Residual neural network with mixed loss based on batch training technique for identification of EGFR mutation status in lung cancer

Liye Jia[1] · Wei Wu[2] · Guojie Hou[1] · Juanjuan Zhao[1] · Yan Qiang[1] · Yanan Zhang[1] · Meiling Cai[1]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Epidermal growth factor receptor (EGFR) is the key to targeted therapy with tyrosine kinase inhibitors in lung cancer. Traditional identification of EGFR mutation status requires biopsy and sequence testing, which may not be suitable for certain groups who cannot perform biopsy. In this paper, using easily accessible and non-invasive CT images, the residual neural network (ResNet) with mixed loss based on batch training technique is proposed for identification of EGFR mutation status in lung cancer. In this model, the ResNet is regarded as the baseline for feature extraction to avoid the gradient disappearance. Besides, a new mixed loss based on the batch similarity and the cross entropy is proposed to guide the network to better learn the model parameters. The proposed mixed loss utilizes the similarity among batch samples to evaluate the distribution of training data, which can reduce the similarity of different classes and the difference of the same classes. In the experiments, VGG16Net, DenseNet, ResNet18, ResNet34 and ResNet50 models with the mixed loss are trained on the public CT dataset with 155 patients including EGFR mutation status from TCIA. The trained networks are employed to the collected preoperative CT dataset with 56 patients from the cooperative hospital for validating the efficiency of the proposed models. Experimental results show that the proposed models are more appropriate and effective on the lung cancer dataset for identifying the EGFR mutation status. In these models, the ResNet34 with mixed loss is optimal (accuracy = 81.58%, AUC = 0.8861, sensitivity = 80.02%, specificity = 82.90%).

✉ Juanjuan Zhao
  zhaojuanjuan@tyut.edu.cn

[1] College of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China

[2] Department of Physiology, Shanxi Medical University, Taiyuan 030051, China

1 Introduction

Lung cancer is one of the malignant tumors with the highest mortality rate in the world [30]. The study of epidermal growth factor receptor (EGFR) mutation in cancer driver genes makes targeted therapy a relatively effective treatment [33]. EGFR mutation status is involved in the occurrence, development, invasion and metastasis of lung cancer [22]. The detection of EGFR mutation status is crucial in first-line therapies [3], because EGFR tyrosine kinase inhibitors can target specific mutation within EGFR gene and improve the prognosis of lung cancer patients with EGFR mutation [53]. Biopsy sequencing is the gold standard for gene mutation detection. Due to the widespread heterogeneity of lung tumors, biopsy sequencing need to locate tissue regions for measuring EGFR mutation status. Its applicability is limited due to the difficulty in obtaining tissue samples and repeatedly sampling tumor, relatively high cost and poor DNA quality [31]. In addition, biopsy increases the potential risk of cancer metastasis [25]. In these cases, a non-invasive and easy-to-use method is necessary to identify EGFR mutation status.

Computed tomography (CT), as a non-invasive routine diagnostic technique, can be used in the analysis of lung cancer [20, 52]. Recent studies have shown that the features extracted from CT images of lung cancer are related to gene expression patterns [1, 5, 17, 55], and show the ability to identify EGFR mutation status [23, 28, 35, 48, 54]. Although image-based evaluation cannot replace biopsy, it can be regarded as supplementary information of biopsy [15, 31]. For example, CT imaging can provide some information of the tumor heterogeneity such as tumor density, activity and microenvironment, allowing us to identify the EGFR mutation status [41, 47]. In addition, CT imaging is low-cost and easy to obtain throughout the treatment process. Therefore, it is promising for CT imaging as an alternative method to detect EGFR mutation status.

In recent years, researchers have predicted gene mutation based on CT images mainly by traditional radiomics, machine learning or statistical methods. Liu Y et al. [24] adopted radiomic method to extract features such as size, edge, transparency and uniformity from CT images for identifying EGFR mutation status. Velazquez et al. [31] developed a radiomic model based on CT image features and clinical data to distinguish between EGFR- and EGFR+, KRAS+ and KRAS-. Zhang et al. [50] also developed a radiogenomic model based on CT image features to predict EGFR mutation status in patients with lung adenocarcinoma. Jia T Y et al. [16] extracted radiomic features and adopted random forest model to identify EGFR mutation status in lung adenocarcinoma based on non-invasive imaging. Morgado J et al. [27] utilized a variety of linear, nonlinear, and ensemble predictive classification models, along with several feature selection methods, to classify the binary outcome of wild type or mutant of EGFR. In order to make the model performance better for disease prediction, radiomics methods are also gradually improved in various aspects, such as feature selection, data processing, classification algorithm. For example, Mandal M et al. [26] proposed a feature selection framework based on three-stage wrapper filter for disease detection, such as arrhythmia, leukemia, DLBCL and prostate cancer. Ijaz M F et al. [14] proposed a cervical cancer prediction model (CCPM) using risk factors as input, which removes abnormal data by using outlier detection method, increases the number of cases for balance and finally adopts random forest classifier to achieve good accuracy. Srinivasu P N et al. [39] proposed a computationally efficient anisotropic weighted-heuristic algorithm for real-time image segmentation (AW-HARIS) algorithm to automatically segment CT images for identifying the abnormalities of human liver. However, radiomic methods need to rely on manual labels with accurate tumor

boundaries [6, 10]. Since radiological features are only calculated within the tumor area [49], the microenvironment and tumor attachment tissues are easily overlooked, resulting in poor specificity of result prediction.

To solve these problems, a large number of end-to-end deep learning models have been proposed and successfully applied to image classification, object detection and image segmentation, such as CNN [21], AlexNet [18], VGGNet [36], ResNet [8] and DenseNet [11] models. These models can alleviate these problems by self-learning technique without accurate tumor boundary annotation [19, 34], and can automatically learn features from image data for specific clinical analysis [44]. VGGNet is a multi-layer depth network model, which is proposed by Simonyan K et al. [36]. This model can achieve high accuracy on multiple image recognition datasets, especially VGG-16 and VGG-19 models. Based on VGG-16 model, Chen K et al. [2] make a precision efficiency trade-off for a variety of structured model pruning methods on CIFAR-10 and ImageNet datasets. This method improves memory usage and speed of model on TPUs. ResNet models including ResNet18, ResNet34, ResNet50, ResNet101 and ResNet152 are proposed by He K et al. [8] to address the degradation problem. These networks are easier to optimize and can gain accuracy from considerably increased depth. DenseNet proposed by Huang G et al. [11] is a dense convolution network. The network enhances the network effect and reduces the use of parameters by reusing the extracted features and bypass. Because these deep learning models possess high accuracy, high efficiency and high reliability, they are widely used in all kinds of medical image research, such as skin disease classification [4, 40], eye disease diagnosis [42] and non-invasive liver disease prediction [45]. Srinivasu P N et al. [40] proposed a computerized process of classifying skin disease based MobileNet V2 and Long Short Term Memory (LSTM), which is proved to be efficient in maintaining stateful information for precise predictions.

In addition, many deep learning models perform well in assisting lung cancer analysis [43, 46], and have been gradually applied to the study of image-based gene mutation prediction. Wang S et al. [47] firstly proposed an end-to-end deep learning model that uses CT images to predict the EGFR mutation status in lung adenocarcinoma. Song K et al. [38] proposed a joint network named segmentation-based multi-scale attention model (SMSAM) to predict the mutation status of KRAS gene in rectal cancer. Qin R et al. [29] proposed a hybrid network combining 3D CNN and RNN to design multi-type features and analyze their dependencies for the prediction of EGFR mutation status. However, there are not many studies on identifying EGFR mutation status of lung cancer by images based on deep learning methods, and extracting effective discriminant features for the non-invasive prediction of EGFR mutation status is still a great challenge.

