RESEARCH ARTICLE

Predicting EGFR Mutation in Lung Adenocarcinoma: Development and Validation of the EGFR Mutation Predictive Score (EMPS) in Bali, Indonesia

Edwin Nugroho Njoto[1,2]*, Ida Ayu Jasminarti Dwi Kusumawardani[3], Ida Bagus Ngurah Rai[3]

Abstract

Introduction: The examination of epidermal growth factor receptor (EGFR) mutations may not be routinely available to all patients due to the limited availability and the expensive price of the examination, especially in area with limited resources such as in Indonesia. Therefore, we aimed to build a nomogram to predict the EGFR mutation in patients with lung adenocarcinoma by incorporating significant clinical and radiological parameters. **Methods:** We conducted an age-matched case–control study using 160 treatment-naïve patients [80 patients with EGFR-mutated (*EGFR^mut*) and 80 with *EGFR*-wild-type (*EGFR^wt*)] with pathologically confirmed lung adenocarcinomas with tumor specimens available for genetic analysis taken from 2017 through 2021 in Bali, Indonesia. Radiomics features were extracted from contrast CT images. The cut-off of the tumor diameter was defined using Receiver Operating Characteristic Curve. A conditional logistic regression model was constructed to identify significant risk factors, and a nomogram was developed for predicting the risk of EGFR mutation. A cohort was done to validate the nomogram. **Result:** Being female, never-smoker, having a smaller tumor diameter (<48.5mm), located in the upper lobe, have bubble-like lucency and air-bronchogram in the chest CT scan were identified as independent risk factors of EGFR mutation at the multivariate logistic regression model. The forming normogram model produced an area under the curve of 0.993 (95 % CI = 0.98−1.00) and 0.91 (95 % CI = 0.84−0.99) in development and validation group, respectively. The calibration curve showed good agreement between predicted and actual probability. At the cut-off point of the normogram score 246 shows a sensitivity of 97.5%, a specificity of 98.8%, a positive predictive value of 99.0%, and a negative predictive value of 96.8%. **Conclusion:** Our study indicated that the EGFR Mutation Normogram could provide a non-invasive way to predict the risk of EGFR mutation in patients with lung adenocarcinoma in clinical practice. This normogram need to be validated in other area in Indonesia.

Keywords: Nomograms- adenocarcinoma of Lung- EGFR mutation- Indonesia

Introduction

Lung cancer is the leading cause of cancer death worldwide. Among various types of lung cancers, adenocarcinoma is the most common histologic subtype of lung cancer in men and women. The success of epidermal growth factor receptor (*EGFR*) inhibitor therapies in patients with lung adenocarcinoma who have *EGFR* mutations initiated the era of targeted therapy for advanced cancers and shifted treatment from platinum-based chemotherapy to molecularly targeted therapy.

EGFR is a trans-membrane signaling receptor responsible for cell growth and survival. In normal condition, *EGFR* work only when it binds to ligands or external stimuli such as epidermal growth factor. However, *EGFR* mutated tumor constantly activates tumor growth and metastasizes regardless of any stimuli (Sukauichai et al., 2022). Mutation in *EGFR* is accompanied with a dysregulation in *EGFR* tyrosine kinase (TK) activity which is closely associated with the late disease stage and poor progression of NSCLC (Omar et al., 2022). Several mutations clustered around the TK domain, such as point mutation G719X (G719A, G719C, G719S) in exon 18, in-frame deletions and in-frame insertions in exon 19, point mutation T790M and insertions in exon 20, and point mutations (L858R and L861Q) in exon 21,

[1]Medical Study Program, Faculty of Medicine and Health, Sepuluh November Institute of Technology, Surabaya, East Java, Indonesia. [2]Department of Internal Medicine, Faculty of Medicine, Udayana University / Prof IGNG Ngoerah General Hospital, Denpasar, Bali, Indonesia. [3]Department of Pulmonary Medicine, Faculty of Medicine, Udayana University / Prof IGNG Ngoerah General Hospital, Denpasar, Bali, Indonesia. *For Correspondence: edwin_n_njoto@yahoo.com

have been found as the most commonly found mutations in the *EGFR* Gene.

In 2004, *EGFR* mutations were found to be sensitive to targeted therapies called tyrosine kinase inhibitors (TKIs). Therefore, chimeric monoclonal antibodies (panitumumab and cetuximab) and TKIs (gefitinib, erlotinib, and afatinib) have been developed strategies that target *EGFR*. In particular, these treatment are able to compete with ATP (adenosine triphosphate) binding site inhibitors at the active site of *EGFR* kinase, therefore preventing and blocking vital *EGFR* pathways.(Liu et al., 2017) TKI has been a proved treatment for NSCLC with *EGFR* mutation in Indonesia (Kementerian Kesehatan, 2017).

