Development and validation of artificial intelligence models for prognosis prediction of juvenile myoclonic epilepsy with clinical and radiological features

Kyung Min Kim

The Graduate School
Yonsei University
Department of Medicine

Development and validation of artificial intelligence models for prognosis prediction of juvenile myoclonic epilepsy with clinical and radiological features

A Dissertation Submitted
to the Department of Medicine
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Medical Science

Kyung Min Kim

December 2024

This certifies that the Dissertation of Kyung Min Kim is approved

Thesis Supervisor	Won-Joo Kim
Thesis Committee Member	Seung-Koo Lee
Thesis Committee Member	Min Kyung Chu
Thesis Committee Member	Sung Soo Ahn
Thesis Committee Member	Ji Hyun Kim

The Graduate School
Yonsei University
December 2024

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my dissertation supervisor, Dr. Won-Joo Kim. This work would not have been possible without his unwavering encouragement and insightful guidance. His dedication to my academic and personal growth has been invaluable. I am also profoundly thankful to Dr. Seung-Koo Lee for generously sharing his expertise in neuroimaging and for meticulously reviewing this thesis despite his demanding schedule. His contributions have significantly shaped the direction and quality of this work. My heartfelt thanks go to Dr. Min Kyung Chu, whose guidance on the epilepsy research has been instrumental. His unwavering support and the time he afforded me to focus on my research were crucial to successfully completing this dissertation. I am deeply indebted to Dr. Sung Soo Ahn for introducing me to the fascinating world of Radiomics and for her valuable advice on the experimental aspects of this work. Her insights were key in expanding the scope of my research. I also thank Dr. Ji Hyun Kim for his constant encouragement and sincere advice throughout this journey. His support has been a source of motivation and reassurance. The factors influencing the prognosis of drug treatment in epilepsy patients are still not fully understood, and research utilizing artificial intelligence is in its nascent stages with many unexplored areas. This journey would not have been possible without the unwavering support of my loving family. I am eternally grateful to my parents for their lifelong support, which has been the foundation of all my achievements. Above all, I would like to express my deepest appreciation to my wife, Hyunseong, for her boundless love and unwavering support. Her sacrifices, late nights, early mornings, and tireless care of our family during my frequent absences made it possible for me to complete this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Development and validation of artificial intelligence models for prognosis prediction of juvenile myoclonic epilepsy with clinical and radiological features

Introduction: Juvenile myoclonic epilepsy (JME) is a prevalent form of epilepsy with varying prognoses based on clinical and radiological factors. While many studies have explored JME's clinical aspects, the integration of these factors into a predictive model for prognosis has been limited. This study aims to develop and validate machine learning models that combine clinical and radiological features to predict prognosis in JME patients.

Methods: We conducted a retrospective study including 125 patients diagnosed with JME. Clinical data were collected, including demographic information, seizure history, and treatment details. MRI data were analyzed using volumetric and cortical thickness measurements, as well as radiomics features. Machine learning models - including logistic regression, random forest, extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), support vector machine (SVM), and artificial neural network (ANN) - were developed and evaluated using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUROC) metrics. The models were trained on an internal dataset and validated on an independent external dataset.

Results: The analysis identified that male gender, volumes of the left amygdala and right hippocampus, and cortical thickness of the bilateral temporal poles, left entorhinal cortex, fusiform gyrus, and right inferior and middle temporal cortex were significantly associated with favorable prognosis. Models combining clinical, volumetric, cortical thickness, and Radiomics data outperformed those relying on a single data type. The best-performing machine learning model was random forest, which achieved an AUROC of 0.923. The integrated model demonstrated superior predictive performance, underscoring the importance of a multimodal approach. Brain structures such as the thalamus and hippocampus, known to be involved in JME's pathophysiology, were identified as critical features in the prognostic prediction.

Conclusion: This study highlights the potential of using machine learning models that integrate clinical and radiological data to predict prognosis in JME. The findings suggest that multimodal data models are more effective than those based on single data types, offering a promising approach for improving prognostic accuracy in JME. The integration of advanced imaging features with clinical variables could enhance decision-making in epilepsy management, providing valuable tools for clinicians in the prediction of treatment outcomes. Further research is needed to validate these findings in larger, more diverse populations and to explore the inclusion of additional radiological modalities such as diffusion tensor image and functional MRI.

Key words: Juvenile myoclonic epilepsy, Prognosis, Machine learning, Radiomics, Multimodal integration

I. INTRODUCTION

Epilepsy is a common chronic brain disorder that affects people of all ages and can be found in any country and in any race.[1] Conceptually, epilepsy can be defined as a long-lasting brain condition characterized by a tendency to have recurrent epileptic seizures. Globally, the prevalence of active epilepsy is 6.4 per 1000 (95% Confidence Interval (CI) 5.6-7.3), and the incidence is 61.4 per 100,000 person-years, and according to epidemiological studies of epilepsy in Korea, the incidence and prevalence of epilepsy in Korea are increasing, and as of 2017, the prevalence of epilepsy in Korea was 4.8 per 1000 and the incidence was 35.4 per 100,000 person-years, making it one of the important brain diseases.[2]

Juvenile myoclonic epilepsy (JME) is a prevalent adolescent epilepsy, accounting for approximately 10% of all cases.[3,4] Characterized by recurrent myoclonic seizures, primarily in the shoulders and limbs, JME typically begins around puberty and is often accompanied by generalized tonic-clonic seizures and, in some cases, absence seizures.[3-5] Classified under idiopathic generalized epilepsy (IGE), the prognosis for JME with pharmacological treatment is relatively favorable, with nearly 60% of patients achieving five years or more of seizure freedom on medication and about 25% maintaining remission without medication.[6,7] However, many patients experience relapse upon discontinuation of medication, necessitating lifelong treatment. Meta-analyses suggest that around 35% of patients exhibit drug-refractory epilepsy.[8]

Clinical factors influencing the prognosis with antiseizure medication include female gender, younger age of onset, history of absence seizures, praxis-induced seizures, childhood absence epilepsy, comorbid psychiatric disorders, family history, epileptiform asymmetries on EEG, and absence of photoparoxysmal response.[9,10] While brain MRI in JME patients appears visually normal, recent advancements in quantitative and functional MRI analysis reveal differences in structural and functional connectivity compared to healthy controls, along with widespread neocortical thinning as the disease progresses.[11,12]

Despite the progress in imaging techniques, there is a lack of studies analyzing quantitative imaging features for prognosis in JME. Recent efforts have utilized machine learning and deep learning to develop diagnostic and prognostic models for various neurological diseases, including epilepsy.[13,14] Some studies have combined clinical information and brain MRI images to predict drug

response, but these models often underperform due to the insufficient integration of comprehensive variables.[15,16]

To address this gap, our study aims to develop and validate artificial intelligence models that combine clinical and imaging variables to predict the prognosis of drug treatment response in JME.

II. METHODS

2.1. Subjects

2.1.1. Inclusion and exclusion criteria

This retrospective study included consecutive patients who presented with seizures and visited the epilepsy clinic. Patients diagnosed with JME by epilepsy specialists were included in the study. The initial diagnosis of JME was confirmed by reviewing the medical records of neurologists at the institution, based on clinical and EEG features established by the International League Against Epilepsy. Patients included in the study experienced both generalized onset motor myoclonic seizures and generalized onset motor tonic-clonic seizures. EEG findings consistently demonstrated generalized polyspikes or spike-and-wave complexes in all patients diagnosed with JME. MRI readings were confirmed to be normal by board-certified neuroradiologists. Exclusion criteria for the study were as follows: 1) patients with less than a 3-year follow-up, and 2) participants who underwent a 2D protocol MRI.

2.1.2. Internal data set

The internal data set was used to train this study's prognostic prediction model. The clinical data utilized were retrospectively collected from the medical records of epilepsy patients who visited the epilepsy clinic of a single physician (K. Heo) at Severance Hospital between January 2000 and August 2022. Clinical and imaging data were obtained for research by comprehensively reviewing patients' medical records who met the inclusion criteria.

2.1.3. External data set

The external data set was used to test the prognostic prediction model in this study. Clinical data were collected from patients diagnosed with JME by neurologists at five university hospitals in

South Korea (Severance Hospital [K.M. Kim], Wonju Severance Christian Hospital, Chung-Ang University Gwangmyeong Hospital, Gangneung Asan Hospital, and Incheon St. Mary's Hospital) between January 2000 and July 2024. The same clinical variables as those obtained from the internal data set were used, and brain MRI data were acquired and analyzed.

