without MG group (n = 101). Correlation between the different baseline characteristics and MG status (thymoma with or without MG) was summarized in Table 1. In training cohort, a significant different ratio of histologic classification was found between two groups ($P < 0.001$): TAMG patients with a lower A + AB ratio (22.2% vs. 40.6%) and higher $B_1 + B_2 + B_3$ ratio (77.8% vs.50.5%), comparing to patients without MG. TAMG was not found in Thymic carcinoma (TC), which is consistent with previous reports[16]. In addition, TAMG patients showed significant association with younger age (47.5 ± 12.1 years vs. 51.5 ± 13.1 years, $P = 0.035$) and relatively earlier thymoma Masaoka staging ($P = 0.006$). There were no significant differences between the two groups in terms of gender and smoking history. However, it is worth noting that there was obvious correlation between MG status and postoperative pneumonia ($P = 0.003$) with the presence of TAMG in thymoma as a high-risk factor.

Associations between General CT Image Characteristics and status of TAMG

Ten common variables were used to describe the CT imaging features of thymomas included in this study (Table 2). The statistical differences between two groups were found in necrosis/cystic component rate (P = 0.029), contours (smooth/ irregular, P = 0.030), shape (P = 0.027), adjacent organ invasion (P < 0.001), pleural/pericardial effusion (P = 0.028) and lymphadenopathy (P = 0.030). In general, thymoma patients with TAMG tend to have less enhancement heterogeneity, less lobulated shape and lower rate of adjacent organ invasion.

Detection of TAMG by Radiomics analysis and 3D DenseNet DL model

For the radiomics analysis and deep learning (DL) analysis, a total of 1390 radiomic features were extracted from the Routine contrast enhanced chest CT image data. After applying Variance Threshold, K-best and LASSO methods, the remaining features after application of each method were 499, 300, and 16, respectively. The 16 features finally selected were listed in table S2. To decipher the relationship between features, correlation analysis using the Pearson method was applied and a heatmap was constructed for visualization (Figure S4).

Five machine learning models with radiomics features (RF, XGBoost, SVM, MLP, LR) and 3D-DenseNet-DL model were established to detect the status of TAMG, and the values of each metric were shown in Fig. 2. Comparing with the other five machine learning models, the DL model showed the most favorable results with AUC 0.734, accuracy (ACC) 0.724, sensitivity (SN) 0.787 and specificity (SP) 0.672, respectively.

Building of the 3D-DenseNet-DL based multi-model for TAMG detection

With the multivariable logistic regression analysis, only the shape of thymoma (P = 0.031), the invasion rate of adjacent organ (P = 0.001) and DL score (P < 0.001) qualified as independent predictable factors (Table 3). To optimize the effectiveness of TAMG-detecting model, we further built 3D-DenseNet-DL based multi-model (DL plus two general CT features).With ROC curve analysis, the AUC of DL model, general CT features model (the shape and the invasion rate of adjacent organ) and 3D-DenseNet-DL based multi-model were 0.740, 0.677and 0.766, respectively (Fig. 3A and B), suggesting that the 3D-DenseNet-DL based multi-model demonstrated better performance for detecting TAMG in thymoma patients.

Table 3
Significant correlations of MG with clinical information, routine CT imaging features and DL score using Logistic Regression Forward Stepwise (*Likelihood Ratio*) method.

Characteristic	P &	P #	OR # (95% CI)
Shape	0.031	0.032	1.59 (1.04–2.43)
Adjacent organ invasion	0.001	0.007	0.11 (0.02–0.54)
DL score	0.000	0.000	147.84 (9.15-1238.51)

Note: OR, odd ratio; CI, confidence interval; DL, deep learning; #, The P value was calculated by multivariable logistic regression analysis adjusted for age and gender;&, Unadjusted P value; * P < 0.05 was considered as statistically significant.

