Multi-modality and Multi-view 2D CNN to Predict Locoregional Recurrence in Head & Neck Cancer

Jinkun Guo
Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence
Xidian University
Xi'an 710071, China
guojinkun@foxmail.com

Rongfang Wang
Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence
Xidian University
Xi'an 710071, China
rfwang@xidian.edu.cn

Zhiguo Zhou
School of Computer Science and Mathematics
University of Central Missouri
Warrensburg, MO 64093, United States of America
zzhou@ucmo.edu

Kai Wang
Advanced Imaging and Informatics for Radiation Therapy (AIRT) Laboratory
University of Texas Southwestern Medical Center
Dallas, TX 75390, United States of America
kai.wang@utsouthwestern.edu

Rongbin Xu
Engineering Research Center of Big Data Application in Private Health Medicine, School of Information Engineering
Putian Unviersity
Putian 351100, China
xurongbing@ptu.edu.cn

Jing Wang*
Advanced Imaging and Informatics for Radiation Therapy (AIRT) Laboratory
University of Texas Southwestern Medical Center
Dallas, TX 75390, United States of America
jing.wang@utsouthwestern.edu

2021 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-3900-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/IJCNN52387.2021.9533703

***Abstract*—Locoregional recurrence (LRR) remains one of leading causes in head and neck (H&N) cancer treatment failure despite the advancement of multidisciplinary management. Accurately predicting LRR in early stage can help physicians make an optimal personalized treatment strategy. In this study, we propose an end-to-end multi-modality and multi-view convolutional neural network model (mMmV-CNN) for LRR prediction in H&N cancer. In mMmV, a dimension reduction operator is designed, projecting the 3D volume onto 2D images in different directions, and a multi-view strategy is used to replace the original 3D method, which reduces the complexity of the algorithm while preserving important 3D information. Meanwhile, multi-modal data is used for the classification by making full use of the complementary information from cross modality data. Furthermore, we design a multi-modality deep neural network which is trained in an end-to-end manner and jointly optimize the deep features of CT, PET and clinical features. A H&N dataset which consists of 206 patients was used to evaluate the performance. Experimental results demonstrated that mMmV-CNN can obtain an AUC value of 0.81 and outperform a state of the art CNN-based method.**

Keywords—head and neck cancer, multi-view convolutional neural network, local recurrence, outcome prediction

I. INTRODUCTION

Head and neck (H&N) cancer was the seventh most common cancer worldwide in 2018 [1]. Radiation therapy plays an essential role in H&N cancer management [2]. However, even after the therapy with curative intent, 15% to 50% of H&N cancer patients still experience locoregional recurrence (LRR), mostly within 3 years after treatment [3]-[4]. If the prognosis of patients can be predicted based on certain indicators in the early stage, an optimal personalized treatment plan can be developed. For example, patients at high-risk for LRR could be assigned to a more aggressive treatment regimen, potentially improving the treatment outcome. Similarly, low-risk patients may receive more conservative treatment, leading to reduced side effects, improved quality of life and reduced medical cost [5].

With the development of artificial intelligence and machine learning, many prediction models have been proposed for cancer patient management. Wang et al. segmented lesions and extracted features on CT images, and used the random forest (RF) algorithm to build a radiological model for predicting lymph node metastasis in gastric cancer [6]. Shanthi et al. proposed a new feature selection algorithm based on wrapper by using an improved stochastic diffusion search (SDS) algorithm, and use Naive Bayes and decision trees to classify lung cancer [7]. K. Wang et al. established a multi-classifier, multi-objective, multi-modality model, using clinical features and radiomics features extracted from CT and PET images to predict the outcome of LRR in patients with H&N cancer [4]. However, all the above methods only based on traditional machine learning classifiers. In recent years, deep learning as a powerful data analysis tool has been widely used in medical prognosis prediction [8]. In deep learning-based methods, some are based on 2D CNNs by processing each 2D slice independently [9]-[11], which is considered as a non-optimal use of the volumetric medical image data. Meanwhile, 3D CNNs uses a substantially increased number of parameters, requiring significant amount of memory and computational resources, as well as training data.

This work was supported by Engineering Research Center of Big Data Application in Private Health Medicine, Fujian Province University (No. KF2020005).