Since the detection of *EGFR* mutations is crucial in the treatment of lung adenocarcinoma, this examination is routinely examined in all lung adenocarcinoma patients in Indonesia. However, the examination of *EGFR* mutations both in tissue and in circulation requires special technology such as Polymerase Chain Reaction (PCR) examination which is not easy to obtain in areas with limited resources in Indonesia. In addition, the cost of the examination is paid by pharmacy and patients. These limitations shows the importance of finding an easy, cheap, precise method or scoring in detecting *EGFR* mutations in lung adenocarcinoma patients. With regard to cost effectiveness and early planning of treatment, prediction of gene mutation from patients' characteristics and CT findings may be valuable. In this study, we aimed to develop and to validate normogram prediction EMPS (*EGFR* Mutation Predictive Score) that utilize the differences in patients' characteristics and CT findings between *EGFR*-mutated and nonmutated lung adenocarcinoma.

Participants
Study Group (EMPS model)

A total of 409 patients with lung adenocarcinoma who underwent *EGFR* mutation tests between January 2017 and January 2021 in Prof Dr. IGNG Ngoerah Hospital (Denpasar, Bali) were initially enrolled in this study, from whom 80 cases (patients with lung adenocarcinoma with *EGFR* mutation) were selected according to the following inclusion criteria: patients with pathologically confirmed diagnosis of lung adenocarcinoma and has previously underwent *EGFR* mutation test; patients with thin-section Computed Tomography (CT) images with contrast accessible; and has not received any chemotherapy or radiotherapy or any therapy. Then, 80 out of 329 patients with wild type (negative *EGFR* mutation status) lung adenocarcinoma were individually selected and matched with each case for age (± 5 years) and geographical location (regency) of the case's residence. These 80 patients were then defined as controls. Finally, 160 ethnically Balinese patients (80 as cases and 80 as controls) were reserved for analysis (Supplement 1).

Validation Cohort Group

We conducted a cohort study in June 2023 at Prof Dr. IGNG Ngoerah Hospital, Denpasar, Bali. 48 patients were selected according to the following inclusion criteria: patients with pathologically confirmed diagnosis of lung adenocarcinoma and has previously underwent *EGFR*

mutation test in June 2023; patients with thin-section Computed Tomography (CT) images with contrast accessible; and has not received any chemotherapy or radiotherapy or any therapy. From the 48 patients, we found 25 patients with *EGFR* mutation and 23 patients without *EGFR* mutation (Supplement 1).

Materials and Methods

Smoking status was defined as never smoker and ever smoker. All CT clinical findings, such as the largest tumor diameter, air-bronchogram, bubble-like lucency, were being measured or determined by two pulmonologist from the thorax CT-Scan with contrast. The cut-off for the tumor diameter is determined using a receiver operating characteristic (ROC) curve and the Youden's index (Liu et al., 2016, Zou et al., 2017). Tumor location in upper right, middle right, and upper left lobes were categorized as upper lobe. Those in lower right and lower left lobes were categorized as lower lobe. Brain metastasis was determined using brain CT-scan with contrast.

The calibration of the EMPS was assessed by comparing the actual observed risk and the average probability of *EGFR* mutation by the score. The Hosmer-Lemeshow test was used to assess the corresponding goodness-of-fit. The Harrell's concordance index (C-index) was used to assess the score's discrimination ability. Since the EMPS was derived based on a conditional logistic regression model, C-index values and the corresponding 95% confidence intervals (CIs) were estimated for each groups.

For external validation, the EMPS was computed with the validation data, and score performance was assessed and compared to newly diagnosed adenocarcinoma patients who will have their *EGFR* status being checked. To corroborate the results observed in the derivation and validation sets, a confirmatory analysis was carried out by estimating the Area Under Curve (AUC) of the ROC curve (AUROC) of the EMPS for predicting *EGFR* Mutation. This also allowed us to identify the best score cut-points, which maximized the sensitivity, specificity and predictive values. In addition, calibration plots were being drawn to provide a detailed view on calibration by comparing observed and predicted outcomes among patients with the same estimated risk. The observed outcome proportions and estimated risks by a particular time point of interest are plotted against each other, with deviations from the diagonal signalling miscalibration.

