Clinical decision support system to predict the efficacy for EGFR-TKIs based on artificial neural network

Xiao Liang
The First Affiliated Hospital of Nanjing Medical University

Runwei Guan
Xi’an Jiaotong Liverpool University

Jiamin Zhu
Jiangyin People’s Hospital

Yue Meng
Affiliated hospital of Nantong University

Jing Zhu
The First Affiliated Hospital of Nanjing Medical University

Yuxiang Yang
Northeastern University

Yanan Cui
The First Affiliated Hospital of Nanjing Medical University

Jiali Dai
The First Affiliated Hospital of Nanjing Medical University

Jun Shao
The First Affiliated Hospital of Nanjing Medical University

Weidong Mao
Jiangyin People’s Hospital

Liting Lv
Affiliated hospital of Nantong University

Dong Shen
Jiangyin People’s Hospital

Renhua Guo (✉ rhguo@njmu.edu.cn)
The First Affiliated Hospital of Nanjing Medical University

Research Article

Keywords: NSCLC, EGFR-TKIs, NGS, ANN
Abstract

Background

The efficacy of epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) was affected by numerous factors. We developed and validated an artificial neural network (ANN) system based on clinical characteristics and next-generation sequencing (NGS) to support clinical decisions.

Methods

A multicenter retrospective non-interventional study was conducted. 196 untreated patients from three hospitals with advanced non-small cell lung cancer (NSCLC) and EGFR mutation were tested by NGS before the first treatment. All patients received formal EGFR-TKIs treatment. Five different models were individually trained to predict the efficacy of EGFR-TKIs based on an independent cohort. Others from two independent cohorts were collected for external validation.

Results

Compared with logistic regression, four machine learning methods show better predicting abilities for EGFR-TKIs. The inclusion of mutations further improved the predictive power in all models. ANN performed best on the dataset with mutations TP53, rb1 and pik3ca. The prediction accuracy on the test set reached 84%, the recall on the test set was about 95% for poor efficacy and ROC was 0.67. In the external validation, ANN still showed the best performance and differentiated patients with poor outcomes. Finally, a clinical decision support software based on ANN was developed and provides a visualization interface for clinicians.

Conclusion

This study provides an approach to assess the efficacy of NSCLC patients with first-line EGFR-TKI treatment. A software is developed to support the clinical decision.

Introduction

The public health burden of lung cancer is continuously growing worldwide which is the leading cause of cancer mortality with an estimated 1.8 million deaths in 2020[1, 2]. Chemotherapy has become the standard treatment for advanced lung cancer since the twentieth century. The Iressa Pan-Asia Study (IPASS) in 2009 marked a watershed for the therapy in non-small cell lung cancer (NSCLC)[3]. Before the IPASS study, patients with advanced NSCLC received unselected chemotherapy. The 1-year survival rate of advanced lung cancer patients with chemotherapy was low. Only 35–40% of these patients were alive at 1 year. IPASS study laid the foundation for epidermal growth factor receptor (EGFR)-tyrosine kinase
inhibitors (TKIs) as a standard first-line regimen for the treatment of advanced lung cancer patients with EGFR mutation. EGFR-TKIs have extended the survival for NSCLC.

Drug resistance greatly limits the usefulness of EGFR-TKIs\cite{4}. Although the response rate of EGFR-TKIs as first-line treatment was even up to 70% and the progression-free survival (PFS) was around 10 months\cite{5}, EGFR-TKIs are still less effective or even ineffective for some patients. EGFR-TKIs combined with chemotherapies or anti-angiogenic therapies have been proved to prolong the PFS and increase toxicity in some patients\cite{6, 7}. The PFS in some studies has also translated into an increased overall survival (OS)\cite{7}. Screening ineffective patients and bringing earlier interventions have great clinical implications.

Next-generation sequencing (NGS) is widely available. It is revealing the high heterogeneity of lung cancer at a molecular level. With massive sequencing, more resistance mutations in EGFR mutated patients were discovered. Many mutations have been proved to be associated with poor efficacy of EGFR-TKIs\cite{8}.

