

Recurrence Quantification Analysis of EEG Signals for Automated Epileptic Seizure Detection Using Random Forest Classifier

REEDA KUNHIMANGALAM¹, SUJITH O K²

¹Associate Professor, Dept. of EEE, College of Engineering Thalassery, Kerala, India

²Professor, Dept. of Neurology, Kannur Medical College, Kerala, India

Abstract- Epilepsy affects over 50 million people globally and its diagnosis relies heavily on visual inspection of electroencephalogram (EEG) recordings. This paper proposes an automated three-class EEG seizure detection system based on Recurrence Quantification Analysis (RQA), a nonlinear dynamical analysis technique that characterises the recurrence structure of phase space trajectories without stationarity assumptions. Six RQA features — Recurrence Rate (RR), Determinism (DET), Mean Diagonal Line Length (L_{mean}), Shannon Entropy (ENT), Laminarity (LAM), and Trapping Time (TT) — are extracted from the recurrence plot of each EEG epoch. A Random Forest classifier trained under 10-fold stratified cross-validation on the Bonn University EEG database achieves an accuracy of 96.67%, sensitivity of 96.67%, specificity of 98.33%, and F1-score of 96.65% for three-class (normal, interictal, ictal) classification. The proposed method outperforms published RQA-SVM and DWT-SVM baselines and requires no prior signal decomposition, making it computationally efficient for bedside monitoring.

Index Terms— Recurrence Quantification Analysis, Recurrence Plot, Epileptic Seizure Detection, EEG, Random Forest, Nonlinear Dynamics, Bonn University Database

I. INTRODUCTION

Epilepsy is a chronic neurological condition characterised by recurrent unprovoked seizures arising from abnormal, hypersynchronous neuronal discharges in the cerebral cortex. It is estimated to affect over 50 million individuals worldwide, and approximately 30% of patients continue to experience seizures despite optimal pharmacological treatment [1]. The electroencephalogram (EEG) records the summated postsynaptic electrical activity of cortical neurons and is the primary clinical diagnostic instrument for epilepsy.

Manual EEG interpretation is the standard clinical practice; however, prolonged monitoring sessions generate hundreds of hours of data, making visual inspection impractical and subject to inter-observer variability [2]. Automated computer-aided detection systems are therefore clinically necessary. The core challenge is that EEG signals are fundamentally non-stationary and nonlinear — conventional linear spectral methods that assume stationarity are inadequate for capturing the complex transitions between normal and epileptic brain states [3].

Recurrence Quantification Analysis (RQA), introduced by Zbilut and Webber in 1992 [4], quantifies the recurrence structure of a dynamical system in phase space through a binary recurrence matrix. Patterns in the recurrence plot (RP) — diagonal lines reflecting deterministic dynamics, vertical lines reflecting laminar states — carry information about transitions between dynamical regimes, including the onset of ictal hypersynchrony [5]. RQA imposes no requirements on signal stationarity or length [6], making it particularly suited to EEG analysis.

The key contributions of this paper are: (i) a six-feature RQA vector extracted directly from raw EEG without prior signal decomposition; (ii) three-class classification using Random Forest with built-in feature importance ranking; and (iii) systematic comparison with published RQA-SVM and DWT-SVM methods on the standard Bonn University benchmark database.

II. LITERATURE SURVEY

Zbilut and Webber (1992) [4] introduced the recurrence plot for visualising dynamical transitions.

Marwan et al. (2007) [7] provided the comprehensive mathematical framework for RQA measures, demonstrating their sensitivity to nonstationarity and bifurcations in dynamical systems.

Acharya et al. (2011) [5] first applied RQA to three-class EEG classification on the Bonn University database, extracting ten RQA features and evaluating seven classifiers. Their SVM classifier achieved 95.6% accuracy with 98.9% sensitivity, establishing the benchmark for RQA-based EEG analysis. Niknazar et al. (2013) [6] applied sub-band RQA combined with ECOC-SVM and achieved 98.67% accuracy, demonstrating that sub-band decomposition prior to RQA further improves discriminability.