In this work, we developed the ResNet with mixed loss based on batch training technique (ResNet-MLB) to extract CT image features and identify EGFR mutation status. The proposed models trained on the public dataset can be effectively transferred to another datasets from different hospital, which shows the good applicability and effectiveness in identifying EGFR mutation status. The proposed models can automatically learn the relevant features of EGFR mutation from CT images, which only requires manual selection of image blocks containing tumor regions in CT images and does not require precise tumor boundary segmentation or human-defined features. This study is a non-invasive auxiliary detection method, which is suitable for avoiding invasive injury when surgery and biopsy are inconvenient. Meanwhile, it can help the clinician to make treatment decisions for the patient and it is of positive significance to reduce the burden of doctors and promote the development of medicine.

The main contributions of this paper are as follows:

1. The ResNet-MLB is proposed to identify the EGFR mutation status through extracting more relevant features from CT images, which is a non-invasive and easy-to-implement method for detecting gene status.
2. A novel mixed loss based on the batch similarity and cross entropy is introduced, which can be easily integrated into some existing CNN models, such as VGGNet, DenseNet and ResNet.
3. The combination of mixed loss and batch training strategy is firstly applied to the VGGNet, DenseNet and ResNet models to recognize gene mutation status by images.

The organizational structure of this paper is described as follows. Section 2 introduces the overall architecture of models based on the batch training strategy, the details of the designed mixed loss, and the computational complexity of the mixed loss with batch training strategy. Section 3 describes the experiments to demonstrate the effectiveness of the batch training technique and the mixed loss of the proposed model. The conclusion is given in Section 4.

2 Methods

2.1 Overall architecture

ResNet-MLB is proposed to identify the EGFR mutation status through extracting more discriminative features from CT images. The overall architecture mainly includes two parts: feature extractor and classifier. This paper mainly regards ResNet as a baseline for feature extraction which is mainly composed of the residual block and the jump connection between the block and block. The residual block consists of a series of convolutional layers, batch normalization and Relu activation layers. The jump connection makes the gradient of back propagation better by shortening the distance between non-adjacent layers. In addition, it also enables the network to automatically learn the path of feature motions without affecting the performance of the network, thereby enhancing the generalization ability of the network. A fully connected layer is used in the classifier, and the classification is achieved through softmax. The input dimension of the classifier is fixed as 512, and the output dimension is determined as the number of classes. For example, the status of EGFR genes can be classified as the wild type and mutant, so the output dimension of the classifier is set as 2. The overall architecture is shown in Fig. 1.
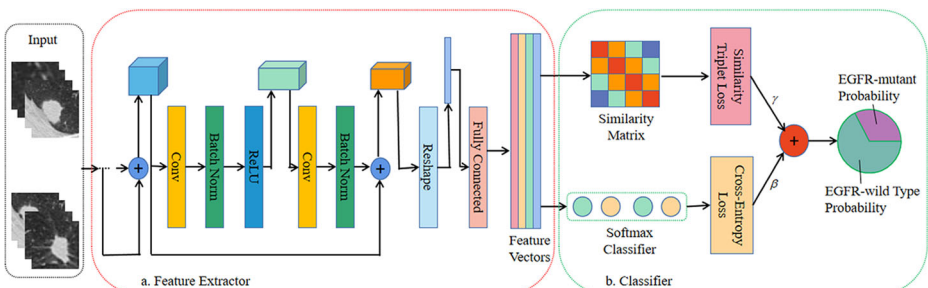

Fig. 1 The overall architecture of the proposed model

In our framework, the feature extractor $F(*)$, the classifier $C(*)$, and the class probability $P(*)$ are defined as follows,

$$F_i = F(x_i), C_i = C(F_i), P_i = P(C_i), i = 1, 2, \cdots, N, \tag{1}$$

where x_i represents the i-th image sample and N is the number of samples.

The feature $F_i \in R^{1 \times l}$ from the image x_i is extracted by the feature extractor $F(x_i)$, in which l represents the length of a feature vector. Then, the classification result $C_i \in R^{1 \times C}$ of the feature F_i is given by the classifier, in which C_i is the number of classes. The class probability $P_i \in R^{1 \times C}$ is obtained through the softmax layer, that is, the class probability is the final prediction result of EGFR mutation status based on the lung cancer.

2.2 Mixed loss

In the research field of medical image classification, the cross-entropy loss (CL) is widely used in CNN models to train networks. However, it is not able to measure the similarity of intra-class and inter-class of samples [12], which prevents CL from learning discriminative features of the samples. Therefore, several other loss functions are proposed in deep learning models, such as the contrastive loss, the triplet loss, the triplet lifted structure loss and the triplet hard loss, which are able to learn discriminative features, suppress intra-class change [51], and maximize the gap between different classes [13]. However, they also have some drawbacks. For example, the contrastive loss [7] needs to construct sample pairs to train the model. Although there are a considerable number of potential sample pairs in the training set, only a small number of sample pairs are usually sampled during the model training phase, which also results in a substantial loss of useful information. The triplet loss [32], triplet lifted structure loss, and triplet hard losses all need to construct triplets from training samples. Although the triplet lifted structure loss [37] considers all possible pairs, it is not smooth, and its smooth upper bound needs to be optimized. The triplet hard loss [9] selects only the hardest pair, which will filter out outliers and make the network unable to learn normal relationships.

In each iteration of the error back propagation algorithm, a batch input X with n_b samples is fed into a CNN model for training. For any sample $x_i \in X$, $i = 1, 2, \cdots, n_b$, we can get the feature vectors $F_i = F(x_i) \in R^{1 \times l}$, $F_{ij}^+, j = 1, 2, \cdots, n_i^+$ and $F_{ik}^-, k = 1, 2, \cdots, n_i^-$, where F_{ij}^+ denotes the feature vector corresponding to the samples of the same class as x_i, and F_{ik}^- denotes the opposite. Here, n_b, n_i^+ and n_i^- represent the number of batch samples, samples of the same class as sample x_i and samples of a different class from sample x_i, respectively, and they satisfy $n_b = n_i^+ + n_i^- + 1$. As mentioned above, the ideal triplet loss function is effective for clustering images of the same classes and separating images of the different classes. Therefore, the regular triplet loss is used and can be expressed as,

$$L_t = \max\left(0, \alpha + d\left(F_i, F_{ij}^+\right) - d\left(F_i, F_{ik}^-\right)\right),$$
$$i = 1, \cdots, n_b; j = 1, \cdots, n_i^+; k = 1, \cdots, n_i^-, \tag{2}$$

where $d(*, *)$ represents the distance measure of two vectors. α is a threshold that represents the minimum interval between the distance of positive sample pairs and the distance of negative sample pairs, which is an important index to measure similarity. The cosine distance is used to measure the similarity between samples in this work. Hence, the corresponding

triplet loss based on similarity for the feature vectors $\left(F_i, F_{ij}^+, F_{ik}^- \right)$, $i = 1, \cdots, n_b; j = 1, \cdots,$ $n_i^+; k = 1, \cdots, n_i^-$, is redefined as

$$L_{trip}\left(s_{ij}^+, s_{ik}^- \right) = \max\left(0, q - s_{ij}^+ + s_{ik}^- \right), \tag{3}$$

where $q \in [0, 1]$ is a threshold to measure cosine similarity between positive and negative sample pairs. s_{ij}^+ and s_{ik}^- are the cosine similarity between the feature vectors F_i and F_{ij}^+, and that between the feature vectors F_i and F_{ik}^-, respectively. They can be calculated by.

$$s_{ij}^+ = \frac{F_i^T F_{ij}^+}{\|F_i\|_2 \|F_{ij}^+\|_2}, \text{ and } s_{ik}^- = \frac{F_i^T F_{ik}^-}{\|F_i\|_2 \|F_{ik}^-\|_2}, \tag{4}$$

in which F_i^T denotes the transposition of F_i and $\|F_i\|_2$ represents the 2 norm of F_i. By reducing the loss L_{trip}, s_{ij}^+ will be close to 1 and s_{ik}^- will be close to 0. Note that due to the diversity and complexity of triple inputs, that is to say, the calculation of the triplet loss of the feature vector F_i needs reference vectors F_{ij}^+ and F_{ik}^-, it is not conducive to training.