Statistical Analysis

All statistical analyses were performed using SPSS 16.0. For bivariate analysis, McNemar Test was used for categorical variables and those variables have P > 0.25 was excluded from the analysis. Subsequently, multivariate analysis was performed to establish a conditional logistic regression model with stepwise selection of variables. As per stepwise selection, effects were entered into and removed from the model. Thus, one or more backward elimination steps could follow each forward selection step. At each forward selection step, if it was significant at the P=0.05 level, the corresponding effect was added to the model. Meanwhile, results of the Wald test for

individual parameters were examined at each backward elimination step. The least significant effect not meeting the P=0.05 level was removed. The stepwise selection process terminated when no further effect could be added to the model or when the current model was identical to a previously visited model. Then, the predictive normogram EMPS (*EGFR* Mutation Predictive Score) was created. The cut-off for the EMPS model is determined using a receiver operating characteristic (ROC) curve and the Youden's index.

Results

Study Population

A total of 208 patients were enrolled, including 160 in the EMPS model (median age [minimum-maximum], 58 [23-89] years old, 74 males and 86 females) and 48 in the validation cohort (median age [minimum-maximum], 60,00 [23 -79] years, 27 males and 21 females) (Supplement 1). Baseline characteristics of patients can be seen in Table 1. Patients' characteristics and CT findings being assessed are gender, age group, place of residence, smoking status, cerebral metastases, family history of lung cancer, history of other lung diseases, tumor diameter, tumor location, the presence of bubble-like lucency (BBL) and air bronchogram on computed tomography examination.

This study is an observational analytic study using a case-control approach in pairs (pairs based on the patient's age and district of residence). The majority of research subjects were over 50 years old (83.75%), female (53.75%), never smoked (53.75%) and lived in Denpasar city (31.75%) (Table 1).

The median diameter of the largest dimension of primary lung tumor found was 49.50 mm with the lowest value being 16 mm and the highest value being 140 mm. Receiver Operating Characteristic (ROC) curve analysis and Youden index determination are performed to determine the cut-off point to be used. The value of the area under the curve (AUC) was 0.827 (95% Confidence Index [95%CI]: 0.758-0.869). The Youden index value found is 48.5 mm. Therefore, the diameter of the tumor in this study was divided into two, namely diameter tumors below 48.5 mm and above 48.5 mm. In this study, 54.37% of the research samples were obtained with a diameter greater than 48.5 mm.

Cerebral metastases were found in 13.75% of the samples and a family history of lung cancer was found in 1.25% of the samples. Asthma and COPD were found as the most common pulmonary comorbidities in 1.88% of the sample. On computed tomography examination, the majority of tumors were found in the upper lobe (65.63%), bubble-like lucency (53.13%), and air-bronchogram (53.75%).

This study also discusses the inter-observer comparison of computed tomography interpretations. The concordance between the two observers at the tumor location, the presence of a bubble-like lucency, and the presence of an air-bronchogram were almost perfect with a coefficient of K varying between 0.85 and 1.0.

EMPS Prediction Normogram Construction

Univariate analysis of the EMPS showed that clinical factors with statistically significant differences in *EGFR* mutation status included sex (female) (OR, 28.00; 95% CI: 11.85, 66.15; P < 0.001) and smoking history (never smoker) (OR, 55.75; 95% CI: 20.34, 142.76; P < 0.001), and CT features factors included long- axis diameter (OR, 31.56; 95% CI: 13.04, 76.37; P < 0.001), location (Upper Lobe) (OR, 6.59; 95% CI: 3.10, 14.01, P< 0.001), bubble-like lucency (OR, 11.24; 95% CI: 5.36, 23.56, P < 0.001), and air bronchogram (OR, 69.89; 95% CI: 24.86, 196.51, P < 0.001), whereas cerebral metastasis and family history of lung cancer were not related to *EGFR* mutation status (P > 0.250) (Table 2).

