Predicting local control of brain metastases after stereotactic radiotherapy with clinical, radiomics and deep learning features

Hemalatha Kanakarajan[1*], Wouter De Baene[1*], Patrick Hanssens[2,3], Margriet Sitskoorn[1]

1 Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands
2 Gamma Knife Center, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands
3 Department of Neurosurgery, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands

*Correspondence concerning this article should be addressed to Hemalatha Kanakarajan - H.Kanakarajan@tilburguniversity.edu and Wouter De Baene - w.debaene@tilburguniversity.edu

Abstract

Background and purpose

Timely identification of local failure after stereotactic radiotherapy for brain metastases allows for treatment modifications, potentially improving outcomes. Previous studies showed that adding radiomics or Deep Learning (DL) features to clinical features increased Local Control (LC) prediction accuracy. However, no study has integrated radiomics, DL, and clinical features into machine learning algorithms to predict LC. We examined whether a model using all these features achieves better accuracy than models using only a subset.

Materials and methods

We collected pre-treatment brain MRIs and clinical data for 129 patients at the Gamma Knife Center of the Elisabeth-TweeSteden Hospital. Radiomics features (extracted using the Python radiomics feature extractor) and DL features (extracted using a 3D ResNet model) were combined with clinical features. Performance of a Random Forest classifier was compared across four models trained with: clinical features only; clinical and radiomics features; clinical and DL features; and clinical, radiomics, and DL features.

Results

The prediction model utilizing only clinical variables provided an Area Under the receiver operating characteristic Curve (AUC) of 0.82 and an accuracy of 75.6%. Adding radiomics features increased the AUC to 0.88 and accuracy to 83.3%, while adding DL features resulted in an AUC of 0.86 and accuracy of 78.3%. The best performance came from combining clinical, radiomics, and DL features, achieving an AUC of 0.89 and accuracy of 87%.

Conclusion

Integrating radiomics and DL features with clinical characteristics improves LC prediction after stereotactic radiotherapy for brain metastases. These findings demonstrate the potential for early outcome prediction, enabling timely treatment modifications to improve patient management.

Index Terms: Brain metastases, Deep learning, Local control, Radiomics, Stereotactic radiotherapy

Clinical and Translational Impact Statement: Our study holds great clinical value, as the increased prediction accuracy can lead to tailored and effective interventions, resulting in better outcomes for brain metastases patients treated with stereotactic radiotherapy.

Introduction:

Metastatic brain tumors represent the most prevalent form of intracranial malignancies [1]. Brain metastases manifest in approximately 20 to 40% of individuals diagnosed with cancer [2, 3]. While any tumor has the potential to spread to the brain, the predominant types include lung cancer, breast cancer, melanoma, and gastrointestinal cancers [1]. The prevalence of brain metastases is increasing [4]. The adoption of sophisticated imaging methods for diagnosis, alongside the implementation of innovative chemotherapeutic approaches for systemic cancer treatment, may contribute to the increased likelihood of finding and developing brain metastases [1].

Currently, the prognosis of patients with brain metastases is poor with median overall survival of a few weeks to months in untreated patients [5]. The survival of patients with brain metastases depends upon prompt diagnosis and treatment efficacy. The standard treatment options are surgical resection and radiotherapy [5]. Surgery is recommended for patients with a single large tumor in a reachable location [6]. The three principal modalities of radiotherapy for brain metastases are Whole-Brain Radiation Therapy (WBRT), Single-fraction Stereotactic Radiosurgery (SRS), and hypo-fractionated Stereotactic Radiotherapy (SRT). WBRT was the main treatment in the past for patients with multiple brain metastases [7]. There has been a shift from WBRT to SRT and SRS due to the adverse effects of WBRT, such as fatigue and cognitive decline [8]. Through SRS, multiple nonparallel beams are converged to deliver a single, high radiation dose to a targeted region whereas SRT delivers multiple, smaller doses of radiation over time. In SRS and SRT, the delivered radiation is confined to the lesion and there is a rapid dose fall-off at the edge of the treatment volume. Since the radiation dose is not delivered to the

healthy brain tissue, there is a reduced likelihood of posttreatment neurocognitive decline compared to WBRT [9].

The assessment of Local Control (LC) of brain metastases is an important clinical endpoint. A stable disease after treatment is categorized as LC while a progressive disease indicates a Local Failure (LF) [10]. It may require several months before local changes of the treated lesions become evident on follow-up scans. Considering that the median survival of patients with brain metastases following radiotherapy can range between 5 months and 4 years [11, 12], timely identification of LF subsequent to radiotherapy is crucial as it offers the opportunity for timely tailored treatment modifications, ensuring that patients receive the most effective care and maximizing their chances of a favorable prognosis.

Cancer imaging analysis driven by Artificial Intelligence (AI) has the potential to revolutionize medical practice by revealing previously undisclosed characteristics from routinely obtained medical images [13]. These features can serve as valuable inputs for the development of machine learning models aimed at predicting the treatment response or LC of brain metastases [13]. This is particularly important given the advancement in Graphical Processing Unit (GPU) processing capabilities and the availability of large amounts of training data which have led to a rapid expansion in neural networks and deep learning techniques for regression and classification tasks [14]. Deep learning models have demonstrated significant potential in identifying crucial and unique features within medical image data across a range of applications, including cancer treatment [15, 16, 17]. Deep learning uses artificial neural networks to automatically learn features from raw data. In medical imaging, deep learning methods are applied directly to the images themselves, learning hierarchical representations of the data. Deep learning has been particularly successful in tasks like image classification, object detection, and segmentation [18]. The information extracted by the deep learning models from the tumor images can be used to predict treatment outcome [18, 19, 20]. Jalalifar et al [21] introduced a novel deep learning architecture to predict the outcome of LC in brain metastasis treated with stereotactic radiation therapy using treatment-planning magnetic resonance imaging (MRI) alongside standard clinical attributes [21]. Their findings highlighted that the addition of deep learning features to the clinical features significantly enhanced the prediction accuracy.

Radiomics is another research domain for extracting quantitative features from medical images for different clinical applications [22]. Radiomics focuses on extracting quantitative features from

medical images, such as texture, shape and intensity characteristics. These features are then used to characterize tumors or other abnormalities in the images. While both radiomics and deep learning are used in medical imaging, radiomics focuses on extracting handcrafted features (such as the manually delineated tumor segmentations) from images while deep learning learns features directly from raw data using neural networks [18, 20]. Numerous studies have underscored the efficiency of radiomic-based machine learning algorithms in predicting treatment outcomes across different medical conditions [23, 24]. The radiomic-based machine learning algorithms have also been efficiently applied for the prediction of LC of brain metastases after radiotherapy [25, 26, 27]. Karami et al [25] proposed a radiomics framework to predict the LC in patients with brain metastasis treated with SRT. Based on the radiomics features, Kawahara et al [28] proposed a neural network model for predicting the local response of metastatic brain tumor to Gamma Knife Radiosurgery (GKRS). Liao et al [26] and Andrei et al [27] demonstrated the value of combining radiomic features and clinical features to enhance the prediction of brain metastases responses after GKRS. Their findings show that the addition of radiomic features to the clinical features improved the accuracy of the prediction models for LC of brain metastases.