Recently, deep learning (DL) has been widely used in different fields of medicine. Models based on DL have been demonstrated to predict prognosis and risks of various diseases\cite{9}. At the present stage, there are some excellent predictive models to identify the efficacy of drugs\cite{10, 11}. Most of these models are based on radiomics methods. It increases the difficulty of clinical translation. We first attempt to develop and validate a clinical decision support system of EGFR-TKIs based on machine learning and NGS.

**Methods**

**Eligibility Criteria of patients and Study Design**

A total of 196 patients from 3 institutions between December 2014 to April 2021 were enrolled in this study. All patients underwent NGS tests. 158 patients were from the First Affiliated Hospital of Nanjing Medical University as cohort 1. Two independent cohorts with 38 patients were from Affiliated Hospital of Nantong University and Jiangyin people’s hospital as external validation (cohort 2). The study was approved by the Ethics Committee of these three hospitals.

A detailed study design was shown in Fig. 1A. We assumed that the data in the training set and test set were independent identically distributed, therefore, to make the data randomness and robust, the data were shuffled. Then the data was divided into 2 parts in cohort1, 70% for the training set while 30% for the test set. Besides, the ten-fold cross validation was used to evaluate the generalization of the model.

Five different models were trained to distinguish the efficacy of EGFR-TKIs treatment including logistic regression, support vector machine (SVM), random forest (RF), XGBoost and artificial neural network (ANN). The main hyperparameters of machine learning models after fine-tuning are shown in TableS1-S3.

Under normal circumstances, we think that the generalization of traditional machine learning methods is relatively weak, and the solution of nonlinear discrimination problems is weaker than deep learning. Therefore, we also try to use ANN to predict the efficacy. We designed an ANN similar to a multilayer
perceptron, which is shown in Fig. 1B. The input layer is various discrete data of the patient with a bias value. The hidden layer includes a layer of batch normalization, two layers of neurons with bias, a layer of layer normalization, a layer of ReLU activation function, and a layer of sigmoid activation function. The output layer is two discrete floating-point numbers, used to represent the category results. The cross-entropy function is used to calculate the loss value of the artificial neural network for gradient descent and back propagation. The formula of cross entropy function was shown below.

\[ L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \]

In addition, we use momentum-driven stochastic gradient descent as the gradient descent optimizer to backpropagate the artificial neural network. The stochastic gradient descent with momentum is shown below.

\[ v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta) \]
\[ \theta = \theta - v_t \]

The hyperparameters of stochastic gradient descent formula with momentum are shown in Table S4.

**Metrics and Evaluation**

Clinical data included sex, age at diagnosis, smoking status, size and site of tumors, metastatic burden and treatment regimens. PFS which was defined as survival time from the beginning of first-line EGFR-TKIs to disease progression or death was used to assess the efficacy of EGFR-TKIs. Efficacy evaluation indicators of EGFR-TKIs include poor efficacy and good efficacy. The cut-off value of PFS was 9 months referring to the median PFS of EGFR-TKIs in relevant studies[7].

In the machine learning model, the confusion matrix is the basis for the metrics and evaluation. The confusion matrix is shown in Fig. 1C. Based on the confusion matrix, 2 metrics, accuracy and recall, were used to evaluate the performance of different models. Detailed definitions were below.

\[ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]
\[ \text{Recall} = \frac{TP}{TP + FN} \]

In addition, to evaluate the confidence of the model, Receiver Operating Characteristic (ROC) and Area under Curve (AUC) were considered to measure the model in another dimension. For ROC, different thresholds will cause the probability of predicting true positives and false positives to change, and ROC will traverse all thresholds to find the point that makes the highest true positives and the lowest false positives. The area under the ROC curve, we call it AUC, which represents the probability that any pair of
samples (positive and negative) is taken, and the score of the positive sample is greater than the score of the negative sample.

