Supplementary Material

CT data acquisition

All the CT examinations were performed using GE Discovery CT (GE Healthcare, Waukesha, Wisconsin). Thin-slice contrast-enhanced CT images were collected from all patients. Scanning parameters were as follows: tube voltage, 120 kV, with automatic tube current modulation activated; collimation, GE Discovery CT 64 × 0.6 mm; section thickness, 0.625–1 mm; section intervals, 0.625–1 mm.

Details for regions of interest (ROIs) segmentation and preprocessing

For patients absent of MYCN amplification with more than one primary lesions in CT images, the largest lesion was depicted with consensus of two radiologists. For patients present of MYCN amplification with more than one primary lesions in CT images, the lesion found with MYCN amplification was selected for ROI delineation. With the masked delineation of ROIs for primary tumor, we adjust the size of consecutive axial slices of primary lesions into 224 mm * 224 mm for the input layer of the pre-trained CNNs models using a bounding box covering the whole tumor area. The resized images with one axial slice of the maximum area as image channels would be candidate of the CNNs for feature extraction of 2-dimensonal (2D) DL features. And the resized images of the whole tumor volume as image channels would be candidate of the pre-trained DL algorithms for feature extraction of 3-dimensional (3D) DL features.

Deep learning (DL) features

CNNs architecture

Six pretrained CNN algorithms were applied on contrast-enhanced CT sequences respectively for the extraction of 2D and 3D DL-based features with representativeness, including Xception, VGG16, VGG19, ResNet50, InceptionV3, and InceptionResNetV2. These six CNNs were commonly used and pre-trained by the large-scale and well-annotated ImageNet database. This published research released dataset containing enormous object categories and manually annotated training images, the optimization hyperparameters of which was not tuned permitting a broader generalization on other datasets. After preprocessing, prepared slices of CT images with the maximum axial area or the whole tumor volume of primary lesions would be ready as the input of the pre-trained CNNs to generate DL features. The models are publicly accessed by Keras and TensorFlow open-source code (https://github.com/fchollet/deep-learning-models/).

Elimination of the last fully-connected layer

The convolutional base is connected by a fully-connected layer for the pre-trained models. We removed the last fully-connected layer, and then different CNNs reached various numbers of feature maps (2048 for ResNet50, InceptionV3 and Xception, 512 for VGG16 and VGG19, and 1536 for InceptionResNetV2) from the new output of these models.

Addition of max pooling layer and feature extraction

With the utility of a global pooling window for flatten of feature maps, local data of feature maps would be concentrated into 1D numeric features with a decreased dimensionality. After Step 2.2, for models with more than one-dimensional features, we got feature maps with height and width dimensions corresponding to location invariance in the input layer. After global max pooling, each feature map vector was transformed to a maximal raw value among them. During this step, the feature maps were transformed to numeric values, as representativeness of DL-based features.

Statistical analysis

Using the “survival” R package, logistic regression was used to identify associations with MYCN status, and Cox proportional hazard (CPH) analysis was used to find association between variables and survival outcome. The KM analyses with log-rank tests were used to compare survival differences between dichotomized categories (non-amplified and amplified) from nomogram model-predicted MYCN status or histopathological MYCN status in both training and testing sets by “survminer” R package. The integrated nomogram model and its calibration curves for prediction of MYCN status were depicted by “Hmisc” R package.

The results of MYCN amplification prediction performance in clinical predictors and DL-SVM models

Three significant clinical variables were selected as independent predictors for discrimination between amplified and non-amplified MYCN status in the training cohort, including histological differentiation, presence of infiltrating across midline, and calcification (Table.S1). Their AUC and corresponding 95% confidence interval (CI) were calculated with 0.433 (95%CI: 0.337-0.529) for histological differentiation, 0.581 (95%CI: 0.485-0.678) for presence of infiltrating across midline, and 0.505 (95%CI: 0.408-0.603) for presence of calcification in the testing cohort (Table.2).