Multivariate logistic regression analysis was performed on variables with statistical differences in the univariate analysis and found that being female (OR, 13.22; 95%
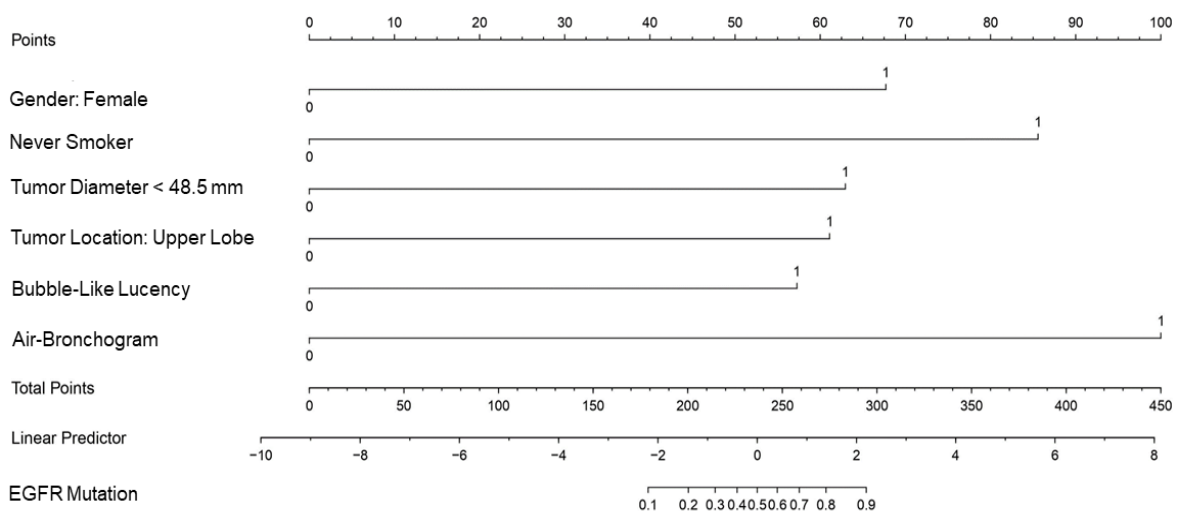

Figure 1. Nomogram for Prediction of EGFR Mutation Status Risk and Its Predictive Performance. The nomogram was constructed in the EMPS model combining the patients' characteristic, including gender and the smoking history, and CT features, including tumor location, tumor diameter, bubble-like lucency, and air-bronchogram. CT, computed tomography; EGFR, epidermal growth factor receptor; EMPS, EGFR Mutation Predictive Score.
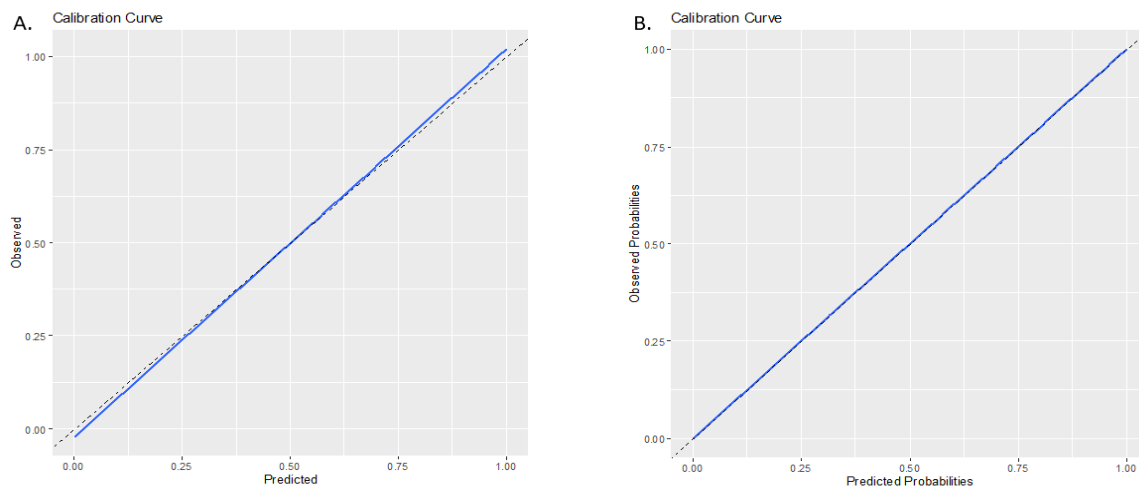
Figure 2. Calibration Curve of the Nomogram in the EMPS (a) and Validation (b) Cohorts. The x-axis represents the predicted EGFR mutation risk. The y-axis represents the actual EGFR mutation rate. The green line represents a perfect estimated mutation rate by an ideal model. The black line represents the performance of the nomogram, in which a closer fit to the blue line represents a better prediction. EGFR, epidermal growth factor receptor; EMPS, EGFR Mutation Predictive Score.

CI: 1.68, 104.11, P=0.014), never smoker (OR, 26.13; 95% CI: 2.76, 247.66; P=0.004), having a smaller tumor diameter (OR, 11.02; 95% CI: 1.12, 108.17, P=0.039), having tumor on upper lobe (OR, 10.27; 95% CI: 1.06, 99.40, P=0.044), having bubble-like lucency (OR, 8.8; 95% CI: 1.21, 65.33, P=0.032), and having air-bronchogram (OR, 45.27; 95% CI: 4.94, 414.66, P=0.001) were independent risk factors related to *EGFR* mutation status (Table 2). Based on the results of multiple logistic regression analysis, we used six independent risk factors to generate an individualized nomogram to predict the *EGFR* mutation status (Figure 1).

