European Journal of Radiology Open 13 (2024) 100601

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Non-invasive, fast, and high-performance EGFR gene mutation prediction method based on deep transfer learning and model stacking for patients with Non-Small Cell Lung Cancer

Anass Benfares [a], Abdelali yahya Mourabiti [b], Badreddine Alami [b], Sara Boukansa [c], Nizar El Bouardi [b], Moulay Youssef Alaoui Lamrani [b], Hind El Fatimi [d], Bouchra Amara [e], Mounia Serraj [e], Smahi Mohammed [f], Cherkaoui Abdeljabbar [g], El affar Anass [h], Mamoun Qjidaa [g], Mustapha Maaroufi [b], Ouazzani Jamil Mohammed [i], Qjidaa Hassan [i,*]

[a] Laboratory of Computer, Signals, Automation and Cognitivism, Dhar El Mehraz Faculty of Sciences, Sidi Mohammed Ben Abdellah University, Fez, Morocco
[b] Radiology Department of University Hospital Center Hassan II Fez, Sidi Mohammed Ben Abdellah University, Fez, Morocco
[c] Laboratory of Anatomic Pathology and Molecular Pathology, University Hospital Center Hassan II, Sidi Mohammed Ben Abdellah University, Fez, Morocco
[d] Anatomopathological Department, University Hospital Center Hassan II, Sidi Mohamed Ben Abdellah University, Fez, Morocco
[e] Pneumology Department, University Hospital Center Hassan II, Sidi Mohamed Ben Abdellah University, Fez, Morocco
[f] Thoracic Surgery Department, University Hospital Center Hassan II, Sidi Mohamed Ben Abdellah University, Fez, Morocco
[g] Laboratoire de Technologies Innovantes, Abdelmalek Essaidi University, Tanger, Morocco
[h] Sidi Mohamed Ben Abdellah University, Fez, Morocco
[i] Laboratory of Intelligent Systems, Energy and Sustainable Development Faculty of Engineering Sciences, Private University of Fez, Fez, Morocco

ARTICLE INFO

Keywords:
EGFR gene mutation
Convolutional Neuron Network
Lung tumor
Deep transfer learning
Stacking model
CT image

ABSTRACT

Purpose: To propose an intelligent, non-invasive, highly precise, and rapid method to predict the mutation status of the Epidermal Growth Factor Receptor (EGFR) to accelerate treatment with Tyrosine Kinase Inhibitor (TKI) for patients with untreated adenocarcinoma Non-Small Cell Lung Cancer.
Materials and methods: Real-world data from 521 patients with adenocarcinoma NSCLC who performed a CT scan and underwent surgery or pathological biopsy to determine EGFR gene mutation between January 2021 and July 2022, is collected. Solutions to the problems that prevent the model from achieving very high precision, namely: human errors made during the annotation of the database and the low precision of the output decision of the model, are proposed. Thus, among the 521 analyzed cases, only 40 were selected as patients with EGFR gene mutation and 98 cases with wild-type EGFR.
Results: The proposed model is trained, validated, and tested on 12,040 2D images extracted from the 138 CT scans images where patients were randomly partitioned into training (80 %) and test (20 %) sets. The performance obtained for EGFR gene mutation prediction was 95.22 % for accuracy, 960.2 for F1_score, 95.89 % for precision, 96.92 % for sensitivity, 94.01 % for Cohen kappa, and 98 % for AUC.
Conclusion: An EGFR gene mutation status prediction method, with high-performance thanks to an intelligent prediction model entrained by highly accurate annotated data is proposed. The outcome of this project will facilitate rapid decision-making when applying a TKI as an initial treatment.

1. Background

Lung cancer is one of the deadliest cancers in the world. The non-small-cell lung cancer (NSCLC) represents 85 % of different types of lung cancer [1–3]. Although chemotherapy is the key to controlling cancer, its success has been limited due to many difficulties [4]. To avoid conventional chemotherapy, researchers have proposed new treatments based on tyrosine kinase inhibitors (TKI) of the epidermal growth factor

* Correspondence to: Laboratory of Intelligent Systems, Energy and Sustainable Development Faculty of engineering sciences, Private University of Fez, lotissement al qaraouine Fez, Morocco.
E-mail addresses: hassan.qjidaa@usmba.ac.ma, qjidah@yahoo.fr (Q. Hassan).

https://doi.org/10.1016/j.ejro.2024.100601
Received 2 July 2024; Received in revised form 9 September 2024; Accepted 10 September 2024

receptor (EGFR). Targeted TKI-EGFR treatment considerably increases the overall survival of patients [5–10]. However, this therapy requires knowledge of the EGFR mutation status, which often involves an invasive surgery or biopsy [11,12].

1.1. Literature review

In literature, the scientific community has moved towards exploiting Artificial Intelligence [13–16] to propose non-invasive methods for predicting EGFR mutations. Thus, several authors [17–20] have developed machine-learning models using radiomic features extracted from lung cancer patient CT images. Other researchers have exploited the many tumor characteristics of clinical and radiomic features to propose intelligent noninvasive methods for predicting EGFR mutation status. Yang et al. [21] use clinical features to establish an explainable machine-learning model to predict epidermal growth factor receptor (EGFR) mutations in lung cancer. However, this model did not exceed the accuracy of 0.771. Chun Sheng Yang et al. [22] propose a method based on least absolute shrinkage and selection operator (LASSO) regression and a 5-fold cross-validation method to obtain radiomic features that predicted EGFR mutation status and the sensitivity of TKIs. However, the performance of their approach did not exceed 0.6713 for the area under the curve (AUC) of the non-improved phases in the training group with EGFR mutation status. Xinna Lv et al. [23] propose four radiomics models for discriminating KRAS from EGFR which all performed well during the training process, but the AUC could not exceed 0.85 during the testing process. Shuo Wang et al. use deep learning with a conventional neural network model to predict EGFR genetic mutations directly from CT images, but the obtained results could not exceed the value of 0.81 for AUC [24]. Other noninvasive methods have shown fairly good performance in automatically predicting EGFR gene mutations [25–27].

1.2. Study rationale

Despite the progress made by these methods, they still have limitations. First, most proposed AI approaches do not present high-performance metrics capable of guaranteeing and securing the clinical applications of these approaches. This pushes the clinical impact of this success on the treatment of oncology patients to remain very limited or even absent [28,29]. This result is far from being achieved considering three 3 obstacles: human errors committed during the annotation of the model training images, the accuracy of the output decision of our model, and the reduction of expenses linked to the complex parameters of the layers of our model.

1.3. Study objectives

The main objective of this study was to propose a noninvasive and high-precision method capable of gaining the trust of specialists and bridging the clinical validation gap. Noninvasiveness is ensured by developing an automatic method based on Deep Learning using patient CT images. High performance was achieved by minimizing human errors during database annotation and by developing a stacking deep transfer model composed of five CNNs, where only the decision with a majority vote is selected as the model's output.

2. Materials and methods

2.1. Ethics

Ethics committee approval for this study was approved by the faculty of Medicine and Pharmacy Ethics Committee of Casablanca, Morocco, according to Helsinki Declaration under reference 17/15. Written informed consent was obtained for all patients and, all their relative data were deidentified.

2.2. Patient cohort and data collection

To collect patient data, a team of three specialist doctors and two IT specialists was established. All team members agreed that this study must comply with international ethical standards and be conducted following the Declaration of Helsinki. Under the suggestions of the Institutional Review Board and the Ethics Committee, it was decided that all data that will be used for the development of the prediction model will be anonymized through the elimination of personal patient information.

The team was interested in patients with lung cancer who had undergone a CT scan in the radiology department of the Hassan 2 University Hospital Center of Sidi Mohammed Ben Abdellah University in February, and who had undergone surgery or a pathological biopsy to determine EGFR gene mutation status between January 2021 and July 2022.

Among the 521 files analyzed, 138 were selected according to the selection criteria shown in Fig. 1. The team included patients with histologically verified primary NSCLC, patients with pathological analysis of tumor tissues with EGFR test results, and patients with preoperative CT images whose interval between CT examination and pathological biopsy was not more than three months. On the other hand, the team excluded patients with missing clinical data such as age, gender, tumor location, tumor size, stage, and patient smoking status, patients who received preoperative treatment, patients without adenocarcinoma cell type, patients with tumor size that exceeds 3 cm, patients with preoperative CT images with large artifacts or poor image quality, patients having neoadjuvant chemotherapy preoperative and patients with no EGFR gene test results.

