UDC 004.855.5(045)
DOI:10.18372/1990-5548.88.20980

[1]Victor Sineglazov,
[2]Maksym Shevchenko

IMPACT OF AUTOMATED SEGMENTATION ON RADIOMICS-BASED GROWTH PREDICTION OF VESTIBULAR SCHWANNOMA

[1]Department of Artificial Intelligence, Institute of Applied Systems Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
[2]Department of Avionics and Control Systems, Faculty of Air Navigation Electronics and Telecommunications, State University "Kyiv Aviation Institute", Kyiv, Ukraine
E-mails: [1]svm@kai.edu.ua ORCID 0000-0002-3297-9060,
[2]maksymshevchenko01@gmail.com ORCID 0009-0004-0540-8213

Abstract*—This paper investigates the impact of nnU-Net-based automatic segmentation on the reproducibility of radiomics features and the quality of vestibular schwannoma growth prediction. The nnU-Net model was trained on 317 T1C image pairs with masks from 155 patients from the Vestibular-Schwannoma-MC-RC2 dataset using 5-fold cross-validation, achieving a mean Dice coefficient of 0.862. ICC analysis showed that 88.8% of wavelet features maintain good or excellent agreement (ICC ≥ 0.75) between manual and automatic masks. Comparison on a subset of 96 patients with growth labels showed: the previously published pipeline (Wavelet + Voting ensemble) with manual masks achieves ROC AUC = 0.742, with automatic masks - 0.639 (−14.0%). Pipeline optimization (ICC filtering, ensemble adaptation to LR + LDA) improves the result to 0.687 (−7.4%). The results determine the cost of automating the radiomics pipeline and provide recommendations for ICC filtering for clinical implementation.*

Keywords—Vestibular schwannoma, automatic segmentation, nnU-Net, radiomics features, reproducibility, ICC, tumor growth prediction, machine learning.

I. INTRODUCTION

Vestibular schwannoma is a benign tumor of the vestibulocochlear nerve with an unpredictable course: some neoplasms remain stable for years, while others demonstrate aggressive growth requiring surgical intervention [1]. The current "wait-and-scan" strategy involves regular MRI examinations over months or years to detect growth dynamics [2], which is a resource-intensive approach, especially considering the increasing incidence of vestibular schwannoma [3].

Radiomics analysis enables quantitative characterization of the internal structure of a tumor through hundreds of texture features that reflect the microstructural heterogeneity of the tissue [4], [5]. In our previous work [6], a method for predicting vestibular schwannoma growth based on a single MRI scan was proposed using wavelet features and a Voting ensemble classifier, achieving ROC AUC = 0.742 ± 0.072. However, practical implementation of this approach requires manual tumor segmentation by a radiologist, which is a time-consuming process and introduces subjective inter-observer variability [7].

Automatic segmentation based on deep learning, particularly the nnU-Net architecture [8], achieves high accuracy for vestibular schwannomas (Dice 0.85–0.93) [9], [10] and can eliminate dependence on manual delineation. However, the impact of automatic segmentation on the reproducibility of radiomics features and the final quality of growth prediction remains unclear. Existing studies evaluate either segmentation quality (Dice) or feature stability (ICC), but not the complete chain from automatic segmentation to clinical outcome prediction [9], [11].

The aim of this work is to evaluate the impact of automatic segmentation on the reproducibility of radiomics features and the quality of vestibular schwannoma growth prediction. The study compares the results of the published radiomics pipeline [6] using manual and automatic segmentation masks, which is a critical question for clinical implementation of automated prediction systems.

II. LITERATURE REVIEW

Research in this area can be divided into three directions: automatic segmentation of vestibular schwannomas, radiomics analysis for predicting their growth, and the impact of the segmentation method on the reproducibility of radiomics features.

A. Automatic segmentation of vestibular schwannomas

Deep learning has become the standard approach for brain tumor segmentation. A systematic review and meta-analysis of 41 studies involving 8028 vestibular schwannoma cases [12] showed that the

best models achieve a pooled Dice coefficient of 0.89 (95% CI: 0.88-0.91). The nnU-Net architecture [8] is the gold standard due to its self-configuration for a specific dataset. The CrossMoDA challenge [9] stimulated the development of domain adaptation methods for schwannoma segmentation between T1-weighted and T2-weighted images. A recent clinical implementation study [11] confirmed that automatic nnU-Net segmentations have clinimetric reliability comparable to the results of experienced radiologists.