2.2. Clinical data

Clinical variables collected from patients included age, sex, age at onset, disease duration, treatment history, number of antiseizure medications, family history of epilepsy, history of febrile seizures, history of absence seizures, and the presence of seizures during the follow-up period. Patients who were seizure-free for 2 years or longer at any time after diagnosis were considered to have a favorable outcome.

2.3. MRI

2.3.1. MRI data acquisition

Brain MRI at Severance Hospital used a 3T MRI system (Achieva or Ingenia, PhiPhilips Healthcare; TrioTim, Siemens) with an encoding head coil of 8-channel sensitivity. MRI data was included if they were deemed visually normal by at least one neuroradiologist. We collected data with T1 3D images. Raw data from other hospitals' MRIs was analyzed by collecting MRIs with T1 3D images.

2.3.2. MRI preprocessing

During the preprocessing stage of the MRI data, the researcher performed the preprocessing steps while blinded to the images of the favorable and poor prognosis groups. Brain region-specific masks were obtained using FreeSurfer 6.0.0 software (https://surfer.nmr.mgh.harvard.edu). The preprocessing process included motion correction, Talairach transformation, segmentation of subcortical white matter and deep gray matter structures, and intensity normalization. FreeSurfer

was used to resample and obtain masks for each region in a consistent size and position, thereby minimizing variability between processing steps and across the data.

2.3.3. Analysis of MRI data

2.3.3.1. Subcortical structure volume

To quantify the volumes of subcortical structures, we utilized the automated segmentation tool FreeSurfer (version 6.0.0). The T1-weighted MRI scans of all subjects were processed using the standard FreeSurfer pipeline as part of the preprocessing stage. Specifically, the subcortical regions of interest (ROIs) volumes were extracted from the aseg.stats output file, generated by the "recon-all" command. This file provides volumetric data for 29 predefined anatomical labels, including bilateral structures such as the thalamus, caudate, putamen, pallidum, hippocampus, amygdala, nucleus accumbens, ventral diencephalon, choroid plexus, cerebellar cortex, and cerebellar white matter. Additionally, it includes midline structures such as the brainstem, optic chiasm, and segments of the corpus callosum (anterior, mid-anterior, central, mid-posterior, and posterior). The resulting volumes were subsequently used for statistical analyses. All segmentations were visually inspected for accuracy, and any errors were corrected in accordance with FreeSurfer's guidelines to ensure the reliability of the volumetric measurements.

2.3.3.2. Cortical thickness

We utilized FreeSurfer (version 6.0.0) to measure cortical thickness, following the standard processing pipeline applied to T1-weighted MRI scans. Cortical reconstruction and volumetric segmentation were conducted using the "recon-all" command, which includes steps such as intensity normalization, skull stripping, and the generation of cortical surface models. Cortical thickness measurements for each hemisphere were extracted from the lh.aparc.stats and rh.aparc.stats files, corresponding to the left and right hemispheres, respectively. These files provide cortical thickness data for the 34 predefined anatomical ROIs per hemisphere based on the Desikan-Killiany atlas. The regions include bankssts, caudal anterior cingulate, caudal middle frontal, cuneus, entorhinal,

fusiform, inferior parietal, inferior temporal, isthmus cingulate, lateral occipital, lateral orbital frontal, lingual, medial orbital frontal, middle temporal, parahippocampal, paracentral, pars opercularis, pars orbitalis, pars triangularis, pericalcarine, postcentral, posterior cingulate, precentral, precuneus, rostral anterior cingulate, rostral middle frontal, superior frontal, superior parietal, superior temporal, supramarginal, frontal pole, temporal pole, transverse temporal, and insula. The extracted thickness values were subsequently used in statistical analyses. All segmentations and surface reconstructions were visually inspected for accuracy, and any errors were corrected in accordance with FreeSurfer's guidelines to ensure the validity and reliability of the cortical thickness measurements.

2.3.3.3. Radiomics

We focused on 22 ROIs, including bilateral cerebral white matter, bilateral thalamus, bilateral caudate, bilateral putamen, bilateral globus pallidus, bilateral hippocampus, bilateral amygdala, bilateral ventral diencephalon, the brainstem, and the corpus callosum (segmented into anterior, mid-anterior, central, mid-posterior, and posterior regions) in this study. These ROIs were selected based on prior studies identifying quantitative and functional differences between individuals with juvenile myoclonic epilepsy and healthy controls. To analyze these regions, we employed radiomics, which involves extracting many quantitative features from medical images that capture the texture, shape, and intensity patterns within the ROIs. Specifically, we used PyRadiomics (http://www.radiomics.io/pyradiomics.html), an open-source software platform, to extract a comprehensive set of radiomics features from each ROI. These features were then used in subsequent statistical and machine-learning analyses to investigate potential biomarkers and differences between the study groups.

2.4. Artificial intelligence method

2.4.1. Machine learning models

We implemented and evaluated six different machine learning models to classify and predict

outcomes based on the extracted features in this study: Logistic regression, Random Forest, Extreme gradient boosting (XGBoost), Light gradient boosting machine (LightGBM), Support vector machine (SVM), and Artificial neural network (ANN).

The model employs weighted scaling and the synthetic minority oversampling technique (SMOTE) for further performance improvement. Additionally, the LASSO feature reduction was applied first. The models were trained using the training sets, and hyperparameter tuning was performed. Hyperparameter tuning was done by a 5-fold cross-validation in the training set.

2.4.1.1. Logistic regression

Logistic regression was employed as a baseline model in our study due to its simplicity and interpretability. This model is a linear classifier that estimates the probability of a binary outcome based on the logistic function. The logistic regression model was trained using the maximum likelihood estimation method to minimize the difference between the predicted and actual outcomes. Despite being a linear model, it can effectively capture the relationship between the features and the outcome when the data is linearly separable. Regularization techniques, such as L2 regularization, were applied to prevent overfitting by penalizing large coefficients.

2.4.1.2. Random Forest

Random Forest, an ensemble learning method, enhanced predictive performance by combining multiple decision trees. Each tree in the Random Forest was trained on a bootstrapped subset of the data, and a random subset of features was considered for splitting at each node, which increases diversity among the trees and improves generalization. The final prediction was obtained by aggregating the predictions from all individual trees, typically through majority voting for classification tasks. This model is particularly advantageous because it is robust to overfitting and can capture non-linear relationships in the data. Hyperparameters such as the number of trees, maximum depth, and minimum samples per leaf were tuned to optimize model performance.

2.4.1.3. Extreme gradient boosting (XGBoost)

XGBoost was implemented as a more sophisticated gradient boosting framework, known for its speed and performance in machine learning competitions. This model builds trees sequentially, where each new tree attempts to correct the errors made by the previous ones. The model minimizes a differentiable loss function using gradient descent and incorporates a regularization term to prevent overfitting. XGBoost also supports parallel processing, which accelerates training, and includes built-in handling of missing values. Hyperparameters such as the learning rate, maximum depth of trees, and number of boosting rounds were carefully tuned using cross-validation to achieve the best performance.

2.4.1.4. Light gradient boosting machine (LightGBM)

LightGBM, another gradient-boosting framework, was chosen for its efficiency and scalability, especially with large datasets. It uses a histogram-based algorithm to discretize continuous features into bins, which significantly speeds up the training process and reduces memory usage. LightGBM grows trees leaf-wise rather than level-wise, allowing it to capture complex patterns in the data more effectively. The model was tuned by adjusting hyperparameters such as the number of leaves, learning rate, and the number of boosting iterations. LightGBM's ability to handle large-scale data with faster training times made it a valuable model in our analysis.

2.4.1.5. Support vector machine (SVM)

SVM with a radial basis function (RBF) kernel was employed to handle the non-linear relationships in our dataset. SVM works by finding the optimal hyperplane that maximizes the margin between the classes in a high-dimensional space. The RBF kernel maps the input features into a higher-dimensional space, where a linear separation between classes may be possible. The key hyperparameters, such as the regularization parameter (C) and the kernel coefficient (gamma), were optimized to balance model complexity and accuracy. SVM is particularly effective in scenarios where the feature space is high-dimensional and the margin between classes is distinct.