978-0-7381-3366-9/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

Recently, different standard-of-care images, such as computed tomography (CT) images and positron emission tomography (PET) images have been explored for treatment outcome prediction. T. Lo et al. proposed a support vector machine (SVM)-based method to predict the distant metastasis (DM) of malignant tumors from the CT images of patients [12]. Sekaran et al. proposed a deep learning method based on CT scans to predict the percentage of cancer spread in the pancreas [13]. EI naqa et al. extract features from PET images to predict the patient's treatment outcome for cervical and H&N cancer [14]. However, these methods only use single modal data, which does not take advantage of complimentary information extracted from multiple modalities [15]. In treatment outcome prediction, data from different modalities can provide complementary information [9], [16]. For example, the CT images can provide anatomical structure and attenuation coefficient of patients, while the PET can show metabolism information about the lesion. In addition, clinical data such as age, primary site, and T-N stage can provide patient-specific characteristics, which can further improve the performance of the prediction model [5], [17].

As such, we proposed an end-to-end multi-modality and multi-view 2D convolutional neural network method (mMmV-CNN) for LRR prediction of H&N cancer. In mMmV-CNN, we design a dimension reduction operator by rotating the 3D CT and PET data along the vertical axis and then calculate the average projection along horizontal axis to obtain 2D image in different directions to make full use of the information of 3D data [18]. The multi-view strategy [19] is then adopted to improve the performance of 2D CNN for extracting image feature by aggregating features extracted from different views. Through the above operations, while greatly reducing the number of network parameters, the spatial 3D context information can be effectively used by the model. Furthermore, we design a multi-modality deep neural network which can be trained in an end-to-end manner and jointly optimize the deep features of CT, PET and clinical parameters. The experimental results demonstrated that proposed mMmV-CNN can obtain accurate prediction results and outperform a 2D CNN method and state of the art 3D method under the same structure.

II. Materials and Pre-processing

A. Materials

The dataset in this study includes the image and clinical parameters of patients with H&N cancer received radiation treatment from September 2005 to November 2015 at the University of Texas Southwestern Medical Center (UTSW) (Dallas, TX, USA). The median follow-up duration is 37 months. Clinical tumor volumes (CTVs) were manually contoured on CT in Velocity AI (Varian Medical systems, Palo Alto, CA, USA) by a radiation oncologist under the guidance of PET. Among these patients, 57 experienced LRR. Clinical information such as age, gender, tumor T-stage, tumor N-stage and disease site statues were also collected from patient chart to build the model. After excluding patients with less than one year follow-up, the filtered UTSW data set we finally used contains 206 patients, including 49 with LRR and 157 without LRR.

B. Pre-processing

Clinical data: The used clinical features are 1) age, 2) primary sites (Larynx, Nasopharynx, Oropharynx, Hypopharynx), 3) T-stages (T1-T4), 4) N-stages (N0-N3), 5) HPV status (+, -, N/A) and 6) therapy (radiation, chemo radiation). The K-NearestNeighbor (KNN) algorithm [20] is used to fill in the missing values, where K is set as 3. The sample is randomly divided into 5 folds, and each time four folds are taken as the training set and one fold is used as the testing set. In each cross-validation, the missing values in the training set are determined by using the remaining training set samples, while the test set is the same. The clinical features can be divided into three types according to the data type: 1) *Numerical feature*, such as age. This kind of data can directly use its value and encode it as a one-bit feature. 2) *Nominal feature*, such as primary sites, HPV status and therapy. This type of feature is represented by one hot encoding, which can make the value of the non-partial order variable not reflect the order relationship, and to a certain extent also plays the role of expanding the feature. According to the number of categories, they are coded into 4, 3, and 2 bit feature values respectively. 3) *Ordinal feature*, such as T-Stage and N-Stage. For the purpose of reflecting the order information, label encoding is used, and 4 consecutive numbers starting from 1 are used to represent the different values of the two features. We divided the age feature value by 100, and divided the T-stage and N-stage feature values by 4, then normalize to the interval [0, 1]. All codes are spliced together, and finally a total of 12 clinical feature codes are generated. A coding example of one patient is shown in Table 1.

TABLE I.

EXAMPLE OF THE CLINICAL FEATURE CODING.