The studies that used either radiomics or deep learning features together with the clinical features to predict LC of brain metastases after SRT showed that the addition of either radiomics or deep learning features increased the prediction accuracy of the models. But to the best of our knowledge, there is no study that combined both radiomics and deep learning features together with clinical features to develop machine learning algorithms to predict LC of brain metastases. A model trained with all these combined features might predict LC with a higher accuracy than the other models trained with a subset of these features, offering a more comprehensive understanding of treatment response, potentially leading to more tailored and effective interventions which may result in improved treatment outcomes, prolonged patient survival, and enhanced quality of life. Hence, the objective of this study was to predict the LC of brain metastases after SRT using the combination of deep learning, radiomics and clinical features.

Methods

Data collection

We retrospectively collected the clinical data from 199 brain metastases patients from the Gamma Knife Center of the Elisabeth-TweeSteden Hospital (ETZ) at Tilburg, The Netherlands. This study was approved by the ETZ science office and by the Ethics Review Board at Tilburg University. The patients underwent GKRS at the Gamma Knife Center. After excluding the patients with incomplete data, we included 129 patients. For these 129 patients, pre-treatment contrast-enhanced (with triple dose gadolinium) brain MRIs were collected using a 1.5T Philips Ingenia scanner (Philips Healthcare, Best, The Netherlands) with a T1-weighted sequence (TR/TE: 25/1.86 ms, FOV: 210x210x150, flip angle: 30°, transverse slice orientation, voxel size: 0.82 x 0.82 x 1.5mm). These high-resolution whole-brain planning scans were made as part of clinical care at the Gamma Knife Center of the ETZ between 2015 and 2021. For all patients, the segmentations of the baseline Ground Truth (GT) were manually delineated by expert oncologists and neuroradiologists at ETZ. These 129 patients were randomly divided into 103 patients for training and 26 patients for testing. At ETZ, the FU MRI scans were made at 3, 6, 9, 12, 15, and 21 months after treatment. A tumor was defined as progressive if there was a relative increase in tumor volume on any of these follow-up MRIs compared to pretreatment MRI. A stable tumor after treatment is categorized as LC while a progressive tumor indicates a LF. The pre-processing, feature extraction, model training, and evaluation were performed in Python (version 3.11).

Preprocessing

As a first preprocessing step, all the MRI scans were registered to standard MNI space using Dartel in SPM12 (Wellcome Trust Center for Neuroimaging, London, UK), implemented in Python using the Nipype (Neuroimaging in Python–Pipelines and Interfaces) software package (version 1.8.6) [29]. The voxel size of the normalized image was set to 1*1*1 mm^3. For all other normalization configurations, the default values offered by SPM12 were used. One other preprocessing step was to combine the GT labels for patients with more than one brain metastasis in one single GT mask. FSL library (Release 6.0) was used for this integration. Pre-processing was applied to improve the reliability of radiomics and deep learning feature extraction [30].

Clinical features

The list of clinical factors that we collected from the Gamma Knife Center of ETZ were gender, survival status, diagnosis of brain metastases within 30 days after diagnosis of primary tumor, prior brain treatment, prior SRS, prior WBRT, prior surgery, prior systemic treatment, presence of extracranial

metastases, presence of lymph node metastases, presence of seizure, number of metastases at diagnosis, Karnofsky Performance Status score (KPS), occurrence of new metastases after GKRS, presence of extracranial metastases, primary tumor type, age at diagnosis of brain metastases, age at diagnosis of primary tumor, presence of local recurrence, tumor volume and treatment dose. For the treatment dose, we took the average value from the dose range. We extracted the tumor volume from the segmentations of the baseline GT and added it to the clinical data. We took the total tumor volume across the metastases for patients with more than one brain metastasis**.** The clinical data was converted to a python dataframe.

Radiomics features

The segment-based radiomics features were extracted from the T1 weighted pre-treatment MRI scans using the radiomics feature extractor of the python radiomics package. The seven groups of features extracted from the Region Of Interest (ROI) of the tumor segmentations were shape-based features (14 features), first-order features (18), Gray Level Cooccurrence Matrix (GLCM) (24) features, Gray Level Dependence Matrix (GLDM) (14) features, Gray Level Run Length Matrix (GLRLM) (16) features, Gray Level Size Zone Matrix (GLSZM) (16) features, and Neighbouring Gray Tone Difference Matrix (NGTDM) (5) features. The resulting 107 radiomics features were considered in this study. The list of radiomics features extracted are listed in the appendix. The mathematical definitions of these radiomics features are given in the Pyradiomics feature documentation (https://pyradiomics.readthedocs.io/en/latest/features.html). The radiomics features were then combined with the clinical features to form a combined python dataframe.

Deep learning features

A 3D ResNet model [31, 32] pre-trained on the ImageNet challenge dataset [33] was used to extract the deep learning features from the manually segmented masks. Prior to input, the images were rescaled to 256 × 256 × 256 using spline interpolation order 3, improving the accuracy of the model [34]. Additionally, the pixels were also sample-wise scaled between -1 and 1. These preprocessing steps contribute to optimizing the performance of the 3D ResNet model in extracting meaningful features from the images [35].

A 3D convolution was applied on the training data. This convolutional layer was designed with 50 filters and a large kernel size of (7, 7, 7), while also employing a stride of (2, 2, 2) for down sampling. The purpose of this stride is to efficiently reduce the spatial dimensions of the input data, capturing broader information across the dataset while managing computational complexity [36].

To fine-tune this convolutional layer, we applied batch normalization and Rectified Linear Unit (ReLU) activation. Batch normalization helps the model to adapt to our dataset, improving its performance, stability and ability to generalize. ReLU is like a simple on/off switch for a neuron in a neural network. ReLU helps the neural networks learn by letting positive signals pass through unchanged while ignoring negative ones and thus enables neural networks to learn complex patterns effectively. ReLU transforms the features from the images to be compatible with the pretrained model, maintain consistency with the original training, and facilitate efficient gradient propagation during the fine-tuning process [37, 38].

