



Circulating Tumor Cell Identification Based on Deep Learning

Zhifeng Guo, Xiaoxi Lin, Yan Hui, Jingchun Wang, Qiuli Zhang and Fanlong Kong*

Department of Oncology, Chifeng Municipal Hospital, Chifeng, China

OPEN ACCESS

Edited by:

Min Tang,
Jiangsu University, China

Reviewed by:

Jiasheng Yang,
Changsha Medical University, China
Lan Yu,
Inner Mongolia People's Hospital,
China

*Correspondence:

Fanlong Kong
dinosaurand@163.com

Specialty section:

This article was submitted to
Cancer Imaging and
Image-directed Interventions,
a section of the journal
Frontiers in Oncology

Received: 27 December 2021

Accepted: 21 January 2022

Published: 16 February 2022

Citation:

Guo Z, Lin X, Hui Y, Wang J,
Zhang Q and Kong F (2022)
Circulating Tumor Cell Identification
Based on Deep Learning.
Front. Oncol. 12:843879.
doi: 10.3389/fonc.2022.843879

As a major reason for tumor metastasis, circulating tumor cell (CTC) is one of the critical biomarkers for cancer diagnosis and prognosis. On the one hand, CTC count is closely related to the prognosis of tumor patients; on the other hand, as a simple blood test with the advantages of safety, low cost and repeatability, CTC test has an important reference value in determining clinical results and studying the mechanism of drug resistance. However, the determination of CTC usually requires a big effort from pathologist and is also error-prone due to inexperience and fatigue. In this study, we developed a novel convolutional neural network (CNN) method to automatically detect CTCs in patients' peripheral blood based on immunofluorescence *in situ* hybridization (imFISH) images. We collected the peripheral blood of 776 patients from Chifeng Municipal Hospital in China, and then used Cytel to delete leukocytes and enrich CTCs. CTCs were identified by imFISH with CD45+, DAPI+ immunofluorescence staining and chromosome 8 centromeric probe (CEP8+). The sensitivity and specificity based on traditional CNN prediction were 95.3% and 91.7% respectively, and the sensitivity and specificity based on transfer learning were 97.2% and 94.0% respectively. The traditional CNN model and transfer learning method introduced in this paper can detect CTCs with high sensitivity, which has a certain clinical reference value for judging prognosis and diagnosing metastasis.

Keywords: circulating tumor cells, detection, count, convolutional neural network, transfer learning

INTRODUCTION

Circulating tumor cells (CTC) are all kinds of tumor cells in peripheral blood (1). Most of the CTCs undergo apoptosis or phagocytosis after entering the peripheral blood, while a minority of CTCs develop into metastasis and undergo for a period of dormancy, and lead to metastatic tumor (2, 3). Cancer recurrence and metastasis are the main causes of death in cancer patients (4, 5). A large number

of experiments on esophageal squamous cell carcinoma (6), breast cancer (7, 8), prostate cancer (9) and lung cancer (10) have proved that CTCs were closely related to the prognosis of patients with advanced cancer. As a simple blood test, CTCs detection has the advantages of high safety, low cost and repeatability, which is available at any time to evaluate the prognosis and recurrence risk of patients (11, 12). Many experiments used liquid biopsy to monitor the CTCs response in patients with malignant tumors to evaluate the therapeutic response (13–15). Many studies have shown that CTCs count is closely related to prognosis, which has an important reference value for determining clinical results and recurrence risk (16–18). The fluid biopsy can predict disease progression in real time to assess tumor heterogeneity, and it was possible to detect single CTCs or cell clusters (13, 19–21). Immune enrichment with multiparameter flow cytometric is the gold standard of CTCs detection (22), but this method was limited due to the lack of tumor-specific markers, in this case, multi-label immunofluorescence staining was essential. Epithelial cell adhesion molecule (EpCAM) was often used to detect cancer cells in the blood because it mediates contact with homotype cells in epithelial tissue (23–25). The methods of CD45⁺, DNA fluorescence *in situ* hybridization (FISH) of the centromere of chromosome 8 probe (CEP8⁺)/chromosome 17 centromere duplication (CEP17⁺) have been widely used to identify CTCs (26, 27).

In recent years, rapid and automatic identification of CTCs is becoming more and more important, and the research on the automatic identification process of CTCs was also accelerating (28, 29), such as cell search system to obtain digital images (30), rare event imaging system (REIS) (31), microfluidic platform composed of multi-functional microfluidic chip and unique image processing algorithm (32). However, due to the heterogeneity of CTCs, these classification methods were often subjective. Therefore, under certain conditions, test results vary from examiner to examiner. The development of artificial intelligence (AI) has accelerated scientists' research on machine learning. Machine learning has been widely used in medical research because of its advantages of objectivity, rapidity, and overcoming noise (33–35), especially in medical images (36, 37). As the classical algorithms of machine learning, deep learning and convolutional neural network (CNN) have made outstanding contributions in promoting medical research (38–40). Anthimopoulos et al. proposed the first problem specific deep CNN for classification of interstitial lung diseases (ILD), the results showed that (classification performance~85.5%) CNN has potential in analyzing ILD (41). Poplin et al. used the Inception-v3 neural network structure to predict potential cardiovascular risk factors in retinal fundus images (42). Le et al. constructed a deep neural network to classify Rab protein molecules through two-dimensional CNN, which provided a valuable reference for biological modeling using deep neural network (43). At present, CNN has been widely used to promote biomedical image analysis and successfully applied in cancer diagnosis and tissue identification (44, 45). Compared with the traditional machine learning methods, CNN-based automatic image processing method has the advantage of eliminating the bias caused by personal subjectivity (46). Negative enrichment

combined with immuno fluorescence *in situ* hybridization (imFISH) to detect CTCs has been proven to be feasible and clinically valuable (47, 48).

In this study, we applied deep learning to identify CTCs to reduce the subjective error. ImFISH was used to detect patients' CTCs, each image contains positive CTCs nucleus and negative control to segment the images of circulating tumor cells. CNN deep learning network was used to identify circulating tumor cells and count CTCs.