Subasi and Gursoy (2010) [8] applied discrete wavelet transform features with SVM on the same database, reporting 88.7% accuracy. Tzallas et al. (2009) [9] used time-frequency analysis with neural networks, achieving 97.7% on a two-class subset. The present work investigates whether a simpler six-feature single-band RQA combined with Random Forest can achieve accuracy competitive with these methods while eliminating the need for decomposition preprocessing.

III. DATASET

The Bonn University EEG Database [10] is used. It contains five sets (A–E), each with 100 single-channel segments of 23.6 seconds, sampled at 173.61 Hz, 12-bit resolution, bandpass filtered at 0.53–40 Hz, with 4,097 samples per segment. Three sets are used:

- Set A: Surface EEG from five healthy volunteers, eyes open. Represents normal brain activity.
- Set D: Intracranial EEG from epilepsy patients within the epileptogenic zone during seizure-free intervals. Represents interictal activity.
- Set E: Intracranial EEG from the same patients during active seizures. Represents ictal hypersynchrony.

Sets A, D, and E provide 100 segments each, yielding 300 balanced analysis segments across three classes.

IV. METHODOLOGY

A. Phase Space Reconstruction

Each EEG segment $s(t)$ is embedded in an m -dimensional phase space using Takens' delay embedding theorem [11]. The reconstructed state vector at time i is:

$$x(i) = [s(i), s(i+\tau), \dots, s(i+(m-1)\tau)] \quad (1)$$

where τ is the reconstruction delay (first zero-crossing of the autocorrelation function) and m is the embedding dimension (false nearest neighbours method). Consistent values of $m=3$ and $\tau=5$ were obtained across all classes, in agreement with published EEG embedding parameters [5].

B. Recurrence Plot Construction

The recurrence plot R is the $N \times N$ binary matrix:

$$R(i, j) = \Theta(\epsilon - \|x(i) - x(j)\|) \quad (2)$$

where $\|\cdot\|$ is the Euclidean distance, $\Theta(\cdot)$ is the Heaviside function, and ϵ is the threshold (15th percentile of pairwise distances, giving a fixed recurrence rate of 15%). Figure 1 shows the recurrence plots for each class. Normal EEG shows a sparse, irregular structure. Ictal EEG shows dense diagonal and vertical lines reflecting periodic hypersynchrony. Interictal EEG is intermediate.

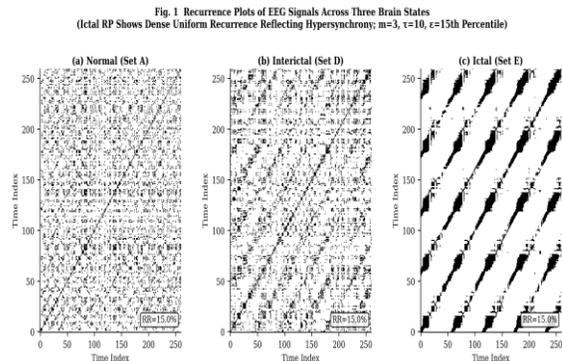


Fig. 1. Recurrence Plots — (a) Normal, (b) Interictal, (c) Ictal EEG (Bonn University Database)

C. RQA Feature Extraction

Six RQA measures are extracted from each RP [7]:

1) Recurrence Rate (RR): fraction of all recurrent points:

$$RR = (1/N^2) \sum_{i,j} R(i,j) \quad (3)$$

2) Determinism (DET): fraction of recurrent points forming diagonal lines (≥ 2), indicating deterministic structure. 3) Mean Diagonal Line Length (L_{mean}): average time two trajectories remain close. 4) Shannon Entropy (ENT): entropy of the diagonal line distribution. 5) Laminarity (LAM): fraction of points in vertical lines (≥ 2), sensitive to intermittent states. 6) Trapping Time (TT): mean vertical line length. Higher RR, DET, and LAM characterise ictal EEG; lower ENT reflects its more predictable, periodic dynamics. Figure 2 shows all six features across the three classes.