Following this, we incorporated three fine-tuned ResNet blocks into the model. After adding the ResNet blocks, we applied global average pooling to reduce the spatial dimensions to 1x1x1. Finally, a dense layer with softmax activation was added for classification. The deep learning features were extracted using this fine-tuned 3D ResNet model and then combined with the clinical and radiomics features to form a combined python dataframe.

The complete process of pre-processing and feature extraction is summarized in Figure 1.

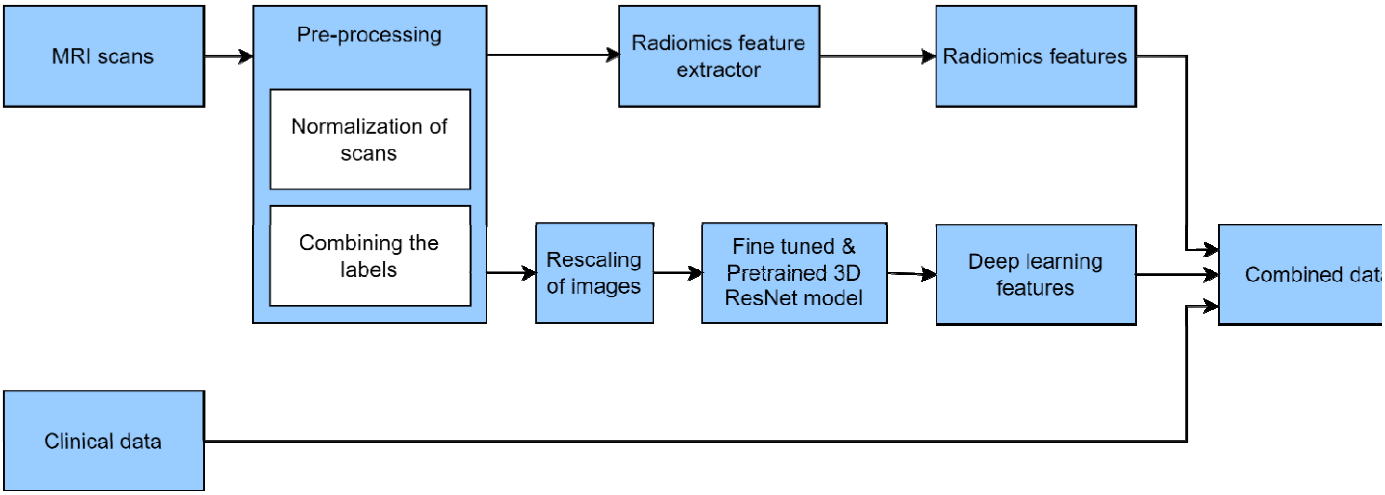

Figure 1: The process of pre-processing, feature extraction, and combining the data.

Model training

The features with low variance (<0.01) were determined and excluded from the combined dataset to improve prediction accuracy. The list of the excluded features is included in the Appendix. The data was then normalized and was supplied to the Random Forest classifier. Experimental results from Chen et al. [39] demonstrated that the Random Forest machine learning algorithm achieves a better classification performance compared to other classification algorithms. Hence, we choose the Random Forest machine learning algorithm to predict LC from the combined data. The model was trained with the training data set and then tested with the test data set. The process of training and evaluation of the models is shown in Figure 2. The binary outcome used in training and validation was the LC after treatment taken from the list of clinical features. The different models that we created were:

1. Random Forest classifier trained with clinical features only.
2. Random Forest classifier trained with the combination of clinical and deep learning features.
3. Random Forest classifier trained with the combination of clinical and radiomics features.
4. Random Forest classifier trained with the combination of clinical, radiomics and deep learning features.

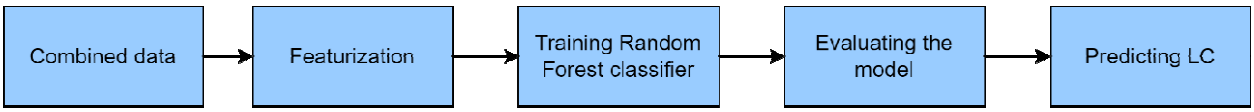

Figure 2: Model training, evaluation and prediction.

Model evaluation

The performance of the model was evaluated by measuring the following metrics: classification accuracy, precision, F1 score, recall and AUC. The classification accuracy is the ratio of the number of correct predictions to the total number of input samples. The precision is the ability of the classifier not to label as positive a sample that is negative and recall is the ability of the classifier to find all the positive samples. In other words, precision is the ratio of true positive predictions to the total number of positive predictions made by the model, while recall is the ratio of true positive predictions to the total

number of actual positives in the dataset. The F1 score in percentage gives the balance between how often the model is correct (precision) and how well it finds all the positive instances (recall). A Receiver Operating Characteristic (ROC) is a graphical plot which is created by plotting the true positive rate vs the false positive rate at various threshold settings. The AUC computes the area under the ROC curve. By doing so, the curve information is summarized in one number. Similar to the F1 score, the AUC reaches its best value at 1.

A K-fold cross validation was applied on the model. The different values that we used for K during cross-validation were 2, 3, and 5. The average accuracy and other metrics across the different folds was calculated. From the trained models, we also extracted the importance of the various factors for predicting the LC.

Results:

Patient characteristics

Table 1 shows the characteristics of patients included in our study. Among the 129 patients, 42% were male and 58% were female. The patients had an average age of 63 and an average tumor volume of 17,445 mm^3. Sixty-nine % of the patients had a primary lung cancer and 94% of the patients had less than 10 brain metastases.

Gender	
Male	54 (42%)
Female	75 (58%)
Extracranial tumor activity	
Yes	39
No	90
Diagnosis of brain metastases within 30 days after diagnosis of primary tumor	
Yes	39
No	90
Prior brain treatment	
Yes	23
No	106
Prior SRS	
Yes	15
No	114
Prior WBRT	
Yes	7
No	122
Prior surgery	
Yes	8
No	121
Prior systemic treatment	

Yes	76
No	53
Presence of extracranial metastases	
Yes	54
No	75
Presence of lymph node metastases	
Yes	68
No	61
Presence of seizure	
Yes	18
No	111
KPS score	
60	3
70	14
80	31
90	38
100	43
Occurrence of new metastases after GKRS	
Yes	62
No	67
Presence of local recurrence	
Yes	40
No	89
Total tumor volume (mm^3)	
Average	17,445 (88-88,029)
Treatment dose (Gy)	
Average (minimum – maximum)	22 (14 – 25)
Age at diagnosis of brain metastases (years)	
Average (minimum – maximum)	63 (36 – 85)
Primary tumor type	
Lung	89
Melanoma	8
Breast	3
Others	29
Number of brain metastases	
1	30
2 -3	50
4-10	41
>10	8

Table 1: Patient characteristics.