To confirm the EGFR gene mutation, these experts used three methods: a method based on an analysis of the cellular tissue of the tumor by a surgical procedure or a biopsy, a method based on the nomogram making it possible to predict the status of EGFR gene mutation from clinical factors [30] and radiomics based method using feature engineering [31,32]. Thus, of the 138 patients collected, only 40 had been selected with a positive EGFR gene mutation status confirmed by all three methods, and 98 patients had wild-type (WT) EGFR. This allowed us to provide a benchmark dataset that was annotated precisely.

2.3. Clinical characteristics of patients

The clinical characteristics recorded are sex, age, and smoking status. The class of smoking was divided into two modes: smoker or non-smoker as for the EGFR gene mutation, it was also divided into two modes: mutated or non-mutated. Following the NCCN Clinical Practice Guidelines for Non-Small Cell Lung Cancer, the tumor stage class has been divided into four modes ranging from Stage 1 to Stage 4. Statistics for these clinical features are presented in Table 1.

2.4. Tomodensitometric characteristics

To identify Tomodensitometric features showing abnormalities, radiologists relied on guidelines published by the Fleischner Society [33]. Thus, the radiologists recorded several characteristics such as the presence or not of an attack of the pulmonary lobes as well as its distribution, the presence or not of ground glass opacities, the presence or not of consolidation, the presence or not of fibrosis, calcification, or emphysema and finally the presence or absence of pleural effusion or pericardial effusion. The statistics of these clinical and CT characteristics are presented in Table 1.

2.5. Acquisition of CT images

CT scans were performed using the SOMATOM Definition as 'Siemens Healthcare GmbH' under free-breathing conditions. The voxel spacing, which is associated with scans in the xy plane, ranged from

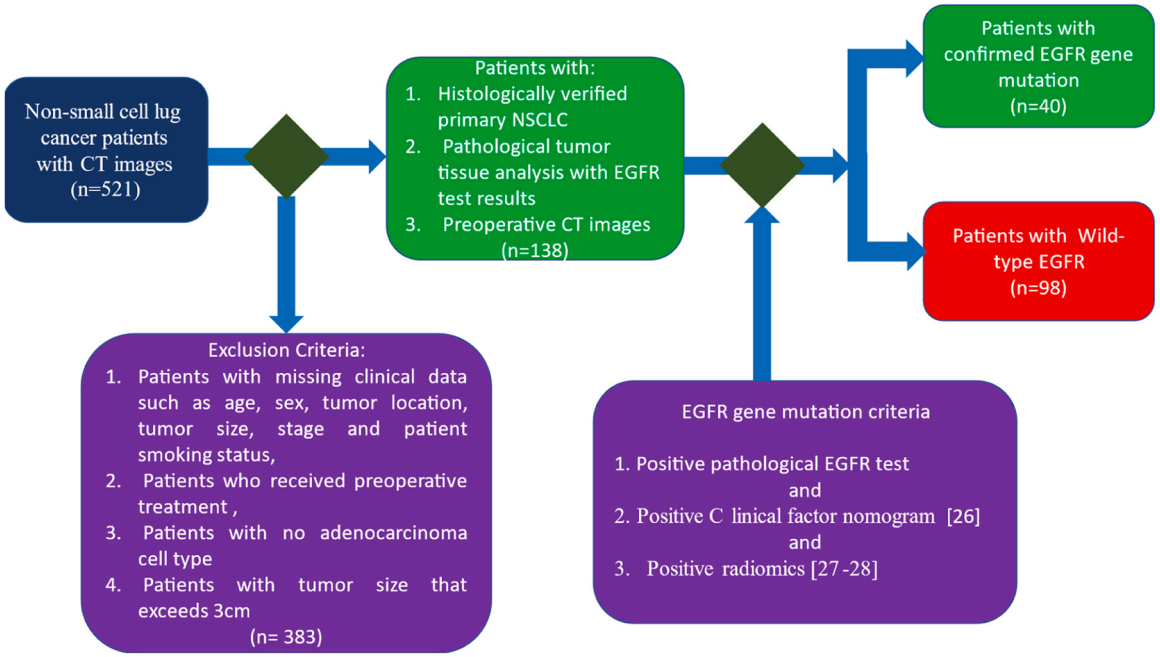

Fig. 1. Patient selection criteria in the study.

0.724 to 0.976 mm, which is much smaller than the Z dimension of 3 mm. We used the Elastix module (version 5.0.1, Linux Foundation, San Francisco, CA, USA, https://elastix.lumc.nl, accessed 20 July 2021) in 3D-Slicer to align the CT images before proceeding to feature extraction. To ensure a good comparison of the characteristics obtained from the different CT images, we normalized the images and corrected the inhomogeneity of the images [34].

2.6. Image preprocessing

Since the images of the selected dataset are in DICOM format, we will use the Python library 'Pydicom' to extract this data. The image pixel values were converted to the Hounsfield Unit (HU) which is a relative quantitative measurement of radio density used by radiologists in the interpretation of computed tomography (CT) images. HU values are obtained using the formula:

$$HU = \text{pixel} \times \text{slope} + \text{intercept} \quad (1)$$

where both slope and intercept were obtained in the DICOM data. The HU values of all slices were truncated to [− 2000, 0] and then were normalized to a value between 0 and 1. The respective binary masks were extracted from the corresponding RTSTRUCT files. The CT volume was resampled at the same resolution [1 × 1 × 3 mm] and each slice was cropped randomly to 256 × 256 size. To process and resize the image we will use SciPy-ndimage packages which provide several general image processing and analysis functions that are designed to operate with arrays 224 × 224 of arbitrary dimensionality. To save the array as a 2D image we will use the Python Imaging Library (PIL) which provides general image handling and lots of useful basic image operations like resizing, cropping, rotating, color conversion, and much more. Fig. 2 illustrates this preprocessing step.

2.7. Method and model

A fully connected CNN network made up of several classifiers is proposed where a prediction vector is obtained at the output of the CNN network. To increase the precision of the output decision of our model proposed model, we opted for a stacking model where only the decision having obtained a majority vote, among the five votes of the general model made up of five CNNs, is selected as the model's output. Indeed, the stacking model solves the problem linked to the choice of network architecture and increases the precision of the proposed model by using several networks as input to the model and choosing to output the solution that has the maximum number of votes. In other words, only the cell that obtained the maximum number of votes by the various networks was declared at the exit, as in cancer cells. For this, we are developing a CNN Deep Transfer Learning Stacking Model (DTLSMCNN) composed of five preformed CNN architectures: MobileNetV2, DenseNet121, InceptionResNetV2, VGG16 [35], and Xception [36], which has already been trained on an ImageNet dataset and will be adapted to our dataset accurately. DTLSMCNN was used to execute the image convolution, ReLU activation, and 2D normalizing processes. As illustrated in Fig. 3, only the prediction with the highest score is selected as the output of the model.

Finally, to ensure rapid convergence and reduce the expenses related to complex layer parameters and in-depth validation processes, we exploited the notion of Transfer Deep Learning. Indeed, the use of deep transfer learning solves the problems associated with the small size of the dataset as well as the lengthy training and validation process [37, 38].

2.8. Evaluation metrics

Several metrics have been proposed in literature to evaluate a Deep Learning model and measure its performance. Among these performance measurement approaches there is the confusion matrix which shows, through a matrix, the number of correct predictions per class given by the used model [39], the accuracy, the precision, the recall, the F1 score, and the area under the curve (AUC) [40].

3. Results

A comparison of our results with those published in the literature, thus providing a reference for other researchers in the field is presented in this section. To support and visualize the results we will present our results in the form of tables, graphs, and diagrams accompanied by statistical analysis [41].

Table 1
Statistics of clinical, CT characteristics and EGFR gene mutation of patients.