B. Radiomics analysis of vestibular schwannomas

Itoyama et al. [13] investigated risk factors for rapid vestibular schwannoma growth based on radiomics features of T1-weighted images, achieving AUC = 0.69. Yang et al. [14] developed a machine learning model for predicting pseudoprogression after radiosurgery, achieving accuracy of 88.4%. Wang et al. [15] proposed a multi-task deep learning model for simultaneous growth prediction and schwannoma segmentation, achieving AUC = 0.77 for the deep learning model and AUC = 0.70 for the radiomics approach. A recent review [16] systematized the application of radiomics for vestibular schwannomas, noting the promise of the approach but the limitations of existing studies with small samples and lack of external validation. In our previous work [6], AUC = 0.742 was achieved using wavelet decomposition and a Voting ensemble of five classifiers on a sample of 96 patients.

C. *Impact of segmentation on radiomics features*

The stability of radiomics features depends on the segmentation method – this is well documented for various tumor types. Parmar et al. [17] showed that segmentation variability is the main source of instability of radiomics features in lung tumors. For quantitative assessment of stability, the intraclass correlation coefficient (ICC) is used: features with ICC ≥ 0.75 are considered stable, with ICC ≥ 0.90 – excellent [18]. A study of feature reproducibility under segmentation variability on CT images [19] showed that texture features are less robust than shape and first-order features. For vestibular schwannomas, Cornelissen et al. [11] confirmed the clinimetric reliability of nnU-Net automatic segmentation, comparable to the results of experienced radiologists. A recent study of radiomics reproducibility with automatic segmentation of coronary arteries [20] confirmed the necessity of ICC filtering for selecting stable features.

However, none of the existing studies evaluated the complete chain: automatic segmentation – ICC analysis – growth prediction. It remains unknown how the decrease in segmentation quality affects not only individual features, but the final clinical outcome – the accuracy of tumor growth prediction.

III. PROBLEM STATEMENT

The task is to evaluate the impact of automatic segmentation on the quality of radiomics-based vestibular schwannoma growth prediction. It is necessary to compare the results of the published pipeline [6] using manual and automatic segmentation masks.

A. Dataset

The public Vestibular-Schwannoma-MC-RC2 dataset from The Cancer Imaging Archive (TCIA) [21] was used. The dataset contains longitudinal MRI scans of patients with unilateral vestibular schwannoma from King's College Hospital, London, UK.

For training the segmentation model, 317 pairs of T1C images with corresponding manual segmentation masks from 155 patients were used. For evaluating growth prediction, a subset of 96 patients who had at least two consecutive MRI scans (T1 with contrast) with an interval of at least 6 months was used.

B. Growth labels

Tumor volume was calculated using the formula:

$$V = N_{\text{voxel}} \times V_{\text{voxel}},$$

where N_{voxel} is the number of voxels in the segmentation mask; V_{voxel} is the volume of a single voxel (mm³).

Relative volume change is determined as:

$$\Delta V = \frac{V_2 - V_1}{V_1} \times 100\%,$$

where V_1 is the volume at the first scan; V_2 is the volume at the second scan.

Class label is assigned according to the rule:

$$y = 1, \text{ if } \Delta V > 10\%, \quad y = 0, \text{ otherwise.}$$

The 10% threshold was chosen based on clinical practice. Class distribution: 40 patients (41.7%) with growth, 56 patients (58.3%) without growth.

C. Published pipeline (baseline)

In the previous work [6], the following pipeline for growth prediction from a single MRI scan was proposed:

- Extraction of 744 wavelet features (8 components × 93 texture features) using PyRadiomics [22];

- Selection of k = 5 most informative features using SelectKBest with ANOVA F-score criterion [23];
- Classification by Voting ensemble (soft) of five classifiers: SVM, logistic regression, k-NN, Random Forest, LDA.

This pipeline achieved ROC AUC = 0.742 ± 0.072 with 5-fold stratified cross-validation.

D. Evaluation metrics

Three levels of metrics were used for comprehensive evaluation:

1) Segmentation quality: Dice coefficient (DSC), 95th percentile Hausdorff distance (HD95), relative volume difference.

2) Feature stability: intraclass correlation coefficient ICC(3,1) between features from manual and automatic masks [18]. Features are classified as: excellent (ICC ≥ 0.90), good (0.75 ≤ ICC < 0.90), moderate (0.50 ≤ ICC < 0.75), poor (ICC < 0.50).

3) Prediction quality: ROC AUC with 5-fold stratified cross-validation.

IV. Proposed Approach

The proposed approach consists of four stages: automatic segmentation, radiomics feature extraction, ICC stability analysis of features, and comparison of growth prediction.