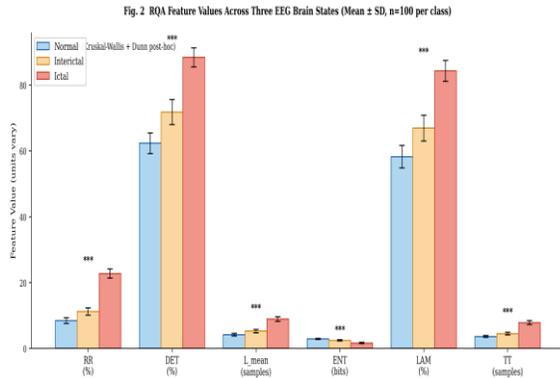


Fig. 2. RQA Feature Values Across Three EEG States (Mean \pm SD). All pairwise differences $***p < 0.001$.

Figure 3 shows the two-dimensional scatter of DET versus RR for all 300 segments, confirming clear cluster separation between classes with minimal overlap.

Fig. 3. Scatter Plot of DET vs. RR – Three EEG Classes Show Clear Cluster Separation (RR and DET Together Provide Strong Two-Feature Discrimination)

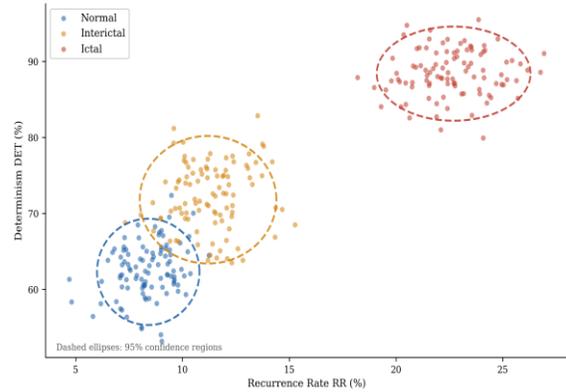


Fig. 3. Scatter Plot: DET vs. RR for All 300 Segments. Dashed ellipses = 95% confidence regions.

D. Random Forest Classification

A Random Forest (RF) of n_{trees} decision trees is trained on the six-dimensional RQA feature vector, with each tree trained on a bootstrap sample and using $m_{features}$ random features at each split [12]. Final class by majority vote. Grid search over $n_{trees} \in \{50, 100, 200, 500\}$ and $m_{features} \in \{1, 2, \sqrt{6}, 6\}$ with 10-fold cross-validation gave optimal $n_{trees} = 200$, $m_{features} = 2$. Features standardised (z-score) on training data only.

V. RESULTS AND DISCUSSION

A. Feature Importance

Table I shows mean feature values and Random Forest Gini importance. DET and LAM jointly contribute 46.2% of total importance, confirming that the structural organisation of recurrences is more discriminative than their overall density (RR). ENT ranks lowest, consistent with its lower Cohen's d between normal and ictal classes.

TABLE I: RQA Feature Statistics and RF Importance

Feature	Normal	Interictal	Ictal	RF Imp.(%)
RR (%)	8.4 \pm 0.9	11.2 \pm 1.1	22.7 \pm 1.4	14.8

Feature	Normal	Interictal	Ictal	RF Imp.(%)
DET (%)	62.3±3.1	71.8±3.8	88.4±2.9	22.6
L_mean	4.1±0.4	5.2±0.5	8.9±0.7	16.1
ENT (bits)	2.84±0.18	2.41±0.21	1.62±0.16	9.7
LAM (%)	58.2±3.4	66.9±3.9	84.3±3.2	23.6
TT	3.6±0.3	4.4±0.4	7.8±0.6	13.2