Characteristics	Mutated n(%)	WT n(%)
sex		
F	18(13.04)	20(14.49)
M	22(15.94)	78(56.52)
Age		
< 60	9(6.52)	47(34.05)
> 60	31(22.46)	51(36.95)
Tobacco		
Yes	13(9.42)	70(50.72)
No	27(19.56)	28(20.28)
Location: central/peripheral		
central	22(15.94)	54(39.13)
peripheral	18(13.04)	42(30.43)
pneumonic form	0	2(1.44)
Lob (LS/LM/LI)		
LI	14(10.14)	23(16.66)
LS	24(17.39)	71(51.44)
LM	0	1(0.72)
LI + LS	2(1.44)	3(2.17)
Contours: regular, irregular		
irregular	35(25.36)	92(66.66)
regular	5(3.62)	4(2.89)
pneumonic form	0	2(1.44)
Speculation: yes no		
Yes	21(15.21)	30(21.73)
No	19(13.76)	66(47.82)
pneumonic form	0	2(1.44)
Shape: round, oval, indefinite		
indefinite	19(13.76)	44(31.88)
oval	6(4.34)	26(18.84)
round	15(10.86)	26(18.84)
pneumonic form	0	2(1.44)
Limit: well or badly limited		
GOOD	22(15.94)	50(36.23)
wrong	18(13.04)	46(33.33)
pneumonic form		2(1.44)
Density: solid, VD, mixed		
mixed	4(2.89)	10(7.24)
solid	36(26.08)	87(63.04)
pneumonic form	0	1(0.72)
Air bronchogram		
Yes	17(12.31)	38(27.53)
No	23(16.66)	60(43.47)
Fissural attachment		
Yes	24(17.39)	62(44.92)
No	16(11.59)	36(26.08)
Pleural attachment		
Yes	24(17.39)	81(58.69)
No	16(11.59)	17(12.31)
Fissural retraction		
Yes	9(6.52)	26(18.84)
No	31(22.46)	72(52.17)
Pleural retraction		
Yes	16(11.59)	40(28.98)
No	24(17.39)	58(42.02)
Enhancement: homogeneous, heterogeneous		
homogeneous	14(10.14)	2(1.44)
heterogeneous	26(18.84)	96(69.56)
Cavitation		
Yes	4(2.89)	10(7.24)
No	36(26.08)	88(63.76)
Emphysema		
Yes	8(5.79)	54(39.13)
No	32(23.18)	44(31.88)
Calcification		
Yes	11(7.97)	22(15.94)
No	29(21.01)	76(55.07)
Fibrosis		
Yes	1(0.72)	7(5.072)
No	39(28.26)	91(65.94)
Maximum size: < 3 cm, > 3 cm		
> 3 cm	36(26.08)	84(60.86)
< 3 cm	4(2.89)	12(8.69)
pneumonic form	0	2(1.44)
Pulmonary nodule in the same lobe		
Yes	24(17.39)	36(26.08)
No	16(11.59)	62(44.92)
Nodule in another lobe		
Yes	21(15.21)	39(28.26)
No	19(13.76)	59(42.75)
Mediastinal ADP		
Yes	26(18.84)	64(46.37)
No	14(10.14)	34(24.63)
Ipsilateral pleural effusion		
Yes	18(13.04)	37(26.81)
No	22(15.94)	61(44.20)
Contralateral pleural effusion		
Yes	3(2.17)	4(2.89)
No	37(26.81)	94(68.11)
Distant metastasis		
Yes	27(19.56)	53(38.40)
No	13(9.42)	45(32.60)

3.1. Statistical analysis

Table 1 summarizes the clinical and CT characteristics of the study population, comprising a total of 138 patients. EGFR gene mutations were detected in 40 (28.98 %) patients, and mutations were wild-type (WT) in 98 (71.01 %) patients. In addition, a description of the characteristics of patients and tumors according to their EGFR status is given in Table 1.

3.2. Prediction performance

TensorFlow r1.9 [Apache 2.0 license] was used to write the software code for this study, which was created in Python 3.6. The neural networks were trained on a graphics processing unit [GPU] optimized workstation with four GeForce GTX 1080 Ti cards [11 GB, Pascal microarch] and a graphics processing unit [GPU] optimized workstation. [NVIDIA, Santa Clara, California; 11 GB, Pascal microarch]. Statistics on inference speed were computed using a single GPU.

3.2.1. Training and testing data set distribution

The 138 patients of our dataset benchmark were randomly divided into two sets: a training set composed of 110 patients (32 with EGFR gene mutation and 78 with wild-type EGFR) and a test set consisting of the remaining 28 patients (8 with EGFR gene mutation and 20 with wild-type EGFR) who were left out to be used later to test our model, as shown in Table 2.

Thus, for each CT image of a patient, we only selected 2D images (sections) containing tumor cells, considering that the number of images present in each class is balanced. Thus, our model was trained using 7752 2D images carefully selected from the CT images: 3776 images belonging to the first class, and 3776 images belonging to the second class. In the training validation process, we selected 1938 2D images, with 969 images belonging to the first EGFR gene mutation type class and 969 belonging to the EGFR wild-type class. The proposed model was trained and validated using five-fold cross-validation. Finally, in the testing process, we used 2350 2D images, with 1175 images belonging to the first class and 1175 images belonging to the second class.

All lung tumor images from the same CT scan were used for randomization. The neural network was trained and validated using five-fold cross-validation. Thus, the training data set was divided into five equal data intervals of which four were used for training and the fifth data interval for validation. This allowed us to give the measurements of the different performances of our model in the form of an average value.

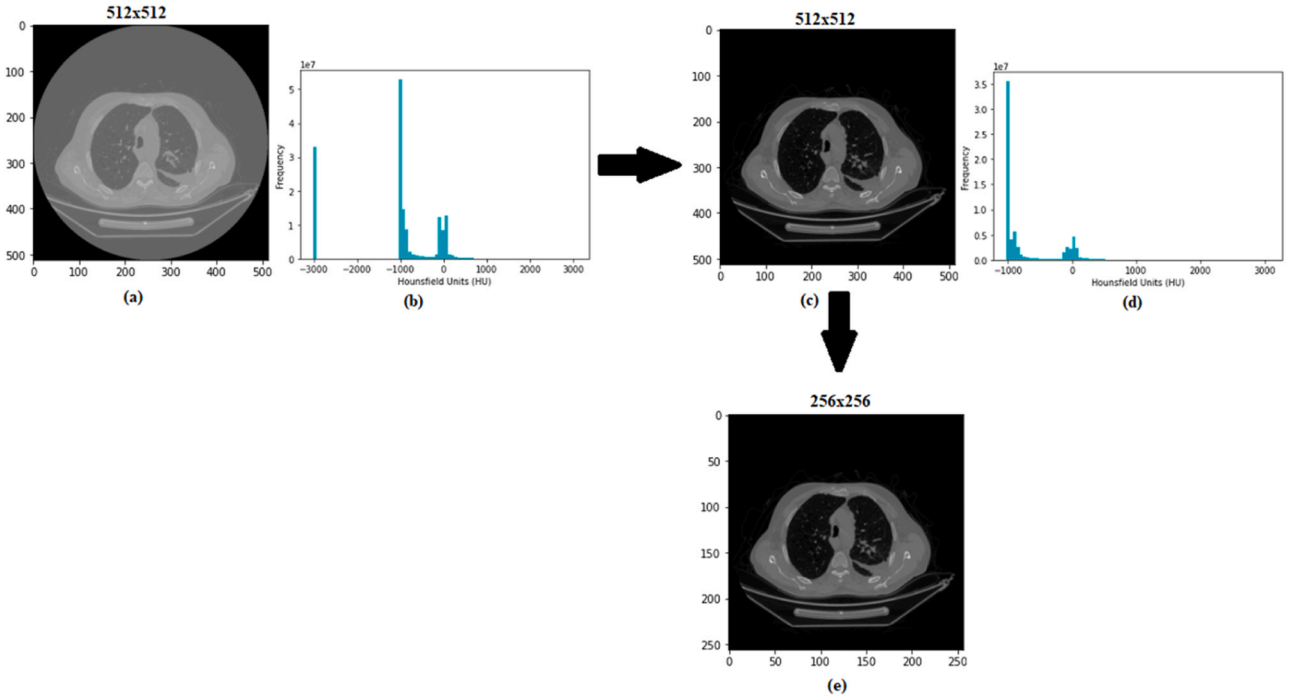

Fig. 2. Preprocessing step (a) The original image and (b) the corresponding Hounsfield Unit (HU) histogram, (c) Image result after application of Hounsfield Unit of [-2000,0] and (d) the corresponding Hounsfield Unit (HU) histogram, (e) Image result after resizing.