	Age	**Primary Site**	**T-Stage**	**N-Stage**	**HPV Status**	**Therapy**
Value	82	Oropharynx	T2	N2	N/A	chemo radiation
Coding	0.82	0 0 1 0	0.50	0.75	0 0 1	0 1

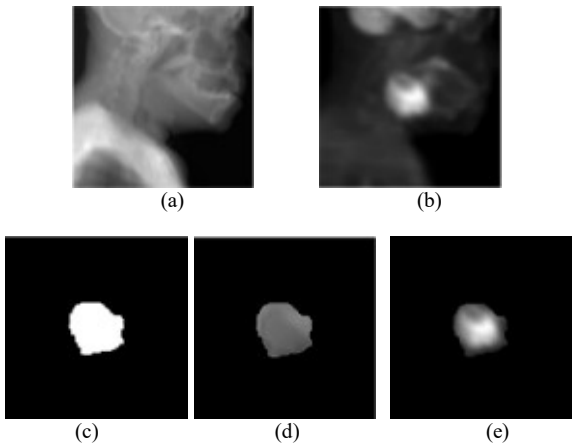

Fig. 1. (a) average slice of CT; (b) average slice of PET; (c) average slice of contour; (d) tumor sample of CT; (e) tumor sample of PET.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

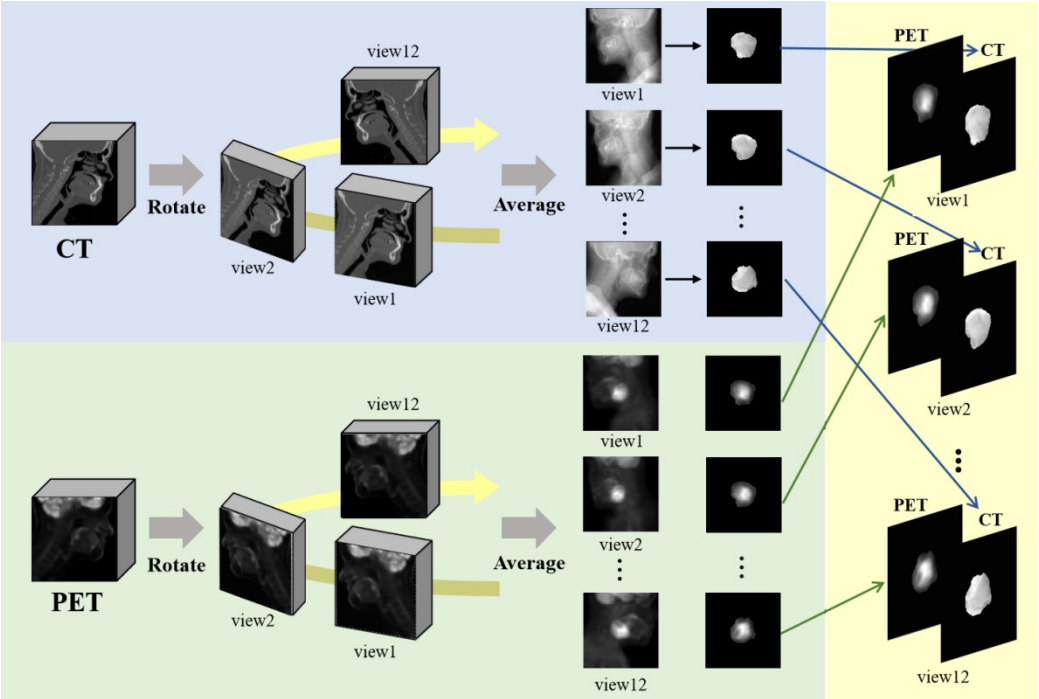

Fig. 2. The construction of multi-modality and multi-view samples.

Image data: We resample CT and PET data to the same resolution $1 \times 1 \times 1mm^3$ according to the standard process of CT-PET data analysis [21]-[23].

Then, we convert the pixel value v_p of the CT and PET image to CT value and standard uptake value (SUV) value, respectively. The calculation equation of CT value is (1):

$$CT = v_p \times rescale_{slope} + rescale_{intercept} \tag{1}$$

The calculation equation of SUV value is (2):

$$SUV = (a \times v_p + b) \times \frac{W}{D \times e^{-log2 \times \frac{\Delta T}{half\_T}}} \tag{2}$$

where, a is the recalculation slope, b is the recalculation intercept, v_p is the gray value of pixel, W is the patient's weight, D is the total dose of radionuclide, ΔT is the delay between the injection time and the start time of scanning, $half\_T$ is the half-life of radionuclide. To eliminate the influence of singular samples, the HU value and SUV value are normalized to the interval [0, 1].