In order to address these problems, a new mixed loss based on batch similarity and cross entropy is proposed to guide the network to better learn the model parameters. The mixed loss function is defined on the input batch samples for training the network. The similarities of all possible sample pairs in the batch is stored in the batch similarity matrix $S \in R^{n_b \times n_b}$, in which the element S_{ij} can be calculated by

$$S_{ij} = \widetilde{F}_i \widetilde{F}_j^T, i, j = 1, 2, \cdots, n_b, \tag{5}$$

where S_{ij} represents the similarity of the i-th feacture vector $\widetilde{F}_i$ and the j-th feature vector $\widetilde{F}_j$, in which $\widetilde{F}_i = F_i / \|F_i\|_2, i = 1, 2, \cdots, n_b$ with the size $1 \times l$. n_b is the number of the input batch samples.

To analyze the similarity matrix S, a binary matrix $B \in R^{n_b \times n_b}$ corresponding the ground truth is constructed to distinguish the similarity between positive and negative pairs. Their element B_{ij} can be calculated by

$$B_{ij} = y_i y_j^T, i, j = 1, 2, \cdots, n_b, \tag{6}$$

where y_i is a row vector of all zeros, except the c_i-th element is one, corresponding to the i-th sample with the ground truth label c_i, in which $y_i \in R^{1 \times C}$, $c_i \in \{1, 2, \cdots, C\}$ and C is the number of classes.

From Eqs. (5) and (6), the discriminative similarity matrix D can be constructed as follows,

$$D = S \odot (2B - 1), \tag{7}$$

where $\mathbf{1} \in R^{n_b \times n_b}$ is a matrix whose elements are all 1 and the symbol $\odot$ denotes the Hadamard product. Note that the similarities of positive pairs are greater than 0, and the similarities of negative pairs are smaller than 0 in the matrix D. Moreover, the diagonal elements of D and S will be set to zeros, i.e. $d_{ii} = 0$ and $s_{ii} = 0$, because they represent the similarity between the vectors F_i and themselves, the data distribution cannot be evaluated. The process of constructing the required discriminant matrix to evaluate the similarities among all samples is shown in Fig. 2.
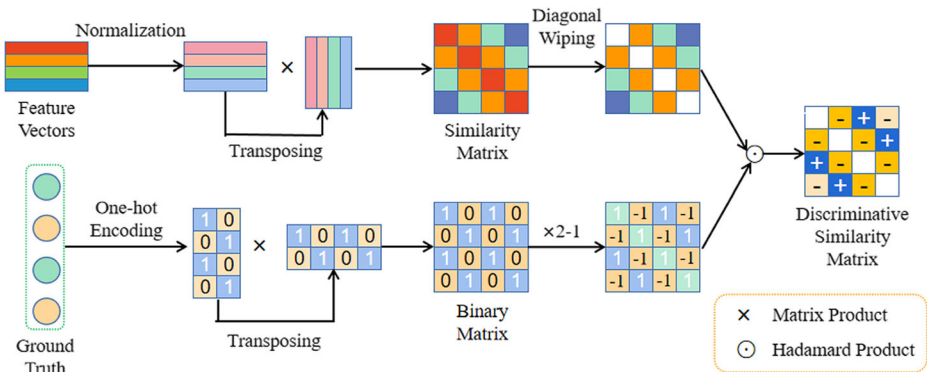
Fig. 2 Construction process of discriminative similarity matrix

The loss of each sample $x_i \in X$ in input batch can be evaluated by the i-th row of the above discriminant similarity matrix D. The value of $d_{ij}(d_{ij} \in D)$ represents the similarity between the feature vector F_i and other vectors in the batch samples (except for the diagonal element $d_{ii} = 0$). Obviously, the positive pairs and negative pairs in the i-th row may have multiple similarities. For convenience, for the i-th row, we re-express the similarity of positive pairs as $d_{ij}^+, j = 1, 2, \cdots, n_i^+$, and the similarity of negative pairs as $d_{ik}^-, k = 1, 2, \cdots, n_i^-$, and they satisfy that $n_b = n_i^+ + n_i^- + 1$. The triplet loss of x_i based on the batch similarity is defined as

$$L_{s-trip}(D, x_i) = \max\left(0, q - \frac{1}{n_i^+} \sum_{j=1}^{n_i^+} d_{ij}^{+2} + \frac{1}{n_i^-} \sum_{k=1}^{n_i^-} d_{ik}^{-2}\right) \tag{8}$$

where the similarities of the positive pairs and the negative pairs are replaced by the average square similarity, respectively. Since $|d_{ij}| \leq 1$, the square of d_{ij} is used instead of the linear function, so that the loss has a smoother gradient and more easily converges to the optimal solution. Therefore, the average batch similarity loss of a batch is expressed as

$$\overline{L_{s-trip}} = \frac{1}{n_b} \sum_{i=1}^{n_b} L_{s-trip}(D, x_i) \tag{9}$$

It is worth noting that the new batch similarity loss based on the triplet loss can further guide the model learning, so that samples of the same class are closer and the differences between samples of different class are more obvious, as shown in Fig. 3.

In order to better achieve the classification of EGFR mutation status and better evaluate the ability of classification, the softmax classifier and the CL function L_{ce} are also used in the structure of model. The CL based on the softmax probability is defined as,

$$L_{ce} = -\frac{1}{n_b} \sum_{i=1}^{n_b} \sum_{j=1}^{C} 1_{c_i = j} \log \frac{e^{w_j^T F_i^T + b_j}}{\sum_{j=1}^{C} e^{w_j^T F_i^T + b_j}}, \tag{10}$$

where F_i represents the extracted feature vector corresponding to the i-th image sample; w_j and b_j are the parameters of the classifier corresponding to the j-th class, which is composed of a
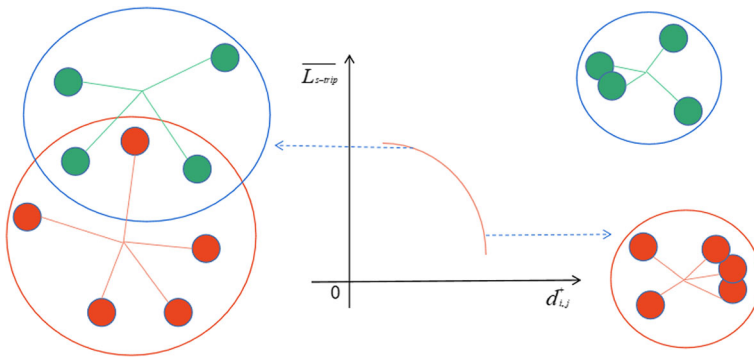
Fig. 3 Clustering process based on batch similarity loss

fully connected layer. c_i denotes the class of the i-th image. The value of $1_{c_i=j}$ is 1 when the ground truth c_i is equal to j.

Therefore, based on the average batch similarity loss in Eq. (9) and the CL in Eq. (10), the new mixed loss (ML) can be defined as,

$$L = \beta L_{ce} + \gamma \overline{L_{s-trip}}, \tag{11}$$

where β and γ are the weight parameters of the cross-entropy L_{ce} and the average batch similarity loss $\overline{L_{s-trip}}$, respectively. The ML can better classify and obtain the discriminative features of samples and Algorithm 1 shows the implementation procedure of the mixed loss.

Algorithm 1 The implementation procedure of the mixed loss

Input: $c_i, F_i, \beta, \gamma, i = 1, 2, \cdots, n_b$

c_i: the label of the image sample x_i. $y \in R^{n_b}$, where n_b represents the number of samples of a training batch.

F_i: the feature vector of the image sample x_i, and $F_i \in R^{1 \times l}$, in which l represents the length of the feature vector of the image.

β, γ: a constant between 0 and 1.

Output: the mixed loss L

---

1. Calculate the discriminative similarity matrix $D \in R^{n_b \times n_b}$, using the similarity matrix $S \in R^{n_b \times n_b}$ in Eq.(5) and the binary matrix $B \in R^{n_b \times n_b}$ in Eq. (6).