The average performance metrics for the four models across the cross-validation datasets are shown in Table 2. In bold is the value of the best score for the corresponding performance metric.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC
RF trained with clinical features only	75.6	83.6	71.0	75.3	0.82
RF trained with the combination of clinical and deep learning features	78.3	77.6	86.3	80.6	0.86
RF trained with the combination of clinical and radiomics features	83.3	86.6	82.0	84.0	0.88
RF trained with the combination of clinical, radiomics and deep learning features	**87.0**	**90.0**	**86.6**	**87.6**	**0.89**

Table 2: Average performance of the models on the validation datasets.

The accuracy of the Random Forest (RF) model trained with clinical features only was 75.6%. The model trained with the combination of clinical and deep learning features had an improved accuracy of 78.3%. The model trained with the combination of clinical and radiomics features was even higher at 83.3%. But the model with the highest prediction accuracy of 87% was the model trained with the combination of clinical, radiomics and deep learning features. This model also achieved the highest precision of 90%, recall of 86.6%, F1 score of 87.6% and the best AUC of 0.89. The model trained with the combination of clinical and radiomics features showed the second highest accuracy of 83.3%, precision of 86.6%, F1 score of 84% and an AUC of 0.88. Interestingly, this model had a recall of 82% which is lower than the recall score (86.3%) of the model trained with the combination of clinical and deep learning features. However, the F1 score (which is the balance between the precision and recall) was higher for the model trained with clinical and radiomics features when compared with the model trained with the combination of clinical and deep learning features. The model with the lowest prediction accuracy of 75.6% was the model trained with clinical features only. This model also achieved the lowest recall of 71%, F1 score of 75.3% and the lowest AUC of 0.82. Though this model had better

precision than the model trained with the combination of clinical and deep learning features, it still had a lower F1 score when compared to the model trained with the combination of clinical and deep learning features. The ROC curve of the model trained with the combination of clinical, radiomics and deep learning features is shown in Figure 3. The ROC curve of the other three models were added to the Appendix.

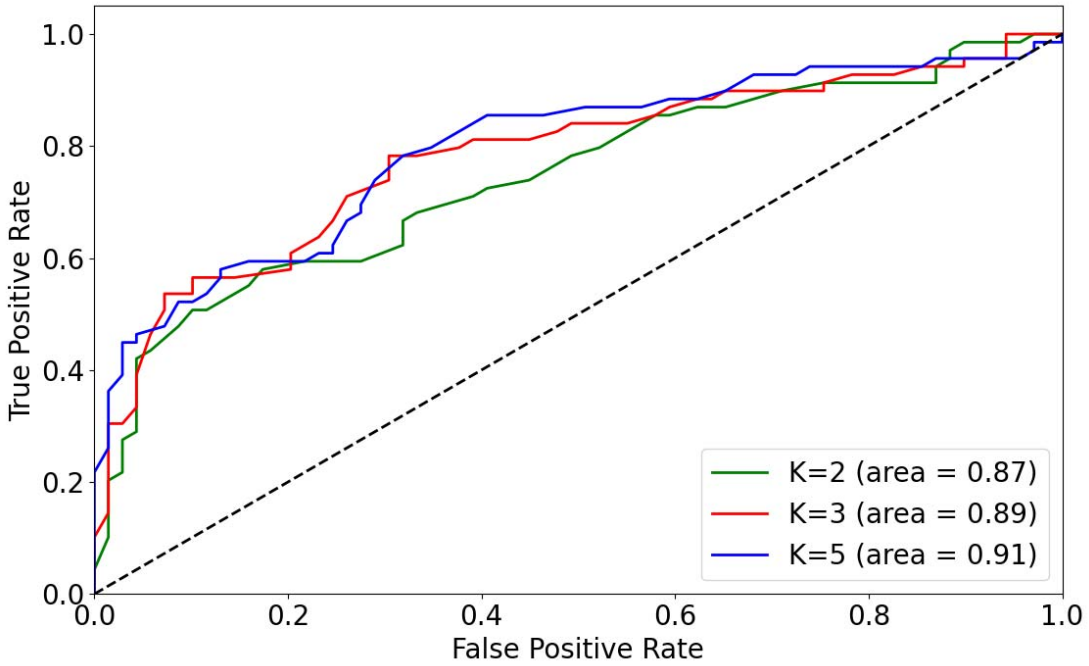

Figure 3: ROC curve of model trained with clinical, deep learning and radiomics features.

The variable importance of the top 10 factors for predicting the LC for each of the models were added to the appendix. For the model trained with the combination of clinical, radiomics and deep learning features, the top 10 features associated with the prediction of LC were original_glrlm_LowGrayLevelRunEmphasis, original_firstorder_Entropy, original_shape_LeastAxisLength, original_gldm_LargeDependenceHighGrayLevelEmphasis, original_shape_Elongation, original_firstorder_skewness, age at brain metastases diagnosis, original_shape_Maximum3DDiameter, age at primary tumor diagnosis, and original_glcm_SumSquares. Except for age at brain metastases diagnosis and age at primary tumor diagnosis which are clinical features, the rest of these top 10 features are radiomics features. There were no deep learning features in this list.

Discussion

Brain metastases patients treated with SRT are at risk of developing local failure. Prompt diagnosis of local failure might increase treatment options and hence improve treatment outcomes. In this study, we trained and tested machine learning algorithms to predict the local failure using clinical features and T1-weighted MR imaging features. Four distinct models were developed: One model was trained with clinical features only, one with the combination of clinical and deep learning features, one with the combination of clinical and radiomics features and one with the combination of clinical, radiomics and deep learning features. Notably, the model incorporating all three feature sets outperformed the others, demonstrating the significance of integrating clinical, radiomics and deep learning features into predictive models.

Jalalifar et al [21] introduced a novel deep learning architecture to predict the outcome of LC in brain metastasis treated with stereotactic radiation therapy using treatment-planning magnetic resonance imaging and standard clinical attributes. The accuracy of the model developed with only the clinical features was 67.5%, but the addition of deep learning features to the clinical features increased the prediction accuracy to 82.5%. Our model trained with only the clinical and deep learning features provided a prediction accuracy of 78.3% (vs. 75.6% with clinical features only). Kawahara et al [28] proposed a neural network model using only the radiomics features for predicting the local response of metastatic brain tumor to GKRS. This proposed network provided a prediction accuracy of 78%. Karami et al [25] produced a radiomics framework to predict the outcome of SRT for brain metastases using the clinical and radiomics features. The framework predicted LC with an accuracy of 82%. Our model trained with only the clinical and radiomics features provided a prediction accuracy of 83.3%. To the best of our knowledge, the current study marks the first attempt to combine both radiomics and deep learning features with clinical features to predict LC of brain metastases. The model incorporating all these features predicted LC with an accuracy of 87%, surpassing all models trained with a subset of these features. This increased prediction accuracy holds promises for improved treatment outcomes for the patients. Furthermore, insights into the variable importances provided by the model could offer valuable insights into the features associated with the LC of brain metastases, potentially guiding future research and clinical decision-making.