A. Automatic segmentation

The nnU-Net v2 architecture [8] in the 3d_fullres configuration was used for automatic segmentation. The model was trained on 317 pairs of T1C images with corresponding manual segmentation masks from 155 patients. To prevent data leakage, 5-fold cross-validation was applied: data was divided into 5 folds by patients (scans of the same patient did not appear in different folds), the model was trained on 4 folds and auto-segmented the 5th fold. Thus, each patient was segmented by a model that had not seen them during training.

The network architecture is PlainConvUNet with 6 stages (32, 64, 128, 256, 320, 320 channels), Conv3d + InstanceNorm3d + LeakyReLU. Training was performed for 1000 epochs per fold with SGD optimizer and poly learning rate schedule. Training was conducted on an NVIDIA GTX 1080 GPU (8 GB VRAM) with adapted parameters: batch size = 1, patch size = 32 × 192 × 160.

B. Radiomics feature extraction

For each of the 317 T1C images, radiomics features were extracted twice: with the manual mask and with the automatic mask. The PyRadiomics library [22] was used with identical settings: binWidth = 25, intensity normalization enabled, mask correction enabled.

A total of 744 wavelet features were extracted: Coiflet-1 wavelet decomposition decomposes the image into 8 components (LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH), from each component 93 texture features were extracted (shape features were excluded due to potential data leakage).

One case (schwannoma_174) was excluded due to an empty automatic mask, leaving 316 cases for ICC analysis.

C. ICC stability analysis of features

For each of the 744 features, the intraclass correlation coefficient ICC(3,1) was calculated between values from manual and automatic masks on 316 cases. ICC(3,1) is a two-way mixed model with single measurement [18]:

$$\mathrm{ICC}(3,1) = \frac{\mathrm{MS_R} - \mathrm{MS_E}}{\mathrm{MS_R} + (k-1) \times \mathrm{MS_E}},$$

where $\mathrm{MS_R}$ is the mean square between rows (cases); $\mathrm{MS_E}$ is the residual mean square; k is the number of segmentation methods ($k = 2$).

Additionally, a stratified analysis was performed: ICC was calculated separately for three groups of cases by segmentation quality (Dice ≥ 0.85; Dice 0.70 – 0.85; Dice < 0.70).

D. Comparison of growth prediction

For 96 patients with growth labels, comparison was performed in three scenarios:

Scenario A (baseline): published pipeline [6] with manual masks – wavelet (744 features), SelectKBest (k = 5), Voting ensemble (SVM + LR + RF + kNN + LDA), soft voting.

Scenario B: the same pipeline with automatic masks – without any changes, for direct comparison with the published result.

Scenario C: optimized pipeline for automatic masks – additionally extracted LoG features (465) were combined with wavelet features (744), yielding a total of 1209 features, followed by ICC filtering (ICC ≥ 0.75), optimization of number of features and ensemble composition.

All scenarios were evaluated using 5-fold stratified cross-validation [24] (with randomized fold assignment and fixed random seed for reproducibility) with the ROC AUC metric.

Classifier hyperparameters for Scenarios *A* and *B*: SVM (C = 1, kernel = RBF), logistic regression (C = 1, solver = saga), Random Forest (n_estimators = 100, max_depth = 5), k-NN (k = 5, weights = distance), LDA (default).

For Scenario C, a systematic search for optimal parameters was performed: testing all 26 ensemble combinations of 5 classifiers, optimization of k from 3 to 25, comparison of ICC thresholds (0.50; 0.75; 0.90). The overview of the proposed approach is presented in Fig. 1.