MATERIALS AND METHODS

A Framework for Identifying CTCs

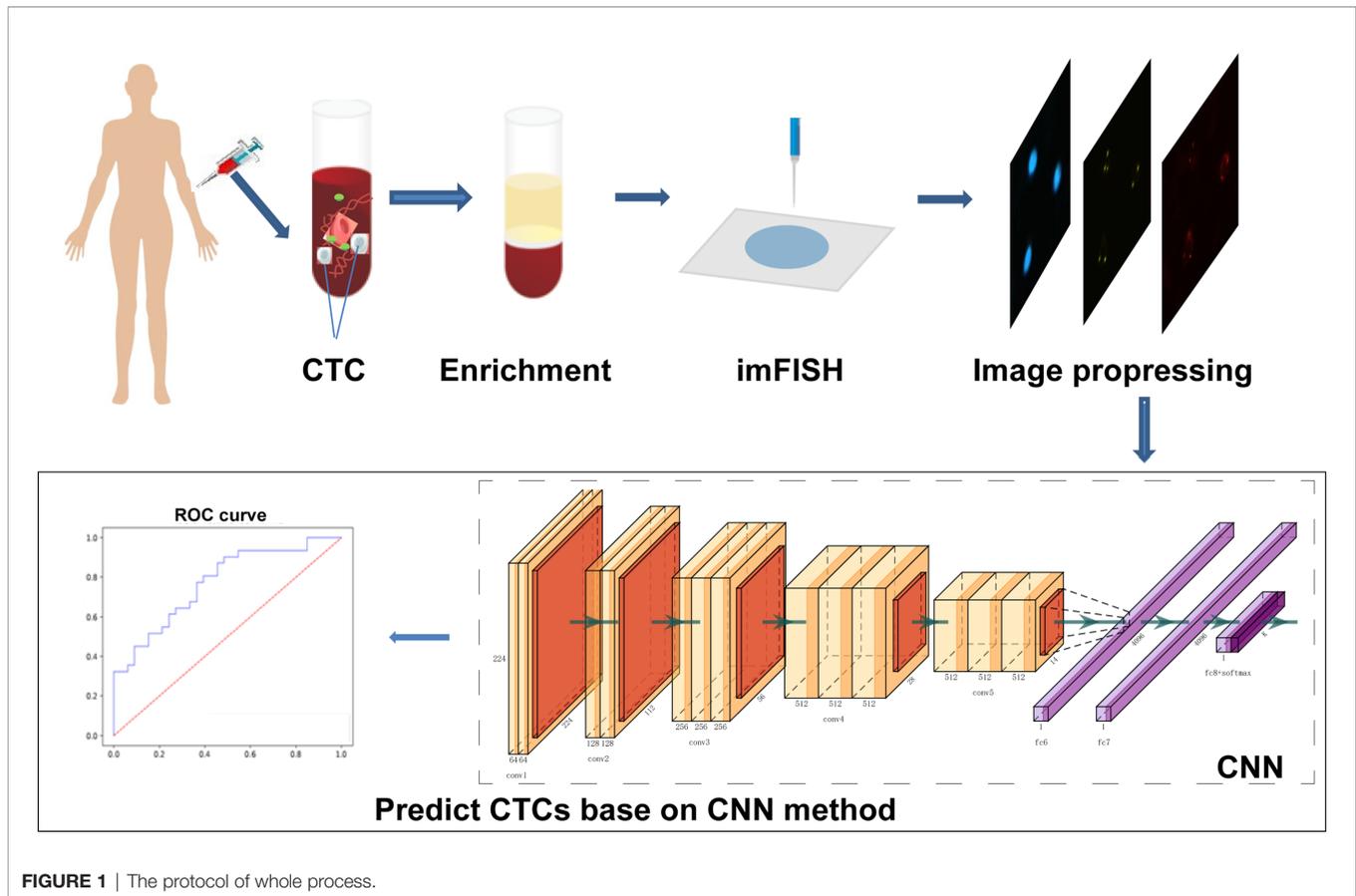
The complete process of identifying CTCs based on CNN was shown in **Figure 1**. Specifically, the peripheral blood of 776 cancer patients in Chifeng Municipal Hospital was collected firstly. Then the blood samples were processed by the Cytel method. Based on the principle of immunology and with the help of magnetic particle technology, CTCs were enriched by gradually removing the components of plasma, red blood cells and white blood cells, and CTCs were processed by imFISH (26). Finally, after preprocessing the images, stratified sampling was used to divide the data, 80% of the images were used for training, and a deep learning algorithm based on CNN was used to train the model. The remaining 20% were used as separate test set. The prediction performance of the model was evaluated with the results of a 5-fold cross validation (CV).

Samples Preparation

We conducted a retrospective study using plasma samples from the Chifeng Municipal Hospital. A total of 776 patients were enrolled from 2017 to 2019. Cancer types include lung cancer, liver cancer, gastrointestinal cancer, breast cancer, carcinoma of thyroid, NPC and others. After puncture for each patient, discarding the first 2ml blood sample to avoid skin epithelial cell pollution, then routinely collect 4ml peripheral venous blood, these samples were placed in a blood collection vessel containing acid citrate dextrose (ACD), gently inverted and mixed for 8 times before stored at room temperature, and CTCs were enriched within 24 hours after collection. The study was approved by the Ethics Committee of Chifeng Municipal Hospital.

Enrichment of CTCs

The detection method selected in this study was Cytel (49). The collected blood was taken out and put into a centrifuge tube for centrifugation experiment. After centrifugation at 776 g for 5 minutes, the supernatant was discarded to retain the precipitation, washed the precipitate with CS1 buffer (Cytel Biosciences Co., Ltd., Beijing, China), and then the red blood cells were fully dissolved with CS2 buffer. Added anti-CD45⁺ monoclonal antibody binding beads and the mixture was shaken evenly for 20 minutes to fully bind with leukocytes. Another 3 ml of separation medium was added to the centrifuge tube and centrifuged at a gradient of 300 g for 5 minutes. Then the upper rare cell layer was then centrifuged at a gradient of 776 g for 5 minutes, resuspended with CS1 buffer, and



the test tube was placed on a magnetic scaffold for 2 minutes. ImFISH identification was performed within 24 hours after coating, fixing, and drying.

imFISH Identification of CTCs

The samples were fixed with a fixative, dehydrated and dried at room temperature. The slide was coated with 10 μ l CEP8⁺ antibody, sealed and hybridized at 37°C for 1.5 h. After hybridization, removed the cover slide and eluted the probe for 15 min. The slides were washed twice in 2 \times SSC. Then, the prepared CD45⁺ fluorescent antibody was added to the sample area, and the slides were placed in a humid box and incubated in an oven at 33°C for 1 hour. After incubation, fluorescent antibody against CD45⁺ was absorbed and 10 μ l DAPI⁺ was added to the specimen area. Then, CTCs were observed and counted under a fluorescence microscope.

Detection Standard of CTCs

Each cell was divided into three different color channels: blue, orange and red. Among them, the nucleus was shown blue in DAPI⁺ (Figure 2A), and the centromere was shown in orange by CEP8⁺ (Figure 2B), and the white blood cells were stained by CD45⁺ immunofluorescence (Figure 2C). The interpretation criteria of CTCs count are: (1) eliminate the aggregation, superposition and interference of nuclei or impurities,

(2) positive for DAPI⁺, (3) negative for CD45⁺, (4) CEP8⁺ signal points >2. That is, cells are regarded as CTCs if they are CD45⁻/DAPI⁺/CEP8 \geq 3 (50, 51).

Image Preprocessing

The Python package *openCV* was used to handle CTCs images (52), including color and shape conversions. To be specific, the DAPI⁺ channel was first transformed into gray scale, and then the Gaussian filter was used to denoise. After extracting the gradient of the image, the regions with a high horizontal gradient and low vertical gradient were left, and the Gaussian filter was used to denoise. Then, the blurred image was binarization, that is, each pixel was replaced by the average value of the surrounding pixels in order to smooth and replace those regions with obvious intensity changes. Due to the lack of details in the contour of the obtained image, it may interfere with the subsequent contour detection, so it is necessary to be expanded and perform four morphological corrosion and expansion respectively. After the contour of the nuclear region was found, the minimum matrix coordinates of the contour were obtained, and the coordinates were mapped to the CEP8⁺ channel and segmented.