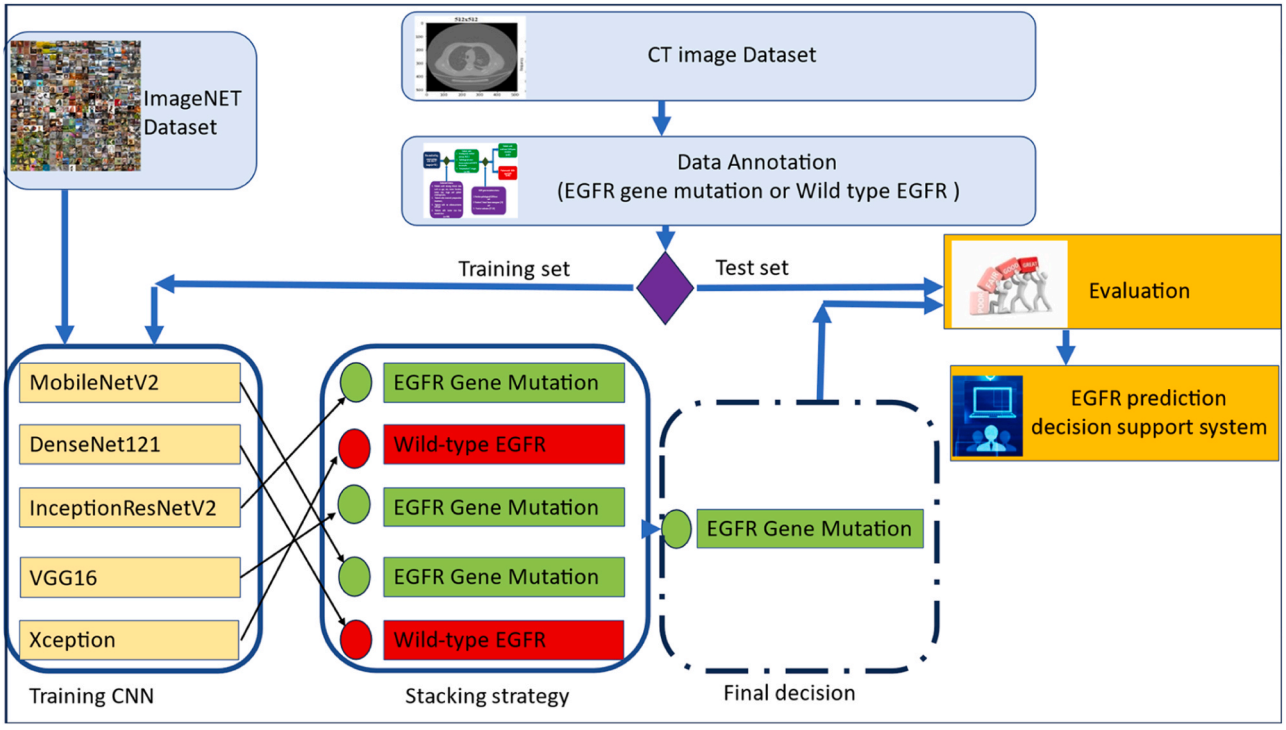

Fig. 3. Proposed Deep Transfer Learning Stacking Model process.

3.2.2. *Characterization of the learning and validation processes*

The images were scaled to the same size and signal intensity was standardized for each subject. The images were normalized before being input into each of the five models for training. The characterization of the learning and validation processes was done using the loss function deduced from cross-entropy as is shown in Fig. 4. The process of optimizing the values of the different weights of our architecture was carried out thanks to the Adam algorithm.

The learning rate starts from the value 0.001 and is reduced after 4 epochs if the loss value doesn't improve with the help of the callbacks function. The models were configured to train for 60 iterations.

3.2.3. *Confusion matrix*

The confusion matrix calculated for each CNN of our proposed

Table 2
Distribution of CT image data used for training, validation, and testing of the proposed model.

Training and validation		Testing
110 patients (80 %) (32 with EGFR gene mutation and 78 with wild-type EGFR)		28 patients (20 %) (8 with EGFR gene mutation and 20 with wild-type EGFR)
Selected 2D images for training process (80 %)	Selected 2D images for the validation process (20 %)	Selected 2D images for the test process
7752	1938	2350

stacking model is given in Fig. 5. This figure shows that the VGG16 model is the one that better discriminated the two classes in our database. Indeed, the VGG16 (Fig. 5.d) correctly classified all the images corresponding to the mutated cases, however, it only incorrectly classified two images among the 1500 images.

3.2.4. Air under the ROC curve (AUC)

Generally, sensitivity, specificity, and positive and negative predictions are used to evaluate the performance of a prediction model. Nevertheless, these evaluation criteria do not make it possible to judge the ability of a model to distinguish between patients with Mutated EGFR and patients with non-mutated EGFR. This is why we used the ROC curve (receiver operation characteristic) and more precisely the air under the ROC curve (AUC) in addition to other criteria such as confusion matrix, precision, recall, F1-Score, Cohen Kappa, and accuracy calculated during a fivefold cross-validation of our model. Given that the populations of the two classes considered (EGFR gene muted and EGFR wild type) are unbalanced, we will use the Air Under Curve (AUC) approach as an evaluation criterion. Indeed, the latter guarantees that a dominant class does not distort the performance of the model in the case where the classes are unbalanced. Additionally, AUC is often used to reliably compare different models while identifying the optimal threshold for each model's classification decisions. Fig. 6 which represents the plot of the receiver operating characteristic (ROC) of our proposed stacking model shows that the AUC of the latter reached a value of 0.98 thus far exceeding all the AUC values of the other CNN models.

3.2.5. Performance of each class prediction

The performance of each class prediction result obtained by our model during the training, validation, and test phase is presented in Tables 3–5 respectively. As is shown in these tables the stacking model has allowed improving considerably the result obtained by each model taken individually.

As we can see in Table 4, VGG16 and InceptionResNetV2 achieved the best test accuracy 95.08 and 92.01, respectively. The result for each neural network model is shown in Table 4. After analyzing the obtaining of the predictions of all models, we combined the results of 5 models: MobileNetV2, DenseNet121, InceptionResNetV2, VGG16 and Xception, we combine the predicted class of each model in a vector, and we take the class which was most frequently predicted by all models. By using this technique of ensemble model, we were able to make our finale classifier achieved the best performance with test accuracy of 95.22, f1-score 96.02, precision of 95.43, Sensitivity 97.01 and Cohen Kappa 92.01 as is presented in Table 5.

4. Discussion

In recent years, the number of articles and databases related to the application of artificial intelligence in the field of oncology has increased considerably. Nevertheless, the clinical impact of this success on the treatment of patients in oncology remains very limited if not absent, thus maintaining a significant gap between scientific work and their direct clinical applications in oncology [42,43].

Among the causes of this problem is that most of the proposed AI approaches do not present strong performance measures that can guarantee and secure the clinical applications of these approaches. Indeed, despite the satisfaction of radiologists with the contributions of AI in the profession of radiologists where 82 % of radiologists think that AI can improve the relevance of clinical decisions [44], the application of an AI tool does not pass the clinical stage only if the performance of the tests carried out presents very high performance and the percentage of false positives and false negatives approaches zero.

In this work, we proposed a high precision, high AUC, a non-invasive, intelligent, automatic, and stable method capable of crossing the clinical stage to define EGFR mutation status in patients with non-small cell lung cancer.

For this, we tackled the three obstacles that limit the high precision of our proposed model, namely: human errors committed during the annotation of the model training images, the precision of the output decision of our model, and the reduction expenses related to the complex layer parameters of our model.

Firstly, since the accuracy of the output data, or the decision made by the AI, depends on the input data, there is always a dependence on humans and the data entered by them, favoring a rate of human error. In this work, to minimize human errors made during the annotation of the learning database, we increased the number of experts for the annotation of the database by choosing three experts each with at least ten years of experience where only an annotation validated by the three experts is taken into consideration. In addition, to confirm the EGFR gene mutation, these experts used three methods: a method based on an analysis of the cellular tissue of the tumor by a surgical procedure or a biopsy, a method based on the nomogram making it possible to predict the of EGFR gene mutation from clinical factors and radiomics-based method using feature engineering. This allowed us to provide a Benchmark dataset annotated in a very precise way.