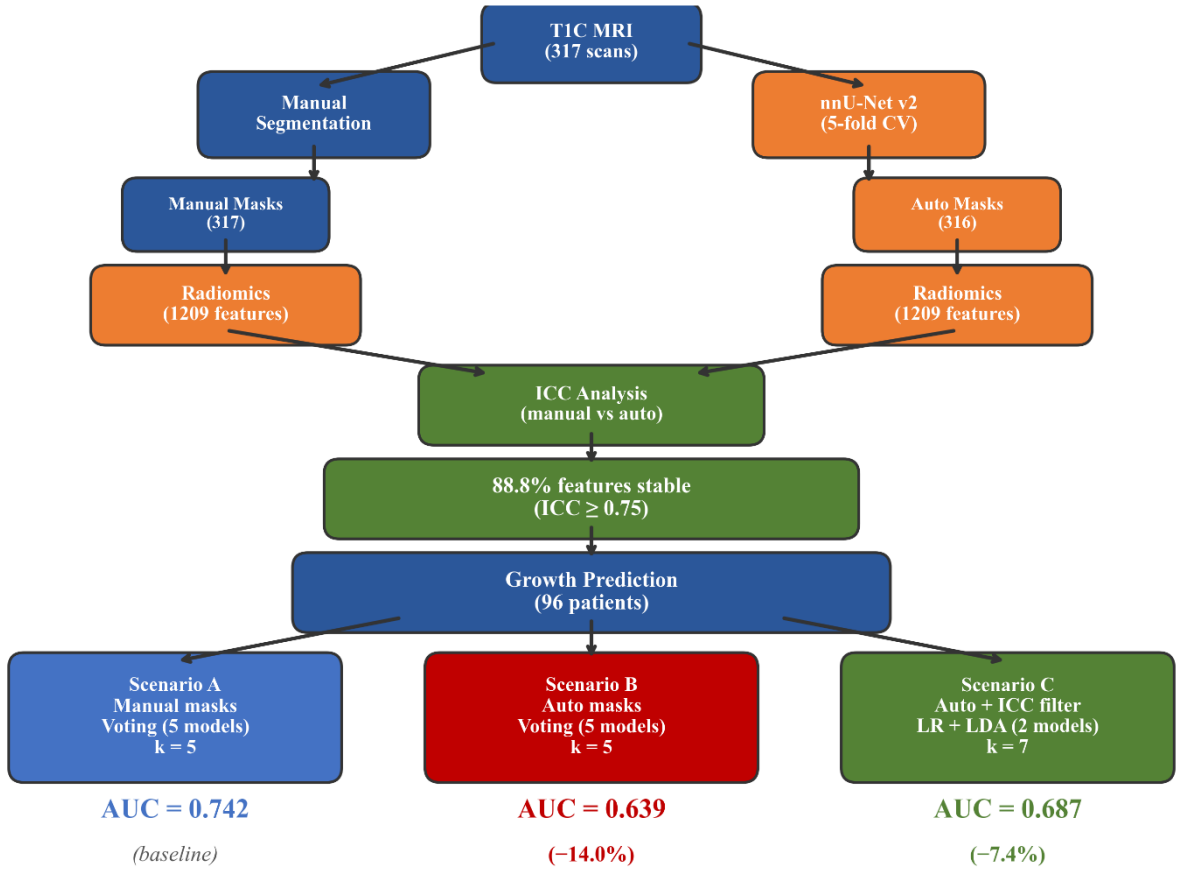

Fig. 1. Overview of the proposed approach

V. Results

The study was implemented in Python 3.10 using PyRadiomics 3.0 [22], scikit-learn 1.3, and nnU-Net v2.6.4 [8] libraries.

A. Automatic segmentation quality

The results of 5-fold cross-validation of nnU-Net on 317 cases are presented in Table I.

Summary segmentation metrics for all 317 cases are presented in Table II.

TABLE I. NNU-NET SEGMENTATION RESULTS (5-FOLD CV)

Fold	EMA Dice	Validation Dice
0	0.899	0.848
1	0.919	0.847
2	0.914	0.879
3	0.904	0.862
4	0.904	0.872
Mean	0.908	0.862

TABLE II. SUMMARY SEGMENTATION METRICS (317 CASES)

Metric	Mean	Median	Min	Max
Dice Score	0.861	0.921	0.000	0.990
HD95 (mm)	10.10	0.39	0.00	112.31
Volume Diff (%)	8.8	1.5	–	–

Distribution by quality: 232 cases (73.2%) with Dice ≥ 0.85; 48 cases (15.1%) with Dice 0.70–0.85; 37 cases (11.7%) with Dice < 0.70, including 3 cases with Dice = 0 (small tumors, model missed) and 1 case with an empty automatic mask.

Analysis by tumor size showed the expected relationship: small tumors (Q1) had a mean Dice of 0.763, while large tumors (Q4) had 0.900.

B. ICC analysis of feature stability

The results of ICC analysis for 744 wavelet features on 316 cases (excluding the empty mask case) are presented in Table III.

TABLE III. ICC DISTRIBUTION FOR WAVELET FEATURES (316 CASES)

ICC Category	Count	Proportion
Excellent (≥ 0.90)	520	69.9%
Good (0.75–0.90)	141	19.0%
Moderate (0.50–0.75)	55	7.4%
Poor (< 0.50)	28	3.8%
Stable (ICC ≥ 0.75)	661	88.8%

Mean ICC = 0.895, median ICC = 0.954. The vast majority of wavelet features (88.8%) maintain good or excellent agreement between manual and automatic masks. The distribution of ICC values is shown in Fig. 2.