The external validation of 3D-DenseNet-DL based multi-model

We further evaluated 3D-DenseNet-DL model and the 3D-DenseNet-DL based multi-model in an external validation set, composed of 48 thymoma patients from another medical center. The results showed a comparable agreement in both datasets for the detection of TAMG, with an AUC of 0.730, ACC of 0.732, SN of 0.700 and SP of 0.690 for 3D-DenseNet-DL based multi-model; and AUC of 0.704, ACC of 0.690, SN of 0.760 and SP of 0.710 for 3D-DenseNet-DL model (Fig. 3C). This favorable result further confirmed the reliability and efficacy of our 3D-DenseNet-DL based multi-model in screening TAMG in patients with thymoma.

Discussion

In this study, we proposed and validated a non-invasive method based on preoperative routine CT imaging of thymoma, referred to as “3D DenseNet deep learning (DL) based multi-model”, to detect TAMG before operation. With this model, we successfully filtered out most of TAMG patients in the training set (n = 182, AUC of 0.766), and further verified its reliability and efficacy in an external validation set (n = 48, AUC of 0.730). These results suggest our 3D-DenseNet-DL based multi-model is an effective and non-invasive method for screening TAMG in patients with thymoma. To our knowledge, this is the first study about the diagnosis of TAMG in thymoma patients by using machine learning based on CT imaging data.

Currently, there are three accepted diagnostic criteria for confirming MG by neurologists: immunological, electrophysiological, and pharmacological approaches. The immunological assay for serum AChR binding antibodies is considered as the most reliable approach to diagnose MG [24, 25]. AChR antibody is found in nearly all of TAMG patients, but the false positive rate was also high [14]. Repetitive nerve stimulation (RNS) [26] and single-fiber electromyography (SFEMG) [27] are widely used in electrophysiological confirmation. However, SFEMG may not provide confirmation of the presence of MG unless weak muscles are tested, and the reliability of results is highly dependent on the experience of the technician [13]. Pharmacological confirmation has long been used for the diagnosis of MG [28]. However, the reported false-positive results [29] and the possible occurrence of potentially lethal vagal bradycardia following Tensilon injection [30], particularly in elderly persons, greatly limit its clinical application for MG confirmation. Therefore, although current diagnostic criteria are widely used for the final diagnosis of MG, some other methods may be used as

a supplement for the initial screening or diagnosis of MG. Our 3D-DenseNet-DL based multi-model is a candidate, and the favorable results indicates that this model can be considered as a complementary method to the conventional diagnostic criteria, especially for screening TAMG before thoracic surgery. Considering the efficacy, safety, minimal-invasiveness and economic cost, we proposed a clinical flow chart for preoperative screening of MG: a combination of clinical symptoms, serum AChR antibody and image-based DL method (Figure S5). This flow chart may be important for necessary clinical management and preoperative risk assessment of the disease.

Nowadays, increasing number of studies are performed to evaluate the potential relationship between image and biological features of solid tumors [31], such as glioblastoma [32], rectal and lung adenocarcimoma[33, 34]. As the most common primary neoplasms of the mediastinum, the prediction of thymoma histology and stage by radiographic criteria have been mentioned in several previous reports. CT findings, such as smooth contours [35], calcification [35, 36], heterogeneous attenunation [36, 37], were interpreted as being of value in differentiating the various histologic subtypes of thymomas. Recently, Angelo Iannarelli and colleagues [17] found the relationship between radiomics parameters, histology and grading of thymic tumors. More importantly, their study also demonstrated that MG syndrome was significantly associated with some parameters in quantitative texture analysis (QTA) [17], which represented an incentive for further evaluation the value of radiographic analysis in detection of MG syndrome in thymoma patients. Unfortunately, their study only included 16 patients (7 patients with TAMG). We therefore proposed a DL model based on preoperative CT imaging for screening TAMG in large cohort of thymoma patients (230 cases, and 95 with TAMG). Moreover, our results further confirmed the superior reliability and efficacy of this developed 3D-DenseNet-DL model comparing to the other five radiomic-based methods. These results also highlight the importance of radiographic analysis as diagnostic tools from the accurate characterization of the lesion itself to the detection of the paraneoplastic syndromes, which is a great stride in the application of AI in the medical field.