Secondly, to increase the precision of the output decision of our model proposed model, we opted for a stacking model where only the decision having obtained a majority vote, among the five votes of the general model made up of five CNNs, is selected as the model's output.

Finally, to ensure rapid convergence and reduce expenses related to complex layer parameters and in-depth validation processes, we exploited the notion of Transfer Deep Learning.

The proposed model showed encouraging results in the test cohort (AUC = 0.98). Thus, our results are valuable and can be distinguished from previous studies as a first attempt to bridge the AI translational gap between initial model development and its clinical application by increasing the accuracy of our model's output decisions. Our proposed model provides an alternative method to non-invasively assess EGFR information and to aid rapid decision-making when applying a TKI as an initial treatment.

Although studies of deep learning models have demonstrated promising performance in aiding lung cancer analysis [45–47], our study differs from previous studies in its design and high-rate output decision accuracy.

Unlike these methods, we proposed a method based on Deep Learning which did not require segmentation of the edges of the tumor which are often very difficult to locate with the naked eye from CT images, nor extraction and selection of features. The big advantage of our method is that it makes it possible to extract abstract characteristics that are surely strongly correlated with the EGFR gene mutation, unlike radiomic or clinical characteristics.

In this study, we only considered a population very restricted to the North African region. It would be preferable to consider data from different sources so that the decision of EGFR gene mutation is not affected by the race of the population considered. Although the performance of our proposed method, based on Deep Learning, has exceeded other Machine Learning methods based on clinical or radiomic characteristics, it would be desirable to propose, in perspective, a model

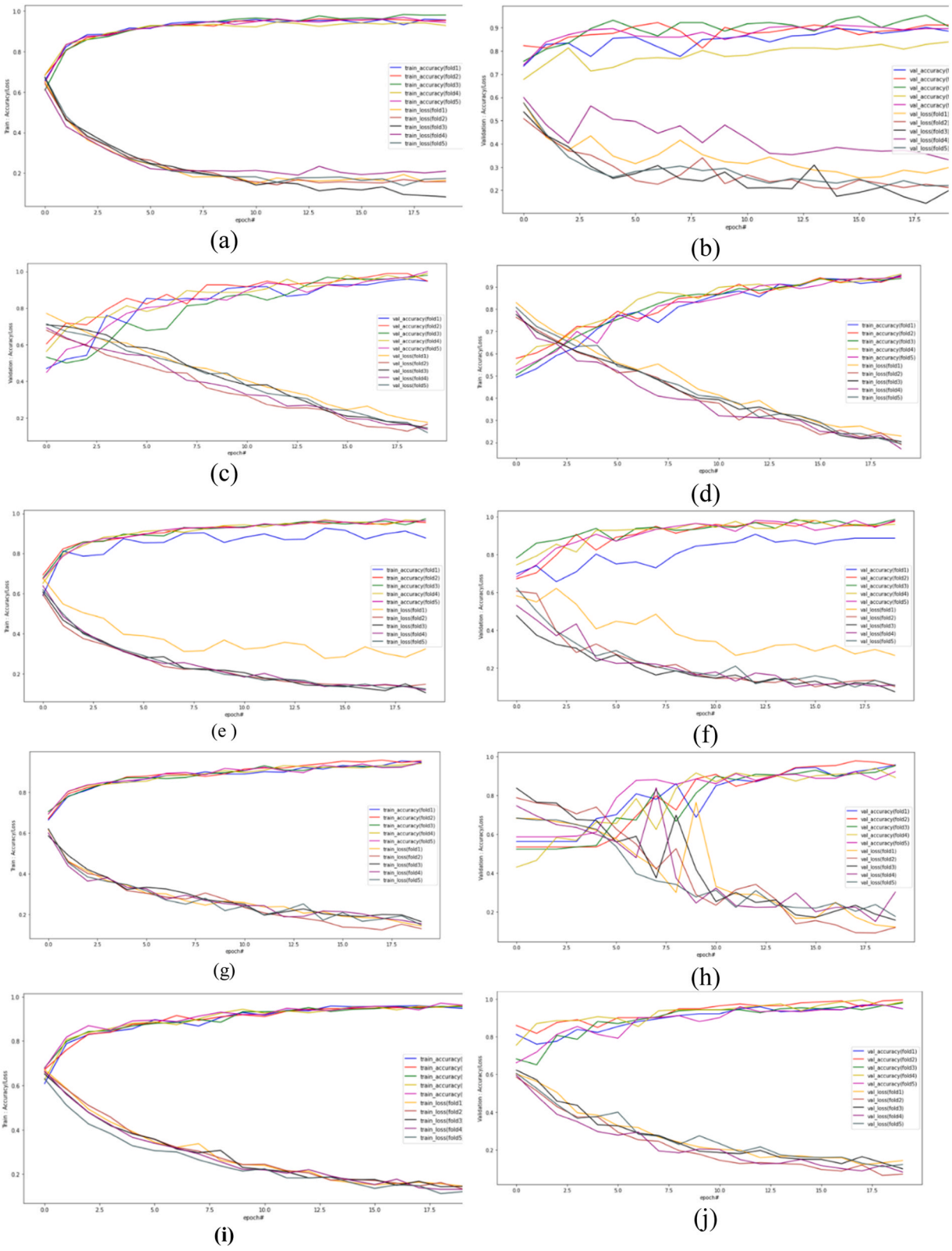

Fig. 4. Train accuracy versus train loss and validation accuracy versus validation loss with 5-fold cross-validation for a five CNN model: (a) Top Train Accuracy versus Train loss (b) Bottom Validation Accuracy versus Validation loss with 5-fold cross-validation for MobileNetV2. (c) Top Train Accuracy versus Train loss (d) Bottom Validation Accuracy versus Validation loss with 5-fold cross-validation for DenseNet121. (e) Top Train Accuracy versus Train loss (f) Bottom Validation Accuracy versus Validation loss with 5-fold cross-validation for InceptionResNetV2 (g) Top Train Accuracy versus Train loss (h) Bottom Validation Accuracy versus Validation loss with 5-fold cross-validation for VGG16, (i)Top Train Accuracy versus Train loss (j) Bottom Validation Accuracy versus Validation loss with 5-fold cross-validation for Xception.