Stratified analysis by segmentation quality (Table IV) revealed a clear dependence of feature stability on Dice.

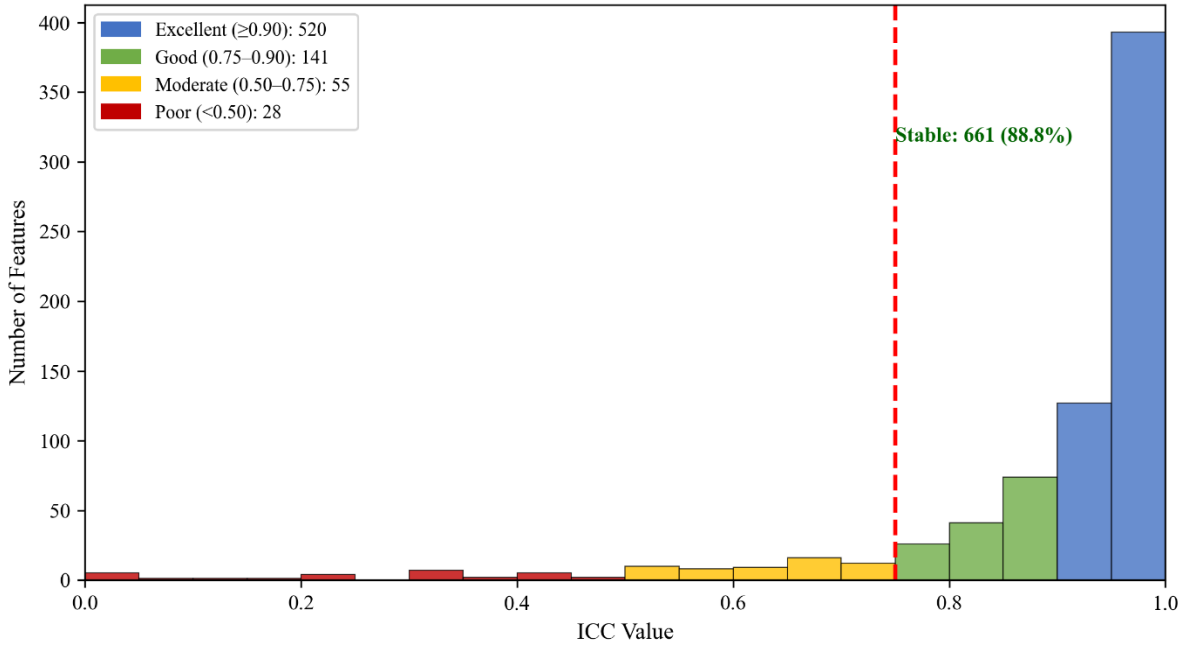

Fig. 2. Distribution of ICC values for 744 wavelet features (316 cases)

TABLE IV. STRATIFIED ICC ANALYSIS BY SEGMENTATION QUALITY

Group	Cases	ICC Median	ICC ≥ 0.75	ICC < 0.50
Dice ≥ 0.85	232	0.992	97.4%	0.9%
Dice 0.70–0.85	48	0.948	88.3%	2.9%
Dice < 0.70	36	0.807	57.0%	18.1%

With good segmentation (Dice ≥ 0.85), virtually all features are stable (97.4% with ICC ≥ 0.75). With poor segmentation (Dice < 0.70), nearly one-fifth of features (18.1%) become unstable. The stratified results are visualized in Fig. 3.

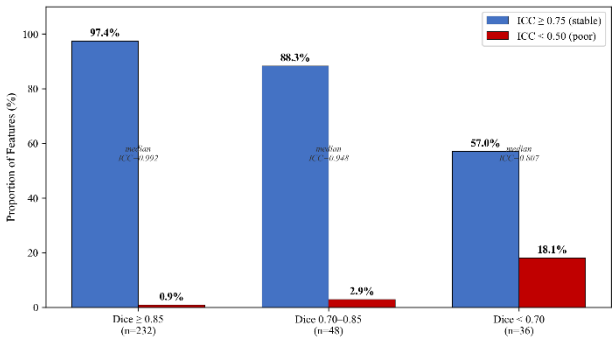

Fig. 3. Feature stability stratified by segmentation quality

C. *Comparison of growth prediction*

The results of three scenarios are presented in Table V.