We rotated the 3D CT/PET data around the Z axis (vertical axis) by every 15 degrees from -75° to 90° to generate the multi-view dataset, take a total of 12 views. Then, to reduce the input parameters without losing useful information, we designed a dimensionality reduction operator, calculated the average CT, PET and contour slice along the Y axis (sagittal axis) of all twelve views. All points that exist in at least 10% of the non-blank sections are considered as contour image. According to the corresponding average contour slice, the tumor part of the CT and PET image is extracted. To unify the size of the input image without losing tumor information, the input image is cropped to a size of 200×200, and the bounding box of the tumor is placed in the center of the image. The average slice and input sample of one patient at the twelfth view (90°) are shown in Fig. 1. Finally, the obtained corresponding CT and PET image is superimposed into a matrix of 200×200×2 as the input of the prediction model. The specific operation process is shown in Fig. 2.

Augmentation and balance operation: For improving the generalization ability and the robustness of the model, image augmentation was adopted. Each average 2D slice was randomly flipped (horizontally and/or vertically), rotated a random amount (from -30° to 30°) and was augmented roughly 30 times. To increase the model's attention to positive samples, we adjust the data input to balance the positive and negative samples. In each epoch, some positive samples are reused to ensure that in each batch, the same number of positive samples and negative samples are used as input, and the number of enhanced data of the input positive and negative samples is also consistent in each batch.

III. METHODOLOGY

A. Framework

The framework of the mMmV-CNN is shown in Fig. 3. The input layer is twelve views of CT and PET dual-channel images with a size of 200*200*2 and clinical features with 12 dimension. The image is fed into Module 1 to extract features based on multiple views. The output features of each view in Module1 are aggregated in the view pooling layer and then used as the input to Module 2. The clinical parameters are used as the input in Module 2 as well. The output of Module 2 is obtained through the SoftMax layer, and the final prediction results can be obtained. MV-CNN is a directed acyclic graph structure,

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

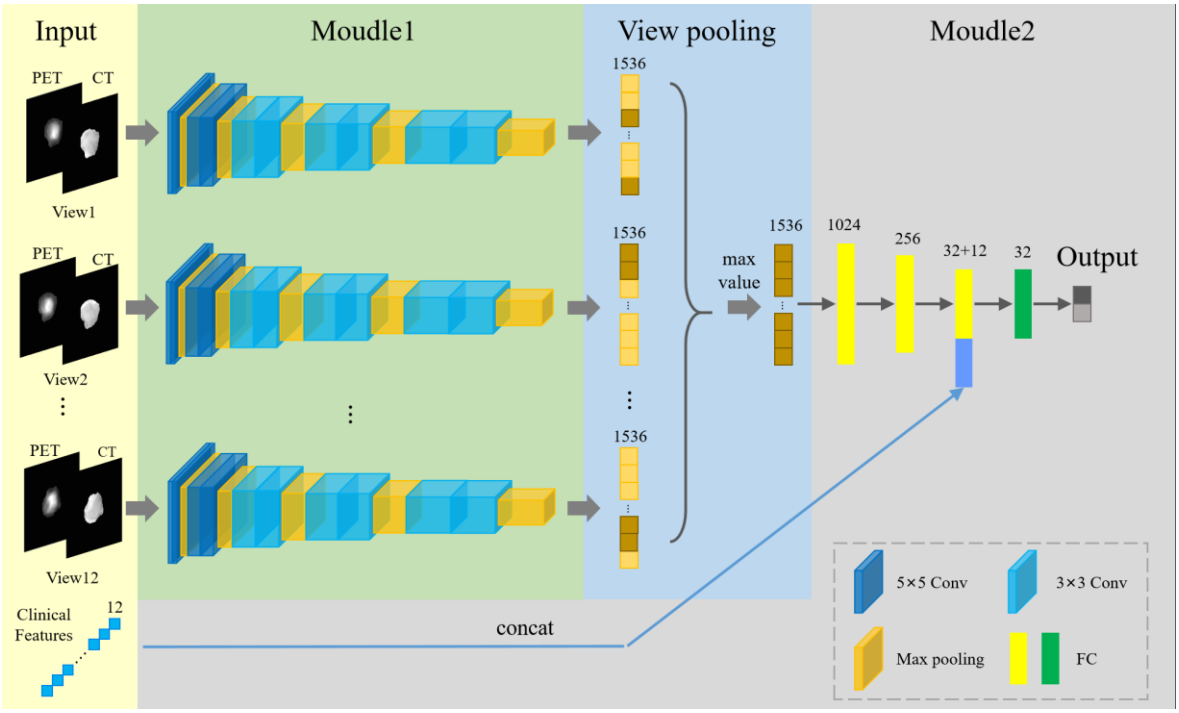

Fig. 3. Illustration of the proposed mMmV-CNN framework.

which can be trained through stochastic gradient descent back propagation in an end-to-end way.

