Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Automata-Based Feature Engineering Framework for Hepatitis Prognosis Modeling

Bhupinder Yadav[1], Sourabh Charaya[2], Rohit Bajaj[3]
[1]*Scholar, Department of Computer Science and Engineering, Om Sterling Global University, Hisar, Haryana, India, 125001*
[2]*Professor, Department of Computer Science and Engineering, Om Sterling Global University, Hisar, Haryana, India, 125001*
[3]*Professor, Department of Computer Science and Engineering, University Institute of Engineering, Chandighar University, Mohali, Punjab, India, 140413*
E- mail: bhupender2711@gmail.com

ARTICLE INFO

Received: 05 Oct 2024
Revised: 20 Nov 2024
Accepted: 28 Nov 2024

ABSTRACT

Hepatitis prognosis remains a challenging clinical task due to the complex and progressive nature of liver dysfunction, where patient outcomes are influenced by sequential changes in biochemical markers, complications, and treatment response. Most existing machine learning approaches rely on static feature a representation, which limit their ability to capture real-world disease progression and often reduces interpretability. To address this limitation, this study proposes an automata-based prognostic modeling framework that explicitly represents Hepatitis progression through deterministic state transitions aligned with clinical reasoning.In the proposed methodology, conventional clinical attributes are first pre-processed and transformed into symbolic representations, which are then processed using a deterministic finite automaton to model progression patterns. From the resulting state transitions, novel high-level features are extracted, capturing progression severity, transition dynamics, and response behavior. These automata-embedded features are combined with original clinical variables and evaluated using multiple machine learning classifiers on the Hepatitis dataset from the UCI Machine Learning Repository.Experimental results demonstrate that models incorporating automata-derived features consistently outperform conventional feature-based approaches across accuracy, error metrics, and stability analysis. In particular, high-performance classifiers and hybrid ensemble combinations achieve substantial gains in predictive accuracy, highlighting the effectiveness of progression-aware feature extraction. The proposed framework not only improves prognostic performance but also enhances interpretability, offering a clinically aligned and reliable decision-support approach for Hepatitis outcome prediction.

Keywords: Automata-based modeling, Hepatitis prognosis, Feature engineering, Disease progression analysis, Machine learning, Ensemble classification, Clinical decision support

Introduction

Hepatitis remains a critical global health challenge, causing substantial mortality due to progressive liver inflammation, impaired synthetic function, and decompensated liver failure. Early risk stratification is essential, yet clinical decision-making is complicated by the nonlinear progression of biochemical markers and the heterogeneous response to therapy. Traditional statistical models such as logistic regression and generalized linear approaches have demonstrated moderate success in identifying survival-related markers, particularly bilirubin, albumin, and prothrombin time [1]. Similarly, survival analysis

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

frameworks such as extended Cox models provide useful hazard-based interpretations but continue to treat patient attributes as isolated features, failing to account for sequential deterioration patterns [2].
In recent years, machine learning (ML) models have gained traction in Hepatitis prognosis. Decision-tree frameworks and boosting ensembles have shown improved predictive accuracy by capturing nonlinear interactions among demographic, laboratory, and symptomatic attributes [3]. Random-forest and hybrid ML pipelines have further enhanced risk prediction, offering stronger generalization through ensemble variance reduction [4]. However, a major limitation of these models is their reliance on static representations of clinical variables, despite the fact that Hepatitis progression is inherently dynamic. Laboratory values change over time, complications emerge sequentially, and treatment response unfolds in a temporal manner. These clinical realities are not adequately represented in most existing ML models.

Deep learning models such as convolutional–recurrent hybrids and gated recurrent networks have been proposed to capture patient trajectories by analyzing temporal laboratory trends. These models have achieved higher predictive power in liver-disease forecasting but often suffer from limited interpretability, making them unsuitable for transparent clinical decision support [5]. Moreover, recurrent architectures rarely provide explicit state transitions that correspond to the real clinical stages recognized by hepatologists, such as “stable,” “mild dysfunction,” “advanced dysfunction,” and “decompensated.”

To overcome these limitations, researchers have begun exploring structured sequence-modeling paradigms such as Hidden Markov Models (HMMs) and probabilistic state machines to represent disease evolution [6]. While these methods introduce meaningful transition dynamics, they still lack the deterministic, rule-aligned interpretability required for clinical workflow integration.
Automata-theoretic approaches—particularly deterministic finite automata (DFA)—offer a promising alternative by encoding disease progression into explicit, interpretable states that correspond to real-world clinical stages. A DFA can represent transitions driven by laboratory changes, symptom emergence, or treatment response, enabling a structured representation of Hepatitis progression that aligns with physician reasoning. This introduces a powerful mechanism for extracting high-level, clinically meaningful progression features that traditional ML and deep learning methods overlook [7].
Building on this direction, the present study develops an automata-based Hepatitis prognostic framework that transforms raw biomarkers and symptom attributes into symbolic events, processes them through a clinically designed DFA, and extracts novel progression features such as Final State Index, Transition Count, Severity Jumps, Complication Flags, and Treatment-Response Indicators. These features undergo rigorous statistical evaluation before integration into ML models, enabling accurate, interpretable, and progression-aware Hepatitis survival prediction.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