Calibration and Validation of the EMPS model

In the EMPS and the validation cohorts, the 95% CI of the calibration curve did not cross the di- agonal line (Figure 2a and 2b). Therefore, the predicted probability of the nomogram model was consistent with the actual probability, which indicated that the model had good consistency. In the EMPS model, the AUC of the EMPS nomogram was 0,993 (95% CI: 0.830, 1.000) (Figure 3a). The AUC of the external validation cohort was 0.919 (95% CI: 0.843, 0.995) (Figure 3b).

Determining the cutoff value for predicting EGFR mutation

The values of sensitivity and specificity and the predictive values of the predicted probability under different cutoff values of the nomogram are shown in Table 3. A lower cutoff value resulted in a higher sensitivity and negative predictive value, and the lower specificity and positive predictive value increased. The diagnostic odds ratios of the nomogram at different cutoff values are shown in Table 3. The cutoff values for good performance of the nomogram ranged between ≥0.1 and ≥0.5. Finally, the optimal cutoff value determined by
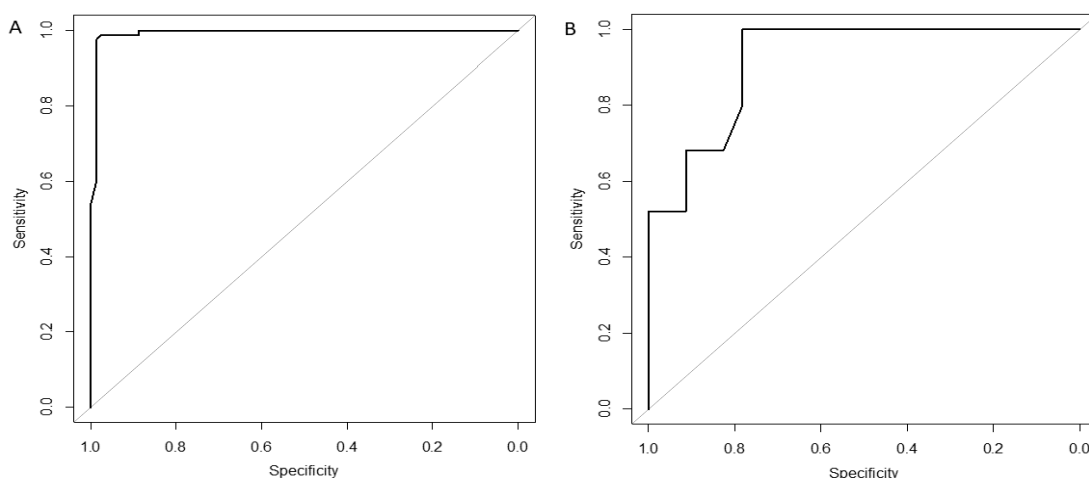

Figure 3. ROC Curves of the Predictive *EGFR* Mutation Status in the EMPS (a) and Validation (b) Cohorts. The AUC of the nomogram in the EMPS was AUC 0,993 (95% CI: 0,83-1,00) (Fig. 3a), and the AUC in the validation cohort was 0.92 (95% CI: 0.84-1.00) (Figure 3b). AUC, area under the curve; EGFR, epidermal growth factor receptor; ROC, receiver operating characteristic.

Table 1. Baseline Characteristics of Patients with *EGFR*-mutated (*EGFR^{mut}*) and 80 with EGFR-wild-type (*EGFR^{wt}*)] Included in the EMPS Study and in the External Validation Datasets