Computational Identification of CTC

With the development of artificial intelligence, deep learning has been widely used in medical image processing. CNN is one of the

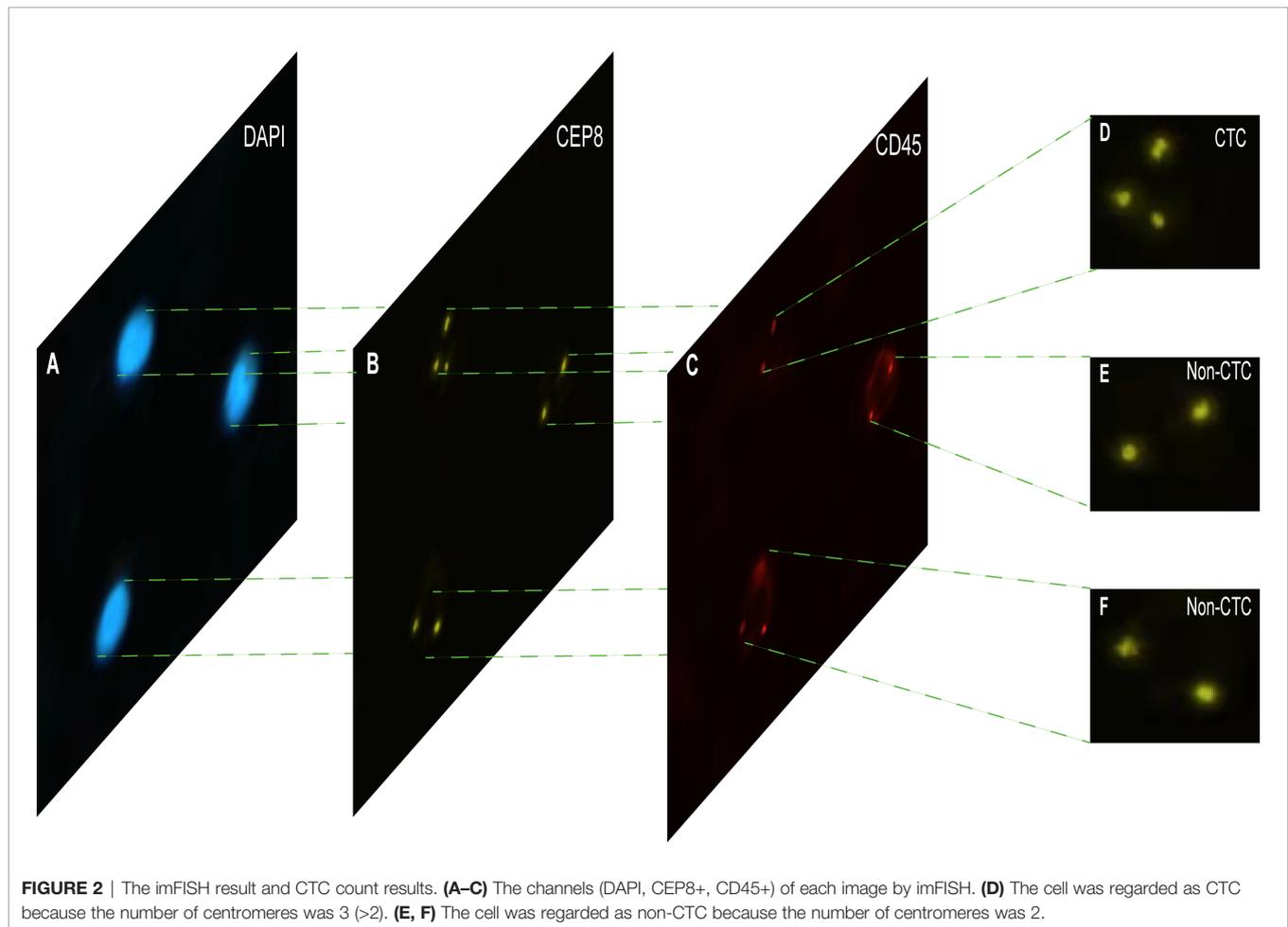


FIGURE 2 | The imFISH result and CTC count results. **(A–C)** The channels (DAPI, CEP8+, CD45+) of each image by imFISH. **(D)** The cell was regarded as CTC because the number of centromeres was 3 (>2). **(E, F)** The cell was regarded as non-CTC because the number of centromeres was 2.

representative algorithms of deep learning, which allows higher-level feature extraction and higher-level data prediction. After images were preprocessed, stratified sampling was used to divide the data, the model was trained by 5-fold CV, and the down sampling method was used for the training set to ensure the balance of positive samples and negative samples. CTCs in a single nucleus were identified through CNN. CNN includes input layer, hidden layer and output layer, the hidden layer includes layer1, layer2 and layer3, and each layer also includes convolution layer, excitation layer and pooling layer. After the images were fed into the input layer, it first enters the first intermediate hidden layer with convolution layer is composed of 32 5x5 convolution cores, and then fed to the pool layer through the ReLU excitation layer for dimensionality reduction. After dimensionality reduction, data was output from the first hidden layer to complete the feature extraction process. Then, all features are extracted through layer2 and layer3 hidden layers in turn. Finally, it enters the output layer and outputs the result of whether it is CTC or not. The CNN in this study involved VGG16, VGG19 (53), ResNet18, ResNet50 (54) and AlexNet (55).

These pre-training models have consumed huge time resources and computing resources when developing neural networks

usually. In recent years, transfer learning has become a new learning framework to solve this problem (56, 57). CNN model is pre-trained using a large number of images, and the trained model is distributed by its inventors for adoption. Transfer learning relies on the pre-trained CNN model to realize the knowledge transfer of different but related tasks, that is, using the existing knowledge learned from the completed tasks to help complete the new tasks. If the knowledge transfer is successful, it will greatly improve the learning efficiency by avoiding expensive data labeling. Transfer learning is defined as follows: a given domain D consists of feature space X and edge probability distribution $P(X)$, a label space y and a prediction function f consist a task T . D_S and D_T represent the source domain and the target domain, respectively, may have different feature spaces or different edge probability distributions, that is, $X_S \neq X_T$ or $P_S(X) \neq P_T(X)$, in addition, task T_S and T_T are subject to different label spaces (58).

Statistical Analysis

Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the model to identify CTC, and the area under the curve (AUC) at the 0.5 cut-off point was used to

measure the prediction accuracy. At the same time, the confusion matrix was used to observe the specificity and sensitivity. All of the analyses were performed using python version 3.6.9 and “sklearn” package version 0.23.2

RESULTS

Patient Characteristics

From Jan. 2017 to Jun. 2019, a total of 776 patients from Chifeng Municipal Hospital were included in this study. All the sample types were peripheral blood, and their ages ranged from 11 to 90 years old, with an average age of 65 years. Among the known cancer types, lung cancer was the most common (20.7%), followed by breast cancer and gastrointestinal cancer, and thyroid cancer was the least with only 2 cases. The clinical characteristic data of the enrolled patients were shown in **Table 1**.

Thousands of CTC and Non-CTC Images Were Segmented by openCV

The imFISH was performed on the samples from 776 patients. In order to avoid the influence of human factors, Python package *OpenCV* was used to process cell images. After the nuclear region

contour was found in the blue channel (DAPI⁺), the minimum matrix coordinates of the contour were obtained, mapped to the orange channel (CEP8⁺) and segmented, and the number of centromeres was observed. If there were more than 2 centromeres, the cell was considered CTC (**Figure 2D**). Otherwise it was non-CTC (**Figures 2E, F**). Finally, we obtained 14166 images, including 694 CTC images and 13472 non-CTC images. The details of data were shown in **Table 2**, in original train set, the number of CTC and non-CTC were 555 and 10777, respectively, ensuring balanced positive and negative samples, we performed down sampling method, at last, the number of CTC and non-CTC after down sampling were 555, respectively.