As for the performance of DL-SVM models for differentiation between amplified and non-amplified MYCN, the AUCs ranged from 0.873 to 0.966 for the training group and from 0.553 to 0.819 for the testing group (Table.S2). A total of 18 DL-SVM models including 6 2D models, 6 3D models, and 6 combined models, were constructed based on selected features from different pre-trained DL-CNNs (Table.S3). The combined model with ResNet50 algorithm achieved the best discriminative performance considering the AUC value of the testing cohort among all DL-SVM models, with an AUC of 0.959, a C-index of 0.959, accuracy of 96.53%, sensitivity of 90.12%, and specificity of 97.34% in the training cohort, and an AUC of 0.819, a C-index of 0.819, accuracy of 74.68%, sensitivity of 74.36%, and specificity of 74.72% in the testing cohort (Table.S2). For ResNet50 CNN model, DL-based features extracted from different layer might exist difference since its complex DL-CNN architecture. We constructed DL-SVM models on DL features from earlier layers of ResNet50 algorithm to explore the superiority of DL features from the last layer before the fully connected layer, Res5c in classification of amplified and non-amplified MYCN (Table.S4). The results demonstrated the advantages of current extraction strategy from the layer of Res5c in ResNet50 algorithm. ResNet50 architecture was applied to generate feature maps highlighting important subregions for generation of the output (Figure.2). As shown in the feature heatmaps, the valuable areas were different in different MYCN status for extraction of specific feature patterns, which were interpreted for further clinical view.

The results of EFS prediction performance in clinical variables

According to the results of multivariate Cox regression, age at diagnosis, INRGSS stage, histological differentiation, presence of infiltrating across midline, presence of MYCN amplification, and nomogram model-predicted probability were considered as independent predictors for EFS (Table.S5). The hazard ratio (HR) and corresponding 95% CT were calculated with 1.011 (95%CI: 1.002-1.021) for age at diagnosis, 2.674 (95%CI: 2.117-3.377) for INRGSS stage, 1.319 (95%CI: 1.046-1.665) for histological differentiation, and 1.541 (95%CI: 1.210-1.963) for presence of Infiltrating across midline (Table.S5). The KM survival analysis for the histopathological non-amplified and amplified MYCN groups showed the mean survival of 83.476 months (95%CI: 79.401-87.550 months) versus 63.938 months (52.568-75.308 months) in the training cohort, and the mean survival of 85.458 months (95%CI: 79.308-91.609 months) versus 82.538 months (66.945-98.130 months) in the testing cohort, with the p value of 0.031 in the log-rank test (Table.3).

Supplementary Table **1 Univariate and multivariate logistic analysis on clinical factors and the DL signature associated with MYCN amplifiaction in the training cohort.**

Characteristic	**Univariate analysis**		**Multivariate analysis**	
	HR (95% CI）	***P* value**	**HR (95% CI)**	***P* value**
Age	0.989 (0.973-1.006)	0.201		
Gender	0.992 (0.562-1.749)	0.977		
INRGSS stage	1.263 (0.795-2.007)	0.324		
Histological differentiation	2.329 (1.460-3.715)	<0.001	2.276 (1.422-3.642)	0.001
Infiltrating across midline	1.662 (1.021-2.706)	0.041	1.638 (1.000-2.684)	0.050
Calcification	4.097 (1.206-13.925)	0.024	3.812 (1.093-13.294)	0.036
DL-based signature	354.096 (64.388-1947.329)	<0.001	294.186 (50.422-1716.428)	<0.001

Abbreviations: CI, Confidence interval; INRGSS, the International Neuroblastoma Risk Grouping Staging System; DL, Deep learning.

Supplementary Table **2 Predictive performance of DL-based models in identification of MYCN amplification of neuroblastomas in the training and testing cohorts.**

Model	**Method**	**Training cohort**				**Testing cohort**			
		AUC	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**
Xception	2D	0.932	94.86	83.95	96.24	0.572	67.21	38.46	71.38
	3D	0.943	95.97	86.42	97.18	0.637	69.48	56.41	71.38
	2D+3D	0.954	95.97	88.89	96.87	0.672	70.13	56.41	72.12
VGG16	2D	0.873	93.75	72.84	96.40	0.696	75.97	58.97	78.44
	3D	0.924	95.28	85.19	96.56	0.685	74.68	61.54	76.58
	2D+3D	0.923	97.36	85.19	98.90	0.699	66.23	66.67	66.17
VGG19	2D	0.912	94.44	80.25	96.24	0.652	70.13	56.41	72.12
	3D	0.898	94.03	75.31	96.40	0.700	74.35	61.54	76.21
	2D+3D	0.915	96.67	81.48	98.59	0.759	70.78	74.36	70.26
ResNet50	2D	0.940	94.31	86.42	95.31	0.613	71.75	51.28	74.72
	3D	0.964	97.08	91.36	97.81	0.813	71.10	79.49	69.89
	2D+3D	0.959	96.53	90.12	97.34	0.819	74.68	74.36	74.72
InceptionV3	2D	0.944	95.28	86.42	96.40	0.649	71.43	51.28	74.35
	3D	0.951	96.53	90.12	97.34	0.561	71.43	41.03	75.84
	2D+3D	0.961	98.61	92.59	99.37	0.621	64.29	53.85	65.80
InceptionResNetV2	2D	0.949	96.25	88.89	97.18	0.627	63.64	56.41	64.68
	3D	0.936	96.39	83.95	97.97	0.553	71.43	38.46	76.21
	2D+3D	0.966	98.19	92.59	98.90	0.631	67.21	56.41	68.77