**Statistical analysis**

Chi-squares and Fisher’s exact test were used to analyze relations between categorical variables. Kaplan–Meier survival curves and log-rank tests were used to compare prognosis differences. COX regression model was used to evaluate the Hazard ratios (HRs) and corresponding 95% confidence intervals (CIs). SPSS 23 was used for statistical analysis. R 4.0.2 and Graphpad Prism 7.0 were used to generate plots. Python 3.7 was used to build deep learning models.

**Data availability**

Data are available upon request but may require data transfer agreements.

**Code availability**

https://github.com/GuanRunwei/A-Clinical-Decision-Support-System-of-EGFR-TKIs

**Result**

**Clinical characteristics**

158 patients from the First Affiliated Hospital of Nanjing Medical University. 70% of patients were included as a training set randomly and others were applied as an internal validation set in cohort 1. All patients were treated with first-line TKI-EGFRs.

**Logistic regression and DL**

For predicting the efficacy of EGFR-TKIs and screening the insensitive populations, five different models based on clinical characteristics were individually trained (Dataset1). For logistic regression, the accuracy and recall were 66 and 65 respectively (Fig. 2A, 2B). Four models including SVM, RF and XGBoost, and ANN were also trained. The accuracy of these four models were 0.71, 0.69, 0.71 and 0.75, respectively (Fig. 2A). The recall of four models were 0.86, 0.88, 0.89 and 0.71, respectively (Fig. 2B). ANN, as a more powerful model, had the highest accuracy among five models and not satisfactory recall. For ANN, early stopping was used to avoid overfitting in training process. It is obvious that there are violent oscillations in convergence curve when training the model and the curve did not reach convergence until the 200th epoch (Fig. 2C). Overall, SVM, RF and XGBoost, and ANN worked better to identify the efficacy for EGFR-TKIs than logistic regression.

**Model optimization based on NGS and selection**

Prediction models based on clinical characteristics were unsatisfactory. Genetic mutations could result in the resistance of EGFR-TKIs[8]. Taking into account the stability and universality of the model, top 10 frequency of mutations in NGS tests of our cohort (TP53, RB1, PIK3CA, CTNMB1, RBM10, APC, ATM, LRP1B, SETD2, CDKN2A) were selected which were demonstrated in Fig. 3A. Mutation of tumor
suppressor genes could influence the choice of treatment. HRs of mutations were calculated after adjustment for the covariate of treatment by Cox regression model and shown in forest plot (Fig. 3B). TP53 was the most common mutation (98/158, 62%) (Fig. 3A) in our cohort and had a close relationship with poor efficacy (HR = 1.6, 95%CI: 1.058–2.418, P = 0.022) (Fig. 3B). Other high frequency and malignant potential mutations included RB1 (14/158, 8.9%) (Fig. 3A) and PIK3CA (12/158, 7.6%) (Fig. 3A). These three mutations were also often reported to be associate with resistance to EGFR-TKIs in previous studies[8, 12, 13]. Whether mutations per se or numbers of these mutations could differentiate patients with poor outcomes (Fig. 3C, Fig. 3D). So TP53, RB1 and PIK3CA were included into our model. Clinical characteristics combined with mutated situation and numbers of mutated genes were included into Dataset2 and Dataset3. Compared with Dataset1, the predictability of models was increased in Dataset2 and Dataset3 (except for ACC of dataset2 for XGboost) (Fig. 3E).

ANN had the highest accuracy among models in Dataset2 and Dataset3. We can also find that the overfitting problem of traditional machine learning models is more serious, and the prediction accuracy of the training set is much greater than that of the test set in XGBoost and RF. The fitting ability and generalization of the ANN are somewhat improved compared with the traditional machine learning model. In addition, the negative sample recall rate of the four algorithm models is relatively high, stable at 85–95%. It means that these four models based on NGS can complete the screen of ineffective patients better. So, ANN was selected for further study.