TABLE V. COMPARISON OF GROWTH PREDICTION (5-FOLD STRATIFIED CV)

Scenario	Pipeline	ROC AUC	± std	Difference
A: Manual masks	Wavelet + Voting5 ($k = 5$)	0.742	0.072	baseline
B: Auto masks	Same pipeline	0.639	0.075	−14.0%
C: Auto masks optimized	ICC ≥ 0.75 + LR+LDA ($k = 7$)	0.687	–	−7.4%

Scenario *A* exactly reproduced the published result (AUC = 0.742). Direct replacement of manual masks with automatic ones (Scenario *B*) led to a decrease in AUC by 14.0%. Optimization of the pipeline for automatic features (Scenario *C*) partially compensated for the loss, reducing the difference to 7.4%. The comparison of all three scenarios is visualized in Fig. 4.

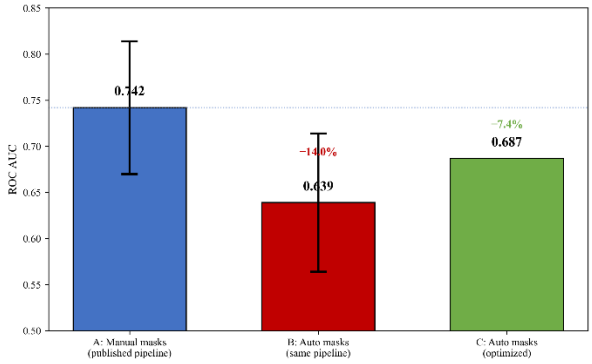

Fig. 4. Comparison of growth prediction quality across three scenarios

D. Optimization for automatic masks

A systematic search for optimal parameters for Scenario *C* included testing all 26 ensemble combinations of 5 classifiers. The top-5 results are presented in Table VI.

TABLE VI. TOP-5 ENSEMBLES FOR AUTOMATIC MASKS (ICC ≥ 0.75, $k = 7$)

#	Combination	ROC AUC
1	LR + LDA	**0.687**
2	SVM + LDA	0.679
3	SVM + LR + LDA	0.679
4	SVM + LR	0.672
5	SVM + RF + LDA	0.667

LDA is present in all top-5 combinations, indicating its key role for classification with automatic features. The optimal ensemble for automatic masks (LR + LDA, 2 models) differs from the published one (5 models) – suggesting that a more compact ensemble of linear classifiers may be more robust to the additional noise introduced by automatic segmentation.

E. Additional experiments

To confirm the robustness of the 0.687 result, three additional experiments were conducted:

1) Wavelet vs LoG separately: Wavelet features (AUC = 0.655) proved more robust to automatic segmentation than LoG features (AUC = 0.537). The combination (AUC = 0.677) was better than each type separately.

2) Combined ICC × F-score selection: Proper CV with feature selection inside each fold yielded AUC = 0.678, Nested CV – 0.677. The method did not improve the result compared to simple SelectKBest.

3) Hyperparameter optimization: Systematic search of classifier parameters yielded AUC = 0.686 – virtually no improvement (−0.1%).

These results confirm that AUC = 0.687 is a stable maximum for automatic masks with the current data.

VI. CONCLUSION

This work investigated the impact of nnU-Net-based automatic segmentation on the reproducibility of radiomics features and the quality of vestibular schwannoma growth prediction. The main results of the study:

1) Automatic nnU-Net segmentation achieves a mean Dice = 0.862 on 317 cases, which corresponds to the level of existing studies (0.85–0.93). At the same time, 73.2% of cases have Dice ≥ 0.85, and the main difficulties arise for small tumors.

2) The vast majority of wavelet features (88.8%) maintain good or excellent agreement (ICC ≥ 0.75) between manual and automatic masks. With high-quality segmentation (Dice ≥ 0.85), this indicator increases to 97.4%.

3) Direct replacement of manual masks with automatic ones in the published pipeline reduces ROC AUC from 0.742 to 0.639 (−14.0%). Pipeline optimization (addition of LoG features, ICC filtering, reduction of the ensemble to LR + LDA, $k = 7$) partially compensates for the loss, achieving AUC = 0.687 (−7.4%).

4) Wavelet features are more robust to automatic segmentation errors than LoG features (AUC 0.655 vs 0.537). LDA is a key classifier for working with automatic features.

The practical significance of this work lies in determining the cost of automating the radiomics pipeline: the transition from manual to automatic segmentation reduces prediction quality by 7–14%, which is acceptable for screening tasks. For clinical implementation, the following is recommended:

a) use ICC filtering for selecting stable features;

b) adapt the ensemble composition to the characteristics of automatic features;

c) with Dice ≥ 0.85, radiomics features can be used without additional filtering.