Patients' Characterictics	Derivation Set (EMPS patients) (n=160)		External Validation Set (n=48)	
	EGFR^{mut}	*EGFR^{wt}*	*EGFR^{mut}*	*EGFR^{wt}*
Age Group				
30-39 years old	2 (2.50%)	2 (2.50%)	2 (8.00%)	0 (0.00%)
40-49 years old	11 (13.75%)	11 (13.75%)	2 (8.00%)	2 (8.70%)
50-59 years old	31 (38.75%)	31 (38.75%)	6 (24.00%)	10 (43.48%)
60-69 years old	20 (25.00%)	20 (25.00%)	11 (44.00%)	8 (34.78%)
70-79 years old	12 (15.00%)	12 (15.00%)	4 (16.00%)	3 (13.04%)
80-89 years old	4 (5.00%)	4 (5.00%)	0 (0.00%)	0 (0.00%)
Total	80	80	25	23
Residence				
Badung	10 (12.50%)	10 (12.50%)	3 (12.00%)	4
Bangli	2 (2.50%)	2 (2.50%)	2 (8.00%)	1 (4.35%)
Buleleng	11 (13.75%)	11 (13.75%)	2 (8.00%)	3 (13.04%)
Denpasar	25 (31.25%)	25 (31.25%)	8 (32.00%)	7 (17.39%)
Gianyar	10 (12.50%)	10 (12.50%)	1 (4.00%)	3 (13.04%)
Jembrana	5 (6.25%)	5 (6.25%)	1 (4.00%)	1 (4.35%)
Karangasem	7 (8.75%)	7 (8.75%)	0 (0.00%)	3 (13.04%)
Klungkung	1 (1.25%)	1 (1.25%)	4 (16.00%)	0 (0.00%)
Tabanan	9 (11.25%)	9 (11.25%)	4 (16.00%)	1 (4.35%)
Total	80	80	25	23
Cerebral Metastasis				
Yes	16 (20%)	6 (7.50%)	0 (0.00%)	0 (0.00%)
No data available	64 (80%)	74 (92.50%)	25 (100.00%)	23 (100.00%)
Total	80	80	25	23
Pulmonary Comorbidity				
Ashtma	1 (1.25%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
COPD	1 (1.25%)	1 (1.25%)	0 (0.00%)	0 (0.00%)
None	78 (97.5%)	79 (98.75%)	25 (100.00%)	23 (100.00%)
Total	80	80	25	23
Gender				
Female	70 (87.50%)	16 (20.00%)	11 (44.00%)	10 (43.48%)
Male	10 (12.50%)	64 (80.00%)	14 (56.00%)	13 (56.52%)
Total	80	80	25	23
Smoking Status				
Never smoker	73 (91.25%)	13 (16.25%)	21 (84.00%)	1 (4.35%)
Ever smoker	7 (8.75%)	67 (83.75%)	4 (16.00%)	22 (95.65%)
Total	80	80	25	23
Tumor diameter: Median (Min-Max)	50.6 mm (16-140)			
< 48.5 mm	64 (80.00%)	9 (11.25%)	15 (60.00%)	8 (34.78%)
> 48.5 mm	16 (20.00%)	71 (88.75%)	10 (40.00%)	15 (65.22%)
Total	80	80	25	23
Tumor Location				
Upper Lobe	68 (85.00%)	37 (46.25%)	20 (80.00%)	14 (60.87%)
Lower Lobe	12 (15.00%)	43 (53.75%)	5 (20.00%)	9 (39.13%)
Total	80	80	25	23

Min, Minimum; Max, Maximum; *EGFR^{mut}*, EGFR mutated; *EGFR^{wt,}* EGFR wild type; COPD, Chronic Obstructive Pulmonary

Table 1. Continued

Patients' Characterictics	Derivation Set (EMPS patients) (n=160)		External Validation Set (n=48)	
	EGFR^{mut}	*EGFR^{wt}*	*EGFR^{mut}*	*EGFR^{wt}*
Bubble-like lucency				
Yes	64 (80.00%)	21 (26.25%)	24 (96.00%)	7 (30.43%)
No	16 (20.00%)	59 (73.75%)	1 (4.00%)	16 (69.57%)
Total	80	80	25	23
Air-Bronchogram				
Yes	74 (92.50%)	12 (15.00%)	23 (92.00%)	7 (30.43%)
No	6 (7.50%)	68 (85.00%)	2 (8.00%)	16 (69.57%)
Total	80	80	25	23
Family history of Lung Cancer				
Yes	2 (2.50%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
No	78 (97.50%)	80 (100.00%)	25 (100.00%)	23 (100.00%)
Total	80	80	25	23

Min, Minimum; Max, Maximum; *EGFR^{mut}*, EGFR mutated; *EGFR^{wt}.* EGFR wild type; COPD, Chronic Obstructive Pulmonary

Youden's method was 246 (sensitivity: 97.5%, specificity: 98.8%, positive predictive value: 99.0%, negative predictive value: 96.8%) (Table 3).