2.4.1.6. Artificial neural network (ANN)

An ANN was constructed to model complex, non-linear relationships in the data. The ANN architecture consisted of an input layer, multiple hidden layers with neurons, and an output layer. Each neuron applied a non-linear activation function to a weighted sum of inputs, enabling the network to learn hierarchical representations of the data. The network was trained using backpropagation, where the weights were adjusted to minimize the loss function via gradient descent. Hyperparameters such as the number of hidden layers, the number of neurons per layer, the learning rate, and the type of activation function were tuned to optimize the network's performance. Dropout regularization was also applied to prevent overfitting by randomly dropping neurons during training.

2.4.2. Performance evaluation

The performance of the artificial intelligence models was evaluated using a comprehensive set of metrics, including the area under the receiver operating characteristic curve (AUROC), recall, precision, accuracy, and F1 score. AUROC measures the model's ability to discriminate between classes, representing the trade-off between sensitivity (true positive rate) and specificity (false positive rate) across different threshold settings. An AUROC value of 1 indicates perfect discrimination, while a value of 0.5 suggests no better performance than random chance. Recall, also known as sensitivity or true positive rate, is the proportion of actual positives correctly identified by the model. It is calculated as the number of true positives divided by the sum of true positives and false negatives. High recall is important in contexts where minimizing false negatives is critical. Precision is the proportion of correct positive predictions, calculated as the number of true positives divided by the sum of true and false positives. Precision is particularly relevant in scenarios with the high cost of false positives. Accuracy is the overall proportion of correct predictions (both true positives and true negatives) out of all predictions made by the model. It provides a general measure of the model's correctness but can be misleading in imbalanced datasets where one class is much more frequent than the other. The F1 score is the harmonic mean of precision and recall, providing a metric that balances both concerns. The F1 score is instrumental when the dataset is imbalanced, giving a more nuanced view of the model's performance than accuracy alone. By presenting these

metrics, we aimed to provide a thorough assessment of the model's performance, ensuring that it is evaluated in terms of its overall correctness and ability to balance the trade-offs between different types of errors.

2.4.3. Feature importance and model interpretability

We employed Shapley Additive Explanations (SHAP) to interpret the artificial intelligence models to identify and present important features that contributed to the model's predictions. SHAP values are derived from cooperative game theory and provide a consistent way to assign importance scores to each feature for individual predictions. Specifically, SHAP calculates the contribution of each feature by considering all possible combinations of features and their impact on the model's output. This method allows for a clear understanding of how each feature influences the prediction, offering both global and local interpretability. Globally, SHAP values indicate the overall importance of each feature across the entire dataset, while locally, they explain the contribution of each feature to a specific prediction. By using SHAP, we could identify and visualize the most critical features that the model relied upon, enhancing the transparency and interpretability of the machine learning model's decision-making process.

2.5. Statistical analysis

The statistical analysis method used in the variable description and selection stage utilizes Pearson correlation coefficient and variance inflation factor to prevent over-selection of variables including multicollinearity, and Student t-test and Pearson chi-square test to compare variables. All statistical analyses were performed using R version 4.11, and machine learning and deep learning were performed using Python version 3.6.8. Statistical significance was determined based on a two-sided p-value <0.05.

III. RESULTS

3.1. Clinical characteristics

The process of selecting the study participants is illustrated in Figure 1. For the internal dataset, we enrolled JME patients treated by a single physician (K. Heo) at Severance hospital between January 2000 and August 2022. Among the 210 patients with available clinical data and brain MRI, we excluded 37 patients who had less than 3 years of follow-up. Additionally, 48 patients without 3D MRI were excluded, followed by the exclusion of 26 patients who had received prior treatment before visiting the hospital, resulting in the absence of pre-treatment MRI. Consequently, the final internal dataset included clinical variables and 3D T1 MRI images from 99 patients. For the external dataset, we enrolled 76 patients from five hospitals who had at least 3 years of follow-up after disease onset between January 2000 and July 2004. We excluded 22 patients who had only 2D MRI scans and 28 patients who received treatment before their first MRI scan, resulting in a final external dataset consisting of clinical variables and 3D T1 MRI images from 26 patients from 5 university hospitals (Severance Hospital [K.M. Kim] – n=12, Wonju Severance Christian Hospital, n=3, Chung-Ang University Gwangmyeong Hospital – n=3, Gangneung Asan Hospital – n=5, and Incheon St. Mary's Hospital – n=3)

The clinical characteristics of all included patients are summarized in Table 1. Among the 125 patients from the combined internal and external datasets, 85 were seizure-free for more than 2 years, indicating a favorable prognosis, while 40 patients did not achieve this outcome, representing a poor prognosis. A significantly higher proportion of males was observed in the favorable prognosis group compared to the poor prognosis group (60.0% vs. 40.0%, $p = 0.036$). No significant differences were found between the two groups in terms of age, age at onset, or duration of epilepsy. Additionally, there were no significant differences in family history, history of febrile seizures, or the presence of absence seizures. Regarding treatment, the groups did not differ in the number of antiseizure medications used, or the use of specific medications, including valproic acid, lamotrigine, levetiracetam, and topiramate. The follow-up duration was also similar between the two groups.

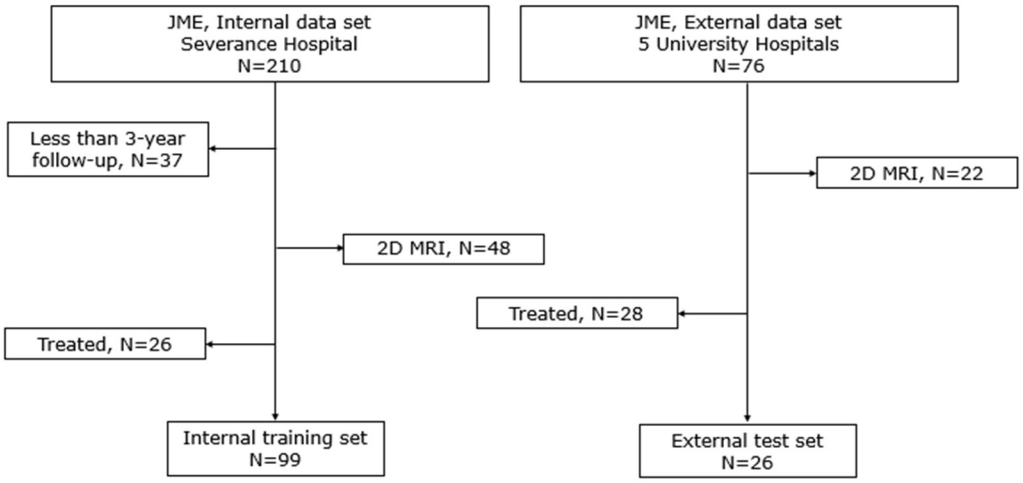

Figure 1. Flow chart depicting the study subjects. JME, juvenile myoclonic epilepsy

Table 1. Demographic and clinical characteristics of juvenile myoclonic epilepsy patients in the study

	Favorable prognosis (n = 85)	Poor prognosis (n = 40)	*p*-value
Age (years)	23.5 ± 8.7	23.2 ± 7.7	0.843
Male sex, n (%)	51 (60.0)	16 (40.0)	0.036
Onset age (years)	15.2 ± 4.5	15.2 ± 4.8	0.991
Epilepsy duration (years)	8.3 ± 9.4	8.0 ± 8.2	0.888
Family history, n (%)	16 (18.8)	4 (10.0)	0.209
Febrile seizure history*, n (%)	9 (10.6)	5 (12.5)	0.767
Absence seizure, n (%)	31 (36.5)	14 (35.0)	0.873
Number of ASMs	2 (1-3)	2 (1-3)	0.577
VPA, n (%)	65 (76.5)	30 (75.0)	0.857
LTG, n (%)	39 (45.9)	21 (52.5)	0.490
LEV, n (%)	40 (47.1)	24 (60.0)	0.177
TPM, n (%)	17 (20.0)	6 (15.0)	0.501
Follow-up duration (years)	13.9 ± 6.7	11.1 ± 7.2	0.178

Data are presented as the number of patients (percentage) or as the mean ± standard deviation. ASM, antiseizure medication; VPA = valproic acid; LTG, lamotrigine; LEV, levetiracetam; TPM = topiramate.

* Fisher's exact test was used.