The Computational Method Performed Well in Identifying CTC

The CNN method was used to identify the segmented cell images. The whole process was shown in **Figure 3**. Firstly, the hierarchical sampling method was adopted for all images, 80% of the data were used for training and 20% of the data were used for testing. The CNN-based methods were used to train the model, including VGG16, VGG19, ResNet18, ResNet50, and AlexNet. Specifically, the traditional CNN model and transfer learning model were respectively trained on the training set based on 5-fold CV. The transfer learning model relied on the pre-trained CNN model to realize the task of CTC recognition. The trained traditional CNN model and the transfer learning model were used to test sets and output the final results. The results of 5-fold CV based on the trainset were shown in **Figure 4A**, the best performance was based on ResNet18 with AUC was about 0.98, in which 90.26% of non-CTCs were successfully identified, 9.74% were incorrectly identified as CTCs, while only 5.41% of CTCs are misclassified (**Figure 4B**), the sensitivity and specificity were 95.3% and 91.7% respectively. After training the model, we used ResNet18 with the best performance on the test data set, its ROC curve was shown in **Figure 4C** with AUC was 0.988, and the confusion matrix also shown that ResNet18 performed well (**Figure 4D**). In addition, in order to improve the prediction performance and save computing resources, transfer learning was also used to train the model. The results of 5-fold CV based on transfer learning in train data set was shown in **Figure 5A**, the results of confusion matrix showed that the sensitivity and specificity of transfer learning were 97.2% and 94.0% respectively (**Figure 5B**). After training the

TABLE 1 | Summary of the general clinical information of patients.

Characteristics	No. (%) of Participants
Age	
0-39	30 (3.9)
40-69	313 (40.3)
>70	103 (13.3)
Unknown	330 (42.5)
Gender	
Male	248 (42.0)
Female	199 (25.6)
Unknown	329 (42.4)
CTC number	9.9(0-318)
Cancer type	
Lung cancer	161 (20.7)
Liver cancer	18 (2.3)
Gastrointestinal cancer	107 (13.8)
Breast cancer	91 (11.7)
Carcinoma of thyroid	2 (0.3)
NPC	30 (3.9)
Other	367 (47.3)

TABLE 2 | The number of images.

	Original			Down sampling	
	Train set	Test set	Total	Train set	Test set
No. of CTC	555	139	694	555	139
No. of Non-CTC	10777	2695	13472	555	2695
Total	11332	2834	14166	1110	2834