Abbreviations: DL, Deep learning; AUC, area under the receiver operating characteristic curve.

Supplementary Table **3 Number of features selected and used in construction of DL-based models for identification of MYCN amplification in neuroblastomas.**

Feature extractor	Method	Number of selected features
	2D	10
Xception	3D	10
	2D+3D	11
	2D	10
VGG16	3D	10
	2D+3D	9
	2D	10
VGG19	3D	9
	2D+3D	8
	2D	9
ResNet50	3D	10
	2D+3D	10
	2D	8
InceptionV3	3D	8
	2D+3D	9
	2D	10
InceptionResNetV2	3D	9
	2D+3D	11

Abbreviations: DL, Deep learning.

Supplementary Table **4 Predictive performances of the DL-based models constructed by features extracted from different layers of Resnet50 for identification of MYCN amplification in neuroblastomas.**

Layer	**Method**	**Training cohort**				**Testing cohort**			
		AUC	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**
Res2b	2D	0.867	88.89	70.37	91.24	0.557	72.73	38.46	77.70
	3D	0.894	92.78	75.31	94.99	0.586	69.48	46.15	72.86
	2D+3D	0.929	93.75	82.72	95.15	0.613	71.75	53.85	74.35
Res3d	2D	0.920	94.72	82.72	96.24	0.556	67.53	41.03	71.38
	3D	0.954	97.64	88.89	98.75	0.654	73.05	48.72	76.58
	2D+3D	0.934	94.03	86.42	94.99	0.686	70.45	53.85	72.86
Res4f	2D	0.968	97.36	91.36	98.12	0.669	68.51	61.54	69.52
	3D	0.979	96.39	95.06	96.56	0.649	67.86	53.85	69.89
	2D+3D	0.968	95.97	92.59	96.40	0.715	69.16	58.97	70.63
Res5c	**2D**	**0.940**	**94.31**	**86.42**	**95.31**	**0.613**	**71.75**	**51.28**	**74.72**
	3D	**0.964**	**97.08**	**91.36**	**97.81**	**0.813**	**71.10**	**79.49**	**69.89**
	2D+3D	**0.959**	**96.53**	**90.12**	**97.34**	**0.819**	**74.68**	**74.36**	**74.72**
FC1000	2D	0.917	92.36	79.01	94.05	0.538	67.86	41.03	71.75
	3D	0.926	92.08	80.25	93.58	0.595	71.10	35.90	76.21
	2D+3D	0.880	85.14	71.60	86.85	0.706	76.62	46.15	81.04

Abbreviations: DL, Deep learning; AUC, Area under the receiver operating characteristic curve.

Supplementary Table **5 Results of multivariable Cox regression analysis for prediction of EFS in neuroblastomas**

Variable	***p* value**	**Coefficient**	**SE**	**Wald x2**	**HR**	**95%CI**
Age	0.013	0.011	0.005	6.156	1.011	1.002-1.021
Gender	0.462	0.106	0.144	0.542	1.111	0.839-1.473
INRGSS stage	<0.001	0.984	0.119	68.232	2.674	2.117-3.377
Histological differentiation	0.019	0.277	0.119	5.473	1.319	1.046-1.665
Infiltrating across midline	<0.001	0.433	0.123	12.311	1.541	1.210-1.963
Calcification	0.058	0.681	0.359	3.596	1.975	0.977-3.992
MYCN amplification	0.009	0.438	0.167	6.878	1.550	1.117-2.151
Nomogram model-predicted probability	0.003	0.023	0.008	8.693	1.023	1.008-1.039

Abbreviations: EFS, Event-free survival; INRGSS, the International Neuroblastoma Risk Grouping Staging System; SE, Standard error; HR, Hazard ratio; CI, Confidence interval.