The input data for Module 1 contains the average slices of twelve views from different angles of the original 3D volumetric data. Module 1 consists of convolutional layers and max-pooling layers, used to extract the features of the input image, and all views use the same network architecture. The network structure uses a stack structure which is similar to VGGNet [24]. It adopts multiple convolutional layers with smaller convolution kernels to replace one convolutional layer with larger convolution kernels. This multi-layer convolutional stacking method increases nonlinearity, which can express the characteristics of the input better than the structure of a single convolutional layer using fewer parameters. Selecting a small stride can prevent a larger stride from causing loss of detailed information. Meanwhile, as the number of network layers is deepened, and the extracted features are deeper. During training, even if the weight of a layer changes slightly, it will continue to be enlarged in deep network, so that the network will continue to adapt to the new input distribution [23]. In each view, 1536 features are extracted. The view-pooling layer uses element-wise maximum operation across the views, which reduces the amount of parameters while keeps the extracted texture information as much as possible. We expand all view elements into a one-dimensional array, and take the maximum value of the corresponding position of all views in each position. The newly composed 1536-dimensional vector is input into the fully connected layer. Module 2 consists of four fully connected layers. The clinical parameter is concatenated with the third fully connected layer in Module 2. The fc3 layer contains 32 features, which are slightly different from the clinical data feature dimensions and can increase the degree of influence of clinical features on the classification results. The output layer is the probability value of two categories: with or without LRR.

B. *Network Settings*

TABLE II.

MMMV- CNN ARCHITECTURE.

Module	Layer	Input Size	Output Size	Filter	Stride
Module1	Conv1	200×200×2	200×200×32	5×5	1
	Conv2	200×200×32	200×200×32	5×5	1
	Pool1	200×200×32	50×50×32	2×2	4
	Conv3	50×50×32	50×50×48	5×5	1
	Conv4	50×50×48	50×50×48	5×5	1
	Pool2	50×50×48	25×25×48	2×2	2
	Conv5	25×25×48	25×25×64	3×3	1
	Conv6	25×25×64	25×25×64	3×3	1
	Pool3	25×25×64	13×13×64	2×2	2
	Conv7	13×13×64	13×13×80	3×3	1
	Conv8	13×13×80	13×13×80	3×3	1
	Pool4	13×13×80	7×7×80	2×2	2
	Conv9	7×7×80	7×7×96	3×3	1
	Conv10	7×7×96	7×7×96	3×3	1
	Pool5	7×7×96	4×4×96	2×2	2
12views					
View pooling	View_pool	12×4×4×96	1536	-	-
Module2	Fc1	1536	1024	-	-
	Fc2	1024	256	-	-
	Fc3	256	32	-	-
	Clinical	32+12	44	-	-
	Fc4	44	32	-	-
	Fc5	32	2	-	-

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

The structure of the network is summarized in Table 2. Module 1 includes a stack of 10 convolutional layers and 5 max-pooling layers, and Module 2 includes 5 fully connected layers. The rectified linear units (ReLU) is employed as non-linear activation functions in the convolutional layer and the fully connected layer, the SoftMax function is used in the output layer. The network weights are optimized using Adam algorithm based on the cross entropy loss function. The fixed learning rate is 1×10^{-5} and the mini-batch size is 40. The maximum epoch is 300. The L2 regularization term is added to the loss function to prevent overfitting and improve the generalization ability of the model.

IV. Experiments

A. Experimental setup

We tested our mMmV-CNN performance using different combinations of input, including CT, PET, CT+PET and CT+PET+Clinical data in our experiment. When CT, PET or CT+PET was used as the input, we used the multi-view 2D CNN prediction model with the same parameters. When only clinical data were used as the input, a deep neural network (DNN) [25] was used.

The area under the receiver operating characteristic curve (AUC) is used as the objective function to construct the predictive model. The model with the highest AUC value is considered as the final model [26]. Sensitivity (SEN= $\frac{TP}{TP+FN}$), specificity (SPE= $\frac{TN}{TN+FP}$), accuracy (ACC= $\frac{TP+TN}{TP+FN+TN+FP}$) and AUC were used to evaluate the performance of different models. We performed five-fold cross-validation to evaluate the performance of all methods. No changes to the hyper-parameters were made between any of the folds.