2. Find the similarity of the same class as x_i in the i-th row of the matrix D, which is marked as d_{ij}^+, i.e. positive sample pairs.

3. Find the similarity of the different class as x_i in the i-th row of the matrix D, which is marked as d_{ik}^-, i.e. negative sample pairs.

4. Calculate the batch similarity loss as Eq. (9).

5. Calculate the cross-entropy loss L_{ce} as Eq. (10).

6. Calculate the mixed loss L as Eq. (11).

Return L

2.3 Batch training strategy

The batch training (BT) strategy is adopted in the structure of model. It enables gradient descent to act on each batch and the amount of samples in each batch is small, so that models can be trained in limited memory. Meanwhile, BT can also be used for the distributed training to make the convergence faster. But the size of batch affects the stability of the training process. So it is significant to choose appropriate batch size for improving running efficiency and memory utilization. In order to obtain a better data distribution of the training data set, a new construction scheme of batch samples is used as Table 1. Table 1 gives the specific steps of construction of batch samples. Then, based on the batch samples, the ML is calculated and the ResNet-MLB model is trained. It is worthy note that this strategy is only used during training phase. In the testing phase, the batch similarity loss and the CL will be ignored, and the the softmax function is used to output the prediction results.

Taking the EGFR status of lung cancer as an example, the number of samples is recorded as $X = \{x_1, x_2, \cdots, x_N\}$, where N is the total number of samples. Using the construction scheme of batch samples in Table 1, the samples can be firstly divided into EGFR-wild type denoted as $\left\{x_1^-, x_2^-, \cdots, x_{N_w}^-\right\}$ and the EGFR-mutant denoted as $\left\{x_1^+, x_2^+, \cdots, x_{N_m}^+\right\}$ based on the sample labels. Note that $N = N_w + N_m$, where N_w and N_m denote the number of EGFR-wild type and the EGFR-mutant, respectively. Then randomly divide the EGFR-wild type into three groups by clustering: $\left\{x_1^-, x_2^-, \cdots, x_i^-\right\}$, $\left\{x_{i+1}^-, x_{i+2}^-, \cdots, x_j^-\right\}$, $\left\{x_{j+1}^-, x_{j+2}^-, \cdots, x_{N_w}^-\right\}$. Similarly, EGFR-mutant type is also randomly divided into three groups: $\left\{x_1^+, x_2^+, \cdots, x_m^+\right\}$, $\left\{x_{m+1}^+, x_{m+2}^+, \cdots, x_r^+\right\}$, $\left\{x_{r+1}^+, x_{r+2}^+, \cdots, x_{N_m}^+\right\}$. Finally, randomly select CT images in each group proportionally to form an input batch.

2.4 Computational complexity

In this subsection, the computational complexity (C_{cpl}) of the mixed loss based on batch training strategy in each iteration is performed at the cost of the required number of floating-point operations (FLOPs). Assuming that the number of training data set is N_{train}, the computational complexity can be calculated by

$$C_{cpl} = C_{BT} + C_{ML}, \tag{12}$$

where C_{BT} and C_{ML} are the computational complexity of executing batch training strategy and mixed loss in each iteration, respectively.

The computational complexity of executing batch training C_{BT} can be obtained through complexity of four steps in Table 1,

$$C_{BT} = N_{train} + O(clustering) + N_{group} + n_b, \tag{13}$$

Table 1 The construction procedure of the batch samples

(1). The original CT images are divided into different classes according to the ground truth;
(2). The CT images in each class are randomly divided into several equal groups by clustering;
(3). Count the number of images in each group and determine the proportion of images in the group;
(4). Proportionally randomly select CT images in each group as an input batch.

where N_{train} denotes the complexity of the first step of BT strategy. $O(clustering)$ is the required number of FLOPs for the utilized clustering method in the second step. N_{group} is the number of groups of training images and N_{group} FLOPs is taken to get the proportion of images in the group in the third step. The fourth step requires n_b FLOPs to build an input batch. It is noted that the first three steps of the proposed BT strategy are only executed once before the training process, while the fourth step is executed once per iteration in the training process.

The computational complexity of executing mixed loss C_{ML} can be obtained by Eqs. (5)–(9),

$$C_{ML} = \left(l^2 + C^2 + l + C + 3\right)n_b^2 + (2l + 5)n_b, \tag{14}$$

where Eq. (5) requires ln_b multiplication and $(l-1)n_b$ addition for normalization of embedded vectors, i.e. $(2l-1)n_b$ FLOPs. The calculation of the similarity matrix S needs $\left(l^2 + l - 1\right)n_b^2$ FLOPs in Eq. (5). In addition, the ground truth encoded as a one-hot vector y_i needs n_b index and allocation operations, which is calculated as $2n_b$ FLOPs for convenience in this paper. Meanwhile, the calculation of the binary matrix B in Eq. (6) is $\left(C^2 + C - 1\right)n_b^2$ FLOPs. For the discriminant similarity matrix D, the cost is $3n_b^2 + n_b$ FLOPs, which is calculated every time and shared each sample in the batch, and it is obtained by setting the diagonal value of the matrix on the right of Eq. (7) to 0. Eq. (8) needs to be repeatedly calculated n_b times, with a total cost of $2n_b^2 + 2n_b$ FLOPs. In Eq. (9), n_b FLOPs is performed for the average operation of the batch similarity loss.

3 Results and discussion

3.1 Datasets and details

3.1.1 Clinical characteristics of patients

Some experiments are conducted on the public dataset NSCLC Radiogenomics as the training set and the cooperative hospital dataset from Shanxi Province as the validation set. The NSCLC Radiogenomics dataset is downloaded from the TCIA website (https://wiki.cancerimagingarchive.net). The institutional review board of Shanxi cancer hospital has approved this retrospective study and abandoned the need to obtain patient informed consent. Meanwhile, patients from the public and the cooperative hospital need to meet the following inclusion criteria:

(1).　Primary lung cancer confirmed by histology;
(2).　Pathological examination of tumor specimens to confirm EGFR mutation status;
(3).　Preoperative enhanced CT data.

Besides, in the training and validation datasets, patients will be excluded with the following situation, such as (1) the lack of clinical data (age, gender, stage); (2) receiving preoperative treatment; or (3) exceeding 1 month from CT examination to postoperative operation.

The lesion areas in all CT images from 155 patients of the public dataset and 56 patients of the cooperative hospital are marked by these experienced radiologists (lung imaging practice for 12 years) in the partner hospital. Based on these marked lesion areas, the dataset in the
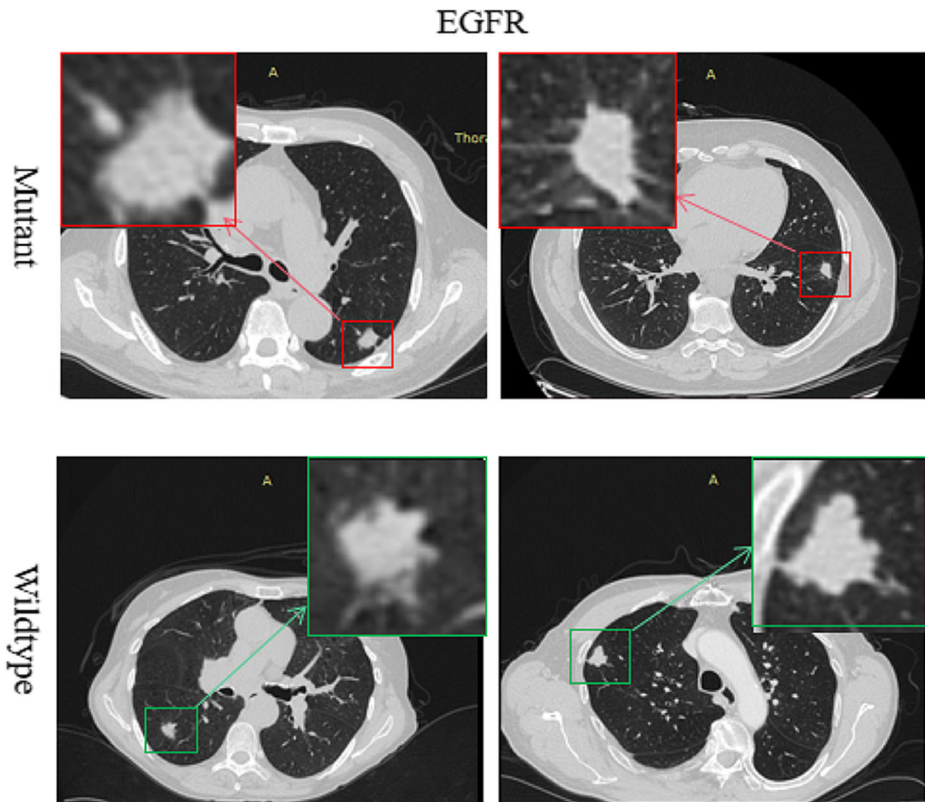
Fig. 4 Lung cancer CT images including the EGFR mutant and EGFR wild image samples