3.2. Volumetric analysis of subcortical structures

The volumetric analysis is summarized in Table 2. A comparison of the volume data between the two groups revealed that the volumes of the left amygdala (1739.9 ± 263.4 mm^3 vs. 1601.7 ± 358.2 mm^3, p = 0.017) and the right hippocampus (4396.6 ± 417.7 mm^3 vs. 4128.8 ± 825.7 mm^3, p = 0.017) were significantly smaller in the poor prognosis group compared to the favorable prognosis group.

However, no significant differences were observed between the groups in the volumes of other bilateral subcortical structures, including the thalamus, caudate, putamen, pallidum, nucleus accumbens, ventral diencephalon, choroid plexus, cerebellar cortex, and cerebellar white matter. Additionally, the brainstem, corpus callosum, and total intracranial volume did not show significant volumetric differences between the two groups.

Table 2. Volumetric analysis of brain subcortical structures based on prognostic outcomes

	Favorable prognosis (n = 85)	Poor prognosis (n = 40)	*p*-value
Left			
Thalamus	8054.0 ± 844.4	7748.2 ± 1393.7	0.131
Caudate	3569.1 ± 433.8	3458.2 ± 643.6	0.259
Putamen	5075.5 ± 619.2	4902.2 ± 895.1	0.211
Pallidum	2073.8 ± 244.5	2017.6 ± 353.9	0.303
Hippocampus	4180.6 ± 415.2	4027.1 ± 630.2	0.107
Amygdala	1739.9 ± 263.4	1601.7 ± 358.2	0.017
Nucleus accumbens	511.4 ± 98.5	286.6 ± 133.8	0.246
Ventral diencephalon	4175.7 ± 457.7	4003.7 ± 621.1	0.084
Choroid Plexus	437.0 ± 163.7	427.5 ± 159.2	0.765
Cerebellum – cortex	56595.1 ± 5734.4	54113.8 ± 8452.0	0.057
Cerebellum – white matter	14800.9 ± 1846.9	14133.2 ± 2293.1	0.084
Right			
Thalamus	7574.4 ± 775.1	7219.0 ± 1275.0	0.056
Caudate	3637.9 ± 444.1	3544.3 ± 567.1	0.318
Putamen	5130.4 ± 617.7	4985.8 ± 809.3	0.273
Pallidum	1977.6 ± 238.8	1942.0 ± 284.8	0.467
Hippocampus	4396.6 ± 417.7	4128.8 ± 825.7	0.017
Amygdala	1848.7 ± 279.3	1744.6 ± 361.0	0.080
Nucleus accumbens	577.2 ± 105.1	554.5 ± 117.2	0.280
Ventral diencephalon	4173.3 ± 444.1	4017.0 ± 562.6	0.095
Choroid Plexus	430.6 ± 154.9	422.9 ± 188.5	0.809
Cerebellum – cortex	56266.6 ± 5838.5	53800.4 ± 8392.4	0.059
Cerebellum–white matter	14242.4 ± 2001.1	13540.7 ± 2200.4	0.079
Midline			

Brainstem	21182.7 ± 2266.9	20565.8 ± 3631.8	0.248
Optic-chiasm	154.3 ± 58.7	140.1 ± 60.8	0.213
Corpus callosum			
Anterior	862.7 ± 141.4	836.8 ± 158.5	0.360
Mid-anterior	669.4 ± 181.2	666.1 ± 171.8	0.922
Central	688.8 ± 172.1	660.2 ± 171.5	0.388
Mid-posterior	552.8 ± 103.9	560.6 ± 128.1	0.720
Posterior	990.4 ± 178.0	982.0 ± 205.2	0.815
Total intracranial volume	1581418.5 ± 175060.8	1502151.6 ± 250957.1	0.077

The presented values represent the volume of each brain region, with units in mm^3. Data are presented as the mean ± standard deviation.

3.3. Cortical thickness analysis

The cortical thickness analysis between the favorable prognosis and poor prognosis groups revealed significant differences in several brain regions (Table 3). The left entorhinal cortex exhibited a statistically significant difference, with the favorable prognosis group having a mean cortical thickness of 3.370 mm, compared to 3.235 mm in the poor prognosis group ($p = 0.046$). Additionally, the left fusiform gyrus was thinner in the poor prognosis group (2.761 mm) than in the favorable prognosis group (2.832 mm), with a p-value of 0.018. Similarly, the left temporal pole was thinner in the poor prognosis group (3.513 mm) compared to the favorable prognosis group (3.656 mm), with a p-value of 0.044.

In the right hemisphere, significant differences were also observed. The inferior temporal cortex was thinner in the poor prognosis group (2.738 mm) compared to the favorable prognosis group (2.813 mm), with a p-value of 0.020. Additionally, the right middle temporal cortex (poor prognosis: 2.922 mm vs. favorable prognosis: 2.996 mm, $p = 0.007$) and the right temporal pole (poor prognosis: 3.361 mm vs. favorable prognosis: 3.682 mm, $p = 0.004$) were also significantly thinner in the poor prognosis group.

In contrast, no statistically significant differences in cortical thickness were observed between the favorable and poor prognosis groups in other examined brain regions, including the bankssts, caudal anterior cingulate, caudal middle frontal, cuneus, inferior parietal, isthmus cingulate, lateral occipital, lateral orbital frontal, lingual, medial orbital frontal, parahippocampal, paracentral, pars opercularis, pars orbitalis, pars triangularis, pericalcarine, postcentral, posterior cingulate, precentral, precuneus, rostral anterior cingulate, rostral middle frontal, superior frontal, superior parietal, superior temporal, supramarginal, frontal pole, transverse temporal, and insula cortex.

Table 3. Cortical thickness analysis based on prognostic outcomes

	Favorable prognosis (n = 85)	Poor prognosis (n = 40)	*p*-value
Left			
Bankssts	2.636 ± 0.199	2.615 ± 0.221	0.587
Caudalanteriorcingulate	2.816 ± 0.203	2.792 ± 0.218	0.547
Caudalmiddlefrontal	2.671 ± 0.182	2.651 ± 0.189	0.583
Cuneus	1.901 ± 0.176	1.883 ± 0.238	0.625
Entorhinal	3.370 ± 0.322	3.235 ± 0.399	0.046
Fusiform	2.832 ± 0.145	2.761 ± 0.173	0.018
Inferiorparietal	2.546 ± 0.170	2.575 ± 0.206	0.409
Inferiortemporal	2.803 ± 0.146	2.745 ± 0.192	0.061
Isthmuscingulate	2.471 ± 0.209	2.492 ± 0.255	0.634
Lateraloccipital	2.180 ± 0.164	2.198 ± 0.238	0.625
Lateralorbitofrontal	2.758 ± 0.164	2.735 ± 0.198	0.487
Lingual	2.051 ± 0.159	2.029 ± 0.170	0.487
Medialorbitofrontal	2.562 ± 0.137	2.551 ± 0.237	0.739
Middletemporal	2.961 ± 0.157	2.919 ± 0.264	0.267
Parahippocampal	2.720 ± 0.258	2.735 ± 0.367	0.800
Paracentral	2.539 ± 0.165	2.550 ± 0.222	0.768
Parsopercularis	2.703 ± 0.186	2.675 ± 0.197	0.456
Parsorbitalis	2.882 ± 0.246	2.833 ± 0.229	0.286
Parstriangularis	2.644 ± 0.183	2.631 ± 0.198	0.716
Pericalcarine	1.584 ± 0.161	1.567 ± 0.294	0.683
Postcentral	2.109 ± 0.179	2.079 ± 0.359	0.534
Posteriorcingulate	2.668 ± 0.177	2.660 ± 0.189	0.810
Precentral	2.623 ± 0.189	2.614 ± 0.257	0.824
Precuneus	2.469 ± 0.157	2.476 ± 0.162	0.812