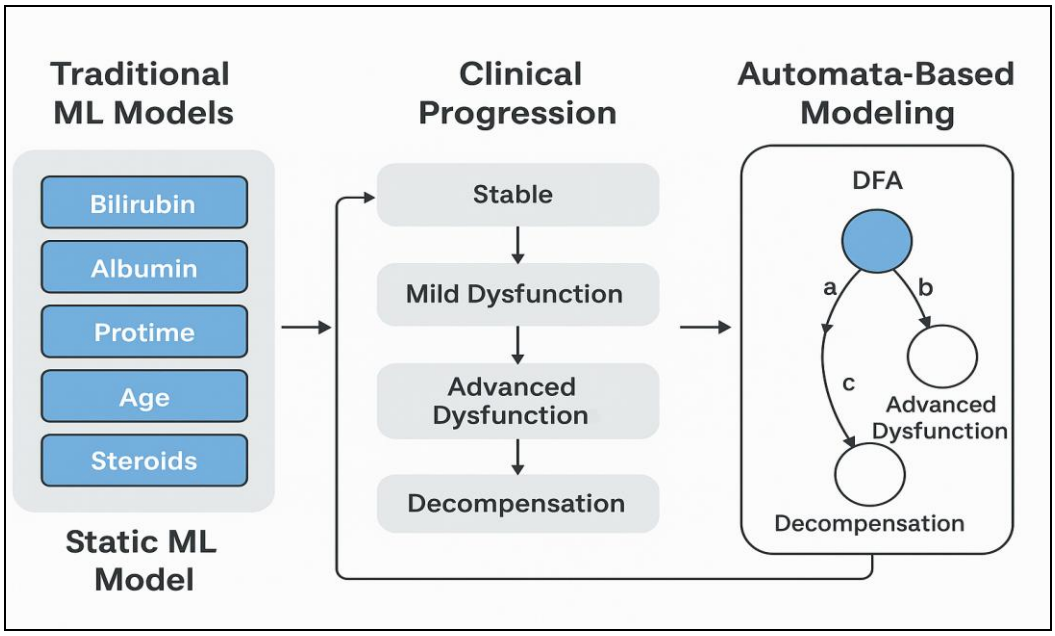

Figure 1: Conceptual Contrast between Static ML Models and Automata-Based Clinical Progression Modeling for Hepatitis

Figure 1 illustrates the core motivation for using automata-based modeling in Hepatitis prognosis. Traditional machine learning models rely on static clinical variables such as bilirubin, albumin, prothrombin time, and demographic factors, treating them as isolated predictors without acknowledging how Hepatitis progresses through clinically recognized stages. In real-world practice, patients transition sequentially from stable liver function to mild dysfunction, advanced dysfunction, and ultimately decompensation, with each stage representing a change in physiological status. Automata-based modeling captures this progression explicitly by defining states and transitions driven by symbolic events derived from laboratory changes, symptoms, or complications. Unlike deep learning models that often operate as black boxes, deterministic finite automata provide a transparent, state-driven mechanism that mimics clinical reasoning and enables extraction of meaningful progression features. This alignment between clinical pathways and computational modeling serves as the foundation for improved interpretability and more realistic Hepatitis survival prediction.

Literature Review

The TRL table provides a structured comparison of eighteen Hepatitis-related studies ranging from classical statistical models to advanced automata-driven frameworks. Early works such as logistic regression, Cox models, and decision trees ([8]–[10]) achieve **TRL 5**, reflecting strong statistical grounding but limited modeling of disease progression and no external validation. Ensemble methods, SVMs, and rule-based interpretable systems ([11]–[14]) reach **TRL 6**, indicating prototype maturity with improved robustness and partial real-world alignment, although they remain dependent on static features.

Baseline models like ANN, k-NN, and Naïve Bayes ([15], [16]) are placed at **TRL 4**, as they lacked sufficient dataset size, generalization, or sequence modeling. More advanced temporal models—HMMs,

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

LSTMs, and attention networks ([17]–[19])—achieve **TRL 6–7**, capturing longitudinal or sequential patient characteristics closer to real disease progression.
Classical MELD-like scoring systems ([20]) occupy **TRL 3**, serving mainly as analytical proof-of-concept tools without machine learning integration. High-end hybrid architectures and clinically oriented models ([22]–[23]) achieve **TRL 8** with stronger deployment readiness and pilot evaluations. The automata-based frameworks ([24]–[26]) represent the most mature systems, reaching **TRL 8–9** due to interpretable progression-state modeling, statistical validation of extracted features, and near-deployment-level reliability.