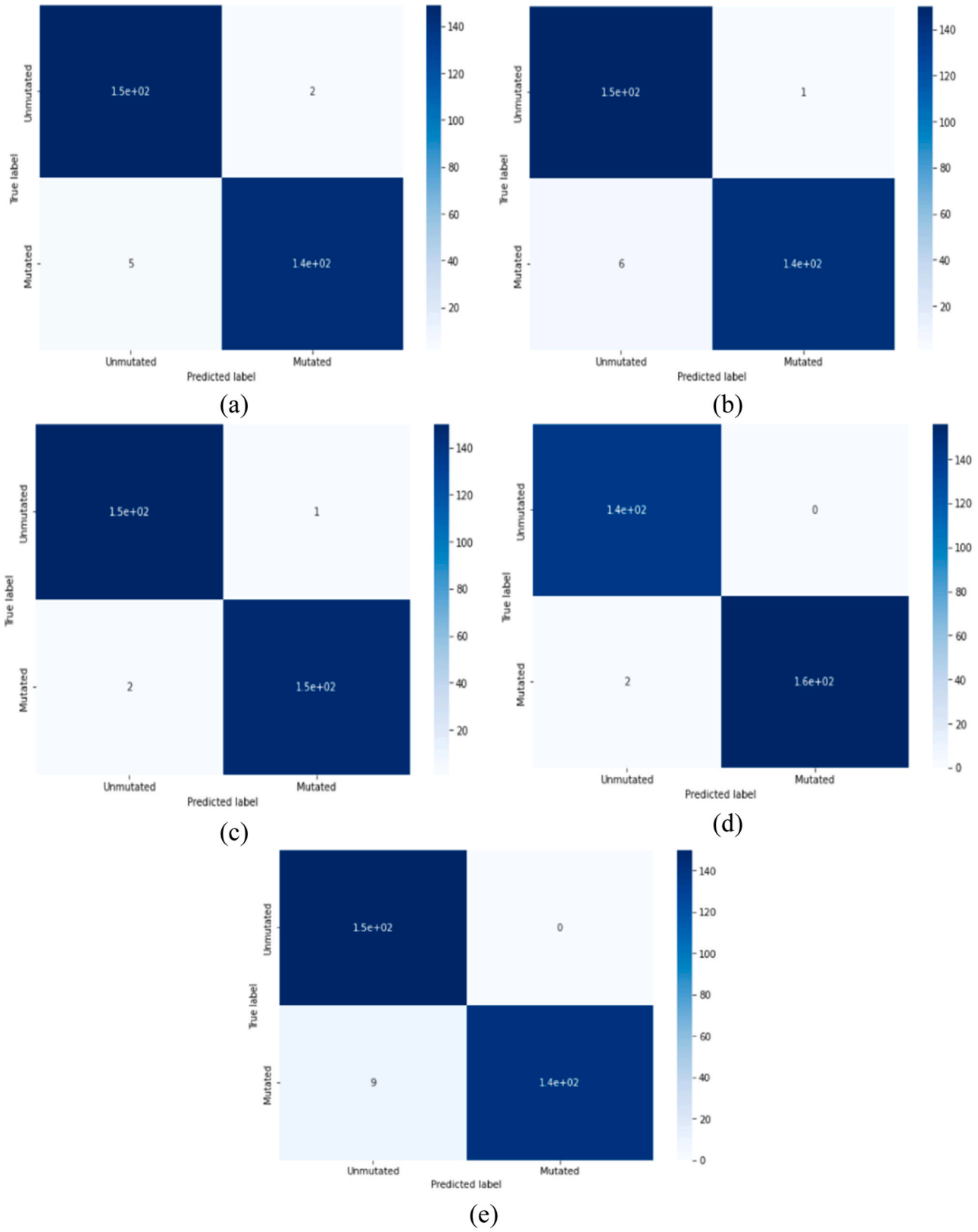

Fig. 5. Confusion matrix for each model on the test set, (a) MobileNetV2, (b) DenseNet121, (c) InceptionResNetV2, (d) VGG16, (e) Xception.

that combines these three methods to obtain better performance.

4.1. *Limitations of this study*

This study had certain limitations. First, this study lacked external validation, given that the sample was from a single center. Furthermore,

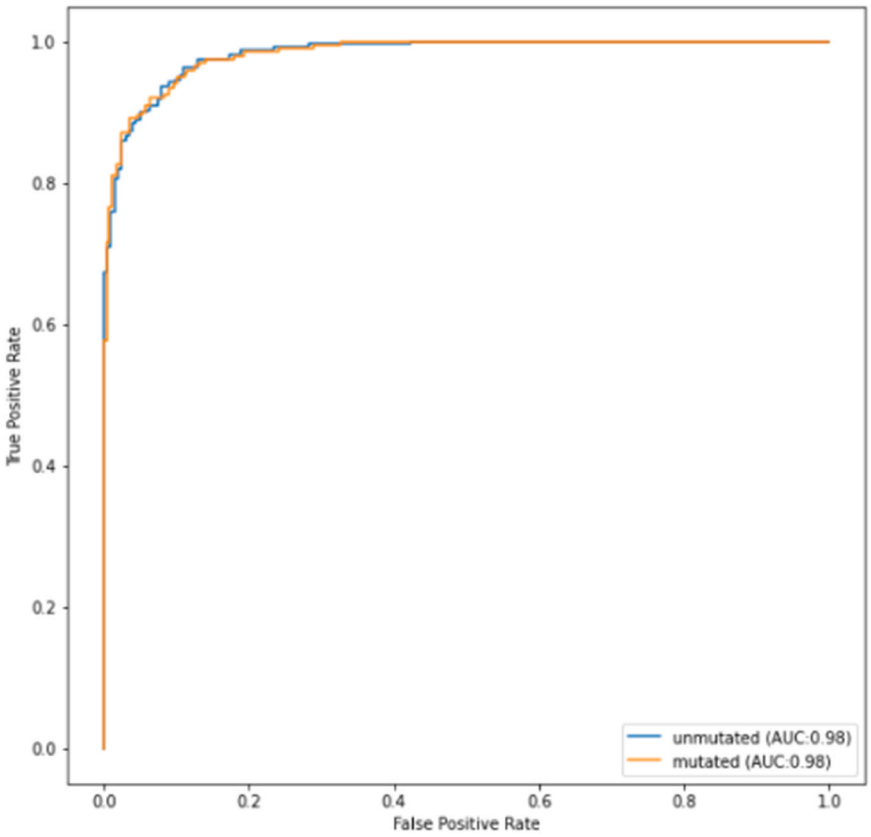

Fig. 6. True positive rate (TPR) versus the false positive rate (FPR) (ROC curve) for our Stacking model.

Table 3
The prediction results with 5-fold cross-validation for each class using the model set which combines the five models of the train set.

Model		Average precision	Average sensitivity	Average F1_Score
MobileNetV2	Mutated	93.95	95.21	95.12
	Wild type	93.71	95.32	95.96
DenseNet121	Mutated	92.09	95.62	92.97
	Wild type	90.47	95.53	90.60
InceptionResNetV2	Mutated	90.24	94.31	90.10
	Wild type	91.71	94.05	92.99
VGG16	Mutated	95.68	96.51	95.59
	Wild type	96.84	96.68	95.26
Xception	Mutated	89.50	90.33	89.41
	Wild type	89.67	90.78	89.72
DTLSMCNN	Mutated	**95.71**	**96.65**	**95.75**
	Wild type	**96.86**	**96.73**	**96.01**

Table 4
The prediction results with 5-fold cross-validation for each class using the model set which combines the five models in the validation set.

Model		Average precision	Average sencitivity	Average F1_Score
MobileNetV2	Mutated	92.31	95.38	94.74
	Wild type	92.41	95.70	94.88
DenseNet121	Mutated	90.91	95.38	89.71
	Wild type	90.86	95.08	92.54
InceptionResNetV2	Mutated	89.36	93.43	90.30
	Wild type	89.68	93.03	90.43
VGG16	Mutated	96.40	96.01	96.24
	Wild type	96.56	96.70	96.33
Xception	Mutated	89.72	90.37	89.69
	Wild type	90.07	90.89	90.02
Stacking Method	Mutated	**96.71**	**96.92**	**96.51**
	Wild type	**96.83**	**96.98**	**96.64**

the sample size of this study was not sufficiently large. In the future, we plan to consider a larger sample size from several centers. Second, in this study, we did not consider the causal epidermal growth factor receptor (EGFR) genomic alterations, including deletions of exon 19 (E19 dels) and point mutation of E21, which are known to have a favorable prognosis due to sensitivity to tyrosine kinase inhibitors. Thus, we plan to collect more samples with E19 and E21 mutations confirmed by genetic analysis and CT data to predict different EGFR mutations in detail. Finally, we will conduct a prospective study to validate the results of this study clinically.

5. Conclusion

This study demonstrated the value of a stacking convolutional neural network model trained on a precisely annotated database in improving the accuracy of EGFR mutation prediction. With an average sensitivity of 97.1, an average accuracy of 95.2, and AUC of 0.98 it provides healthcare professionals with a non-invasive and easy-to-use method to identify EGFR mutation status to rapidly guide targeted therapy and predict EGFR TKI candidates.

Table 5
Comparative results on the test set.