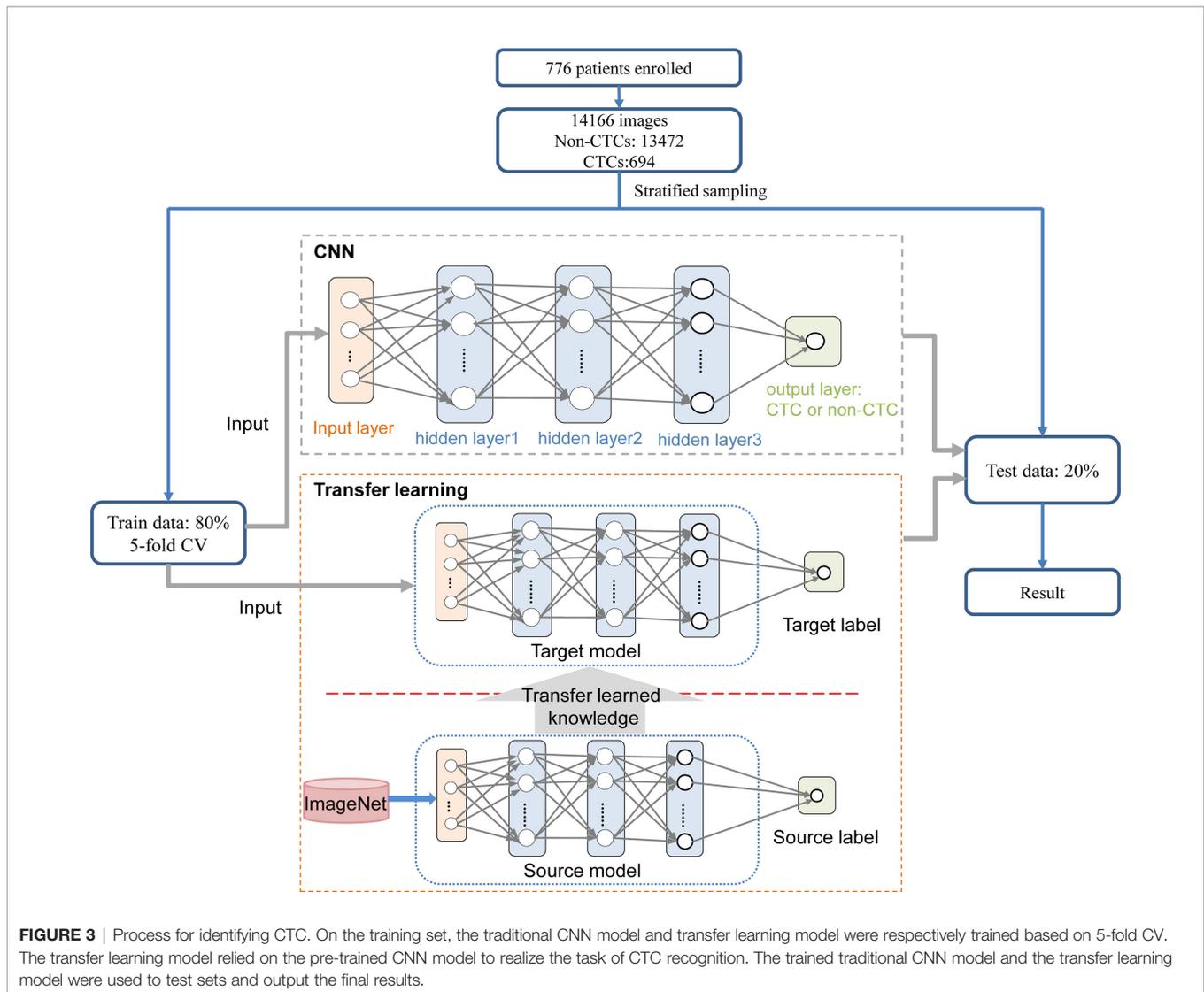


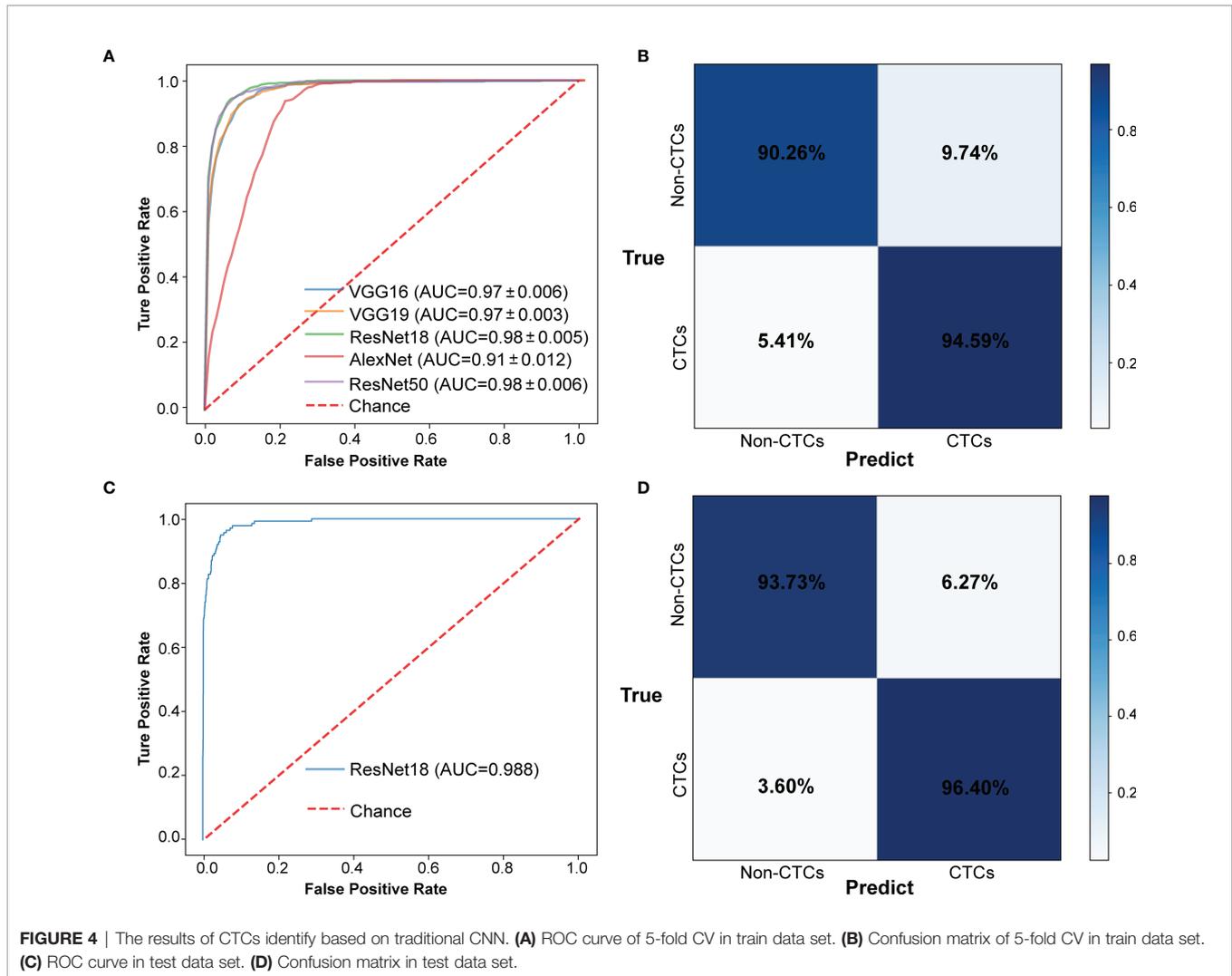
FIGURE 3 | Process for identifying CTC. On the training set, the traditional CNN model and transfer learning model were respectively trained based on 5-fold CV. The transfer learning model relied on the pre-trained CNN model to realize the task of CTC recognition. The trained traditional CNN model and the transfer learning model were used to test sets and output the final results.

model with transfer learning, we used VGG16 with the best performance on the test data set, its AUC was 0.988 (**Figure 5C**), and the sensitivity and specificity were 98.6% and 96.5% respectively (**Figure 5D**).

The experimental results showed that the deep learning method based on CNN can accurately identify CTC and provide a powerful reference for the prognosis of patients. In addition, we also summarized some samples that were discriminated incorrectly, such as samples that were originally non-CTC but were predicted to be CTC (**Figure 6A**), and samples that predicted CTC to be non-CTC (**Figure 6B**). The reason for the misjudgment first considers the artificial or instrumental noise in the process of negative enrichment techniques. Secondly, the exposure during the photographing process after imFISH processing resulted in us not getting the original film data. The third is that the centromere was not completely located in the nucleus due to the platform.

DISCUSSION

More and more evidence showed that CTCs are closely related to the prognosis of patients with advanced cancer. It has an important reference value for determining the clinical results and recurrence risk. In recent years, blood testing has been widely used to monitor the CTC response of patients with malignant tumors and evaluate the prognosis and recurrence risk of patients since its high safety and low cost. It reduces errors caused by manually setting interpretation standards and save time and labor costs. The importance of CTC rapid automatic recognition is increasing, and the research of the automatic recognition process of CTCs is also accelerating. Deep learning has been proved to be suitable for detecting CTC due to its high sensitivity and specificity in CTC counting. In addition, image interpretation using machine learning can capture important image features.



In this study, we developed a CTCs recognition method based on deep learning. After collecting the blood samples from Chifeng Municipal Hospital, we conducted CTCs enrichment and imFISH experiments on the samples, and screened the fluorescent images according to the image quality. A total of 14166 images were used for downstream analysis, including 694 CTC images and 13472 non-CTC images. 80% of the images were used for training models and 20% for test. In order to reduce the error caused by manual intervention, we used machines instead of manual screening. Firstly, images were segmented by using the Python package *openCV*, and the coordinate information of the nucleus was obtained after image preprocessing. Then, we used CNN models such as VGG16, VGG19, ResNet18, ResNet50 and AlexNet to identify CTCs. The results of 5-fold CV showed that their AUC reached 0.98, and the sensitivity and specificity were 95.3% and 91.7%, respectively. In order to overcome the shortcomings of consuming a lot of time and computing resources when

developing neural networks, transfer learning was used to train the model. Finally, the AUC was improved to 0.99, and the recognition sensitivity and specificity also reached to 97.2% and 94.0% based on transfer learning.

The method of transfer learning was proposed, which can carry out image interpretation, capture important image features, reduce the errors caused by subjective factors in manual interpretation, and save computing time and computing resources. In the process of 5-fold CV, the down-sampling method was used to overcome the serious imbalance between positive samples and negative samples, and the 5-fold CV results of transfer learning shown higher sensitivity and specificity. Nevertheless, this study still has some limitations. The CTC images contained in the enrollment data do not cover the whole film, but focus on a CTC positive area under the microscope. Due to quality issues, some images in the enrollment data are abandoned. How to expand the image scope is the focus of attention in the future.