Limitations of the study include the relatively small sample size for growth prediction (96 patients), the use of a single dataset without external validation, and a single segmentation architecture (nnU-Net). Prospects for further research include validation on independent data, comparison of multiple segmentation architectures, and investigation of the

segmentation quality threshold below which radiomics analysis loses its prognostic value.

REFERENCES

[1] M. L. Carlson and M. J. Link, "Vestibular Schwannomas," *N Engl J Med*., 2021, 384(14):1335–1348. https://doi.org/10.1056/NEJMra2020394

[2] S. E. Stangerup, P. Caye-Thomasen, M. Tos, and J. Thomsen, "The natural history of vestibular schwannoma," *Otol Neurotol*, 2006, 27(4):547–552. https://doi.org/10.1097/01.mao.0000217356.73463.e7

[3] M. Reznitsky, M. M. B. S. Petersen, N. West, S. E. Stangerup, and P. Cayé-Thomasen, "Epidemiology of vestibular schwannomas – prospective 40-year data from an unselected national cohort," *Clin Epidemiol*, 2019, 11:981–986. https://doi.org/10.2147/CLEP.S218670

[4] P. Lambin, E. Rios-Velazquez, R. Leijenaar, et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *Eur J Cancer.*, 2012, 48(4):441–446. https://doi.org/10.1016/j.ejca.2011.11.036

[5] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, 2016, 278(2):563–577. https://doi.org/10.1148/radiol.2015151169

[6] V. Sineglazov and M. Shevchenko, "Prediction of Vestibular Schwannoma Growth Based on Radiomics Features of MRI Images Using Ensemble Machine Learning Methods," *Electronics and Control Systems*, 2026, No. 1(87): 50–56. https://doi.org/10.18372/1990-5548.87.20884

[7] H. McGrath, P. Li, R. Dorent, et al., "Manual segmentation versus semi-automated segmentation for quantifying vestibular schwannoma volume on MRI," *Int J Comput Assist Radiol Surg*., 2020, 15(9):1445–1455. https://doi.org/10.1007/s11548-020-02222-y

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, 2021, 18(2):203–211. https://doi.org/10.1038/s41592-020-01008-z

[9] R. Dorent, A. Kujawa, M. Ivory, et al., "CrossMoDA 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation," *Medical Image Analysis*, 2023; 83:102628. https://doi.org/10.1016/j.media.2022.102628

[10] A. Kujawa, R. Dorent, S. Connor, et al., "Deep learning for automatic segmentation of vestibular schwannoma: a retrospective study from multi-center routine MRI," *Front Comput Neurosci*, 2024, 18:1365727. https://doi.org/10.3389/fncom.2024.1365727

[11] S. Cornelissen, S. M. Schouten, P. P. J. H. Langenhuizen, et al., "Towards clinical implementation of automated segmentation of vestibular schwannomas: a reliability study comparing AI and human performance," *Neuroradiology*, 2025; 67:1049–1059. https://doi.org/10.1007/s00234-025-03611-3

[12] B. Hajikarimloo, I. Mohammadzadeh, P. Shirzadi, et al., "Deep Learning Models for Radiomics-Based Segmentation of Vestibular Schwannoma on Magnetic Resonance Imaging: A Systematic Review and Meta-analysis," *J Imaging Inform Med*, 2025. https://doi.org/10.1007/s10278-025-01757-3

[13] T. Itoyama, T. Nakaura, T. Hamasaki, et al., "Whole Tumor Radiomics Analysis for Risk Factors Associated with Rapid Growth of Vestibular Schwannoma in Contrast-Enhanced T1-Weighted Images," *World Neurosurg*., 2022, 166:e572–e582. https://doi.org/10.1016/j.wneu.2022.07.058

[14] H. C. Yang, C. C. Wu, C. C. Lee, et al., "Prediction of pseudoprogression and long-term outcome of vestibular schwannoma after Gamma Knife radiosurgery based on preradiosurgical MR radiomics," *Radiother Oncol*., 2021, 155:123–130. https://doi.org/10.1016/j.radonc.2020.10.041

[15] K. Wang, N. A. George-Jones, L. Chen, et al., "Joint Vestibular Schwannoma Enlargement Prediction and Segmentation Using a Deep Multi-task Model," *The Laryngoscope*, 2023, 133(10):2754–2760. https://doi.org/10.1002/lary.30516