However, despite its satisfactory outcomes, this study has some limitations. First, given the retrospective nature of this analysis, a selection bias was unavoidable. Second, patients were not stratified into more detailed clinical status categories due to limited sample size. Third, the status of serum AChR binding antibodies was important for TAMG diagnosis, but the absence of such information in certain cases restrained further analysis. Therefore, a perspective, multi-center clinical trial with larger cohort would be indispensable to further confirm and optimize the screening model for MG patients.

In conclusion, with a large sample data for modeling and an independent cohort for external validation, we firstly developed a 3D-DenseNet-DL based multi-model for TAMG screening in thymoma patients based on preoperative CT imaging and achieved favorable results.

Abbreviations

ACC, accuracy; AChR, acetylcholine receptors; AI, artificial intelligence; AUC, area under ROC curve; CNN, convolutional neural network; CT, computed tomography; DenseNet, densely connected convolutional networks; DL, deep learning; MG, masthenia gravis; PNS, paraneoplastic syndromes; ROI, regions of interests; SN, sensitivity; SP, specificity; TC, thymic carcinoma; RF, random forest; LR, logistic regression.

Declarations

Conflict of Interest Statement

The authors declare no potential conflicts of interest.

Conflicts of interests

The authors have declared that no competing interest exists.

Author contributions

Study concept & design, literature search and writing: LZG, ZY and KZF; data collection/interpretation: ZY, YYJ, YL, WKF, and WMH; data analysis: LZG, ZY, YXY, WX, TX, MJC and ZRG; approval of final version of submitted manuscript, all authors; manuscript editing, all authors.

Acknowledgements

The work was supported by grants from Natural Sciences Foundation of Guangdong Province (No.2018A0303130250 to Zhenguo Liu). Furthermore, special thanks to Dr. Huiyu Feng, a neurologist, for reviewing and revising the clinical flow chart in this study.

References

1. Detterbeck FC, Parsons AM. Thymic tumors. Ann Thorac Surg. 2004;77(5):1860–9.
2. Lewis JE, et al. Thymoma. A clinicopathologic review. Cancer. 1987;60(11):2727–43.
3. Drachman DB. Myasthenia gravis. N Engl J Med. 1994;330(25):1797–810.
4. Gritti P, et al. A standardized protocol for the perioperative management of myasthenia gravis patients. Experience with 110 patients. Acta Anaesthesiol Scand. 2012;56(1):66–75.
5. Kanai T, et al. A clinical predictive score for postoperative myasthenic crisis. Ann Neurol. 2017;82(5):841–9.
6. Lacomis D. Myasthenic crisis. Neurocrit Care. 2005;3(3):189–94.
7. Watanabe A, et al. Prognostic factors for myasthenic crisis after transsternal thymectomy in patients with myasthenia gravis. J Thorac Cardiovasc Surg. 2004;127(3):868–76.
8. Ando T, et al. Predictive factors of myasthenic crisis after extended thymectomy for patients with myasthenia gravis. Eur J Cardiothorac Surg. 2015;48(5):705–9.
9. Juel VC. Myasthenia gravis: Management of myasthenic crisis and perioperative care. Semin Neurol. 2004;24(1):75–81.