Discussion

In this study, we developed and validated a nomogram based on patients' characteristic, such as gender and smoking status, and CT features including tumor diameter,

Table 2. Univariate and Multivariate Conditional Logistic Regression Models, Including 80 Patients with *EGFR*-mutated (*EGFR^{mut}*) and 80 with EGFR-Wild-Type (*EGFR^{wt}*)

Characteristics	*EGFR^{mu}*	*EGFR^{wt}*	Univariate analysis OR (95%CI)	Multivariate analysis OR (95%CI)
Gender				
Female	70	16	28.00 (11.85-66.15)*.a	13.22 (1.68-104.11)**
Male	10	64	Reference	Reference
Smoking status				
Never smoker	73	13	55.75 (20.34-142.76)*.a	26.13 (2.76-247.66)**
Ever smoker	7	67	Reference	Reference
Cerebral Metastasis				
Yes	16	6	N/A	N/A
No data available	64	74	N/A	N/A
Tumor Diameter				
< 48.5 mm	64	9	31.56 (13.04-76.37)*	11.02 (1.12-108.17)**
> 48.5 mm	16	71	Reference	Reference
Tumor Location				
Upper Lobe	68	37	6.59 (3.10-14.01)*	10.27 (1.06-99.40)**
Lower Lobe	12	43	Reference	Reference
Bubble-like lucency				
Yes	64	21	11.24 (5.36-23.56)*	8.88 (1.21-65.33)**
No	16	59	Reference	Reference
Air-Bronchogram				
Yes	74	12	69.89 (24.86-196.51)*	45.27 (4.94-414.66)**
No	6	68	Reference	Reference
Family History of Lung Cancer				
Yes	2	0	1 (0.06-16.27)	N/A
No	78	80	Reference	N/A

OR, Odd Ratio ; 95% CI, 95% Confidence Interval; *EGFR^{mut}*, EGFR mutated; *EGFR^{wt}*, EGFR wild type; * P < 0.25 and McNemar ; ** P < 0.05 and Conditional Logistic Regression Test

Table 3. Accuracy Test of Nomogram Generated from Multivariate Conditional Logistic Regression Models, Including 80 Patients with *EGFR*-mutated (*EGFR^mut*) and 80 with *EGFR*-wild-type (*EGFR^wt*)

Point	Prediction	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
162.5	<10%	100.00	88.80	92.00	100.00
237	49%	98.80	97.50	98.10	98.40
246	59%	97.50	98.80	99.00	96.80
284	86%	90.00	98.80	98.90	87.40
360	>90%	75.00	98.80	98.70	69.90
408	>90%	42.50	100.00	100.00	36.40

tumor location, bubble-like lucency, and air-bronchogram, for personalized prediction of *EGFR* mutation status. The nomogram development case control included 160 patients with lung adenocarcinoma. To further validate the performance of this model, we evaluated it in an independent external validation cohort that included 48 cases. The AUC of the model in the development and validation cohorts was 0.993 and 0.919, respectively.

Even though *EGFR* examination is crucial in the management of all lung adenocarcinoma patients in Indonesia, the detection of *EGFR* mutations in practive has various obstacles due to the limited material or adequate *EGFR* test samples (Zhao et al., 2017), the need of advanced examination that may not available in areas with limited resources in Indonesia, and its expensive cost. Although the treatment of tyrosine kinase inhibitor (TKI) is borne by the BPJS, currently the cost of examining the detection of *EGFR* mutations is still being borne by the pharmacy and the patient at the Prof. IGNG Ngoerah Central General Hospital (RSUP). This limitation shows the importance of finding an easy, cheap, accurate method or scoring in detecting *EGFR* mutations in lung adenocarcinoma patients. One of them is by analyzing risk factors in predicting *EGFR* mutation status and developing a normogram of significant factors. By focusing on the distinction between between *EGFR* wild types and *EGFR*-mutated types, developing normogram should be helpful and useful for clinical treatment.

For the construction of the nomogram, we carried out multiple logistic regression analyses on the variables that showed statistical differences in univariate analysis. We then selected the variables that showed significant differences in the univariate analysis in the final model. Initially, we used 8 candidate variables, including 5 CT features and 3 clinical variables. Due to inadequacy of data on cerebral metastasis, this variable was excluded from the univariate analysis. Hasegawa et al., (2016), Zhao et al., (2017), Zhang et al., (2019) and Liu et al., (2016) used the clinical and pathological data of patients with lung adenocarcinoma to establish a nomogram model to predict the *EGFR* mutation status, and the AUC of the model in the independent validation dataset varied between 0.741-0.784. However, all patients included in this study were from outside Indonesia, so this model may not suitable for Indonesian patients. By combining the variables contained in the model that were easily obtained before operation and observed by the naked eye from CT scan result. Therefore, our model is easier to apply to clinical practice in Indonesia.