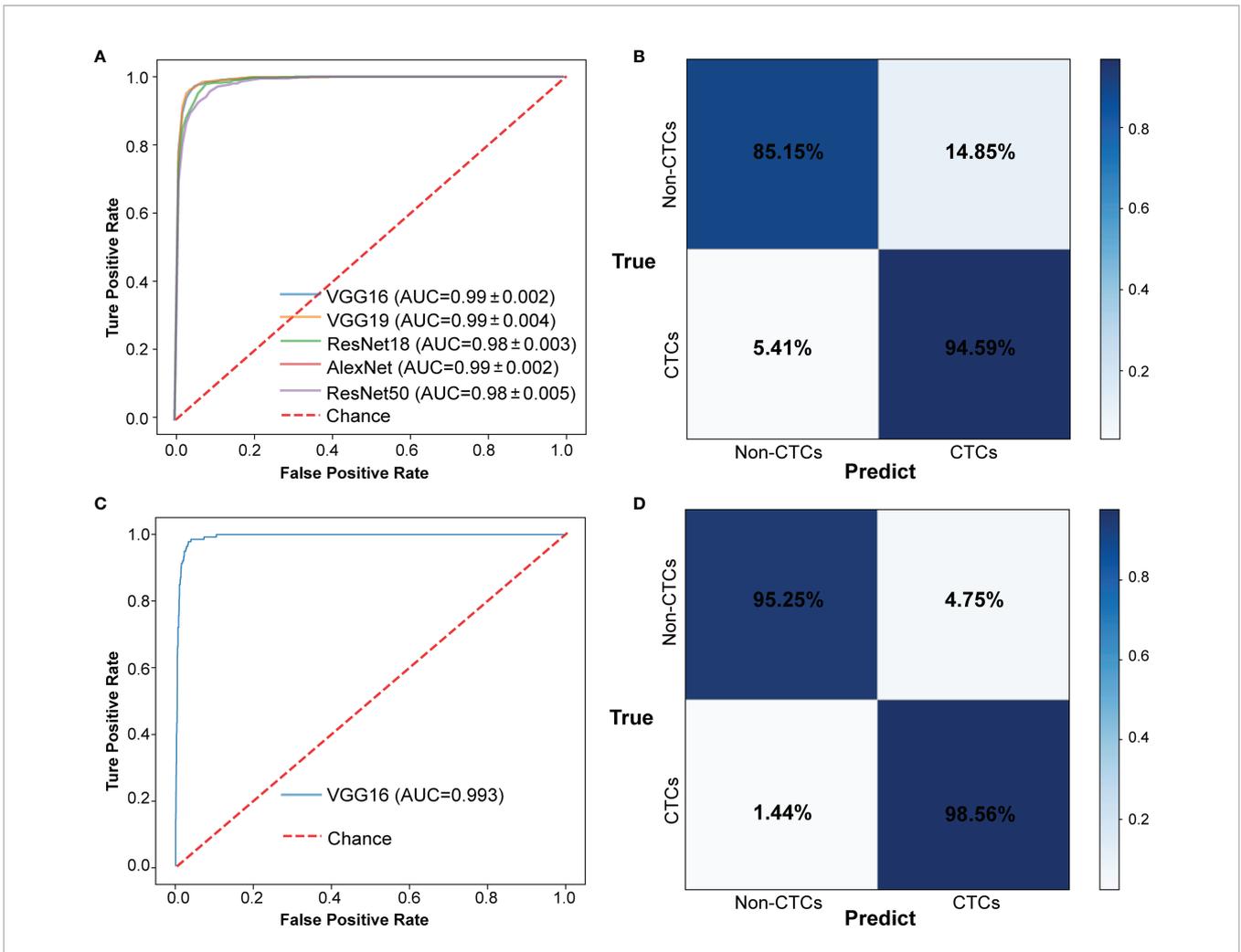


FIGURE 5 | The results of CTCs identify based on transfer learning. **(A)** ROC curve of 5-fold CV in train data set. **(B)** Confusion matrix of 5-fold CV in train data set. **(C)** ROC curve in test data set. **(D)** Confusion matrix in test data set.

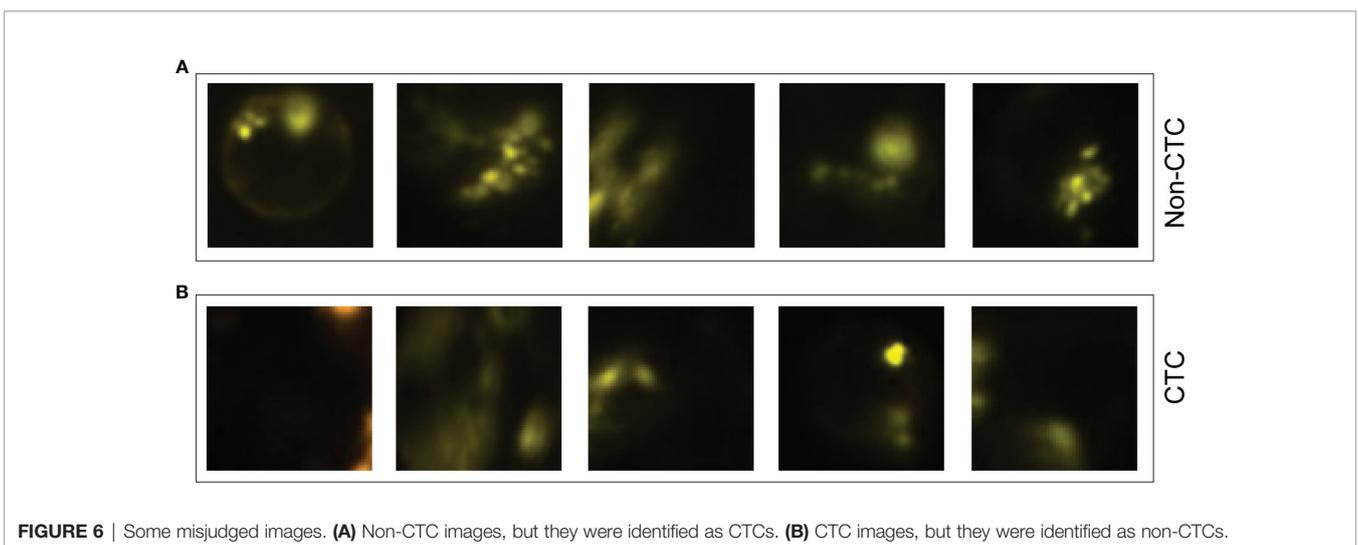


FIGURE 6 | Some misjudged images. **(A)** Non-CTC images, but they were identified as CTCs. **(B)** CTC images, but they were identified as non-CTCs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/benstevan2/CTC_project.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Chifeng Municipal Hospital. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