10. Network NCC NCCN Clinical Practice Guidelines in Oncology. Thymomas and Thymic Caricinomas. 2019 [cited 2019 March 11]; Available from: https://www.nccn.org/professionals/physician_gls/pdf/thymic.pdf.
11. Cata JP, et al., Myasthenia Gravis and Thymoma Surgery: A Clinical Update for the Cardiothoracic Anesthesiologist. J Cardiothorac Vasc Anesth, 2018.
12. Evoli A, Iorio R, Bartoccioni E. Overcoming challenges in the diagnosis and treatment of myasthenia gravis. Expert Rev Clin Immunol. 2016;12(2):157–68.
13. Keesey JC. Clinical evaluation and management of myasthenia gravis. Muscle Nerve. 2004;29(4):484–505.
14. Fujii Y. Thymus, thymoma and myasthenia gravis. Surg Today. 2013;43(5):461–6.
15. Kondo K, et al. WHO histologic classification is a prognostic indicator in thymoma. Ann Thorac Surg. 2004;77(4):1183–8.
16. Radovich M, et al. The Integrated Genomic Landscape of Thymic Epithelial Tumors. Cancer Cell. 2018;33(2):244–58 e10.
17. Iannarelli A, et al. Analysis of CT features and quantitative texture analysis in patients with thymic tumors: correlation with grading and staging. Radiol Med. 2018;123(5):345–50.
18. Yanagawa M, Tomiyama N. Prediction of thymoma histology and stage by radiographic criteria. Thorac Surg Clin. 2011;21(1):1–12, v.
19. Bi WL, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin. 2019;69(2):127–57.
20. Choi JY. Radiomics and Deep Learning in Clinical Imaging: What Should We Do? Nucl Med Mol Imaging. 2018;52(2):89–90.
21. Yushkevich PA, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116–28.
22. van Griethuysen JJM, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104–7.
23. Huang G, et al., Densely Connected Convolutional Networks. 30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017), 2017: p. 2261–2269.
24. Lindstrom JM, et al. Antibody to acetylcholine receptor in myasthenia gravis. Prevalence, clinical correlates, and diagnostic value. Neurology. 1976;26(11):1054–9.
25. Oger J, Kaufman R, Berry K. Acetylcholine receptor antibodies in myasthenia gravis: use of a qualitative assay for diagnostic purposes. Can J Neurol Sci. 1987;14(3):297–302.
26. Keesey JC, Minimonograph AAEE. #33: electrodiagnostic approach to defects of neuromuscular transmission. Muscle Nerve. 1989;12(8):613–26.
27. Sanders DB. Clinical impact of single-fiber electromyography. Muscle & Nerve, 2002: p. S15-S20.
28. Pascuzzi RM. The edrophonium test. Semin Neurol. 2003;23(1):83–8.
29. Dirr LY, et al. A false-positive edrophonium test in a patient with a brainstem glioma. Neurology. 1989;39(6):865–7.

30. Gould L, Zahir M, Gomprecht RF. Cardiac arrest during edrophonium administration. Am Heart J. 1971;81(3):437–8.
31. Colen R, et al. NCI Workshop Report: Clinical and Computational Requirements for Correlating Imaging Phenotypes with Genomics Signatures. Transl Oncol. 2014;7(5):556–69.
32. Jain R, et al. Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor. Radiology. 2014;272(2):484–93.
33. Rizzo S, et al. CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. Eur Radiol. 2016;26(1):32–42.
34. De Cecco CN, et al. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. Invest Radiol. 2015;50(4):239–45.
35. Tomiyama N, et al. Using the World Health Organization Classification of thymic epithelial neoplasms to describe CT findings. AJR Am J Roentgenol. 2002;179(4):881–6.
36. Tomiyama N, et al. Invasive and noninvasive thymoma: distinctive CT features. J Comput Assist Tomogr. 2001;25(3):388–93.
37. Han J, et al. Thymic epithelial tumors classified according to a newly established WHO scheme: CT and MR findings. Korean J Radiol. 2003;4(1):46–53.