Our model is built using the Tensorflow framework based on Python3. All experiments are trained and tested on a NVIDIA GTX 1080 graphics processing unit (GPU).

B. Multi-modal Data

TABLE III.

THE PREDICTION RESULTS OF MMMV-CNN ON DIFFERENT INPUT DATA.

Method	Modality	SEN	SPE	ACC	AUC
mMmV-CNN	Clinical	0.6531	0.6369	0.6408	0.6947
	CT	0.5714	0.7707	0.7233	0.7286
	PET	0.5714	0.7834	0.7330	0.7000
	CT+PET	0.5918	**0.8535**	**0.7913**	0.7717
	CT+PET+Clinical	**0.7347**	0.8089	**0.7913**	**0.8052**

Table 3 shows the prediction results of our MmMv-CNN method on different modal data. Each experiment is the result of the best model obtained after training many times. The AUC value of our method under the common input of three modal data experimental results is roughly between 0.75-0.81. The results of our method under different modal data show that combining CT, PET and clinical data can improve the overall prediction effect and achieve the highest AUC value of 0.81.

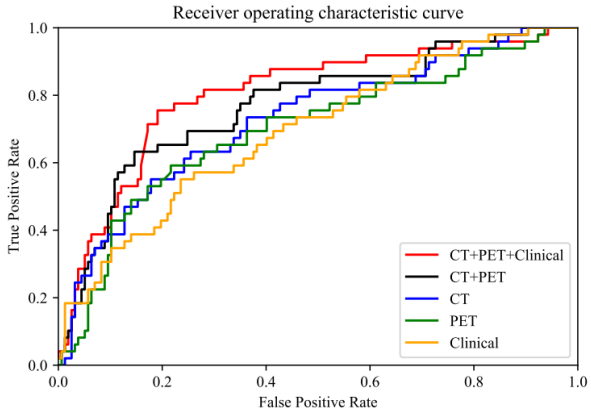

Fig. 4. ROCs of mMmV-CNN on different input data.

Fig. 4 shows the corresponding receiver operating characteristic curves (ROCs) of different approaches. The area under ROC curve (red line) corresponding to CT+PET+ Clinical data was higher than the ROC curves obtained with other input. Athough the sensitivity and AUC values are highest in the model using combined CT, PET and clinical data as input, the specificity after adding the clinical data is not as good as the results of using only CT and PET. One possible reason is that the ability to identify negative samples is poor when only clinical data is used. Due to the large number of negative samples, the accuracy rate is not high as well in the model using clinical data only. As such, specificity in the combined model is reduced after adding clinical data to CT and PET data. However, the sensitivity is greatly improved, which enable the model to better focus on positive samples and find LRR patients, and leading to the overall improvement performance of the model (AUC value).

C. Select the number of views

To analyze the optimal number of views, we compared views taken every 30 degrees from -60° to 90° (6 views are taken in total), views taken every 15 degrees from -75° to 90° (12 views are taken in total), and views taken every 10 degrees from -80°to 90° (take 18 views in total) in three cases. The experiment uses CT, PET and clinical three modal data input at the same time.

For the selection of the number of views, results summarized in Table 4 suggest that the best performance is achieved when the views are taken 15 degrees apart (a total of 12 views are taken). Taking too many or too few views will cause redundancy or lack of information and affect the prediction effect of the model.

TABLE IV.

COMPARISON OF RESULTS BETWEEN DIFFERENT VIEWS.

Views	SEN	SPE	ACC	AUC
6	0.6122	0.7643	0.7282	0.7327
12	**0.7347**	**0.8089**	**0.7913**	**0.8052**
18	0.7142	0.7580	0.7476	0.7776

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

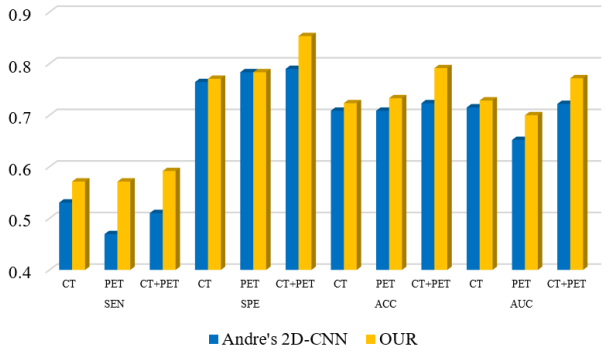

Fig. 5. Comparison result with Andre's 2D-CNN method under three inputs.