Rostralanteriorcingulate	2.966 ± 0.174	2.953 ± 0.266	0.744
Rostralmiddlefrontal	2.513 ± 0.152	2.485 ± 0.155	0.346
Superiorfrontal	2.955 ± 0.155	2.937 ± 0.152	0.550
Superiorparietal	2.272 ± 0.167	2.317 ± 0.232	0.219
Superiortemporal	2.942 ± 0.171	2.922 ± 0.182	0.549
Supramarginal	2.620 ± 0.180	2.623 ± 0.212	0.937
Frontalpole	2.876 ± 0.247	2.883 ± 0.262	0.882
Temporalpole	3.656 ± 0.336	3.513 ± 0.427	0.044
Transversetemporal	2.393 ± 0.258	2.417 ± 0.290	0.641
Insula	3.099 ± 0.190	2.965 ± 0.523	0.037
Right			
Bankssts	2.683 ± 0.180	2.657 ± 0.251	0.517
Caudalanteriorcingulate	2.672 ± 0.221	2.642 ± 0.213	0.474
Caudalmiddlefrontal	2.685 ± 0.176	2.633 ± 0.227	0.161
Cuneus	1.932 ± 0.171	1.903 ± 0.189	0.397
Entorhinal	3.459 ± 0.363	3.347 ± 0.586	0.193
Fusiform	2.803 ± 0.123	2.757 ± 0.224	0.134
Inferiorparietal	2.534 ± 0.162	2.535 ± 0.176	0.970
Inferiortemporal	2.813 ± 0.169	2.738 ± 0.158	0.020
Isthmuscingulate	2.439 ± 0.196	2.441 ± 0.284	0.964
Lateraloccipital	2.255 ± 0.158	2.244 ± 0.169	0.732
Lateralorbitofrontal	2.690 ± 0.199	2.633 ± 0.293	0.197
Lingual	2.072 ± 0.149	2.059 ± 0.192	0.687
Medialorbitofrontal	2.619 ± 0.156	2.577 ± 0.246	0.248
Middletemporal	2.996 ± 0.141	2.922 ± 0.146	0.007
Parahippocampal	2.703 ± 0.245	2.705 ± 0.256	0.967
Paracentral	2.572 ± 0.176	2.557 ± 0.249	0.695
Parsopercularis	2.708 ± 0.172	2.692 ± 0.198	0.642

Parsorbitalis	2.823 ± 0.204	2.783 ± 0.308	0.389
Parstriangularis	2.628 ± 0.199	2.570 ± 0.268	0.178
Pericalcarine	1.626 ± 0.173	1.583 ± 0.170	0.194
Postcentral	2.098 ± 0.183	2.130 ± 0.200	0.378
Posteriorcingulate	2.604 ± 0.158	2.561 ± 0.145	0.147
Precentral	2.569 ± 0.189	2.552 ± 0.234	0.658
Precuneus	2.444 ± 0.168	2.446 ± 0.198	0.946
Rostralanteriorcingulate	3.036 ± 0.242	2.987 ± 0.252	0.298
Rostralmiddlefrontal	2.491 ± 0.156	2.460 ± 0.178	0.314
Superiorfrontal	2.950 ± 0.161	2.900 ± 0.222	0.158
Superiorparietal	2.244 ± 0.173	2.262 ± 0.208	0.607
Superiortemporal	2.996 ± 0.169	2.905 ± 0.388	0.067
Supramarginal	2.603 ± 0.176	2.601 ± 0.254	0.975
Frontalpole	2.807 ± 0.267	2.757 ± 0.523	0.480
Temporalpole	3.682 ± 0.34	3.361 ± 0.880	0.004
Transversetemporal	2.415 ± 0.255	2.306 ± 0.448	0.086
Insula	3.130 ± 0.187	3.011 ± 0.528	0.066

The presented values represent the cortical thickness of each brain region, with units in mm. Data are presented as the mean ± standard deviation.

3.4. Performances of machine learning models

3.4.1. Volumetry model

Table 4 presents the performance of six machine learning models—logistic regression, random forest, XGBoost, LightGBM, SVM, and ANN—trained on the training set and validated on an independent test set. Among these models, XGBoost demonstrated the highest performance with an AUROC of 0.700, LightGBM with an AUROC of 0.618, and random forest with an AUROC of 0.517. The SVM model achieved an AUROC of 0.500, logistic regression an AUROC of 0.431, and ANN an AUROC of 0.425.

Further analysis with the best-performing XGBoost model revealed that when using only clinical data, the AUROC was 0.600. When using only MRI data, the AUROC improved to 0.680. Combining both clinical and MRI data further enhanced the model's performance, achieving an AUROC of 0.700 (Figure 2).

Figure 3 displays the importance of the feature as determined by SHAP values. Among the top ten features, the five most significant MRI variables were the left cerebellum white matter, right thalamus, left globus pallidus, right amygdala, and left caudate. The left nucleus accumbens, right choroid plexus, corpus callosum mid-posterior, onset age, and brainstem followed these.

Table 4. Performances of the combined clinical and volumetric machine learning models on the test set

Models	Accuracy	Precision	Recall	F1-score	AUROC
Logistic Regression	0.600	0.560	0.600	0.565	0.431
Random Forest	0.680	0.664	0.680	0.652	0.580
XGBoost	0.680	0.816	0.680	0.712	0.700
Light GBM	0.560	0.486	0.560	0.505	0.618
SVM	0.640	0.410	0.640	0.500	0.500
ANN	0.600	0.400	0.600	0.480	0.425

AUROC, Area Under the Receiver Operating Characteristic curve; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machine; ANN, artificial neural network.

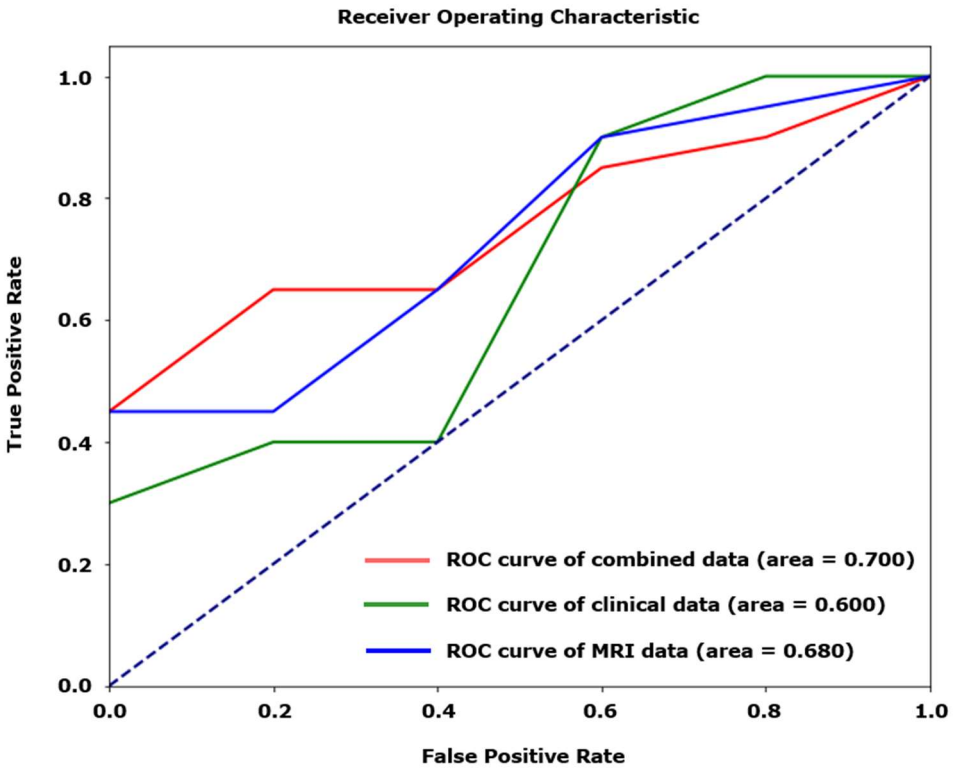

Figure 2. Receiver operating characteristic (ROC) curves for models using clinical, volumetric, and combined variables to predict poor prognosis of juvenile myoclonic epilepsy.