experimentation is constructed, including a total of 16,040 image blocks with size 64*64 in which all marked tumor lesion areas are contained, and each sample image block is classified into EGFR-mutant and EGFR wild type based on the patient's clinical information. Figure 4 shows some CT images including the EGFR-mutant and EGFR-wild image samples.

Table 2 lists the detailed construction of the lung cancer dataset. In the dataset, there are the training sets with 12,835 images including EGFR-mutant images 3310 and EGFR wild images 9525 from the public dataset, and the validation sets with 3205 images including EGFR-mutant images 825 and EGFR wild images 2380 from partner hospitals.

Table 3 presents the clinical characteristics of patients including the number of patients, average age, sex, smoking status, histology and EGFR mutation status in training set and validation set, and the corresponding p value between two datasets. It is from Table 3 that the p values of age, sex, smoking status, histology, and EGFR mutation status are greater than

Table 2 The construction of the lung cancer dataset

Data	EGFR-Mut	EGFR-Wild
Training	3310	9525
Validation	825	2380
Number of images	4135	11,905
Total of images	16,040	

Table 3 The clinical characteristics of patients

Attributes		training set	validation set	p value
The number of Patients		155	56	
Average age		67.9	62.8	0.073
Sex	Female	57 (36.8%)	32 (57.1%)	0.446
	Male	98 (63.2%)	24 (42.9%)	
Smoking status	Nonsmoker	37 (23.9%)	15 (26.8%)	0.797
	Former	95 (61.3%)	30 (53.6%)	
	Current	23 (14.8%)	11 (19.6%)	
Histology	Adenocarcinoma	135 (87.1%)	40 (71.4%)	0.362
	Squamous cell carcinoma	17 (11.0%)	16 (28.6%)	
	NSCLC NOS (not otherwise specified)	3 (1.9%)	0 (0.0%)	
EGFR mutation status	EGFR-wild	115 (74.2%)	36 (64.3%)	0.061
	EGFR-mutant	40 (25.8%)	20 (35.7%)	

0.05, which implies that there are no significant differences in age, sex, smoking status, histology and EGFR mutation status between the training set and the validation set. Note that when the p value is less than 0.05, there are the statistical significant of the corresponding characteristic between the training set and validation set.

3.1.2 Experimental details

Due to the small amount of medical image data, in order to prevent over fitting, some simple data enhancement methods, such as horizontal flip, vertical flip and random rotation, are used to expand the training set for improving the ability of the model classification.

In the experiment, the Adam gradient optimization algorithm is used to optimize the parameters of the model, the weight decay rate is set to 1e-8 and the learning rate is set to 1e-4. The input batch size is set to 36. Moreover, in order to calculate the batch input similarity triplet loss, the dimensionality of the extracted features is reduced and set to 512. In the mixed loss, the weight parameters β and γ are set 1 and 0.5, respectively. The performance of the model is evaluated on the validation set in each epoch.

To test the effectiveness of the proposed mixed loss, we have applied the mixed loss to VGG16Net, ResNet18, ResNet34, ResNet50, DenseNet networks. The number of parameters and computational complexity of these networks are listed in Table 4, where FLOPs are generated when the size of the input image is $64 \times 64 \times 1$. As can be seen from Table 4, DenseNet has the fewest parameters and ResNet18 has the least number of FLOPs.

The experiments in this work are carried out on a workstation with Ubuntu 18.04 LTS, the CPU of the server is 2.90 GHz Intel(R) Xeon(R) W-2102, and the GPU is NVIDIA TITAN XP with CUDA 10.1 for acceleration. Besides, all the deep learning frameworks are realized using Python 3.7.9 with Keras 2.3.1 and TensorFlow 1.15.0.

Table 4 The parameters and computational complexity of the compared networks

	VGG16Net	ResNet18	ResNet34	ResNet50	DenseNet
Parameters(M)	7.18	2.82	3.64	4.06	2.12
FLOPs(M)	415.47	286.73	356.45	395.34	313.25

3.2 Influence of the size of batch samples

In this subsection, consider the influence of the size n_b of the batch samples on the accuracy of the proposed models to choose appropriate batch size for improving running efficiency and memory utilization. Using the construction procedure of the batch samples, the training dataset on the lung cancer in Section 3.1 can be divided into two classes and three groups in each class by the K-means algorithm. Then, using eight different proportions, batch samples of different sizes are constructed, i.e. n_b = 6,12,18,24,30,36,42,48, where each batch of samples is obtained in the 6 groups in the same proportion.

In this experiment, the different batch sizes are used for training ResNet-MLB models, and other parameters are fixed as subsection 3.1.2. The accuracy (ACC) is used as an index to evaluate classification ability of models, which is calculated by:

$$ACC = \frac{TP + TN}{N'}, \tag{15}$$

where TP and TN, respectively, represent the number of samples of the correct prediction in all samples labeling EGFR-mutant (true positive) and that of the correct prediction in all samples labeling EGFR-wild type (true negative). N' is the total number of all images in the validation set.

Figure 5 shows ACC values of models at the different batch size n_b. It can be seen from Fig. 5 that the highest accuracy of 81.58% and the lowest accuracy of 78.52% are obtained at n_b = 36 and n_b = 6, respectively. In our results, the accuracy is positively correlated with the batch size n_b, which is in line with the hypothesis: A larger n_b can ensure that the data distribution of each batch is closer to the overall distribution of the training set. However, considering that a larger n_b will result in insufficient number of iterations for network training within an epoch, n_b = 36 is applied to the rest of the experiments as the default value.
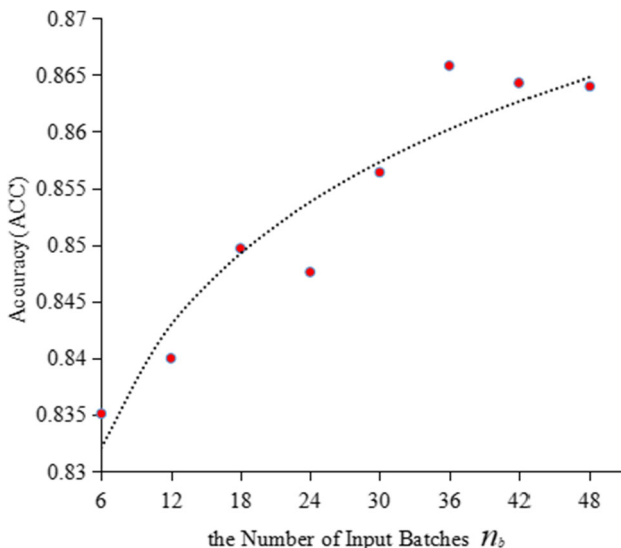