Figures

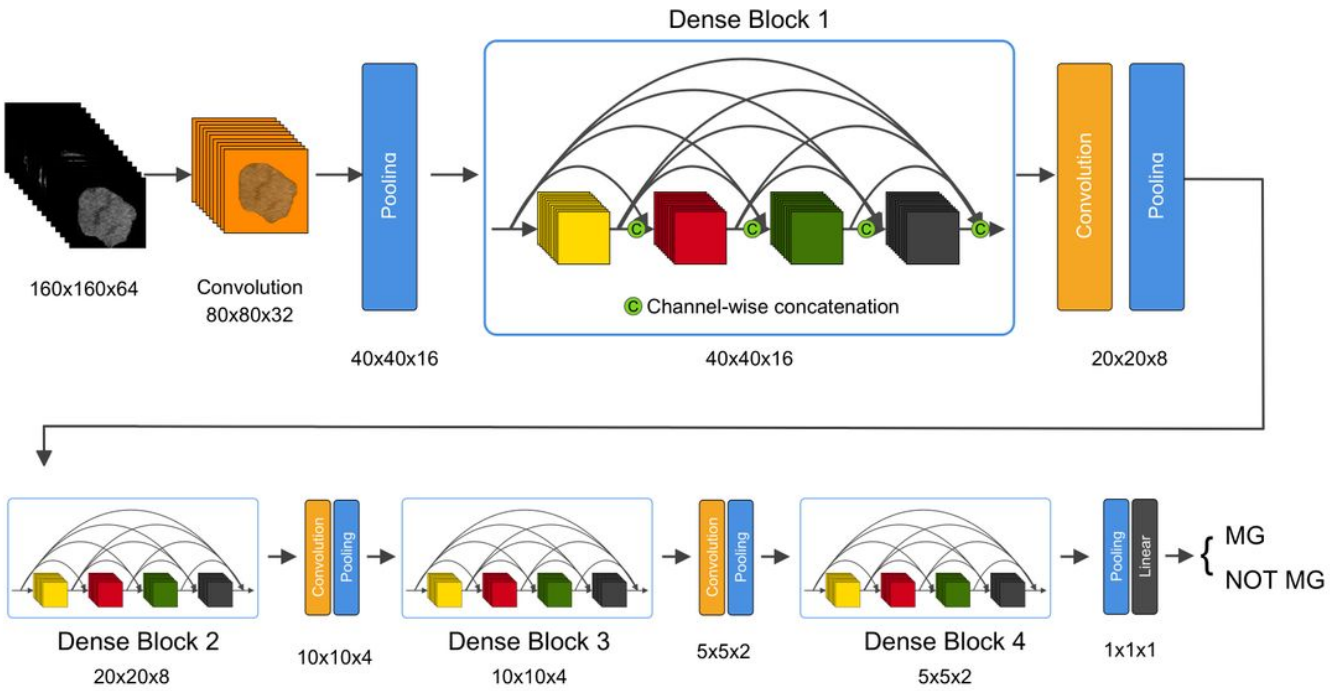