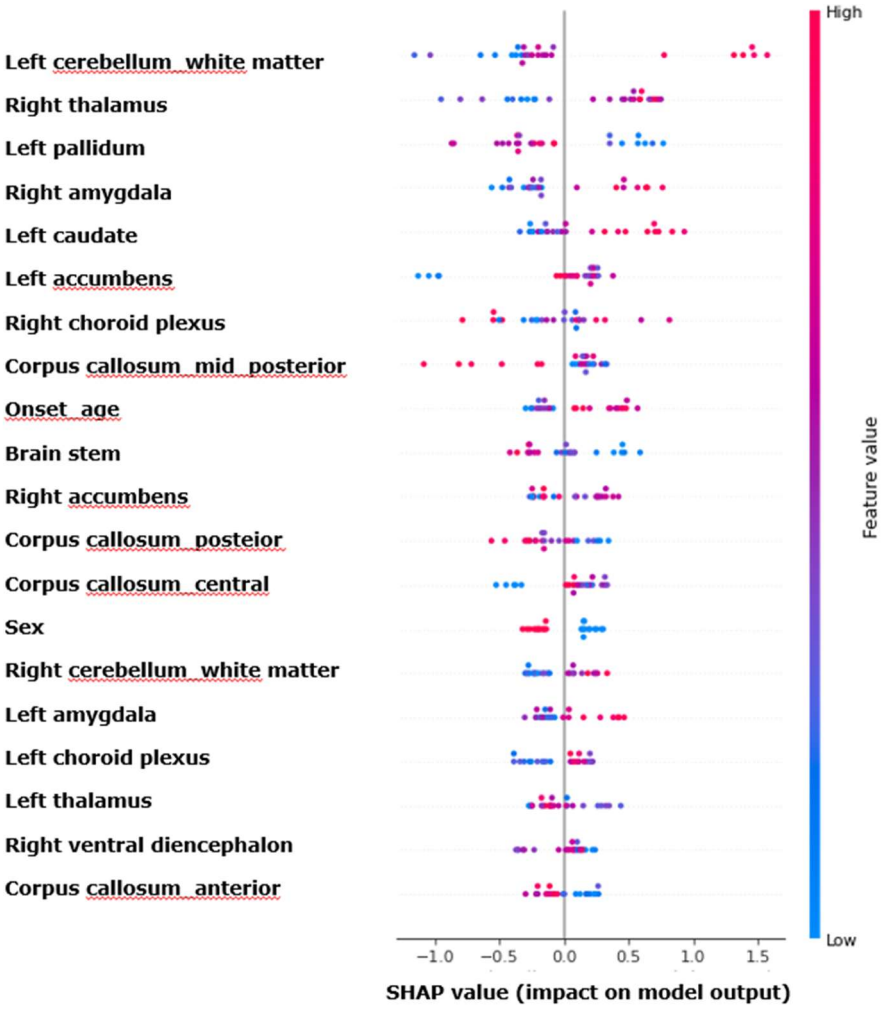

Figure 3. SHAP value summary plot for the combined clinical and volumetric XGBoost model. SHAP, Shapley additive explanations; XGBoost, extreme gradient boosting.

3.4.2. Cortical thickness model

Table 5 presents the performance of six machine learning models—logistic regression, random forest, XGBoost, LightGBM, SVM, and ANN—trained on the training set and validated on an independent test set. Among these models, XGBoost demonstrated the highest performance with an AUROC of 0.676, SVM with an AUROC of 0.603, and ANN with an AUROC of 0.522. The logistic regression model achieved an AUROC of 0.463, LightGBM an AUROC of 0.456, and random forest an AUROC of 0.430. For the best-performing XGBoost model, the accuracy was 0.760, precision was 0.758, recall was 0.760, and the F1-score was 0.733.

Figure 4 displays the importance of the feature as determined by SHAP values. Among the cortical thickness features identified by SHAP values as playing a crucial role in prognosis, the five most important were, in order: right superior parietal, right frontal pole, left caudal middle frontal, right temporal pole, and left parahippocampal cortex.

Table 5. Performances of the combined clinical and cortical thickness machine learning models on the test set

Models	Accuracy	Precision	Recall	F1-score	AUROC
Logistic Regression	0.600	0.533	0.600	0.554	0.463
Random Forest	0.520	0.549	0.520	0.532	0.430
XGBoost	0.760	0.758	0.760	0.733	0.676
LightGBM	0.600	0.533	0.600	0.554	0.456
SVM	0.680	0.462	0.680	0.551	0.603
ANN	0.560	0.574	0.560	0.566	0.522

AUROC, Area Under the Receiver Operating Characteristic curve; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machine; ANN, artificial neural network.

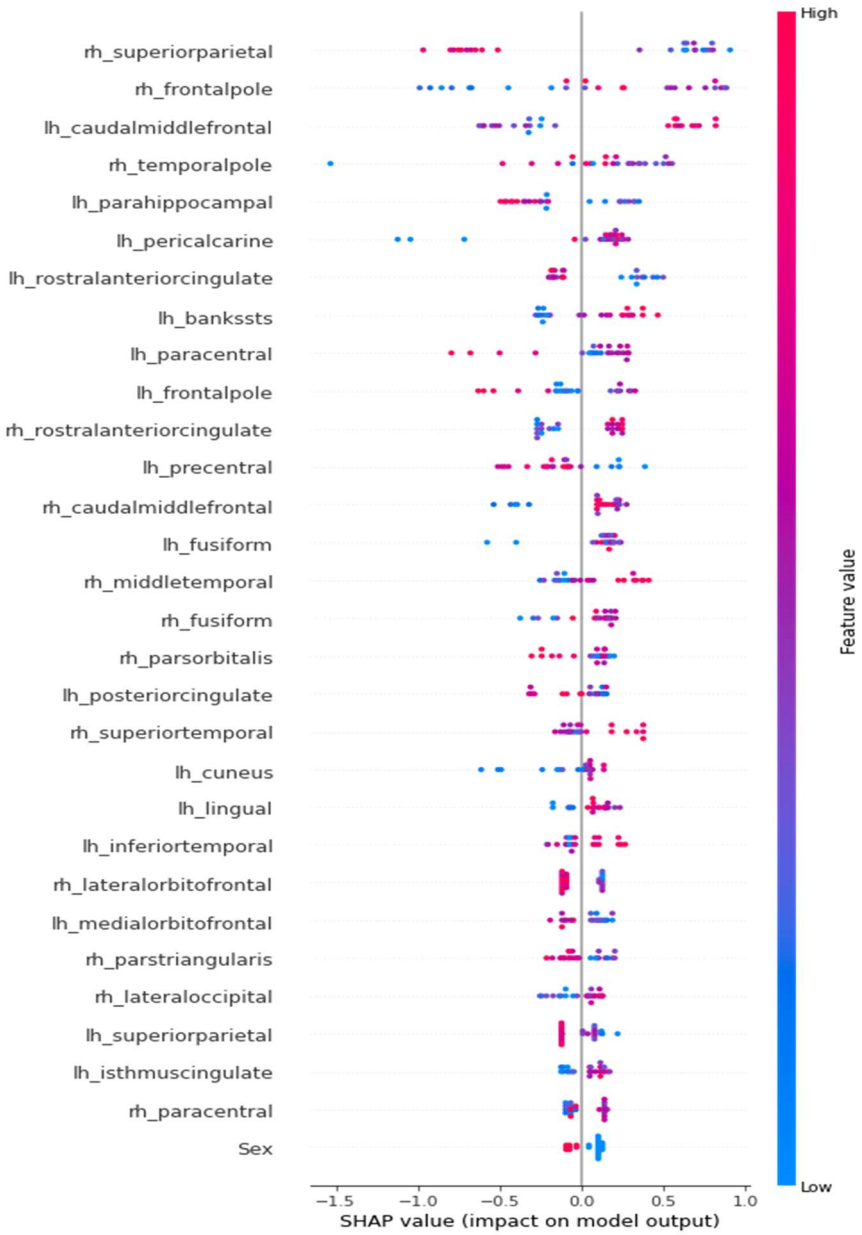

Figure 4. SHAP value summary plot for the combined clinical and cortical thickness XGBoost model. SHAP, Shapley additive explanations; XGBoost, extreme gradient boosting.

3.4.3. Radiomics model

Table 6 presents the performance of six machine learning models—logistic regression, random forest, XGBoost, LightGBM, SVM, and ANN—trained on the training set and validated on an independent test set. Among these models, XGBoost demonstrated the highest performance with an AUROC of 0.824, LightGBM with an AUROC of 0.772, and random forest with an AUROC of 0.772. The logistic regression model achieved an AUROC of 0.691, SVM an AUROC of 0.500, and ANN an AUROC of 0.485.

Table 6. Performances of Radiomics machine learning models on the test set

Models	Accuracy	Precision	Recall	F1-score	AUROC
Logistic Regression	0.720	0.712	0.720	0.715	0.691
Random Forest	0.720	0.802	0.720	0.635	0.768
XGBoost	0.720	0.702	0.720	0.699	0.824
LightGBM	0.760	0.758	0.760	0.733	0.772
SVM	0.680	0.462	0.680	0.551	0.500
ANN	0.680	0.661	0.680	0.666	0.485

AUROC, Area Under the Receiver Operating Characteristic curve; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machine; ANN, artificial neural network.