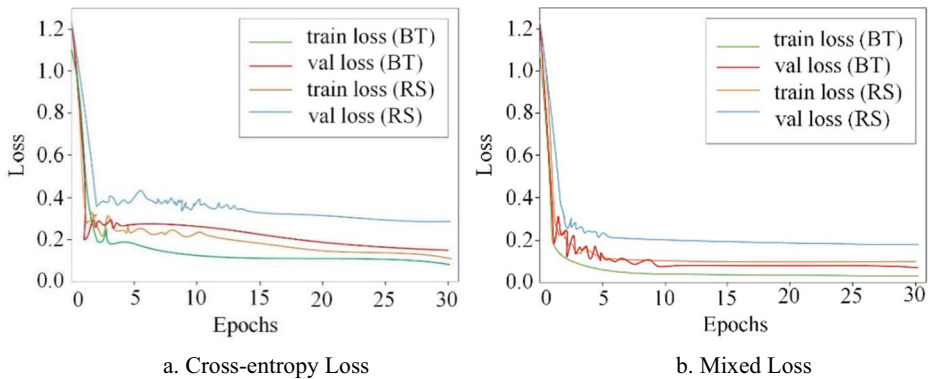Fig. 5 Accuracy analysis based on different input batches n_b of lung cancer dataset

a. Cross-entropy Loss   b. Mixed Loss

Fig. 6 Loss curves of different training strategies for ResNet34

3.3 Influence of batch training strategy

In this subsection, the effect of the BT strategy on the accuracy of the ResNet34 models using the CL (ResNet34-CL) or the ML (ResNet34-ML) is considered through the comparison with the random selection (RS) strategy. Figure 6a shows that the training and validation loss curves of ResNet34-CL models using BT and RS. Figure 6b shows that the training and validation loss curves of ResNet34-ML models using BT and RS. It is seen from these figures that the loss curve of the BT strategy is relatively smoother than that of the RS strategy for all models. This shows that it is easier to train the network using the BT strategy. In addition, we also find that the gap between the training loss and the validation loss is reduced by the BT strategy, which implies that it can alleviate the overfitting problem in the process of model training to some extent.

Table 5 lists the accuracy of ResNet34-CL and ResNet34-ML using BT and RS strategy on the validation set. As listed in Table 5, the accuracy of ResNet34-CL model using the BT strategy is 1.14% higher than that using RS strategy. The accuracy of ResNet34-ML using the BT strategy is improved by 0.93%, compared with ResNet34-ML using the RS strategy. The results indicate that the batch training strategy is beneficial for training models. Here, it can also be seen that the BT strategy is more effective for models based on the CL, meaning that it is able to compensate more significantly for the CL over the data distribution. The reason is that the ML including the batch similarity can evaluate the quality of the training data set distribution to a certain extent, but the CL function does not have this ability.

3.4 Comparison and verification of results

In this subsection, the applicability and effectiveness of ResNet models using the ML is studied. In this experiment, the CL, the CL combined with the triple loss (CTL) in ref. [32], the CL combined with the improved lifted structure loss (CIL) in ref. [37] and the proposed ML

Table 5 Accuracy of different training strategies for ResNet34

Scheme	Cross Entropy (RS)	Cross Entropy (BT)	Mixed Loss (RS)	Mixed Loss (BT)
ACC(%)	78.07	79.21	80.65	81.58

are applied into VGG16Net, ResNet18, ResNet34, ResNet50, DenseNet models for the comparison.

The effectiveness of the new ML is firstly considered by comparing with the CL. Figure 7 shows the curve of training loss and validation loss using different models with the CL and the ML. From these figures, we find that.

(1). The validation loss of the VGG16Net increases as the iteration increases for these models using the CL in Fig. 7b. This implies that the overfitting problem of VGG16Net is more serious than other models.
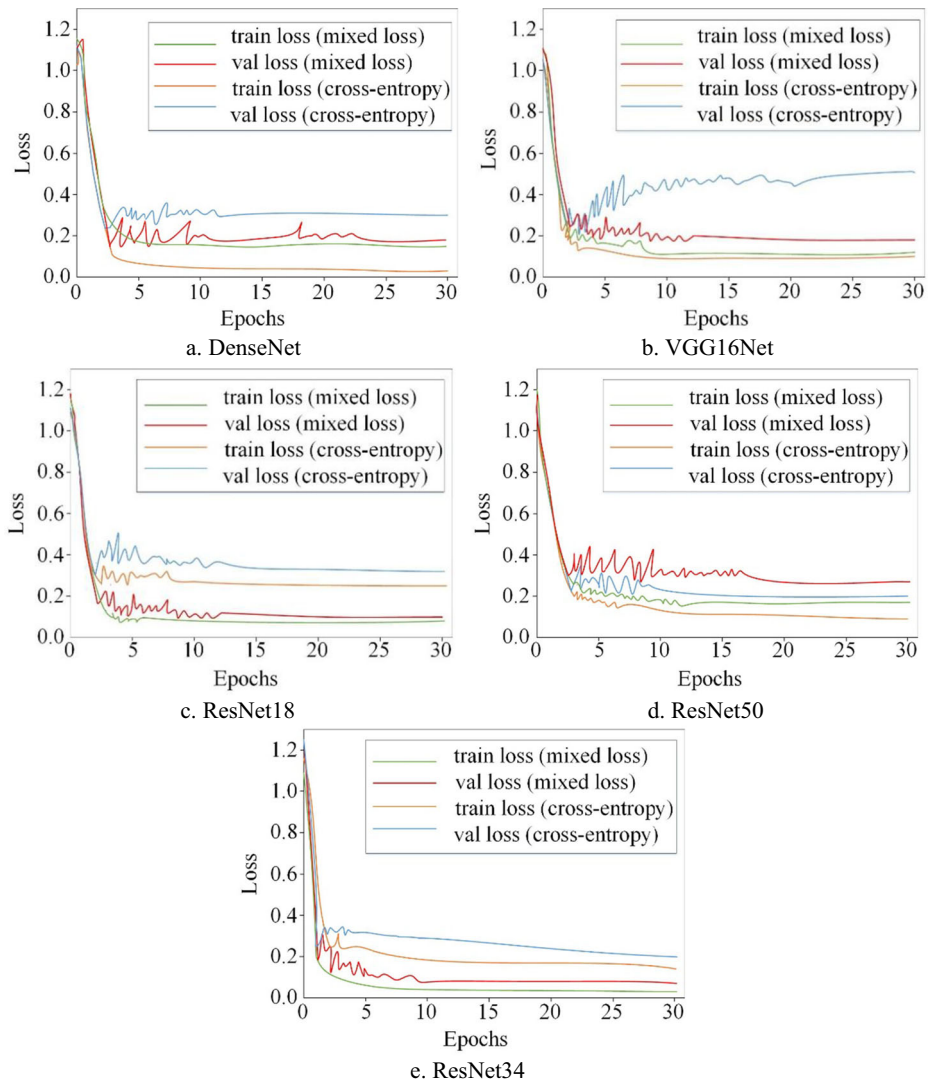(2). These models using ML are smoother, which implies that the overfitting problem can be suppressed.

Fig. 7 Loss curves for the compared networks

(3). The gap between the ML is smaller than that between the CL in all models. This demonstrate that the ML can play a regularizing effect on the CL.

In summary, the mixed loss can suppress the problem of overfitting, indicating that the mixed loss can play a regularizing effect on the CL.

Then, the performance of the models with the new ML is furtherly studied by comparing with models with the CL, CTL and CIL. The identification of the mutation status of EGFR is a binary classification task, and the sensitivity SE and specificity SP are used to evaluate the performance of these models in this experiment. They can be calculated by.

$$SE = \frac{TP}{TP + FN} \text{ and } SP = \frac{TN}{TN + FP}, \tag{16}$$

where FP and FN, respectively, represent the number of samples of the incorrect prediction in all samples labeling EGFR mutation (true positive) and that of the incorrect prediction in all samples labeling EGFR wild type (true negative). The sensitivity SE and specificity SP can measure the ability of models to correctly identify the EGFR-mutant and EGFR-wild type in CT images of lung cancer. In addition, the accuracy (ACC) and the receiver operating characteristic (ROC) area under the curve (AUC) are also used to evaluate the classification ability of the models. The results (including the sensitivity SE, the specificity SP, the ACC and AUC) of VGG16Net, ResNet18, ResNet34, ResNet50, DenseNet models with different loss (CL, CTL, CIL and ML) are listed in Table 6. From the table, we find that:

(1). The accuracy of all models with CIL and ML is higher than all models with CL, which means the improved lifted structure loss and batch similarity loss can improve the optimization ability of these models.