One of the most important reasons for using a nomogram is that it can explain the needs of individual treatment or care needs. However, the clinical consequences of a specific level of discrimination or mis- alignment cannot be captured by performance, discrimination, and cal- ibration of risk prediction (Van Calster and Vickers, 2015, Collins et al., 2015). Therefore, in order to demonstrate the usefulness of a nomogram, a decision curve was used to evaluate whether the decision of a nomogram model would improve the prognosis of patients. This novel approach analyzed clinical consequences based on threshold probabilities and derived net benefits from it (Vickers and Elkin, 2006, Balachandran et al., 2015). The AUC of the ROC curve was 0.993, indicating good predictive ability. We used a cross validation model in a cohort of patients, and the final prediction accuracy was 91%. According to the Youden index and diagnostic odds ratios, we assigned an optimal cutoff value of 246 (Table 3). According to our results, female and not smoking patients who has tumor located in the upper lobe, tumor diameter less than 48.5 mm, air-bronchogram and bubble-like lucency were most likely associated with a higher probability to have *EGFR* mutation.

To the best our knowledge, this is the first study describing a comprehensive scoring system to predict *EGFR* mutation in Bali, Indonesia. Nonetheless, there are some limitations in this study. First, this study is limited to Balinese populations only. Second, patients with other mutation subtypes, such as KRAS mutation or TP53 mutation, were not included in our study, and larger patient cohort studies are required to confirm our observation. Finally, the scoring system obtained from the retrospective analysis in this study needs to be further confirmed by prospective studies including nonsurgical candidates and different ethnic populations to determine whether this model can be used for treatment decision-making without molecular profiling in Indonesia.

In conclusion, we constructed and validated a nomogram for predicting *EGFR* mutation in lung adenocarcinoma patients. The proposed nomogram considered six independent risk factors: female gender, not smoker, a smaller tumor diameter, tumor located in the upper lobe, air-bronchogram, and bubble-like lucency. We have confirmed the precise calibration and excellent discrimination power of our nomogram. The predictive power of this nomogram may be improved by considering other potential important factors that we could not be obtained from are with limited resouces such as Indonesia, and also by external validation.

Author Contribution Statement

Acknowledgements

References

Balachandran VP, Gonen M, Smith JJ, DeMatteo RP (2015). Nomograms in oncology: more than meets the eye. *Lancet Oncol*, **16**, 173-80.

Collins GS, Reitsma JB, Altman DG, Moons KGM (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br Med J*, **350**, g7594.

Hasegawa M, Sakai F, Ishikawa R, et al (2016). CT Features of Epidermal Growth Factor Receptor–Mutated Adenocarcinoma of the Lung: Comparison with Nonmutated Adenocarcinoma. *J Thorac Oncol*, **11**, 819-26.

Kementerian Kesehatan (2017). Panduan Penatalaksanaan Kanker Paru. Kementerian Kesehatan Republik Indonesia. http://www.kanker.kemkes.go.id/guidelines/PNPKParu.pdf.

Liu X, Wang P, Zhang C, Ma Z (2017). Epidermal growth factor receptor (EGFR): A rising star in the era of precision medicine of lung cancer. *Oncotarget*, **8**, 50209-20.

Liu Y, Kim J, Qu F, et al (2016). CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology*, **280**, 271-80.

Omar A, Bahnassy AA, Gaafer R, Desouky EDE, Medhat AM (2022). Assessing Mutations in Treatment Response-related Genes in Egyptian Patients with Non-small Cell Lung Cancer. *Asian Pac J Cancer Biol*, **7**.

Sukauichai S, Maneenil K, Supavavej A, et al (2022). EGFR Mutation-positive Lung Cancer in Real-world Treatment Outcomes: A Multicenter Study from Thailand. *Asian Pac J Cancer Care*, **7**, 643-50.

Van Calster B,Vickers AJ (2015). Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making*, **35**, 162-9.

Vickers AJ, Elkin EB (2006). Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decis Making*, **26**, 565-74.

Zhang H, Cai W, Wang Y, Liao M, Tian S (2019). CT and clinical characteristics that predict risk of EGFR mutation in non-small cell lung cancer: a systematic review and meta-analysis. *Int J Clin Oncol*, **24**, 649-59.

Zhao J, Dinkel J, Warth A, et al (2017). CT characteristics in pulmonary adenocarcinoma with epidermal growth factor receptor mutation. *PLoS One*, **12**, e0182741-e.

Zou J, Lv T, Zhu S, et al (2017). Computed tomography and clinical features associated with epidermal growth factor receptor mutation status in stage I/II lung adenocarcinoma. *Thorac Cancer*, **8**, 260-70.