**Optimization of ANN and external validation**

To better predict the efficacy of EGFR-TKIs, we further optimize ANN model. We compared the performance of ANN in Dataset2 and Dataset3 from multi-dimension. The ANN model in Dataset2 showed higher accuracy, recall and AUC than that in Dataset3 (Fig. 4A, Fig. 4B). Moreover, the ANN model trained in Dataset2 reached convergence at 120th epoch earlier than it in Dataset3 at 140th epoch, while the curve was stable and smooth (Fig. 4C). In summary, due to the good performance of the ANN in Dataset2, we selected clinical characteristics and mutated situation based on ANN model to distinguish the efficacy of EGFR-TKIs. For external validation, the accuracy of SVM, RF, XGBoost and ANN was 0.684, 0.763, 0.763 and 0.79. ANN model still showed best performance. In ANN model, patients with good efficacy differed significantly from those with poor efficacy (P = 0.044) (Fig. 4D).

**Clinical decision support system and visualization**

In order to improve the efficiency of medical staff, a clinical diagnosis decision-making client based on Python GUI was developed which was shown in Fig. 5A. The system uses an artificial neural network model to run in the back-end, and the artificial neural network model will provide corresponding treatment recommendations based on the model discrimination results. The prediction procedure is shown in Fig. 5B. Under the prediction procedure, clinical decision support system can recommend treatment for patients.

**Discussion**
The response rate to first-line EGFR-TKIs approaches 70% which is much higher than chemotherapy. However, it still means that almost 30% of patients cannot benefit from EGFR-TKIs. It has clinical implications to predict the efficacy and screening the less effective patients. There are still many challenges developing a practical clinical tool to assess the efficacy of EGFR-TKIs. Therapeutic resistance is a complex and multi-factorial participating process. At the present stage, models based on imaging have demonstrated the ability to predict the efficacy of cancer treatment[14]. These models often lack generalization and are difficult to transduce. So, we attempted to develop models based on clinical characteristics, genotypes and artificial intelligence.

Many clinical characteristics are predictive for EGFR-TKIs. Larger tumors are always thought to be associated with poor prognosis[15]. Patients with concomitant liver metastases, brain metastases and uncommon mutations are also reported to have bad outcomes[16]. Clinical characteristics could determine a degree of the efficacy for EGFR-TKIs but not sufficient.

Co-mutations played an important role in the drug resistant of EGFR-TKIs[17]. TP53 was the most frequent and impactful concurrent mutation in lung cancer with EGFR mutation. The co-mutation rate of TP53 ranged from 54.6–64.6%[18]. The co-mutation of TP53 influenced the natural history of patients with EGFR-mutant and allowed for the diversification of subclone. It prompted the therapeutic resistance. A large number of studies also identified that TP53 can be a negative prognostic marker for the outcomes following EGFR-TKIs[19–21]. As an important early genetic event of EGFR-mutant LUADs, the inactivation of RB1 often harbors TP53 co-alterations and controls the cell cycle with mutation rate of approximately 10%[18, 22]. TP53 and RB1 co-mutations in EGFR-mutant LUAD also increase the risk of small-cell transformation[23]. ATM alterations, IDH1 mutations and PTEN mutations were also reported to be associated with shorter PFS and OS in patients receiving first-line EGFR-TKIs[24]. LKB1/AMPK pathway was shown to reduce sensitivity of EGFR-TKIs in vitro[25]. After analyzing genomic changes, TP53, RB1 and PI3Ka were selected as the final model. Genotypes of patients were demonstrated to increase the stability and accuracy of models. NGS tests also increase the accessibility of models. With the constantly mature technology and gradually decreasing costs of NGS tests, genotypes of most NSCLC patients can be definitive prior to first-line treatment. In conclusion, Genotypes well assisted in predicting the efficacy for EGFR-TKIs.