Model	Average Accuracy (± SD)	Average F1_score (± SD)	Average Precision (± SD)	Average Sencitivity (± SD)	Average Cohen_Kappa (±SD)
MobileNetV2	89.58(± 0.011)	89.93(± 0.010)	89.52(± 0.013)	89.50(± 0.016)	88.84(± 0.022)
DenseNet121	90.58(± 0.012)	89.64(± 0.013)	90.05(± 0.008)	89.29(± 0.019)	85.60(0.020)
InceptionResNetV2	92.08(± 0.002)	91.17(± 0.003)	92.14(± 0.006)	90.41(± 0.013)	90.36(± 0.007)
VGG16	95.08(± 0.003)	95.86(± 0.008)	95.19(± 0.006)	96.74(± 0.004)	91.74(± 0.009)
Xception	89.2((± 0.005)	89.16(± 0.004)	89.25(± 0.012)	90.21(± 0.010)	88.52(± 0.009)
Stacking Method	**95.22**	**96.02**	**95.43**	**97.01**	**92.01**

CRediT authorship contribution statement

Bouchra Amara: Validation, Formal analysis, Data curation. **Mounia Serraj:** Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Mohammed Smahi:** Validation, Data curation. **Abdeljabbar Cherkaoui:** Writing – original draft, Software, Resources, Data curation. **Mohammed Ouazzani Jamil:** Writing – original draft, Validation, Resources, Conceptualization. **Sara Boukansa:** Resources, Data curation. **Hassan Qjidaa:** Writing – original draft, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Nizar El Bouardi:** Validation, Data curation. **Moulay Youssef Alaoui Lamrani:** Validation, Supervision, Data curation. **Hind El Fatimi:** Validation, Formal analysis, Data curation. **Mustapha Maaroufi:** Writing – original draft, Validation, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Anass El affar:** Software, Resources, Formal analysis, Data curation. **Anass Benfares:** Writing – original draft, Validation, Software, Project administration, Data curation. **Mamoun Qjidaa:** Writing – original draft, Validation, Software, Resources, Investigation, Data curation. **Abdelali yahya Mourabiti:** Validation, Data curation. **Badreddine Alami:** Validation, Resources, Formal analysis, Data curation, Conceptualization.

Author contributions

AYM, BA, MM, and QH take full responsibility for the content of the manuscript, including the data and analysis. AB, NEB, MYAL, HEF, BA, SM, CA and OJM contributed to study concept, design, interpretation, and writing. SB, AB, ELA, and MS contributed to the data acquisition and analysis.

Ethical statement

Ethics committee approval for this study was approved by the faculty of Medicine and Pharmacy Ethics Committee of Casablanca, Morocco, according to Helsinki Declaration under reference 17/15.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that they did not resort to any use of generative AI and AI-assisted technologies in the writing process.

Funding statement

The authors did not receive support from any organization for the submitted work. No funding was received to assist with the preparation of this manuscript. No funding was received for conducting this study.

Authors' Disclosures of Potential Conflicts of Interest

The authors declare that they have no conflicts of interest.

Prior Presentation

This study has not been previously conducted.

Support

This work was not supported by any organization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgements

The authors thank all the members of the Radiology, Pulmonology, and Oncology departments of the Hassan II University Hospital Center who participated directly or indirectly in the collection of patient data used in this study.

References

[1] G.Da Cunha Santos, M.A. Saieg, W. Geddie, N. Leighl, EGFR gene in cytological samples of non-small cell lung carcinoma: controversies and opportunities, Cancer Cytopathol. 119 (2011) 80–91, https://doi.org/10.1002/cncy.20150.
[2] M. Ladanyi, W. Pao, Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond, Mod. Pathol. 21 (Suppl. 2) (2008) 16–22, https://doi.org/10.1038/modpathol.3801018.
[3] R. Herbst, D. Morgensztern, C. Boshoff, The biology and management of non-small cell lung cancer, Nature 553 (2018) 446–454, https://doi.org/10.1038/nature25183.
[4] J. Mondal, A.K. Panigrahi, A.R. Khuda-Bukhsh, Conventional chemotherapy: problems and scope for combined therapies with certain herbal products and dietary supplements, Austin J. Mol. Cell Biol. 1 (1) (2014) 10.
[5] S.S. Ramalingam, J. Vansteenkiste, D. Planchard, B.C. Cho, J.E. Gray, Y. Ohe, et al., Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC, N. Engl. J. Med. 382 (1) (2020) 41–50, https://doi.org/10.1056/NEJMoa19136627.
[6] W. Tang, X. Li, X. Xie, X. Sun, J. Liu, J. Zhang, et al., EGFR inhibitors as adjuvant therapy for resected non-small cell lung cancer harboring EGFR gene mutations, Lung Cancer 136 (2019) 6–14, 10.1016/j.lungcan.2019.08.001 8.
[7] T. Mitsudomi, S. Morita, Y. Yatabe, S. Negoro, I. Okamoto, J. Tsurutani, et al., Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (Wjtog3405): an open label, randomised phase 3 trial, Lancet Oncol. 11 (2) (2010) 21–28, https://doi.org/10.1016/s1470-2045(09)70364-x (1).
[8] Q.M. Guo, L. Wang, W.J. Yu, L.H. Qiao, M.N. Zhao, X.M. Hu, et al., Detection of plasma EGFR gene mutations in NSCLC patients with a validated ddPCR lung cfDNA assay, J. Cancer 10 (2019) 4341–4349, https://doi.org/10.7150/jca.31326.
[9] D.C. Jupiter, assessing diagnostic tests I: you can't be too sensitive, J. Foot Ankle Surg. 54 (2015) 519–520.
[10] F. Skoulidis, J.V. Heymach, Co-occurring genomic alterations in non–small cell lung cancer biology and therapy, Nat. Rev. Cancer 19 (2019) 495–509.
[11] R. Brody, Y. Zhang, M. Ballas, M.K. Siddiqui, P. Gupta, C. Barker, et al., PD-L1 expression in advanced NSCLC: insights into risk stratification and treatment