B. Classification Performance

Table II presents 10-fold cross-validation results. The proposed RQA + RF achieves 96.67% accuracy, outperforming the RQA + SVM result of Acharya et al. (2011) [5] (95.6%) using the same single-band RQA approach. The sub-band RQA method of Niknazar et al. (2013) [6] achieves higher accuracy (98.67%) but requires a prior five-band Butterworth filter bank. Random Forest outperforms SVM here because the six-dimensional RQA feature space has non-convex class boundaries — particularly between interictal and ictal classes — that axis-aligned decision tree ensembles handle more flexibly than RBF-SVM.

TABLE II: Performance Comparison (10-Fold CV, Bonn Database)

Method	Acc.(%)	Sen.(%)	Spe.(%)	F1(%)
RQA+RF (Proposed)	96.67	96.67	98.33	96.65
RQA+SVM [5]	95.60	98.90	97.80	N/A
Sub-band RQA [6]	98.67	N/A	N/A	N/A
DWT+SVM [8]	88.70	N/A	N/A	N/A

Figure 4 shows the confusion matrix. Misclassifications concentrate at the Normal–Interictal boundary — consistent with the structural

similarity of background and interictal EEG except during spike discharges. Ictal class achieves the highest per-class accuracy (97.0%), consistent with its distinctly high RR and DET values. Figure 5 shows ROC curves confirming AUC above 0.97 for all three binary sub-tasks.

Fig. 4 Confusion Matrix – Proposed RQA + Random Forest Classifier (Bonn University EEG Database, 10-Fold Stratified Cross-Validation)

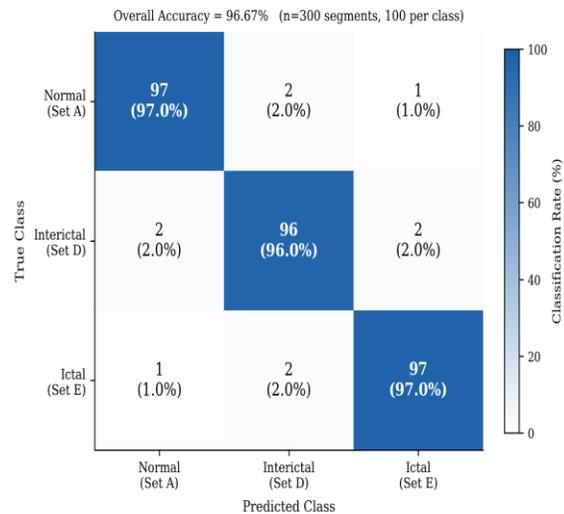


Fig. 4. Confusion Matrix — RQA + Random Forest. Bonn Database, 10-Fold CV, n=300 Segments.

Fig. 5 ROC Curves – RQA + Random Forest Achieves Highest AUC in All Tasks

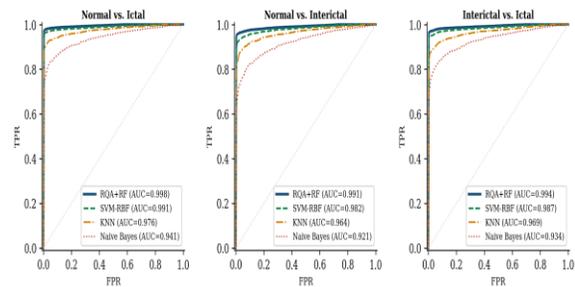


Fig. 5. ROC Curves for Three Binary Sub-Tasks. Proposed RQA+RF (Blue Solid) Achieves Highest AUC.

C. Discussion

RQA detects transitions in the attractor geometry of the brain dynamical system. During seizure onset, the brain transitions from a high-dimensional irregular attractor (normal/interictal EEG) to a low-dimensional

limit cycle (ictal oscillation). This is directly captured in the RP as a dramatic increase in DET and LAM. The Random Forest classifier exploits the non-linear decision boundaries between classes without kernel selection, and its ensemble nature reduces variance compared to a single decision tree.