Table 6 Metrics for the compared networks based on each loss

Model	Loss	SE(%)	SP(%)	ACC(%)	AUC
VGG16Net	Cross-entropy	77.58	78.67	77.87	0.8089
	Cross-entropy+ Triple	78.82	80.59	79.50	0.8273
	Cross-entropy+ Lifted	76.71	79.75	78.40	0.8213
	Mixed loss	**79.07**	**81.53**	**80.32**	**0.8342**
ResNet18	Cross-entropy	76.83	78.21	76.24	0.8041
	Cross-entropy+ Triple	78.53	80.05	78.86	0.8106
	Cross-entropy+ Lifted	79.42	80.87	80.31	0.8191
	Mixed loss	**79.57**	**81.49**	**80.87**	**0.8236**
ResNet34	Cross-entropy	78.48	80.60	79.21	0.8780
	Cross-entropy+ Triple	78.94	81.53	80.85	0.8798
	Cross-entropy+ Lifted	79.17	82.75	81.05	0.8807
	Mixed loss	**80.02**	**82.90**	**81.58**	**0.8861**
ResNet50	Cross-entropy	78.20	79.14	78.31	0.8723
	Cross-entropy+ Triple	78.26	80.10	79.69	0.8783
	Cross-entropy+ Lifted	79.57	82.58	80.47	0.8791
	Mixed loss	**79.65**	**82.60**	**80.87**	**0.8805**
DenseNet	Cross-entropy	78.83	81.23	80.93	0.8714
	Cross-entropy+ Triple	79.68	81.53	80.48	0.8692
	Cross-entropy+ Lifted	80.85	**82.50**	81.06	**0.8778**
	Mixed loss	**81.07**	82.21	**81.25**	0.8741

Table 7 Comparison of the studies

Author	Method	Feature	SE(%)	SP(%)	ACC (%)	AUC
Liu Y et al [24]	SVM	Clinical features	–	–	–	0.709
Gevaert O et al [6]	Decision tree	Semantic features	–	–	–	0.890
Zhang L et al [50]	Statistical analysis	Radiomics features	70.90	79.80	75.60	0.873
Jia T Y et al [16]	Random forest	Radiomics features	75.80	79.10	77.20	0.838
Wang S et al [47]	Deep learning (Improved DenseNet + Cross-entropy)	Depth features	72.27	75.41	73.86	0.810
Dong Y et al.	Deep learning (Inception-resnet-attention+ Cross-entropy)	Depth features+ Clinical features	78.27	81.35	79.43	0.866
Our research	Deep learning (ResNet34+Mixed loss)	Depth features	80.02	82.90	81.58	0.8861

(2). The highest sensitivity, specificity, accuracy and AUC are obtained in VGG16Net, ResNet18, ResNet34 and ResNet50 models with ML, which demonstrates that ML is more effective than CL, CTL and CIL for VGGNet and ResNet models. In addition, the DenseNet with ML and the DenseNet with CIL have higher performance, which means that there is a good robustness of both ML and CIL for the DenseNet model.

(3). ResNet34 provides the highest accuracy (81.58%) for four different losses in all models, which is 2.37% higher than the model with CL, 0.73% higher than the model with CTL and 0.53% higher than the model with CIL. This shows that the ResNet34-ML can better learn the discriminative characteristics of samples, and get better classification ability.

In general, ML has better adaptability and effectiveness in model training, and the ResNet34-ML can obtain best performance.

In the experiment, the task of identifying EGFR mutations in lung cancer is compared with the latest research, and the results are listed in Table 7. Table 7 shows that there is the highest sensitivity, specificity, accuracy and AUC in the ResNet34-ML compared with other studies. It further illustrates that the architecture proposed has a certain degree of improvement ability, and the ResNet34-ML with the BT strategy is effective for identifying the EGFR mutation status.

4 Conclusions

In this work, ResNet-MLB models are proposed using the mixed loss and the batch training technique for identification of EGFR mutation status in lung cancer. In these models, the mixed loss is proposed based on the batch similarity and cross entropy, and the batch training technique is applied, which guide the network to better learn the parameters. Some experiments about the size of batch samples, batch training strategy and various models with different losses are studied on the CT images of lung cancer dataset, and the following conclusions are obtained: (1) The performance of the BT strategy is 0.93% higher than that of a RS strategy for ResNet34-ML models, and the performance of ResNet34-CL using the BT strategy is improved by 1.14%, compared with ResNet34-CL using the RS strategy. Hence, the BT strategy is beneficial for training models, especially for ResNet34-CL. That is because the BT strategy is able to compensate more significantly for the CL over the data distribution,

while the ML including the batch similarity can evaluate the quality of the training data set distribution to a certain extent. (2) For some common models, the proposed mixed loss has superiority in sensitivity, specificity, accuracy and AUC (sensitivity = 80.02%, specificity = 82.90%, accuracy = 81.58%, AUC = 0.8861), compared with other losses, which means ML has better adaptability and effectiveness in model training. (3) The ResNet34-ML with the batch training technique can better learn the discriminative characteristics of samples, and get the best classification performance in all models.

In short, the proposed mixed loss possesses the applicability and effectiveness, and ResNet34-ML with the batch training technique has better identification ability on the CT images of lung cancer dataset. The advantage of our method is that it provides a non-invasive alternative solution for identifying the EGFR mutation status when the patient is not suitable for biopsy, and quickly promotes the clinician to make treatment decisions for the patient.

Although the performance of ResNet-MLB models is encouraging, this study has some limitations. First, our research only focused on the EGFR mutation status of lung cancer. However, the relationship between EGFR mutation and other gene mutations (such as KRAS, ALK) is unconsidered. Secondly, we only consider the identification of the EGFR mutation status of lung cancer based on CT images, the combination of CT and other images (such as PET) is unclear. Therefore, the correlation between EGFR mutations and other gene mutations will be explored by introducing attention mechanism and multi-task learning in the future. Besides, more CT images and other images of lung cancer will be collected to design a fusion strategy which may improve the identifiable performance.

Data availability The public dataset NSCLC Radiogenomics generated during and/or analysed during the current study is available in the [TCIA] repository, [https://www.cancerimagingarchive.net/access-data/]. The cooperative hospital dataset generated during and/or analysed during the current study is available from the corresponding author on reasonable request.

Declarations

Conflict of 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 paper.

References

1. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
2. Chen K, Franko K, Sang R (2021) Structured model pruning of convolutional networks on tensor processing units [J]. arXiv preprint arXiv:2107.04191
3. Eberhard DA, Johnson BE, Amler LC, Goddard AD, Heldens SL, Herbst RS, Ince WL, Jänne PA, Januario T, Johnson DH, Klein P, Miller VA, Ostland MA, Ramies DA, Sebisanovic D, Stinson JA, Zhang YR, Seshagiri S, Hillan KJ (2005) Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non–small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. J Clin Oncol 23(25):5900–5909

4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature. 542:115–118

5. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, Rubin DL, Napel S, Plevritis SK (2012) Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data–methods and preliminary results. Radiology. 264:387–396

6. Gevaert O, Echegaray S, Khuong A et al (2017) Predictive radiogenomics modeling of EGFR mutation status in lung cancer[J]. Sci Rep 7(1):1–8

7. Hadsell R, Chopra S, Lecun Y, et al (2006) Dimensionality reduction by learning an invari- ant mapping[C]. In: IEEE conference on computer vision and pattern recognition (CVPR). pp 1735–1742

8. He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778

9. Hermans A, Beyer L, Leibe B, et al (2017) In defense of the triplet loss for person reidentification arXiv preprint arXiv: 1703.07737