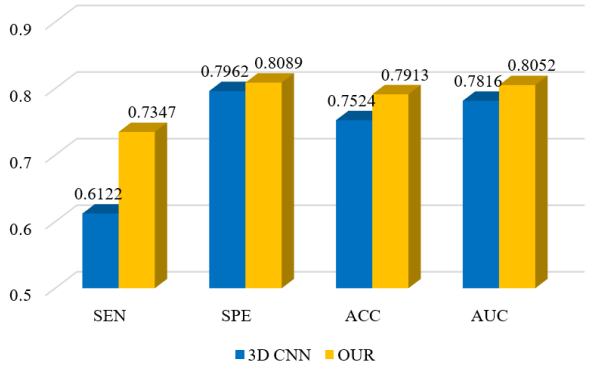

Fig. 6. Comparison result with 3D-CNN method.

D. Comparison with other methods

Our mMmV-CNN was compared with Andre's method (denoted as "Andre's 2D-CNN") [11], which is 2D CNN-based method. In [11], only one central tumor slice of each patient's CT image was used, and the original network is single channel. To compare with our method, we modified it to dual channel network for comparison with inputs with both PET and CT. For fair comparison the input data is the same as our method, using the average projection slice of the 3D data along the vertical axis. It can be seen from Fig. 5 that our model obtains better results in the case of CT, PET and CT+PET as input.

To compare the performance of our method with the 3D method (denoted as "3D-CNN"), we trained a 3D CNN with the same samples and same network structure under one view. The 3D data and clinical data of CT and PET are input, and the 3D data dimension is 135×135×135. As shown in Fig. 6, our method performs better. At the same time, the use of 3D data takes much longer. Table 5 shows the comparison of network parameters and running time (150 epochs) of the two methods.

TABLE V.

COMPARISON OF NETWORK PARAMETERS AND RUNNING TIME.

Method	Parameters	Running time (s)
3D-CNN	4,504,962	31,736
OUR	2,290,306	11,029

V. CONCLUSION

To accurately identify H&N cancer patients at high-risk for LRR after definitive radiation or chemoradiation therapy, a new end-to-end mMmV-CNN model was developed in this study. Compared with the single-view method, the multi-view method makes full use of image data from different angles to obtain more discriminative information. Furthermore, our experimental results demonstrated that mMmV-CNN can obtain better performance by combining extracted CT, PET and clinical features. Once validated in an external patient cohort, the model developed in this work could help physicians develop optimal personalized treatment strategy for H&N cancer patients. In future work, we will add heterogeneous data such as radiomics features, study multi-modal learning methods, and explore the potential correlations between multi-modal data, which can not only reduce the impact of a certain modal data anomaly on the results, but also make the prediction results more comprehensive and of reference value. In addition, it is necessary to improve the method of multi-view image aggregation, reduce the redundancy of information, and further consider the relationship among views.

ACKNOWLEDGMENT

The authors have no relevant financial or non-financial interests to disclose. This work was supported by Engineering Research Center of Big Data Application in Private Health Medicine, Fujian Province University (No. KF2020005).

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 68, no. 6, pp. 394-424, 2018.

[2] J. J. Caudell, J. F. Torres-Roca, R. J. Gillies, H. Enderling, S. Kim, A. Rishi, et al, "The future of personalised radiotherapy for head and neck cancer," The Lancet Oncology, vol. 18, no. 5, pp. e266-e273, 2017.

[3] S. Keek, S. Sanduleanu, F. Wesseling, R. De Roest, M. Van Den Brekel, and M. Van Der Heijden, "Computed tomography-derived radiomic signature of head and neck squamous cell carcinoma (peri) tumoral tissue for the prediction of locoregional recurrence and distant metastasis after concurrent chemo-radiotherapy," PloS one, vol. 15, no. 5, pp. e0232639, 2020.

[4] K. Wang, Z. Zhou, R. Wang, L. Chen, Q. Zhang, D. Sher, and J. Wang, "A multi‐objective radiomics model for the prediction of locoregional recurrence in head and neck squamous cell cancer," Medical physics, vol. 47, no. 10, pp. 5392-5400, 2020.

[5] J. Chang, C. Wu, K. Yuan, A. Wu, and S. Wu, "Locoregionally recurrent head and neck squamous cell carcinoma: incidence, survival, prognostic factors, and treatment outcomes," Oncotarget, vol. 8, no. 33, pp. 55600, 2017.