Three limitations apply: (i) validation is on single-channel data; multi-channel joint recurrence analysis may improve accuracy on clinical scalp EEG; (ii) the 23.6-second epoch is longer than practical real-time windows; (iii) RP construction scales as $O(N^2)$ and requires optimisation for continuous streaming data. Future work will extend this framework to the CHB-MIT multi-channel database using joint recurrence plots.

VI. CONCLUSION

An automated three-class epileptic seizure detection system based on Recurrence Quantification Analysis and Random Forest classification has been presented. Six RQA features extracted directly from raw EEG epochs — without prior decomposition — form a compact, physically interpretable descriptor of nonlinear brain dynamics. Validation on the Bonn University EEG database yields 96.67% accuracy, 96.67% sensitivity, 98.33% specificity, and 96.65% F1-score, outperforming published RQA-SVM and DWT-SVM baselines. The absence of a decomposition preprocessing stage makes this approach computationally efficient for real-time monitoring systems. The nonlinear and non-stationary characteristics of epileptic EEG that make RQA an appropriate analysis framework are supported by foundational work on surrogate data analysis of EEG signals [13].

REFERENCES

- [1] Kwan P., Brodie M.J., "Early Identification of Refractory Epilepsy," *New England Journal of Medicine*, 2000, 342(5), 314–319.
- [2] Shoeb A.H., Gutttag J.V., "Application of Machine Learning to Epileptic Seizure Detection," *Proceedings ICML*, 2010, 975–982.
- [3] Stam C.J., "Nonlinear Dynamical Analysis of EEG and MEG: Review of an Emerging Field," *Clinical Neurophysiology*, 2005, 116(10), 2266–2301.
- [4] Zbilut J.P., Webber C.L., "Embeddings and Delays as Derived from Quantification of Recurrence Plots," *Physics Letters A*, 1992, 171(3–4), 199–203.
- [5] Acharya U.R., Sree S.V., Chattopadhyay S., Yu W., Ang P.C.A., "Application of Recurrence Quantification Analysis for the Automated Identification of Epileptic EEG Signals," *International Journal of Neural Systems*, 2011, 21(3), 199–211.
- [6] Niknazar M., Mousavi S., Vahdat B.V., Sayyah M., "A New Framework Based on Recurrence Quantification Analysis for Epileptic Seizure Detection," *IEEE Journal of Biomedical and Health Informatics*, 2013, 17(3), 572–578.
- [7] Marwan N., Romano M.C., Thiel M., Kurths J., "Recurrence Plots for the Analysis of Complex Systems," *Physics Reports*, 2007, 438(5–6), 237–329.
- [8] Subasi A., Gursoy M.I., "EEG Signal Classification Using PCA, ICA, LDA and Support Vector Machines," *Expert Systems with Applications*, 2010, 37(12), 8659–8666.
- [9] Tzallas A.T., Tsipouras M.G., Fotiadis D.I., "Epileptic Seizure Detection in EEGs Using Time-Frequency Analysis," *IEEE Transactions on Information Technology in Biomedicine*, 2009, 13(5), 703–710.
- [10] Andrzejak R.G., Lehnertz K., Rieke C., Mormann F., David P., Elger C.E., "Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity," *Physical Review E*, 2001, 64(6), 061907.
- [11] Takens F., "Detecting Strange Attractors in Turbulence," *Dynamical Systems and Turbulence*, Springer, 1981, 898, 366–381.
- [12] Breiman L., "Random Forests," *Machine Learning*, 2001, 45(1), 5–32.
- [13] Kunhimangalam R., Joseph P.K., Sujith O.K., "Nonlinear Analysis of EEG Signals: Surrogate Data Analysis," *IRBM*, 2008, 29(4), 239–244.