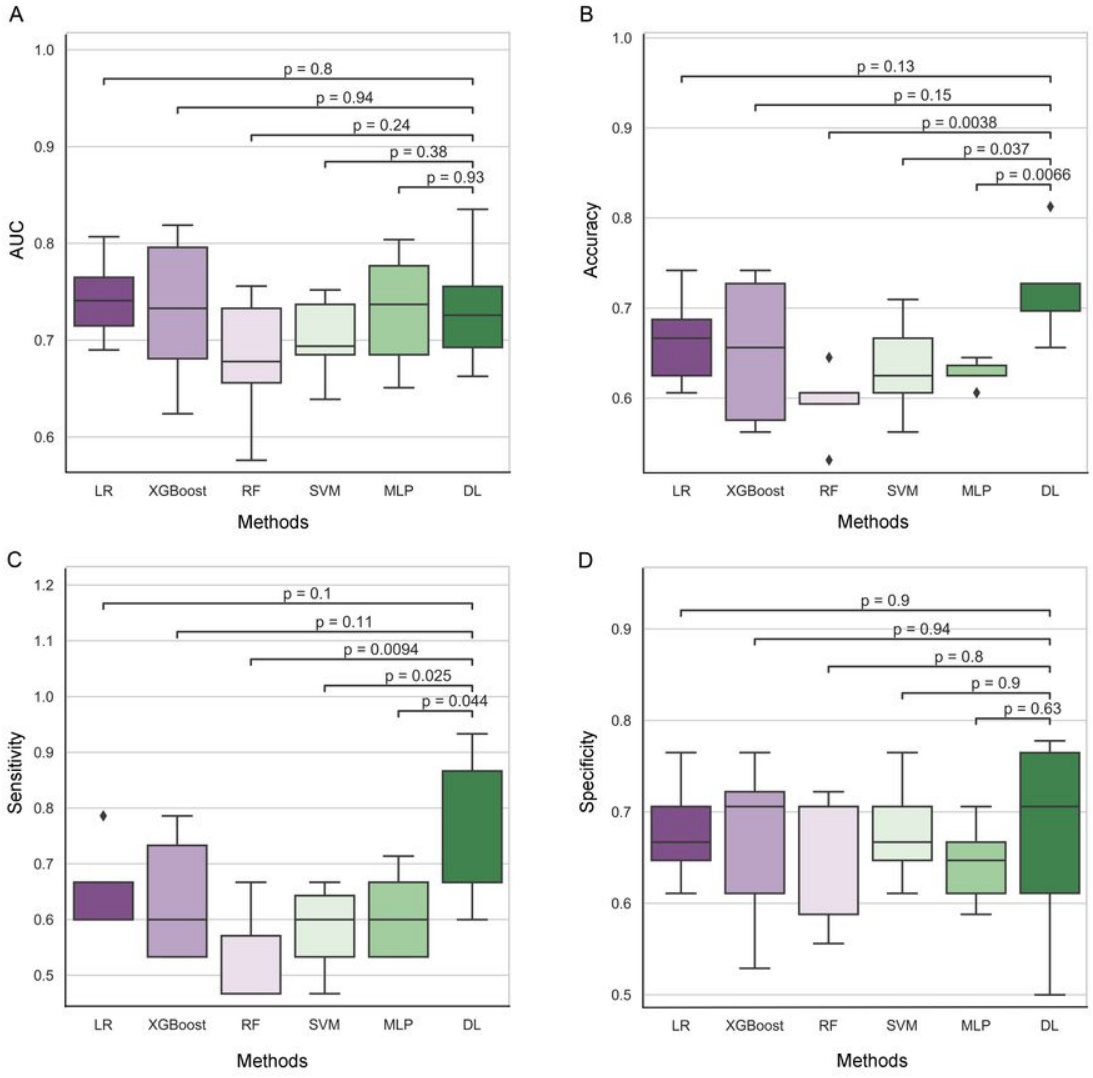