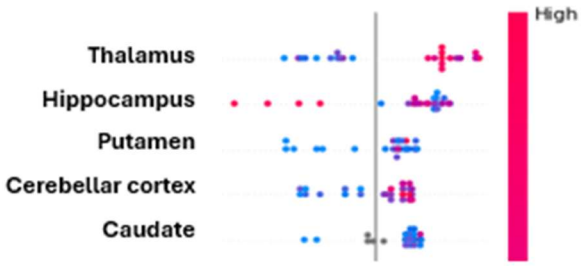

Figure 5. SHAP value summary plot for the XGBoost model. SHAP, Shapley additive explanations; XGBoost, extreme gradient boosting.

3.4.4. Combined model

Table 7 presents the performance of six machine learning models—logistic regression, random forest, XGBoost, LightGBM, SVM, and ANN—trained on the training set and validated on an independent test set, using clinical, volumetric, cortical thickness, and Radiomics data. Random forest demonstrated the highest performance with an AUROC of 0.923, LightGBM with an AUROC of 0.750, and logistic regression with an AUROC of 0.750. The XGBoost model achieved an AUROC of 0.735, ANN an AUROC of 0.559, and SVM an AUROC of 0.500. For the best-performing random forest model, the accuracy was 0.840, precision was 0.859, recall was 0.840, and the F1-score was 0.844.

Table 7. Performance of machine learning models combining clinical, volumetric, cortical thickness, and Radiomics data on the test set

Models	Accuracy	Precision	Recall	F1-score	AUROC
Logistic Regression	0.720	0.712	0.720	0.715	0.750
Random Forest	0.840	0.859	0.840	0.844	0.923
XGBoost	0.760	0.750	0.760	0.750	0.735
LightGBM	0.720	0.708	0.720	0.674	0.750
SVM	0.680	0.462	0.680	0.551	0.500
ANN	0.680	0.704	0.680	0.688	0.559

AUROC, Area Under the Receiver Operating Characteristic curve; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machine; ANN, artificial neural network.

IV. DISCUSSION

This study highlights several key findings that contribute to our understanding of prognostic factors in JME. First, we identified that male gender, the volumes of the left amygdala and right hippocampus, and the cortical thickness of the bilateral temporal poles, left entorhinal cortex, fusiform gyrus, and right inferior and middle temporal cortex were significantly associated with a favorable prognosis. Second, our prognostic prediction models that combined clinical and radiological variables outperformed those based on clinical variables alone. The model that integrated clinical data with volumetric, cortical thickness, and Radiomics data demonstrated the highest predictive performance. Third, we found that brain structures such as the thalamus and hippocampus, which are already known to be involved in the pathophysiology of JME, also emerged as significant features in the prognostic prediction models.

The association of the male gender with better outcomes is consistent with previous studies on JME prognosis.[9,17] This may be attributed to the broader range of antiseizure medications available to males, particularly compared to females of childbearing age, who have more limited treatment options due to teratogenic risks.

In our analysis of brain subcortical volumes, the amygdala and hippocampus—regions known to be involved in emotion and cognition—were identified as important structures associated with prognosis. These findings align with earlier studies that observed differences in these brain regions between JME patients and healthy controls, suggesting that these structures may play a critical role in influencing disease outcomes.[18]

Regarding cortical thickness, our study found that thinning in several regions, including the bilateral temporal poles, fusiform gyrus, entorhinal cortex, and middle and inferior temporal cortex, was associated with prognosis. Previous studies comparing JME patients to healthy controls have reported thinning in the temporal cortex and fusiform gyrus.[12,19] Our findings extend this knowledge by demonstrating a link between these pathological changes and treatment outcomes, suggesting that cortical thinning may serve as a marker for prognosis in JME.

In developing a novel prognostic prediction model for JME, we combined clinical and radiological variables, showing that this multimodal approach provides superior predictive performance compared to models relying solely on either data type. This is consistent with recent advances in artificial intelligence in clinical settings, where the integration of multimodal data is

increasingly recognized as essential for enhancing diagnostic and prognostic capabilities. Our previous research has highlighted the value of radiological variables in diagnosing and classifying JME and other generalized epilepsies, demonstrating that microstructural brain changes observed on MRI can improve diagnostic accuracy.[20,21] Building on these insights, the current study further validates the importance of combining clinical and radiological data to improve prognostic predictions, emphasizing the potential of AI models to assist clinicians—especially non-epileptologists—in making more informed decisions regarding epilepsy management.

In our study, we employed various machine learning models for prognostic prediction, with tree-based models such as XGBoost, Random Forest, and LightGBM generally performing the best. These models likely excelled because they are well-suited for tasks that require classification based on multiple variables. Models combining radiological features—including volumetry, cortical thickness, and radiomics data—outperformed those based solely on clinical variables. The model that combined clinical, volumetric, cortical thickness, and radiomics data achieved the highest predictive performance, suggesting that multimodal data models hold significant promise for clinical applications.

Moreover, the thalamus, a region well-documented to undergo microstructural changes in JME, was confirmed in our study as critically important for prognosis.[22,23] Specific volume alterations in the thalamus, along with disrupted thalamo-frontal connectivity, have been implicated in seizure regulation, further underscoring the thalamus's role in the underlying mechanisms and clinical outcomes of the disease.[24,25]

Despite these significant findings, our study has limitations that warrant consideration. The relatively small sample size, coupled with the large number of variables used to develop the prediction models, may have limited the models' performance on external validation sets. Additionally, our multimodal approach was somewhat restricted, as it relied solely on structural T1-weighted 3D MRI scans and excluded other radiological variables such as diffusion tensor image (DTI) and functional MRI (fMRI) data, which could potentially enhance model accuracy. Future research should aim to address these limitations by utilizing larger, more homogeneous samples and incorporating a broader range of multimodal data.

Despite these challenges, our study presents several notable strengths. It is the first to combine clinical and radiological variables to develop a prognostic model for JME, demonstrating that radiological data significantly enhance the predictive power of clinical variables. We

successfully identified models with appropriate performance levels using advanced machine learning techniques. Additionally, our study confirmed that both clinical and radiological variables previously associated with JME prognosis are indeed significant, thus providing clinical explainability and reinforcing the relevance of these factors in understanding the disease's mechanisms.

V. CONCLUSION

This study demonstrates the utility of machine learning models that integrate clinical and radiological data to predict prognosis in JME. We identified key prognostic factors, including male gender, volumetric differences in the left amygdala and right hippocampus, and cortical thickness in specific brain regions such as the bilateral temporal poles and entorhinal cortex. The combined model, incorporating clinical, volumetric, cortical thickness, and Radiomics data, outperformed models relying on single data types, highlighting the importance of a multimodal approach in predicting treatment outcomes in JME. Additionally, our findings reaffirm the critical role of brain structures such as the thalamus and hippocampus, which are already associated with JME's pathophysiology, in influencing disease prognosis. Despite the limitations related to sample size and data modalities, this study provides a foundation for further research into AI-driven prognostic tools in epilepsy management, with the potential to improve clinical decision-making and patient outcomes.