The ability to predict LC with high accuracy before initiating SRT treatment offers an invaluable opportunity for tailoring treatment strategies for the best outcomes. Providing clinicians with

information on the risk of local recurrence for individual patients empowers them to discuss these risks with patients prior to SRT. The capability to predict LC prior to treatment not only aids in informed decision-making regarding SRT but also opens avenues to consider alternative treatment modalities such as systemic therapy or WBRT. Additionally, it can enable clinicians to explore alternative radiotherapy approaches such as fractionated SRT or SRS with higher dose, depending on the predicted risk of local recurrence. Conversely, in cases where the risk of local recurrence is deemed low, SRT may be favored over other treatment options. Ultimately, pre-treatment prediction of LC serves as a valuable tool for both clinicians and patients, facilitating shared decision-making and optimizing treatment plans tailored to the needs and risk profiles of individual patients.

Although this study focused on creating a model for predicting the LC after SRT, the same approach can be extended to other treatment options and to the prediction of other clinical endpoints like overall survival.

One notable strength of this study lies in the meticulous brain metastases segmentation procedure. Expert oncologists and neuroradiologists at ETZ manually delineated the segmentations of the baseline GT on all the planning MRI scans used in this study. This ensures accurate regions of interest (ROIs) for the extraction of segment-based radiomics features.

It is important to note that in this study, we exclusively used the T1 weighted MRI scans. Exploring additional sequences and extracting radiomics and deep learning features from them could potentially improve the accuracy of the prediction models even more. In addition, for a more rigorous evaluation of the efficacy and robustness of the models, further investigations involving larger patient cohorts, preferably with multi-institutional data are warranted. Furthermore, the inclusion of an external validation dataset could significantly improve the generalizability of the prediction model, strengthening confidence in its clinical applicability across diverse patient populations and healthcare settings.

Conclusion

The findings of this study show that the machine learning model trained with the combination of clinical, radiomics and deep learning features predict LC of brain metastases with high accuracy, outperforming models trained with the subset of these features. The increased prediction accuracy can lead to more tailored and effective interventions, resulting in improved treatment outcomes, prolonged patient survival, and enhanced quality of life.

Data Availability:

The data used for this study is available at ETZ and is accessible after approval from the ETZ Science office.

Funding:

This research is supported by KWF Kankerbestrijding and NWO Domain AES, as part of their joint strategic research programme: Technology for Oncology IL. The collaboration project is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

Acknowledgments:

We would like to acknowledge the support provided by Eline Verhaak for this research and thank her for helping us during the retrospective collection of clinical data from the Gamma Knife Center of the Elisabeth-TweeSteden Hospital at Tilburg, The Netherlands.

Consent for publication:

Not applicable.

Contributions:

All authors contributed to this research.

Competing interests

All authors declare that they have no competing interests.

Ethics approval

This study is part of the AI in Medical Imaging for novel Cancer User Support (AMICUS) project at Tilburg University. This project is approved by the Ethics Review Board at the Tilburg University.

Consent to participate:

The data did not contain any identifiable personal information, therefore the need for consent to participate was waived by the Institutional Review Board Elisabeth-TweeSteden Hospital (ETZ), Tilburg, The Netherlands (Study number : L1267.2021 - AMICUS).

List of abbreviations

SRS	Stereotactic Radiosurgery
LC	Local Control
ETZ	Elisabeth-TweeSteden Hospital
AUC	Area Under the receiver operating characteristic Curve
WBRT	Whole-Brain Radiation Therapy
SRT	hypo-fractionated Stereotactic Radiotherapy
LF	Local Failure
AI	Artificial Intelligence
GKRS	Gamma Knife Radiosurgery
GT	Ground Truth
MRI	Magnetic Resonance Imaging
KPS	Karnofsky Performance Status score
ROC	Receiver Operating Characteristic
RF	Random Forest model
GPU	Graphical Processing Unit
ReLU	Rectified Linear Unit

References

1. L. Nayak, E. Q. Lee, and P. Y. Wen, “Epidemiology of Brain Metastases,” Current Oncology Reports, vol. 14, no. 1, pp. 48–54, Oct. 2011, doi: https://doi.org/10.1007/s11912-011-0203-y.

2. R. A. Patchell, “The management of brain metastases,” *Cancer Treatment Reviews*, vol. 29, no. 6, pp. 533–540, Dec. 2003, doi: https://doi.org/10.1016/S0305-7372(03)00105-1.

3. K. A. Bradley and M. P. Mehta, “Management of brain metastases,” *Seminars in Oncology*, vol. 31, no. 5, pp. 693–701, Oct. 2004, doi: https://doi.org/10.1053/j.seminoncol.2004.07.012.

4. L. R. Rogers, “Neurologic Complications of Cancer, 2nd ed.,” *Neuro-Oncology*, vol. 11, no. 1, pp. 96–97, Feb. 2009, doi: https://doi.org/10.1215/15228517-2008-118.

5. M. Preusser *et al.*, “Brain metastases: pathobiology and emerging targeted therapies,” *Acta Neuropathologica*, vol. 123, no. 2, pp. 205–222, Jan. 2012, doi: https://doi.org/10.1007/s00401-011-0933-9.

6. C. M. Carapella, N. Gorgoglione, and P. A. Oppido, “The role of surgical resection in patients with brain metastases,” *Current Opinion in Oncology*, vol. 30, no. 6, pp. 390–395, Nov. 2018, doi: https://doi.org/10.1097/cco.0000000000000484.

7. P. D. Brown, M. S. Ahluwalia, O. H. Khan, A. L. Asher, J. S. Wefel, and V. Gondi, “Whole-Brain Radiotherapy for Brain Metastases: Evolution or Revolution?,” *Journal of Clinical Oncology*, vol. 36, no. 5, pp. 483–491, Feb. 2018, doi: https://doi.org/10.1200/JCO.2017.75.9589.

8. P. D. Brown *et al.*, “Effect of Radiosurgery Alone vs Radiosurgery With Whole Brain Radiation Therapy on Cognitive Function in Patients With 1 to 3 Brain Metastases: A Randomized Clinical Trial,” *JAMA*, vol. 316, no. 4, pp. 401–409, 2016, doi: https://doi.org/10.1001/jama.2016.9839.