10. Horvat N, Veeraraghavan H, Pelossof RA, Fernandes MC, Arora A, Khan M, Marco M, Cheng CT, Gonen M, Golia Pernicka JS, Gollub MJ, Garcia-Aguilar J, Petkovska I (2019) Radiogenomics of rectal adenocarcinoma in the era of precision medicine: a pilot study of associations between qualitative and quantitative MRI imaging features and genetic mutations [J]. Eur J Radiol 113:174–181

11. Huang G, Liu Z, Van Der Maaten L, et al (2017) Densely connected convolutional networks[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 4700–4708

12. Huang KK, Ren CX, Liu H et al (2020) Hyperspectral image classification via discriminative convolutional neural network with an improved triplet loss[J]. Pattern Recogn 112(2):107744

13. Huang Z, Zhou Q, Zhu X, Zhang X (2021) Batch similarity based triplet loss assembled into light-weighted convolutional neural networks for medical image classification [J]. Sensors. 21(3):764

14. Ijaz MF, Attique M, Son Y (2020) Data-driven cervical cancer prediction model with outlier detection and over-sampling methods [J]. Sensors. 20(10):2809

15. Itakura H, Achrol AS, Mitchell LA et al (2015) Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med 7:303ra138

16. Jia TY, Xiong JF, Li XY, Yu W, Xu ZY, Cai XW, Ma JC, Ren YC, Larsson R, Zhang J, Zhao J, Fu XL (2019) Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling[J]. Eur Radiol 29(9):4742–4750

17. Karlo CA, Di Paolo PL, Chaim J et al (2014) Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations. Radiology. 270:464–471

18. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks [J]. Comm ACM 60(6):84–90

19. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 284:574–582

20. Lambin P, Leijenaar RT, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762

21. LeCun Y, Boser B, Denker J et al (1989) Handwritten digit recognition with a back-propagation network [J]. Adv Neural Inf Proces Syst 2

22. Linardou H, Dahabreh IJ, Kanaloupiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S (2008) Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. Lancet Oncol 9(10):962–972

23. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, Ye Z, Gillies RJ (2016) CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. Radiology. 280: 271–280

24. Liu Y, Kim J, Balagurunathan Y et al (2016) Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. Clin Lung Cancer 17(5):441–8. e6

25. Loughran C, Keeling C (2011) Seeding of tumour cells following breast biopsy: a literature review. Br J Radiol 84:869–874

26. Mandal M, Singh PK, Ijaz MF, Shafi J, Sarkar R (2021) A tri-stage wrapper-filter feature selection framework for disease classification [J]. Sensors. 21(16):5571

27. Morgado J, Pereira T, Silva F, Freitas C, Negrão E, de Lima BF, da Silva MC, Madureira AJ, Ramos I, Hespanhol V, Costa JL, Cunha A, Oliveira HP (2021) Machine learning and feature selection methods for egfr mutation status prediction in lung cancer [J]. Appl Sci 11(7):3273

28. Oh JE, Kim MJ, Lee J, Hur BY, Kim B, Kim DY, Baek JY, Chang HJ, Park SC, Oh JH, Cho SA, Sohn DK (2020) Magnetic resonance-based texture analysis differentiating KRAS mutation status in rectal cancer[J]. Cancer Res Treat Off J Korean Cancer Assoc 52(1):51–59

29. Qin R, Wang Z, Qiao K, Hai J, Jiang L, Chen J, Pei X, Shi D, Yan B (2020) Multi-type interdependent feature analysis based on hybrid neural networks for computer-aided diagnosis of epidermal growth factor receptor mutations[J]. IEEE Access 8:38517–38527

30. Rebecca L, Siegel et al (2017) Cancer statistics [J]. CA Cancer J Clin 67:7–30

31. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos M, Fennessy F, Mak RH, Gillies R, Quackenbush J, Aerts HJWL (2017) Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer Res 77:3922–3930

32. Schroff F, Kalenichenko D, Philbin J, et al (2015) FaceNet: a unified embedding for face recognition and clustering [C]. In: IEEE conference on computer vision and pattern recognition (CVPR). pp 815–823

33. Sequist LV, Yang JC, Yamamoto N, O'Byrne K, Hirsh V, Mok T, Geater SL, Orlov S, Tsai CM, Boyer M, Su WC, Bennouna J, Kato T, Gorbunova V, Lee KH, Shah R, Massey D, Zazulina V, Shahidi M, Schuler M (2013) Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 31:3327–3334

34. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y, Tian J (2017) Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recogn 61:663–673

35. Shiri I, Maleki H, Hajianfar G, et al (2020) Next generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning approaches [J]. Mol Imag Biol 22:1132–1148

36. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556

37. Song HO, Xiang Y, Jegelka S et al (2016) Deep metric learning via lifted structured feature embedding [C]. In: IEEE conference on computer vision and pattern recognition (CVPR). pp 4004–4012

38. Song K, Zhao Z, Wang J, et al (2022) Segmentation-based multi-scale attention model for KRAS mutation prediction in rectal cancer [J]. Int J Mach Learn Cybern 13:1283–1299

39. Srinivasu PN, Ahmed S, Alhumam A et al (2021) An AW-HARIS based automated segmentation of human liver using CT images [J]. Comput Mater Contin 69(3):3303–3319

40. Srinivasu PN, SivaSai JG, Ijaz MF et al (2021) Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM[J]. Sensors. 21(8):2852

41. Thawani R, McLane M, Beig N et al (2018) Radiomics and radiogenomics in lung cancer: a review for the clinician. Lung Cancer 115:34–41

42. Ting DS, Cheung CY, Lim G et al (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 318:2211–2223

43. Wang S, Zhou M, Liu Z, Liu Z, Gu D, Zang Y, Dong D, Gevaert O, Tian J (2017) Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med Image Anal 40:172–183

44. Wang S, Liu Z, Rong Y et al (2018) Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. Radiother Oncol:S0167–S8140

45. Wang K, Lu X, Zhou H, et al (2019) Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multi-centre study [J]. Gut 68(4):729–741

46. Wang S, Liu Z, Chen X, et al (2018) Unsupervised deep learning features for lung cancer overall survival analysis [C]. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society, IEEE. pp 2583–2586

47. Wang S, Shi J, Ye Z, et al (2019) Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning [J]. Europ Respir J 53(3):1800986

48. Yano M, Sasaki H, Kobayashi Y, Yukiue H, Haneda H, Suzuki E, Endo K, Kawano O, Hara M, Fujii Y (2006) Epidermal growth factor receptor gene mutation and computed tomographic findings in peripheral pulmonary adenocarcinoma. J Thorac Oncol 1:413–416

49. Yue Q, Yu Y, Shi Z, Wang Y, Zhu W, du Z, Yao Z, Chen L, Mao Y (2017) Prediction of BRAF mutation status of craniopharyngioma using magnetic resonance imaging features [J]. J Neurosurg 129(1):27–34

50. Zhang L, Chen B, Liu X, Song J, Fang M, Hu C, Dong D, Li W, Tian J (2018) Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell lung Cancer. Transl Oncol 11(1):94–101

51. Zhang J, Lu C, Wang J, Yue XG, Lim SJ, al-Makhadmeh Z, Tolba A (2020) Training convolutional neural networks with multi logize images and triplet loss for remote sensing scene classification [J]. Sensors. 20(4):1188

52. Zhao J, Ji G, Qiang Y, Han X, Pei B, Shi Z (2015) A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm [J]. PLoS One 10(4):e0123694

53. Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, Zhang S, Wang J, Zhou S, Ren S, Lu S, Zhang L, Hu C, Hu C, Luo Y, Chen L, Ye M, Huang J, Zhi X, … You C (2011) Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 12:735–742
54. Zhou J, Zheng J, Yu Z et al (2015) Comparative analysis of clinicoradiologic characteristics of lung adenocarcinomas with ALK rearrangements or EGFR mutations. Eur Radiol 25:1257–1266
55. Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, Berry GJ, Plevritis SK, Rubin DL, Napel S, Gevaert O (2018) Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. Radiology. 286:307–315