- Masuda T, Hayashi N, Iguchi T, Ito S, Eguchi H, Mimori K. Clinical and Biological Significance of Circulating Tumor Cells in Cancer. *Mol Oncol* (2016) 10(3):408–17. doi: 10.1016/j.molonc.2016.01.010
- Arko D, Lim AR, Ghajar CM. Circulating and Disseminated Tumor Cells: Harbingers or Initiators of Metastasis? *Mol Oncol* (2017) 11:40–61. doi: 10.1002/1878-0261.12022
- Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
- He B, Dai C, Lang J, Bing P, Tian G, Wang B, et al. A Machine Learning Framework to trace Tumor Tissue-Of-Origin of 13 Types of Cancer Based on DNA Somatic Mutation. *Biochim Biophys Acta Mol Basis Dis* (2020) 1866:165916. doi: 10.1016/j.bbadis.2020.165916
- He B, Lang J, Wang B, Liu X, Lu Q, He J, et al. TOOMe: A Novel Computational Framework to Infer Cancer Tissue-of-Origin by Integrating Both Gene Mutation and Expression. *Front Bioeng Biotechnol* (2020) 8:394. doi: 10.3389/fbioe.2020.00394
- Qiao Y, Li J, Shi C, Wang W, Qu X, Ming X, et al. Prognostic Value of Circulating Tumor Cells in the Peripheral Blood of Patients With Esophageal Squamous Cell Carcinoma. *Oncotargets Ther* (2017) 10:1363–73. doi: 10.2147/OTT.S129004
- Giuliano M, Giordano A, Jackson S, Hess KR, Giorgi UD, Mego M, et al. Circulating Tumor Cells as Prognostic and Predictive Markers in Metastatic Breast Cancer Patients Receiving First-Line Systemic Treatment. *Breast Cancer Res* (2011) 13:1–9. doi: 10.1186/bcr2907
- Pierga JY, Hajage D, Bachelot T, Delaloue S, Brain E, Campone M, et al. High Independent Prognostic and Predictive Value of Circulating Tumor Cells Compared With Serum Tumor Markers in a Large Prospective Trial in First-Line Chemotherapy for Metastatic Breast Cancer Patients. *Ann Oncol* (2011) 23:618–24. doi: 10.1093/annonc/mdr263
- Stott SL, Lee RJ, Nagrath S, Yu M, Maheswaran S. Isolation and Characterization of Circulating Tumor Cells from Patients with Localized and Metastatic Prostate Cancer. *Sci Transl Med* (2010) 2:25ra23. doi: 10.1126/scitranslmed.3000403
- Maheswaran S, Sequist LV, Nagrath S, Ullus L, Brannigan B, Collura CV, et al. Detection of Mutations in EGFR in Circulating Lung-Cancer Cells. *N Engl J Med* (2008) 359:366–77. doi: 10.1056/NEJMoa0800668
- Lindsay CR, Faugeroux V, Michiels S, Pailler E, Facchinetti F, Ou D, et al. A Prospective Examination of Circulating Tumor Cell Profiles in Non-Small-Cell Lung Cancer Molecular Subgroups. *Ann Oncol* (2017) 28(7):1523–31. doi: 10.1093/annonc/mdx156
- Praharaj PP, Bhutia SK, Nagrath S, Bittling RL, Deep G. Circulating Tumor Cell-Derived Organoids: Current Challenges and Promises in Medical Research and Precision Medicine. *Biochim Biophys Acta (BBA) - Rev Cancer* (2018) 1869(2):117–27. doi: 10.1016/j.bbcan.2017.12.005
- Yan M, Wang L, Song Y, Zhang K. Monitoring Non-Operative Treatment for Advanced Esophageal Cancer by Dynamic Ctcs Count: A Case Report. *J Clin Case Rep* (2017) 9(4):10001227. doi: 10.4172/2165-7920.10001227

AUTHOR CONTRIBUTIONS

FK designed the study. ZG collected, analyzed and interpreted the data, and wrote the manuscript. XL and YH performed the experiment. JW and QZ reviewed and modified the manuscript. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the 2020 Chifeng Natural Science Research project (grant number SZR157).

- Huebner H, Fasching PA, Gumbrecht W, Jud S, Rauh C, Matzas M, et al. Filtration Based Assessment of CTCs and CellSearch based Assessment are Both Powerful Predictors of Prognosis for metastatic Breast Cancer Patients. *BMC Cancer* (2018) 18:204. doi: 10.1186/s12885-018-4115-1
- Yang J, Hui Y, Zhang Y, Zhang M, Ji B, Tian G, et al. Application of Circulating Tumor DNA as a Biomarker for Non-Small Cell Lung Cancer. *Front Oncol* (2021) 11:725938. doi: 10.3389/fonc.2021.725938
- Banyas P, Schneck H, Blassl C, Schultz S, Meier S, Niederacher D, et al. Prognostic Relevance of Circulating Tumor Cells in Molecular Subtypes of Breast Cancer. *Geburtshilfe Frauenheilkd* (2015) 75(3):232–7. doi: 10.1055/s-0035-1545788
- Jingsi D, Zhu D, Tang X, Qiu X, Bingjie D. Detection of Circulating Tumor Cell Molecular Subtype in Pulmonary Vein Predicting Prognosis of Stage I-III Non-small Cell Lung Cancer Patients. *Front Oncol* (2019) 9:1139–9. doi: 10.3389/fonc.2019.01139
- Cristofanilli M, Pierga JY, Reuben J, Rademaker A, Davis AA, Peeters DJ, et al. The Clinical Use of Circulating Tumor Cells (CTCs) Enumeration for Staging of Metastatic Breast Cancer (MBC): International Expert Consensus Paper. *Crit Rev Oncology/Hematol* (2019) 134:39–45. doi: 10.1016/j.critrevonc.2018.12.004
- Wan J, Massie C, Garcia-Corbacho J, Moulriere F, Brenton JD, Caldas C, et al. Liquid Biopsies Come of Age: Towards Implementation of Circulating Tumour DNA. *Nat Rev Cancer* (2017) 17:223. doi: 10.1038/nrc.2017.7
- Marcuello M, Vymetalkova V, Rui P, Duran-Sanchon S, Gironella M. Circulating Biomarkers For Early Detection and Clinical Management of Colorectal Cancer. *Mol Aspects Med* (2019) 69:107–22. doi: 10.1016/j.mam.2019.06.002
- Dbal A, Lc A, Mza D, Tmmab C, Esq A. Liquid Biopsy in Ovarian Cancer Using Circulating Tumor DNA and Cells: Ready for Prime Time? - ScienceDirect. *Cancer Lett* (2020) 468:59–71. doi: 10.1016/j.canlet.2019.10.014
- Racila E, Euhus D, Weiss AJ, Rao C, Mcconnell J, Terstappen L, et al. Detection and Characterization of Carcinoma Cells in the Blood. *Proc Natl Acad Sci* (1998) 95:4589–94. doi: 10.1073/pnas.95.8.4589
- Grover PK, Cummins AG, Price TJ, Roberts-Thomson IC, Hardingham JE. Circulating Tumour Cells: The Evolving Concept and the Inadequacy of their Enrichment by EpCAM-Based Methodology for Basic and Clinical Cancer Research. *Ann Oncol Off J Eur Soc Med Oncol* (2014) 25:1506. doi: 10.1093/annonc/mdu018
- Tellez GM, Rodriguez CL, Antoine C, Benjamin O, Dominique H. Circulating Tumor Cells: A Review of Non-EpCAM-Based Approaches for Cell Enrichment and Isolation. *Clin Chem* (2016) 62(4):571–81. doi: 10.1373/clinchem.2015.249706
- Keller L, Werner S, Pantel K. Biology and Clinical Relevance of EpCAM. *Other* (2019) 3:165–80. doi: 10.15698/cst2019.06.188
- Koudelakova V, Trojanec R, Vrbkova J, Donevska S, Bouchalova K, Kolar Z, et al. Frequency of Chromosome 17 Polysomy in Relation to CEP17 Copy Number in a Large Breast Cancer Cohort. *Genes Chromosomes Cancer* (2016) 55:409–17. doi: 10.1002/gcc.22337
- Xiangqi L, Zhang Z, Yijie B, Chao Z, Shan B. Circulating Tumor Cells Detection in Neuroblastoma Patients by EpCAM-Independent Enrichment and Immunostaining-Fluorescence In Situ Hybridization. *Ebiomedicine* (2018) 35:244–50. doi: 10.1016/j.ebiom.2018.08.005