9. Isabela Peña-Pino and C. C. Chen, “Stereotactic Radiosurgery as Treatment for Brain Metastases: An Update,” *Asian journal of neurosurgery*, vol. 18, no. 02, pp. 246–257, Jun. 2023, doi: https://doi.org/10.1055/s-0043-1769754.

10. P. H. Ko, H. J. Kim, J. S. Lee, and W. C. Kim, “Tumor volume and sphericity as predictors of local control after stereotactic radiosurgery for limited number (1□4) brain metastases from nonsmall cell lung cancer,” *Asia-Pacific Journal of Clinical Oncology*, vol. 16, no. 3, pp. 165–171, Feb. 2020, doi: https://doi.org/10.1111/ajco.13309.

11. P. W. Sperduto *et al.*, “Estimating Survival in Patients With Lung Cancer and Brain Metastases,” *JAMA Oncology*, vol. 3, no. 6, pp. 827–827, Jun. 2017, doi: https://doi.org/10.1001/jamaoncol.2016.3834.

12. P. W. Sperduto *et al.*, “Estimating Survival in Melanoma Patients With Brain Metastases: An Update of the Graded Prognostic Assessment for Melanoma Using Molecular Markers (Melanoma-molGPA),” *International Journal of Radiation Oncology, Biology, Physics*, vol. 99, no. 4, pp. 812–816, Nov. 2017, doi: https://doi.org/10.1016/j.ijrobp.2017.06.2454.

13. Seyed Ali Jalalifar, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, “A Self-Attention-Guided 3D Deep Residual Network With Big Transfer to Predict Local Failure in Brain Metastasis After Radiotherapy Using Multi-Channel MRI,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 13–22, Jan. 2023, doi: https://doi.org/10.1109/jtehm.2022.3219625.

14. Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: https://doi.org/10.1038/nature14539.

15. S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, “A survey: Deep learning for hyperspectral image classification with few labeled samples,” *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021, doi: https://doi.org/10.1016/j.neucom.2021.03.035.

16. D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” *Evolutionary Intelligence*, vol. 15, Jan. 2021, doi: https://doi.org/10.1007/s12065-020-00540-3.

17. M. Wang, Q. Zhang, S. Lam, J. Cai, and R. Yang, “A Review on Application of Deep Learning Algorithms in External Beam Radiotherapy Automated Treatment Planning,” *Frontiers in Oncology*, vol. 10, Oct. 2020, doi: https://doi.org/10.3389/fonc.2020.580919.

18. D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: https://doi.org/10.1146/annurev-bioeng-071516-044442.

19. E. Wetzer, H. Harlin, J. Lindblad, and Natasa Sladoje, “When Texture Matters: Texture-Focused Cnns Outperform General Data Augmentation and Pretraining in Oral Cancer Detection,” *IEEE*, Apr. 2020, doi: https://doi.org/10.1109/isbi45749.2020.9098424.

20. P. Afshar, A. Mohammadi, K. N. Plataniotis, A. Oikonomou, and H. Benali, “From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities,” *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 132–160, Jul. 2019, doi: https://doi.org/10.1109/msp.2019.2900993.

21. Seyed Ali Jalalifar, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, “Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRI based deep learning features,” *Medical Physics*, vol. 49, no. 11, pp. 7167–7178, Jul. 2022, doi: https://doi.org/10.1002/mp.15814.

22. J. P. B. O'Connor *et al.*, "Imaging biomarker roadmap for cancer studies," *Nature Reviews Clinical Oncology*, vol. 14, no. 3, pp. 169–186, Mar. 2017, doi: https://doi.org/10.1038/nrclinonc.2016.162.

23. S. Rizzo *et al.*, "Radiomics: the facts and the challenges of image analysis," *European Radiology Experimental*, vol. 2, no. 1, Nov. 2018, doi: https://doi.org/10.1186/s41747-018-0068-z.

24. S. Bae *et al.*, "Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction," *Radiology*, vol. 289, no. 3, pp. 797–806, Dec. 2018, doi: https://doi.org/10.1148/radiol.2018180200.

25. E. Karami, M. Ruschin, H. Soliman, A. Sahgal, G. J. Stanisz, and A. Sadeghi-Naini, "An MR Radiomics Framework for Predicting the Outcome of Stereotactic Radiation Therapy in Brain Metastasis," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2019, pp. 1022–1025, Jul. 2019, doi: https://doi.org/10.1109/EMBC.2019.8856558.

26. C.-Y. Liao *et al.*, "Enhancement of Radiosurgical Treatment Outcome Prediction Using MRI Radiomics in Patients with Non-Small Cell Lung Cancer Brain Metastases," *Cancers*, vol. 13, no. 16, p. 4030, Aug. 2021, doi: https://doi.org/10.3390/cancers13164030.

27. A.Mouraviev *et al.*, "Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery," *Neuro-Oncology*, Jan. 2020, doi: https://doi.org/10.1093/neuonc/noaa007.

28. D. Kawahara, X. Tang, C. K. Lee, Y. Nagata, and Y. Watanabe, "Predicting the Local Response of Metastatic Brain Tumor to Gamma Knife Radiosurgery by Radiomics With a Machine Learning Method," *Frontiers in Oncology*, vol. 10, p. 569461, 2020, doi: https://doi.org/10.3389/fonc.2020.569461.

29. K. Gorgolewski *et al.*, "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python," *Frontiers in Neuroinformatics*, vol. 5, 2011, doi: https://doi.org/10.3389/fninf.2011.00013.

30. C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Frontiers in Energy Research*, vol. 9, Mar. 2021, doi: https://doi.org/10.3389/fenrg.2021.652801.

31. R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional Neural networks: an Overview and Application in Radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: https://doi.org/10.1007/s13244-018-0639-9.

32. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Jun. 2016, doi: https://doi.org/10.1109/cvpr.2016.90.

33. O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Apr. 2015, doi: https://doi.org/10.1007/s11263-015-0816-y.

34. S. A. Keek *et al.*, “Predicting Adverse Radiation Effects in Brain Tumors After Stereotactic Radiotherapy With Deep Learning and Handcrafted Radiomics,” *Frontiers in oncology*, vol. 12, Jul. 2022, doi: https://doi.org/10.3389/fonc.2022.920393.

35. K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing.*, pp. 630–645, 2016, doi: https://doi.org/10.1007/978-3-319-46493-0_38.

36. R. Riad, Olivier Teboul, D. Grangier, and N. Zeghidour, “Learning strides in convolutional neural networks,” *arXiv (Cornell University)*, Feb. 2022, doi: https://doi.org/10.48550/arxiv.2202.01653.

37. K. Hara, D. Saito, and H. Shouno, “Analysis of function of rectified linear unit used in deep learning,” *2015 International Joint Conference on Neural Networks (IJCNN). IEEE*, pp. 1–8, Jul. 2015, doi: https://doi.org/10.1109/ijcnn.2015.7280578.