[6] Y. Wang, W. Liu, Y. Yu, J. Liu, H. Xue, Y. Qi, et al, "CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer," European radiology, vol. 30, no. 2, pp. 976-986, 2020.

[7] S. Shanthi, and N. Rajkumar, "Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods," Neural Processing Letters, 2020, pp. 1-14.

[8] W. Zhu, L. Xie, J. Han, and X. Guo, "The application of deep learning in cancer prognosis prediction," Cancers, vol. 12, no. 3, pp. 603-621, 2020.

[9] J. Rose, K. Jaspin, and K. Vijayakumar, "Lung Cancer Diagnosis Based on Image Fusion and Prediction Using CT and PET Image," Signal and Image Processing Techniques for the Development of Intelligent Healthcare Systems. Springer, Singapore, 2021, pp. 67-86.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

[10] W. Le, and F. P. Romero, "A Normalized Fully Convolutional Approach to Head and Neck Cancer Outcome Prediction," arXiv preprint arXiv:2005.14017, 2020.

[11] A. Diamant, A. Chatterjee, M. Vallières, G. Shenouda, and J. Seuntjens, "Deep learning in head & neck cancer outcome prediction," Scientific reports, vol. 9, no. 1, pp. 1-10, 2019.

[12] T. Y. Lo, P. Wei, C. Yen, J. F. Lirng, M. Yang, P. Chu, and S. Y. Ho, "Prediction of metastasis in head and neck cancer from computed tomography images," Proceedings of the 2018 4th International Conference on Robotics and Artificial Intelligence, 2018, pp. 18-23.

[13] K. Sekaran, P. Chandana, N. M. Krishna, and S. Kadry, "Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer," Multimedia Tools and Applications, vol. 79, no. 15, pp. 10233-10247, 2020.

[14] I. El Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, et al, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," Pattern recognition, vol. 42, no.6, pp. 1162-1171, 2009.

[15] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE transactions on pattern analysis and machine intelligence, vol. 41, no.2, pp. 423-443, 2018.

[16] W. Lv, S. Ashrafinia, J. Ma, L. Lu, and A. Rahmim, "Multi-level multi-modality fusion radiomics: application to PET and CT imaging for prognostication of head and neck cancer," IEEE journal of biomedical and health informatics, vol. 24, no.8, pp. 2268-2277, 2019.

[17] L. J. Beesley, P. G. Hawkins, L. M. Amlani, E. L. Bellile, K. A. Casper, S. B. Chinn, et al, "Individualized survival prediction for patients with oropharyngeal cancer in the human papillomavirus era," Cancer, vol. 125, no. 1, pp. 68-78, 2019.

[18] T. Zhou, H. Fu, Y. Zhang, C. Zhang, X. Lu, J. Shen, and L. Shao, "M2Net: Multi-modal Multi-channel Network for Overall Survival Time Prediction of Brain Tumor Patients," MICCAI, 2020, pp. 221-231.

[19] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," Proceedings of the IEEE international conference on computer vision, 2015, pp. 945-953.

[20] T. Abeywickrama, M. A. Cheema, M, and D. Taniar, "K-nearest neighbors on road networks: a journey in experimentation and in-memory implementation," arXiv preprint arXiv:1601.01549, 2016.

[21] X. Zhao, L. Li, W. Lu, and S. Tan, "Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network," Physics in Medicine & Biology,vol. 64, no. 1, pp. 015011, 2018.

[22] Z. Zhong, Y. Kim, K. Plichta, B. G. Allen, L. Zhou, J. Buatti, and X. Wu, "Simultaneous cosegmentation of tumors in PET‐CT images using deep fully convolutional networks," Medical physics, vol. 46, no. 2, pp. 619-633, 2019

[23] A. Kumar, M. Fulham, D. Feng, and J. Kim, "Co-learning feature fusion maps from PET-CT images of lung cancer," IEEE Transactions on Medical Imaging, vol. 39, no. 1, pp. 204-217, 2019.

[24] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409. 1556, 2014.

[25] R. Wang, Y. Weng, Z. Zhou, L. Chen, H. Hao, and J. Wang, "Multi-objective ensemble deep learning using electronic health records to predict outcomes after lung cancer radiotherapy," Physics in Medicine & Biology, vol. 64, no. 24, pp. 245005, 2019.

[26] C. Ling, J. Huang, and H. Zhang, "AUC: a better measure than accuracy in comparing learning algorithms," Conference of the canadian society for computational studies of intelligence. Springer, Berlin, Heidelberg, 2003, pp. 329-341.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 08,2022 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.