[16] T. Gill, D. Hamilton, and A. Rajgor, "The application of radiomics in vestibular schwannomas," *J Laryngol Otol.*, 2025, 139(8):647–654. https://doi.org/10.1017/S0022215125000258

[17] C. Parmar, E. Rios Velazquez, R. Leijenaar, et al., "Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation," *PLoS ONE*, 2014, 9(7):e102107. https://doi.org/10.1371/journal.pone.0102107

[18] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *J Chiropr Med*., 2016, 15(2):155–163. https://doi.org/10.1016/j.jcm.2016.02.012

[19] C. Haarburger, G. Müller-Franzes, L. Weninger, et al., "Radiomics feature reproducibility under inter-rater variability in segmentations of CT images," *Sci Rep*, 2020, 10:12688. https://doi.org/10.1038/s41598-020-69534-6

[20] S. Li, J. Kendrick, M. A. Ebert, et al., "Auto-segmentation, radiomic reproducibility, and comparison of radiomics between manual and AI-derived segmentations for coronary arteries in cardiac [18F]NaF PET/CT images," *EJNMMI Physics*, 2025, 12:42. https://doi.org/10.1186/s40658-025-00751-6

[21] J. Shapey, A. Kujawa, R. Dorent, et al., "Segmentation of vestibular schwannoma from MRI, an open annotated dataset and baseline algorithm," *Scientific Data*, 2021, 8:286. https://doi.org/10.1038/s41597-021-01064-w

[22] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, et al., "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Res.*, 2017, 77(21):e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339

[23] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput Electr Eng*., 2014, 40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

[24] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat Surv*., 2010, 4:40–79. https://doi.org/10.1214/09-SS054

Received: February 09, 2026
Accepted: March 03, 2026
Published: April 19, 2026

Sineglazov Victor. ORCID 0000-0002-3297-9060. Doctor of Engineering Science. Professor.
Department of Artificial Intelligence, Institute of Applied Systems Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind / Solar power plant, artificial intelligence.
Publications: over 850 publications.
E-mail: svm@kai.edu.ua

Shevchenko Maksym. ORCID 0009-0004-0540-8213. Postgraduate Student.
Department of Avionics and Control Systems, Faculty of Air Navigation Electronics and Telecommunications, State University "Kyiv Aviation Institute", Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2020).
Research area: Identification of complex systems.
Publications: 6.
E-mail: maksymshevchenko01@gmail.com

В. М. Синєглазов, М. В. Шевченко. Вплив автоматичної сегментації на радіоміксне прогнозування росту вестибулярної шванноми

У роботі досліджено вплив автоматичної сегментації на основі nnU-Net на відтворюваність радіомікснних ознак та якість прогнозування росту вестибулярної шванноми. Модель nnU-Net навчено на 317 парах T1C-зображень з масками від 155 пацієнтів з датасету Vestibular-Schwannoma-MC-RC2 методом 5-кратної крос-валідації, досягнувши середнього коефіцієнта Dice = 0,862. ICC-аналіз показав, що 88,8% wavelet-ознак зберігають хорошу або відмінну узгодженість (ICC ≥ 0,75) між ручними та автоматичними масками. Порівняння на підвибірці з 96 пацієнтів з мітками росту показало: опублікований раніше pipeline (Wavelet + Voting ансамбль) з ручними масками досягає ROC AUC = 0,742, з автоматичними масками – 0,639 (−14,0%). Оптимізація pipeline (ICC-фільтрація, адаптація ансамблю до LR + LDA) підвищує результат до 0,687 (−7,4%). Результати визначають ціну автоматизації радіомікснного pipeline та надають рекомендації щодо ICC-фільтрації для клінічного впровадження.

Ключові слова: вестибулярна шваннома, автоматична сегментація, nnU-Net, радіоміксні ознаки, відтворюваність, ICC, прогнозування росту пухлини, машинне навчання.

Синєглазов Віктор Михайлович. ORCID 0000-0002-3297-9060. Доктор технічних наук. Професор.
Кафедра штучного інтелекту, Інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (1973).
Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки, штучний інтелект.
Кількість публікацій: більше 850 публікацій.
E-mail: svm@kai.edu.ua

Шевченко Максим Валерійович. ORCID 0009-0004-0540-8213. Аспірант.
Кафедра авіоніки та систем управління, Факультет аеронавігації, електроніки і телекомунікацій, Державний університет «Київський авіаційний інститут», Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2020).
Напрям наукової діяльності: ідентифікація складних систем.
Кількість публікацій: 6.
E-mail: maksymshevchenko01@gmail.com