38. A.Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2012, doi: https://doi.org/10.1145/3065386.

39. R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, “Selecting critical features for data classification based on machine learning methods,” *Journal of Big Data*, vol. 7, no. 1, Jul. 2020, doi: https://doi.org/10.1186/s40537-020-00327-4.

Appendix

Full list of Radiomics features:

original_shape_Elongation
original_shape_Flatness
original_shape_LeastAxisLength
original_shape_MajorAxisLength
original_shape_Maximum2DDiameterColumn
original_shape_Maximum2DDiameterRow
original_shape_Maximum2DDiameterSlice
original_shape_Maximum3DDiameter
original_shape_MeshVolume
original_shape_MinorAxisLength
original_shape_Sphericity
original_shape_SurfaceArea
original_shape_SurfaceVolumeRatio
original_shape_VoxelVolume
original_firstorder_10Percentile
original_firstorder_90Percentile
original_firstorder_Energy
original_firstorder_Entropy
original_firstorder_InterquartileRange
original_firstorder_Kurtosis
original_firstorder_Maximum
original_firstorder_MeanAbsoluteDeviation
original_firstorder_Mean
original_firstorder_Median
original_firstorder_Minimum
original_firstorder_Range
original_firstorder_RobustMeanAbsoluteDeviation
original_firstorder_RootMeanSquared
original_firstorder_Skewness
original_firstorder_TotalEnergy
original_firstorder_Uniformity
original_firstorder_Variance
original_glcm_Autocorrelation
original_glcm_ClusterProminence
original_glcm_ClusterShade
original_glcm_ClusterTendency
original_glcm_Contrast
original_glcm_Correlation
original_glcm_DifferenceAverage
original_glcm_DifferenceEntropy
original_glcm_DifferenceVariance
original_glcm_Id
original_glcm_Idm
original_glcm_Idmn
original_glcm_Idn

original_glcm_Imc1
original_glcm_Imc2
original_glcm_InverseVariance
original_glcm_JointAverage
original_glcm_JointEnergy
original_glcm_JointEntropy
original_glcm_MCC
original_glcm_MaximumProbability
original_glcm_SumAverage
original_glcm_SumEntropy
original_glcm_SumSquares
original_gldm_DependenceEntropy
original_gldm_DependenceNonUniformity
original_gldm_DependenceNonUniformityNormalized
original_gldm_DependenceVariance
original_gldm_GrayLevelNonUniformity
original_gldm_GrayLevelVariance
original_gldm_HighGrayLevelEmphasis
original_gldm_LargeDependenceEmphasis
original_gldm_LargeDependenceHighGrayLevelEmphasis
original_gldm_LargeDependenceLowGrayLevelEmphasis
original_gldm_LowGrayLevelEmphasis
original_gldm_SmallDependenceEmphasis
original_gldm_SmallDependenceHighGrayLevelEmphasis
original_gldm_SmallDependenceLowGrayLevelEmphasis
original_glrlm_GrayLevelNonUniformity
original_glrlm_GrayLevelNonUniformityNormalized
original_glrlm_GrayLevelVariance
original_glrlm_HighGrayLevelRunEmphasis
original_glrlm_LongRunEmphasis
original_glrlm_LongRunHighGrayLevelEmphasis
original_glrlm_LongRunLowGrayLevelEmphasis
original_glrlm_LowGrayLevelRunEmphasis
original_glrlm_RunEntropy
original_glrlm_RunLengthNonUniformity
original_glrlm_RunLengthNonUniformityNormalized
original_glrlm_RunPercentage
original_glrlm_RunVariance
original_glrlm_ShortRunEmphasis
original_glrlm_ShortRunHighGrayLevelEmphasis
original_glrlm_ShortRunLowGrayLevelEmphasis
original_glszm_GrayLevelNonUniformity
original_glszm_GrayLevelNonUniformityNormalized
original_glszm_GrayLevelVariance
original_glszm_HighGrayLevelZoneEmphasis
original_glszm_LargeAreaEmphasis
original_glszm_LargeAreaHighGrayLevelEmphasis
original_glszm_LargeAreaLowGrayLevelEmphasis
original_glszm_LowGrayLevelZoneEmphasis

original_glszm_SizeZoneNonUniformity
original_glszm_SizeZoneNonUniformityNormalized
original_glszm_SmallAreaEmphasis
original_glszm_SmallAreaHighGrayLevelEmphasis
original_glszm_SmallAreaLowGrayLevelEmphasis
original_glszm_ZoneEntropy
original_glszm_ZonePercentage
original_glszm_ZoneVariance
original_ngtdm_Busyness
original_ngtdm_Coarseness
original_ngtdm_Complexity
original_ngtdm_Contrast
original_ngtdm_Strength

List of excluded features:
original_firstorder_Uniformity
original_glcm_Id
original_glcm_Idm
original_glcm_Idmn
original_glcm_Idn
original_glcm_Imc2
original_glcm_InverseVariance
original_glcm_JointEnergy
original_glcm_MCC
original_glcm_MaximumProbability
original_gldm_LowGrayLevelEmphasis
original_gldm_SmallDependenceLowGrayLevelEmphasis
original_glrlm_GrayLevelNonUniformityNormalized
original_glrlm_LowGrayLevelRunEmphasis
original_glrlm_RunLengthNonUniformityNormalized
original_glrlm_RunPercentage
original_glrlm_ShortRunEmphasis
original_glrlm_ShortRunLowGrayLevelEmphasis
original_glszm_GrayLevelNonUniformityNormalized
original_glszm_LowGrayLevelZoneEmphasis
original_glszm_SmallAreaEmphasis
original_glszm_SmallAreaLowGrayLevelEmphasis
original_ngtdm_Coarseness
DL_5
DL_13
DL_31
DL_34
DL_43
DL_48

Variable importance:

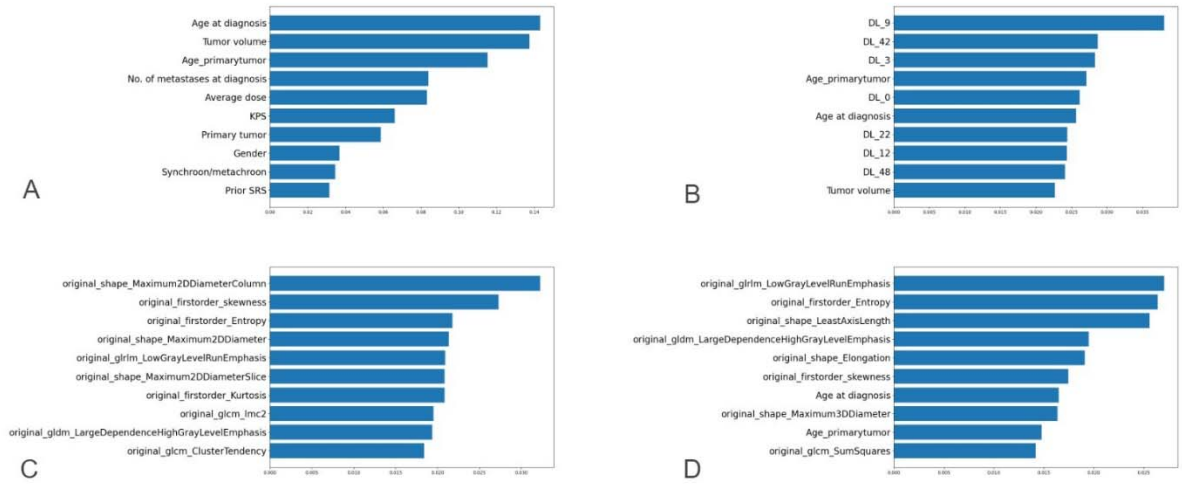

Figure 4. Variable importance for LC in decreasing order of significance for top 10 factors. **A** - model trained with clinical factors only. **B** - model trained with clinical and deep learning factors. **C**- model trained with clinical and radiomics features. **D** - model trained with the combination of clinical, radiomics and deep learning features.

ROC Curves:

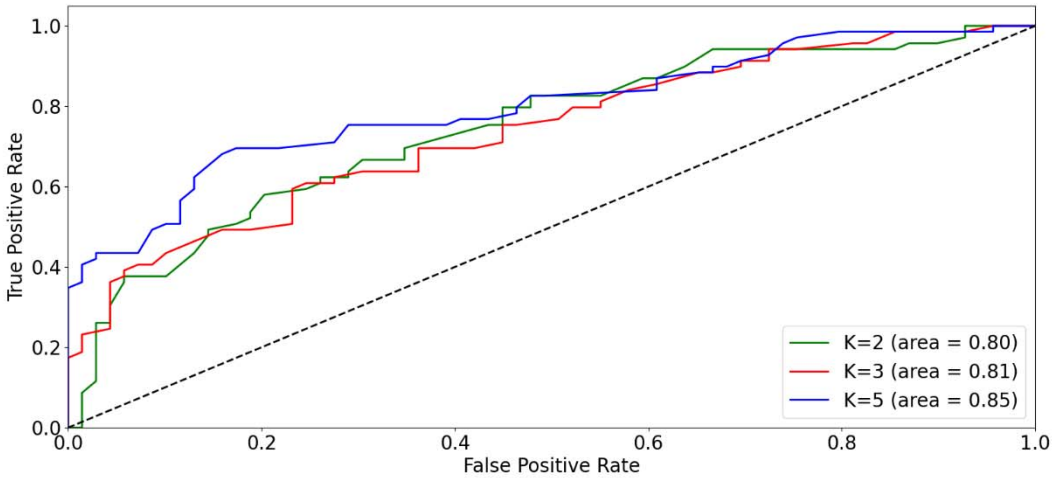

Figure 5: ROC curve of model trained with clinical features only.

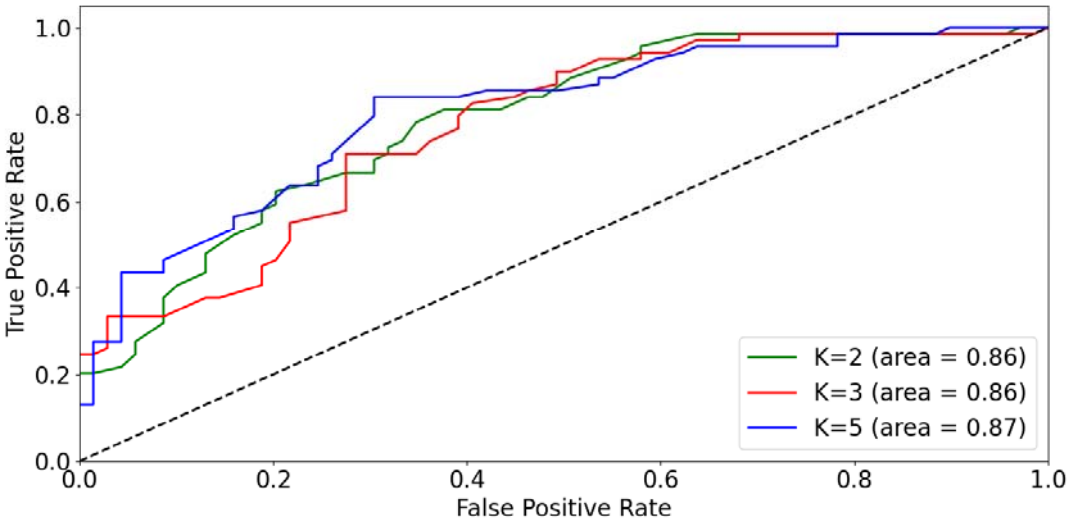

Figure 6: ROC curve of model trained with clinical and deep learning features.

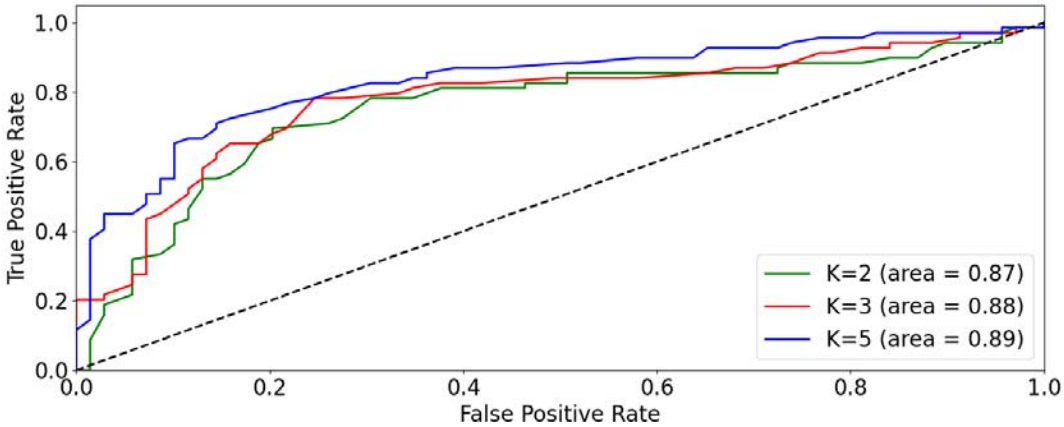

Figure 7: ROC curve of model trained with clinical and radiomics features.