Figure 1

An illustration of the architecture of our 3D DenseNet deep learning model. Images with dimension 160×160×64 pixels are feed into the network, followed by multiple convolution and pooling operations, resulting in probability predictionfor MG. In dense block, features with different levels are concatenated using skip connections. The dimension is halved after each transition layer.



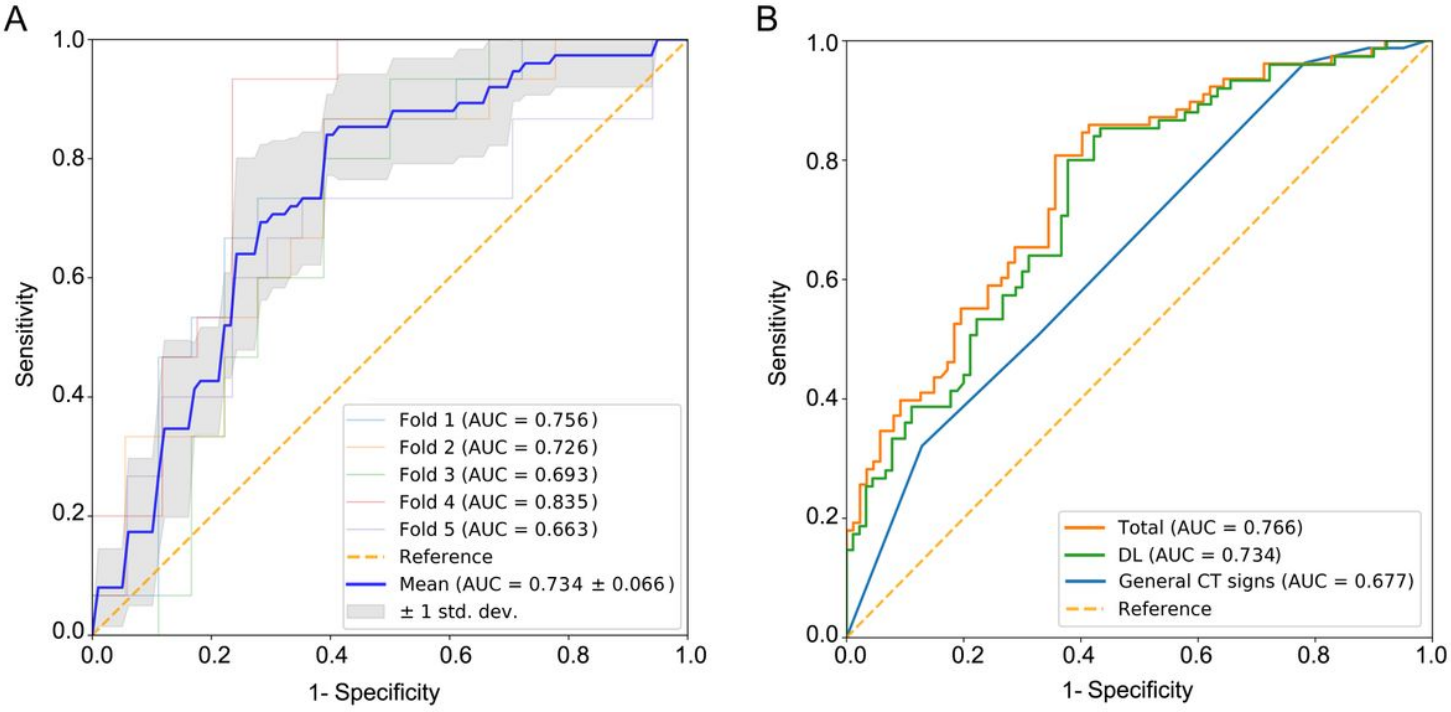
E

Methods	Training Cohort				Validation Cohort			
	AUC	ACC	SN	SP	AUC	ACC	SN	SP
LR	0.834	0.733	0.75	0.718	0.745	0.659	0.637	0.679
XGBoost	1.000	0.998	0.997	1.000	0.731	0.653	0.637	0.667
RF	1.000	0.998	0.997	1.000	0.711	0.64	0.623	0.655
SVM	0.915	0.818	0.78	0.851	0.702	0.634	0.582	0.679
MLP	0.98	0.927	0.919	0.934	0.731	0.628	0.61	0.644
DL	0.892	0.854	0.936	0.754	0.734	0.724	0.787	0.672

Figure 2

Results of Radiomics analysis and 3D DenseNet deep learning model for detecting MG in a cohort of 182 thymoma patients. The performance of five machine learning and 3D-DenseNet-DL model was compared

using Area Under ROC Curve (AUC) (A), accuracy (B), sensitivity (C) and specificity (D). 3D DenseNet deep learning model for detecting MG showed similar results in AUC and specificity, but relatively better results in accuracy and sensitivity comparing to five radiomics analysis models (E). “RF”, “LR” and “DL” refer to “Random Forest”, “Logistic Regression”, and “Deep Learning” respectively; “AUC”, “ACC”, “SN” and “SP” refer to the metrics Area Under ROC Curve, Accuracy, Sensitivity and Specificity, respectively.



C

	Cohort	*AUC*	*ACC*	*SN*	*SP*
3D-DenseNet DL based multi-model	Training	0.766	0.790	0.739	0.801
	External Validation	0.730	0.732	0.700	0.690
3D-DenseNet DL model	Training	0.734	0.724	0.787	0.672
	External Validation	0.704	0.690	0.760	0.710

Figure 3

The prediction metrics of 3D-DenseNet-DL and DL based multi-model. The metrics Area Under ROC Curve (AUC), ACC (accuracy), SN (sensitivity) and SP (specificity) were used to compare the performance of these models.A, the prediction metrics of the deep learning results from training and five-fold cross-validation, a mean AUC of 0.734 ± 0.066 was presented. B, the comparison of three models of general CT signs model, 3D-DenseNet-DLmodel and the comprehensive model (3D-DenseNet-DL based multi-model), with a mean AUC of 0.677, 0.734 and 0.766, respectively. C, values of 3D-DenseNet-DLmodel and 3D-DenseNet-DL based multi-model in external validation, with AUC 0.704, ACC 0.690, SN 0.760 and SP 0.710for DL model, and AUC 0.730, ACC 0.732, SN 0.700 and SP 0.690 for our final 3D-DenseNet-DL based multi-model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiles.docx