References

1. Asadi-Pooya AA, Brigo F, Lattanzi S, Blumcke I. Adult epilepsy. *Lancet* 2023;402:412-424.
2. Jeon J-Y, Lee H, Shin J-Y, Moon H-J, Lee S-Y, Kim J-M. Increasing Trends in the Incidence and Prevalence of Epilepsy in Korea. *J Clin Neurol* 2021;17:393-399.
3. Baykan B, Wolf P. Juvenile myoclonic epilepsy as a spectrum disorder: A focused review. *Seizure* 2017;49:36-41.
4. Grünewald RA, Panayiotopoulos CP. Juvenile myoclonic epilepsy. A review. *Arch Neurol* 1993;50:594-598.
5. Proposal for classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy. *Epilepsia* 1985;26:268-278.
6. Baykan B, Martínez-Juárez IE, Altindag EA, Camfield CS, Camfield PR. Lifetime prognosis of juvenile myoclonic epilepsy. *Epilepsy Behav* 2013;28 Suppl 1:S18-24.
7. Senf P, Schmitz B, Holtkamp M, Janz D. Prognosis of juvenile myoclonic epilepsy 45 years after onset: seizure outcome and predictors. *Neurology* 2013;81:2128-2133.
8. Stevelink R, Koeleman BPC, Sander JW, Jansen FE, Braun KPJ. Refractory juvenile myoclonic epilepsy: a meta-analysis of prevalence and risk factors. *Eur J Neurol* 2019;26:856-864.
9. Giuliano L, Mainieri G, Aguglia U, Bilo L, Durante V, Ermio C, et al. Long-term prognosis of juvenile myoclonic epilepsy: A systematic review searching for sex differences. *Seizure* 2021;86:41-48.
10. Rubboli G, Beier CP, Selmer KK, Syvertsen M, Shakeshaft A, Collingwood A, et al. Variation in prognosis and treatment outcome in juvenile myoclonic epilepsy: a Biology of Juvenile Myoclonic Epilepsy Consortium proposal for a practical definition and stratified medicine classifications. *Brain Commun* 2023;5:fcad182.
11. Guaranha MS, Filho GM, Lin K, Guilhoto LM, Caboclo LO, Yacubian EM. Prognosis of juvenile myoclonic epilepsy is related to endophenotypes. *Seizure* 2011;20:42-48.
12. Wandschneider B, Hong SJ, Bernhardt BC, Fadaie F, Vollmar C, Koepp MJ, et al. Developmental MRI markers cosegregate juvenile patients with myoclonic epilepsy and their healthy siblings. *Neurology* 2019;93:e1272-e1280.
13. Yu Y, Chen Z, Yang Y, Zhang J, Wang Y. Development and validation of an interpretable

machine learning model for predicting post-stroke epilepsy. *Epilepsy Res* 2024;205:107397.

14. Han K, Liu C, Friedman D. Artificial intelligence/machine learning for epilepsy and seizure diagnosis. *Epilepsy Behav* 2024;155:109736.
15. Kaushik M, Mahajan S, Machahary N, Thakran S, Chopra S, Tomar RV, et al. Predicting efficacy of antiseizure medication treatment with machine learning algorithms in North Indian population. *Epilepsy Res* 2024;205:107404.
16. Hakeem H, Feng W, Chen Z, Choong J, Brodie MJ, Fong SL, et al. Development and Validation of a Deep Learning Model for Predicting Treatment Response in Patients With Newly Diagnosed Epilepsy. *JAMA Neurol* 2022;79:986-996.
17. Shakeshaft A, Panjwani N, McDowall R, Crudgington H, Peña Ceballos J, Andrade DM, et al. Trait impulsivity in Juvenile Myoclonic Epilepsy. *Ann Clin Transl Neurol* 2021;8:138-152.
18. Zhang J, Wu D, Yang H, Lu H, Ji Y, Liu H, et al. Correlations Between Structural Brain Abnormalities, Cognition and Electroclinical Characteristics in Patients With Juvenile Myoclonic Epilepsy. *Frontiers in Neurology* 2022;13.
19. Kim SH, Lim S-C, Kim W, Kwon O-h, Jeon S, Lee J-M, et al. Extrafrontal structural changes in juvenile myoclonic epilepsy: A topographic analysis of combined structural and microstructural brain imaging. *Seizure* 2015;30:124-131.
20. Kim KM, Hwang H, Sohn B, Park K, Han K, Ahn SS, et al. Development and Validation of MRI-Based Radiomics Models for Diagnosing Juvenile Myoclonic Epilepsy. *Korean J Radiol* 2022;23:1281-1289.
21. Sim Y, Lee SK, Chu MK, Kim WJ, Heo K, Kim KM, et al. MRI-Based Radiomics Approach for Differentiating Juvenile Myoclonic Epilepsy from Epilepsy with Generalized Tonic-Clonic Seizures Alone. *J Magn Reson Imaging* 2024;60:281-288.
22. Mory SB, Betting LE, Fernandes PT, Lopes-Cendes I, Guerreiro MM, Guerreiro CA, et al. Structural abnormalities of the thalamus in juvenile myoclonic epilepsy. *Epilepsy Behav* 2011;21:407-411.
23. Betting LE, Mory SB, Lopes-Cendes I, Li LM, Guerreiro MM, Guerreiro CA, et al. MRI reveals structural abnormalities in patients with idiopathic generalized epilepsy. *Neurology* 2006;67:848-852.
24. O'Muircheartaigh J, Vollmar C, Barker GJ, Kumari V, Symms MR, Thompson P, et al. Abnormal thalamocortical structural and functional connectivity in juvenile myoclonic epilepsy. *Brain* 2012;135:3635-3644.

25. Gauffin H, Landtblom AM, Vigren P, Frick A, Engström M, McAllister A, et al. Similar Profile and Magnitude of Cognitive Impairments in Focal and Generalized Epilepsy: A Pilot Study. *Front Neurol* 2021;12:746381.

Abstract in Korean

청소년근간대뇌전증의 예후 예측을 위한 임상양상과 영상변수를 결합한 인공지능 모델의 개발과 검증

서론: 청소년근간대뇌전증은 비교적 약물치료에 대한 예후가 좋으나, 35%의 환자는 약물난치성 뇌전증으로 고통받고 있고, 현재까지 다양한 임상적 특성이 예후에 영향을 미치는 것으로 알려져 있다. 청소년근간대뇌전증의 예후 예측에 영향을 미치는 임상적 요인에 대해 많은 연구가 이루어졌지만, 임상적 특성과 영상의학적인 특성을 통합하여 예후를 예측하는 모델에 대한 연구는 제한적이다. 본 연구는 임상양상과 영상변수를 결합하여 청소년근간대뇌전증 환자의 예후를 예측하는 인공지능 모델을 개발하고 검증하는 것을 목표로 한다.

방법: 국내 5개 대학병원에서 청소년근간대뇌전증으로 진단된 전체 125명(내부훈련용데이터 – 1개병원 99명, 외부검증용데이터 – 5개병원 26명)의 환자를 대상으로 후향적 연구를 수행하였다. 인구통계학적 정보, 임상특성, 약물투약, 발작이력의 임상 데이터를 수집하였고, MRI 데이터를 이용하여 피질하 구조물의 부피측정, 대뇌피질두께측정, 라디오믹스 특성 추출등으로 뇌 구조의 정량적인 특성을 추출하였다. 추출된 변수들을 6가지 인공지능 방법 – logistic regression, random forest, XGBoost, LightGBM, SVM, ANN – 을 이용하여, 2년간의 무발작기간을 예측하는 기계 학습 모델을 개발하고, 정확도, 정밀도, 재현율, F1 점수, AUROC 지표를 사용하여 성능을 평가하였다. 모델은 내부 데이터셋에서 학습되었으며 독립적인 외부 데이터셋을 통해 검증되었으며, 모델이 이용하는 특성 중요도도 확인하였다.

결과: 좋은 예후와 관련된 임상변수로는 남성, 영상변수로는 좌측 편도체와 우측 해마의 부피, 피질두께에서는 양측 측두엽극(temporal pole), 좌측 내후각피질(entorhinal cortex), 방추상 회(fusiform gyrus), 그리고 우측 하측(inferior) 및 중간 (middle) 측두엽피질(temporal gyrus)의 두께가 좋은 예후와 유의하게 연관된 것으로 나타났다. 가장 우수한 성능을 보인 기계 학습 모델은 임상 데이터, 부피 측정, 피질 두께, 방사선학적 데이터를 결합한 모델 (AUROC 0.923)로 단일 데이터 유형에 의존한 모델(AUROC 0.600)보다 우수한 성능을 보였다. 또한 시상(thalamus)과 해마(hippocampus)와 같이

청소년근간대뇌전증의 병리생리학적 기전에 관여하는 뇌 구조가 예후 예측에 중요한 특징으로 확인되었다.

결론: 본 연구는 임상양상 및 영상변수를 통합한 기계 학습 모델이 청소년근간대뇌전증의 예후를 예측하는 데 유용할 수 있음을 보여주었다. 다양한 데이터를 결합하는 기반으로 한 모델이 단일 데이터 유형에 기반한 모델보다 더 효과적이라는 점을 시사하며, 이는 뇌전증의 관리에서 치료 결과를 예측하는 데 있어 임상의에게 보조적인 도구로 사용될 수 있다. 이 연구의 결과는 더 큰 규모의 다양한 인구를 대상으로 추가적인 검증이 필요하며, 확산텐서영상 및 기능적 자기공명영상과 같은 추가 방사선학적 기법을 포함하여, 연구가 기대된다.

핵심되는 말 : 청소년근간대뇌전증, 예후, 라디오믹스, 머신러닝