In the last two decades, ML models have been widely applied in medical and health sciences[26]. It is evident that ML models have improved the understanding of cancer progression[27]. It is the first model to predict the efficacy for EGFR-TKIs based on ANN. ANN can handle classification problems satisfactorily and even serve as a gold standard in some tasks. Hidden layers represent the neural connections in mathematical process (Fig. 1C). However, it is also an important drawback it suffered from. It is difficult to explain the classification process as a “black-box” technology. Additionally, the layered structure could be time-consuming and lead to poor performance for some models. So, epoch of convergence is also an important parameter for our model. The model based on dataset2 showed best performance in this regard.
The risk of progression with different generation EGFR-TKIs is an important research direction. The efficacy of first-line osimertinib is much better than previous EGFR-TKIs[28]. In FLAURA study, the median PFS of osimertinib was up to 18.9 months which was prominently longer than gefitinib or erlotinib. Given the excellent efficacy of osimertinib, the number of untreated patients with osimertinib who have completed the follow-up are limited. We decided to exclude patients with osimertinib from our model. With the large-scale use of Osimertinib in untreated EGFR-mutated NSCLC, the comparison between first-generation EGFR-TKIs or second-generation EGFR-TKIs and third-generation EGFR-TKIs will also be included in next improved model after completing the follow-up. Further work will be dedicated to addressing the selection of more specific individualized treatment including the combinations with anti-angiogenic therapies or chemotherapies and the difference between first-generation EGFR-TKIs or second-generation EGFR-TKIs and third-generation EGFR-TKIs.

There are still some limits in our study. Only first-generation EGFR-TKIs and second-generation EGFR-TKIs are involved into the model. Our new model needs validation for more third-generation EGFR-TKIs. The sample size should also be enlarged in the future to make the model more stable.

**Conclusion**

We describe a DL approach to distinguish the efficacy of patients with EGFR-TKI therapy and have completed clinical translation. The approach can work as a clinical decision support system of EGFR-TKI therapy.

**Declarations**

**Acknowledgements**

None.

**Funding**

None

**Authors’ contributions**

Dong Shen and Renhua Guo designed this study and directed the research group in all aspects. Xiao Liang and Yanan Cui drafted the manuscript. Jiamin Zhu, Yue Meng and Jing Zhu collected the data. Runwei Guan and Yuxiang Yang provided the statistical software, performed the data analysis, Jiali Dai and Jun Shao arranged the Figures and Tables. Liting Lv and Weidong Mao revised the manuscript. All authors read and approved the manuscript

**Ethics approval and consent to participate**
The study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University, affiliated hospital of Nantong University and Jiangyin People’s Hospital. All patients gave written consent to participate.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Availability of data and materials

Data are available upon request but may require data transfer agreements. Codes on shared on github(https://github.com/GuanRunwei/A-Clinical-Decision-Support-System-of-EGFR-TKIs).

References


**Figures**

![Diagram A](image1)

**Training set**
- 158 patients from the First Affiliated Hospital of Nanjing Medical University
- NGS
- Clinical Features
- Top 10 frequency of mutations
- Genes associated with resistance to EGFR-TKIs

**Input variables**
- 1. Clinical Features
- 2. Genotypes

**Model training**
- Machine Learning Strategies:
  - Random Forest
  - Support Vector Machine
  - Xgboost
  - ANN

**Round 1**
- Training fields
- Test fields
**Round 2**
- Training fields
- Test fields
**Round 10**
- Training fields
- Test fields

**Optimal Machine Learning Model**

![Diagram B](image2)

**Input Layer**
- Bias
- X
- BatchNorm
- Hidden Layer
- RelU
- LayerNorm
- Output Layer
- Bias
- X
- RelU
- LayerNorm
- Bias
- X
- RelU
- LayerNorm
- Bias
- X

![Diagram C](image3)

<table>
<thead>
<tr>
<th>Result</th>
<th>Good efficacy</th>
<th>Poor efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>True Positive (TP)</td>
<td>True Negative (TN)</td>
</tr>
<tr>
<td>False</td>
<td>False Positive (FP)</td>
<td>False Negative (FN)</td>
</tr>
</tbody>
</table>
Figure 1


Figure 2

Prediction metrics between logistic regression, machine learning and ANN. A: ACC of different models. B: Recall of different models. C: convergence curve of ANN.
Figure 3

Figure 4

Figure 5

Clinical decision support system and visualization. A: Python GUI was developed and the system will provide corresponding treatment recommendations based on the model discrimination results. B: The hidden prediction procedure of the client.

Supplementary Files

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

- TableS1S4.docx