selection from a systematic literature review, Lung Cancer 112 (2017) 200–215, https://doi.org/10.1016/j.lungcan.2017.08.005.
[12] R. Büttner, J.R. Gosne, B.G. Skov, J. Adam, N. Motoi, K.J. Bloom, et al., Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer, J. Clin. Oncol. 35 (34) (2017) 3867–3876, https://doi.org/10.1200/jco.2017.74.7642.
[13] Y. Zhou, X. Xu, L. Song, C. Wang, J. Guo, Z. Yi, et al., The application of artificial intelligence and radiomics in lung cancer, Precis. Clin. Med. 3 (3) (2020) 214–227, https://doi.org/10.1093/pcmedi/pbaa028.
[14] B. Coppin, Artificial Intelligence Illuminated, Jones & Bartlett Learning, Burlington, MA, USA, 2004.
[15] D. Bordoloi, V. Singh, S. Sanober, S.M. Buhari, J.A. Ujjan, R. Boddu, Deep learning in healthcare system for quality of service, J. Health Eng. 8 (2022) 8169203, https://doi.org/10.1155/2022/8169203.
[16] S.K.S. Durai, M.D. Shamili, Smart farming using Machine Learning and Deep Learning techniques, Decis. Anal. J. 3 (2022) 100041, https://doi.org/10.1016/j.dajour.2022.100041 (ISSN 2772-6622).
[17] T. Zhang, Z. Xu, G. Liu, et al., Simultaneous identification of EGFR, KRAS, ERBB2, and TP53 mutations in patients with non-small cell lung cancer by machine learning-derived three-dimensional radiomics, Cancers 13 (8) (2021) 1814, https://doi.org/10.3390/cancers13081814.
[18] W. Tu, G. Sun, L. Fan, et al., Radiomics signature: a potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology, Lung Cancer 132 (2019) 28–35, https://doi.org/10.1016/j.lungcan.2019.03.025.
[19] K.G. Abraham, V.S. Jayanthi, P. Bhaskaran. 10 – Convolutional neural network for biomedical applications, in: Jitendra Kumar Verma, Sudip Paul, Prashant Johri (eds.), Computational Intelligence and Its Applications in Healthcare, Academic Press, 2020, pp. 145–56, ISBN 9780128206041, ⟨https://doi.org/10.1016/B978-0-12-820604-1.00010-8⟩.
[20] S. Nazmus, S. Scandola, C. Kennedy, C. Lummer, J. Chang, R.G. Uhrig, G. Lin, Machine learning classification of plant genotypes grown under different light conditions through the integration of multi-scale time-series data, Comput. Struct. Biotechnol. J. 21 (2023) 3183–3195, https://doi.org/10.1016/j.csbj.2023.05.005.
[21] (a) R. Yang, X. Xiong, H. Wang, W. Li, Explainable Machine Learning Model to Prediction EGFR Mutation in Lung Cancer, Front Oncol. 12 (2022) 924144, https://doi.org/10.3389/fonc.2022.924144;
(b) D. Mazzei, R. Ramjattan, Machine learning for industry 4.0: a systematic review using deep learning-based topic modeling, Sensors 22 (2022) 8641, https://doi.org/10.3390/s22228641.
[22] (a) C. Yang, W. Chen, G. Gong, Z. Li, Q. Qiu, Y. Yin, Application of CT radiomics features to predict the EGFR mutation status and therapeutic sensitivity to TKIs of advanced lung adenocarcinoma, Transl Cancer Res 9 (11) (2020) 6683–6690, https://doi.org/10.21037/tcr-20-1216;
(b) T. Davenport, R. Kalakota R, The potential for artificial intelligence in healthcare, Future Healthc. J. 6 (2) (2019) 94–98, https://doi.org/10.7861/futurehosp.6-2-94.
[23] X. Lv, Y. Li, B. Wang, Y. Wang, Z. Xu, D. Hou, Multisequence MRI-based radiomics signature as potential biomarkers for differentiating KRAS mutations in non-small cell lung cancer with brain metastases, Eur. J. Radiol. Open. 12 (2024) 100548, https://doi.org/10.1016/j.ejro.2024.100548.
[24] S. Wang, J. Shi, Z.X. Ye, D. Dong, D. Yu, M. Zhou, Y. Liu, O. Gevaert, K. Wang, Y. Zhu, H. Zhou, Z. Liu, J. Tian, Predicting EGFR gene mutation in lung adenocarcinoma on computed tomography image using deep learning, Eur. Respir. J. 53 (2019) 1800986, https://doi.org/10.1183/13993003.00986-2018.
[25] H.S. Nguyen, D.K.N. Ho, N.N. Nguyen, H.M. Tran, K.-W. Tam, N.Q.K. Le, Predicting EGFR gene mutation in non–small cell lung cancer using artificial intelligence: a systematic review and meta-analysis, Acad. Radiol. 31 (2) (2024) 660–683, https://doi.org/10.1016/j.acra.2023.03.040.
[26] B.H. Kann, a. Hosny, H. Aerts, Artificial intelligence for clinical oncology, Cancer Cell 39 (7) (2021) 916–927, https://doi.org/10.1016/j.ccell.2021.04.002.
[27] X.Y. Li, J.F. Xiang, T.Y. Jia, T.L. Shen, R.P. Hou, J. Zhao, et al., Detection of epithelial growth factor receptor (EGFR) mutations on CT images of patients with lung adenocarcinoma using radiomics and/or multi-level residual convolutionary neural networks, J. Thorac. Dis. 10 (12) (2018) 6624–6635, https://doi.org/10.21037/jtd.2018.11.03.
[28] K. Chohee, H. h Cho, J.Y. Choi, T.J. Franks, J. Han, Y. Choi, Se-H. Lee, H. Park, K. Soo Lee, Pleomorphic carcinoma of the lung: prognostic models of semantic, radiomics and combined features from CT and PET/CT in 85 patients, Eur. J. Radiol. Open 8 (2021) 100351, https://doi.org/10.1016/j.ejro.2021.100351.
[29] S.H. Park, J. Choi, J.S. Byeon, Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence, Korean J. Radiol. 22 (3) (2021) 442–453, https://doi.org/10.3348/kjr.2021.0048.
[30] N. Girard, C.S. Sima, D.M. Jackman, et al., Nomogram to predict the presence of EGFR activating mutation in lung adenocarcinoma, Eur. Respir. J. 39 (2012) 366–372.
[31] E.R. Velazquez, C. Parmar, Y. Liu, et al., Somatic mutations drive distinct imaging phenotypes in lung cancer, Cancer Res. 77 (2017) 3922–3930.
[32] Y. Liu Y, J. Kim, F. Qu, et al., CT features associated with epidermal growth factor receptor mutation in patients with lung adenocarcinoma, Radiology 280 (2016) 271–280.
[33] P.D. Naidich, J.M. Goo, K.S. Lee, N.C.A. Leung, J.R. Mayo, A.C. Mehta, Y. Ohno, C. A. Powell, M. Prokop, G.D. Rubin, C.M. Schaefer-Prokop, W.D. Travis, P.E. Van Schil, A.A. Bankier, Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner society, Radiology 284 (1) (2007), https://doi.org/10.1148/radiol.2017161659.
[34] S. Roy, A. Carass, P.L. Bazin, J.L. Prince, Intensity inhomogeneity correction of magnetic resonance images using patches, Proc. SPIE Int. Soc. Opt. Eng. 11 (7962) (2011) 79621F, https://doi.org/10.1117/12.877466.
[35] M. Yaqub, J. Feng, M.S. Zia, K. Arshid, K. Jia, Z.U. Rehman, A. Mehmood, State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images, Brain Sci. 10 (2020) 427, https://doi.org/10.3390/brainsci10070427.
[36] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1800–7. ⟨https://doi.org/10.1109/CVPR.2017.195⟩.
[37] W. Dai, Y. Chen, G.R. Xue, Q. Yang, Y. Yu, D. Koller, D. Schuurmans, Y. Bengio, L. Bottou, Translated learning: transfer learning across different feature spaces, in: Advances in Neural Information Processing Systems 21, Proceedings of the Neural Information Processing Systems, Neural Information Processing Systems Foundation, Inc. (NIPS): Vancouver, BC, Canada, 2008, pp. 353–60.
[38] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu, V. Vaidya, Understanding the mechanisms of deep transfer learning for medical images, in: G. Carneiro (ed.), Deep Learning and Data Labeling for Medical Applications;' DLMIA 2016, LABELS 2016, vol. 10008, Springerm Cham, Switzerland, 2016.
[39] M. Ghaffari, A. Sowmya, R. Oliver, Automated brain tumour segmentation using multimodal brain scans, a survey based on models submitted to the BraTS 2012–2018 challenges, IEEE Rev. Biomed. Eng. 13 (2019) 156–168.
[40] G. Rani, P.K. Tiwari, Handbook of Research on Disease Prediction through Data Analytics and Machine Learning, IGI Global, Medical Information Science Reference, Hershey, PA, USA, 2021.
[41] S.H. Park, J. Choi, J.S. Byeon, Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence, Korean J. Radiol. 22 (3) (2021) 442–453, https://doi.org/10.3348/kjr.2021.0048.
[42] K. Sheikh, D.P.I. Capaldi, D.A. Hoover, D.A. Palma, B.P. Yaremko, G. Parraga, Magnetic resonance imaging biomarkers of chronic obstructive pulmonary disease prior to radiation therapy for non-small cell lung cancer, Eur. J. Radiol. Open 2 (2015) 81–89, https://doi.org/10.1016/j.ejro.2015.05.003.
[43] S.H. Park, J. Choi, J.S. Byeon, Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence, Korean J. Radiol. 22 (3) (2021) 442–453, https://doi.org/10.3348/kjr.2021.0048.
[44] French Society of Radiology, Radiologists are optimistic about the impact of new technologies on their profession, 2021. ⟨https://www.caducee.net/actualite-medicale/13662/les-radiologues-sont-optimistes-sur-l-impact-des-nouvelles-technologies-sur-leur-metier.html#:~:text=Les%20radiologues%20interrog%C3%A9s%20estiment%20que,patients%20(82%20et%2080%25⟩. (Accessed 24 November 2021).
[45] W. Zhao, J. Yang, B. Ni, D. Bi, Y. Sun, M. Xu, et al., Toward automatic prediction of EGFR gene mutation in pulmonary adenocarcinoma with 3D deep learning, Cancer Med. 8 (7) (2019) 3532–3543, https://doi.org/10.1002/cam4.2233.
[46] S. Wang, J. Shi J, Z. Ye, D. Dong, D. Yu, M. Zhou, et al., Predicting EGFR gene mutation in lung adenocarcinoma on computed tomography image using deep learning, Eur. Respir. J. 53 (3) (2019) 1800986, https://doi.org/10.1183/13993003.00986-2018.
[47] J.F. Xiong, T.Y. Jia, X.Y. Li, W. Yu W, Z.Y. Xu, X.W. Cai, et al., Identifying epidermal growth factor receptor mutation in patients with lung adenocarcinoma by three-dimensional convolutional neural networks, Br. J. Radiol. 91 (1092) (2018) 20180334, https://doi.org/10.1259/bjr.20180334.