28. Nagrath S, Sequist LV, Maheswaran S, Bell DW, Irimia D, Ulkus L, et al. Isolation of Rare Circulating Tumour Cells in Cancer Patients by Microchip Technology. *Nature* (2007) 450:1235–9. doi: 10.1038/nature06385
29. Yang C, Zhang N, Wang S, Shi D, Zhang C, Kan L, et al. Wedge-Shaped Microfluidic Chip for Circulating Tumor Cells Isolation and Its Clinical Significance in Gastric Cancer. *J Trans Med* (2018) 16:1–12. doi: 10.1186/s12967-018-1521-8
30. Lighthart ST, Coumans F, Attard G, Cassidy AM, Bono J, Terstappen L. Unbiased and Automated Identification of a Circulating Tumour Cell Definition That Associates with Overall Survival. *PLoS One* (2011) 6:e27419. doi: 10.1371/journal.pone.0027419
31. Kraeft S, K. Reliable and Sensitive Identification of Occult Tumor Cells Using the Improved Rare Event Imaging System. *Clin Cancer Res* (2004) 10:3020–8. doi: 10.1158/1078-0432.CCR-03-0361
32. Zhou M, Hui Z, Wang Z, Li R, Liu X, Zhang W, et al. Precisely Enumerating Circulating Tumor Cells Utilizing a Multi-Functional Microfluidic Chip and Unique Image Interpretation Algorithm. *Theranostics* (2017) 7:4710–21. doi: 10.7150/thno.20440
33. Liu C, Wei D, Xiang J, Ren F, Huang L, Lang J, et al. An Improved Anticancer Drug-Response Prediction Based on an Ensemble Method Integrating Matrix Completion and Ridge Regression. *Mol Ther Nucleic Acids* (2020) 21:676–86. doi: 10.1016/j.omtn.2020.07.003
34. Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator Regularized Non-Negative Matrix Factorization Method-Based Drug Repurposing for COVID-19. *Front Immunol* (2020) 11:603615. doi: 10.3389/fimmu.2020.603615
35. Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of Aging and Age-Related Diseases. *Geroscience* (2020) 42:353–72. doi: 10.1007/s11357-019-00106-x
36. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics* (2017) 37:160130. doi: 10.1148/rg.2017160130
37. Ma X, Xi B, Zhang Y, Zhu L, Yang J. A Machine Learning-based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr Bioinf* (2019) 14:349–58. doi: 10.2174/1574893614666191017091959
38. Lundervold AS, Lundervold A. An Overview of Deep Learning in Medical Imaging Focusing on MRI. *Z für Medizinische Physik* (2018) 29:102–27. doi: 10.1016/j.zemedi.2018.11.002
39. Mcbee MP, Awan OA, Colucci AT, Ghobadi CW, Auffermann WF. Deep Learning in Radiology. *Acad Radiol* (2018) 25:1472–80. doi: 10.1016/j.acra.2018.02.018
40. Maier A, Syben C, Lasser T, Riess C. A gentle Introduction to Deep Learning in Medical Image Processing. *Z für Medizinische Physik* (2019) 29:86–101. doi: 10.1016/j.zemedi.2018.12.003
41. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging* (2016) 35:1207–16. doi: 10.1109/TMI.2016.2535865
42. Poplin R, Varadarajan AV, Blumer K, Liu Y, Mcconnell MV, Corrado GS, et al. Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning. *Nat Biomed Eng* (2017) 2(3):158–64. doi: 10.1038/s41551-018-0195-0
43. Le N, Quang-Thai H, Ou YY. Classifying the Molecular Functions of Rab GTPases in Membrane Trafficking Using Deep Convolutional Neural Networks. *Anal Biochem* (2018) 555:33–41. doi: 10.1016/j.ab.2018.06.011
44. Dominguez C, Heras J, Pascual V. IJ-OpenCV: Combining ImageJ and OpenCV for Processing Images In Biomedicine. *Comput Biol Med* (2017) 84:189–94. doi: 10.1016/j.combiomed.2017.03.027
45. Le N, Huynh TT, Yapp E, Yeh HY. Identification of Clathrin Proteins by Incorporating Hyperparameter Optimization in Deep Learning and PSSM Profiles. *Comput Methods Programs Biomed* (2019) 177:81–8. doi: 10.1016/j.cmpb.2019.05.016
46. Xing F, Xie Y, Yang L. An Automatic Learning-Based Framework for Robust Nucleus Segmentation. *IEEE Trans Med Imaging* (2016) 35:550–66. doi: 10.1109/TMI.2015.2481436
47. Copie-Bergman C, Gaulard P, Leroy K, Briere J, Baia M, Jais JP, et al. Immuno-Fluorescence *in situ* Hybridization Index Predicts Survival in Patients With Diffuse Large B-Cell Lymphoma Treated With R-CHOP: a GELA Study. *J Clin Oncol* (2009) 27:5573–9. doi: 10.1200/JCO.2009.22.7058
48. Tamminga M, Groen HH, Hiltermann TJ. Investigating CTCs in NSCLC—a Reaction to the Study of Jia-Wei Wan: A Preliminary Study on the Relationship Between Circulating Tumor Cells Count and Clinical Features in Patients with Non-Small Cell Lung Cancer. *J Thorac Dis* (2016) 8(6):1032–6. doi: 10.21037/jtd.2016.04.17
49. Yu H, Ma L, Zhu Y, Li W, Ding L, Gao H. Significant Diagnostic Value of Circulating Tumor Cells in Colorectal Cancer. *Oncol Lett* (2020) 20:317–25. doi: 10.3892/ol.2020.11537
50. Li J, Wang Z, Chong T, Chen H, Li H, Li G, et al. Over-Expression of a Poor Prognostic Marker in Prostate Cancer: AQP5 Promotes Cells Growth and Local Invasion. *World J Surg Oncol* (2014) 12:284. doi: 10.1186/1477-7819-12-284
51. Qian Y, Wu Y, Yuan Z, Niu X, He Y, Peng J, et al. The Frequency of Circulating Tumour Cells and the Correlation with the Clinical Response to Standard Chemoradiotherapy in Locally Advanced Nasopharyngeal Carcinoma: A Prospective Study. *Cancer Manag Res* (2019) 11:10187–93. doi: 10.2147/CMAR.S222916
52. Bradski G. The OpenCV Library. *Dr Dobb's J: Software Tools Prof Programmer* (2000) 25:120–6. doi: 10.1002/1096-9128(200011)12:13<1317::AID-CPE536>3.0.CO;2-K
53. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Comput Sci* (2014) 1409:1556.
54. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vision* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
55. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification With Deep Convolutional Neural Networks. *Commun ACM* (2017) 60:84–90. doi: 10.1145/3065386
56. Pan SJ, Qiang Y. A Survey on Transfer Learning. *IEEE Trans Knowledge Data Eng* (2010) 22:1345–59. doi: 10.1109/TKDE.2009.191
57. Vilalta R, Giraud-Carrier C, Brazdil P, Soares C. *Inductive Transfer*. Springer US (2017). doi: 10.1007/978-1-4899-7687-1
58. Weiss K, Khoshgoftaar TM, Wang DD. A Survey of Transfer Learning. *J Big Data* (2016) 3:1–40. doi: 10.1186/s40537-016-0043-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Guo, Lin, Hui, Wang, Zhang and Kong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.