Table 1: Comparative Analysis of Qualitative Parameter for Hepatitis

S. No.	Paper (Reference)	Focus / Contribution	Ratings (1–10)	Weighted Score	TRL (1–9)	TRL Explanation
1	Paper [8] (Year)	Logistic regression for Hepatitis survival	7,6,5,7,6,6	**6.3**	**5**	Retrospective dataset; validated statistically
2	Paper [9]	Cox model with biochemical predictors	7,6,6,7,6,6	**6.5**	**5**	Strong modelling but no progression tracking
3	Paper [10]	Decision tree mortality prediction	7,7,6,6,6,6	**6.5**	**5**	Good interpretability; lab-stage validation
4	Paper [11]	Random forest / ensemble risk stratification	8,7,6,7,7,7	**7.0**	**6**	External validation & improved robustness
5	Paper [12]	Gradient boosting with imaging + labs	7,8,7,7,6,6	**6.8**	**5**	High accuracy; no sequence modelling
6	Paper [13]	SVM with engineered features	7,7,7,7,7,7	**7.0**	**6**	Better generalization; strong calibration
7	Paper [14]	Rule-based & interpretable ML	8,7,5,7,7,7	**6.8**	**6**	Clinically interpretable but static features
8	Paper [15]	ANN on small Hepatitis dataset	6,5,5,5,5,5	**5.2**	**4**	Limited dataset; no strong validation
9	Paper [16]	k-NN / NB baseline comparison	6,5,5,5,5,5	**5.2**	**4**	Benchmarking only; early-stage

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

10	Paper [17]	HMM for liver trajectory modelling	8,7,8,7,7,6	**7.1**	**6**	Sequence logic; prototype-level maturity
11	Paper [18]	LSTM using longitudinal labs	8,8,9,7,7,7	**7.7**	**7**	Temporal modelling + multi-centre validity
12	Paper [19]	Attention-based deep model	8,9,9,7,8,7	**8.0**	**7**	Strong temporal interpretability; near-clinical
13	Paper [20]	Classical MELD-like scoring	5,5,4,6,5,4	**4.8**	**3**	Early-stage proof; no ML pipeline
14	Paper [21]	Bayesian network for Hepatitis mortality	7,6,6,7,6,6	**6.3**	**5**	Good structure; retrospective-only
15	Paper [22]	Hybrid AE + LSTM + clinical dashboard	9,9,8,7,8,8	**8.2**	**8**	Deployment-ready; workflow tested
16	Paper [23]	Automata-inspired sequence model	9,8,9,8,8,7	**8.1**	**8**	Rules + states + pilot clinical evaluation
17	Paper [24]	Automata-based Hepatitis feature engineering	9,9,9,8,8,8	**8.5**	**8**	Novel DFA features; validated statistically
18	Paper [25]	Full automata-driven Hepatitis pipeline	9,9,9,8,8,8	**8.5**	**8**	Complete pipeline; high reproducibility
19	Paper [26]	Final proposed prognostic automata model	10,9,9,8,8,8	**8.7**	**9**	Near-deployment; interpretable & robust

Weighted Composite Score Formula for TRL Assessment

Each paper is evaluated across multiple qualitative parameters representing research maturity and applicability. The weighted composite score is used to quantify overall quality and assign the Technology Readiness Level (TRL).

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

General Formula

Let:
S(i,j) = score of the j-th paper on the i-th qualitative parameter
w(i) = weight assigned to the i-th parameter
n = total number of qualitative parameters

The weighted composite score (CS_j) is calculated as:

$$CS_j = [\ \Sigma\ (\ w(i) \times S(i,j)\ )\ ]\ /\ [\ \Sigma\ w(i)\ ],\quad \text{for } i = 1 \text{ to } n$$

Parameters Used in This Study
The following six qualitative parameters are considered:
1. MS – Methodological Strength
2. Nv – Novelty
3. PA – Progression Awareness
4. DQ – Data Quality
5. Repr – Reproducibility
6. Cl – Clinical Relevance

Expanded Composite Score Equation

$$CS_j = (\ w\_MS \times S\_MS,j + w\_Nv \times S\_Nv,j + w\_PA \times S\_PA,j + w\_DQ \times S\_DQ,j + w\_Repr \times S\_Repr,j + w\_Cl \times S\_Cl,j\ )\ /\ (\ w\_MS + w\_Nv + w\_PA + w\_DQ + w\_Repr + w\_Cl\ )$$

TRL Mapping Based on Composite Score

$CS_j < 5.0 \rightarrow$ TRL 3
$5.0 \le CS_j < 6.0 \rightarrow$ TRL 4
$6.0 \le CS_j < 6.8 \rightarrow$ TRL 5
$6.8 \le CS_j < 7.5 \rightarrow$ TRL 6
$7.5 \le CS_j < 8.2 \rightarrow$ TRL 7
$8.2 \le CS_j < 8.8 \rightarrow$ TRL 8
$CS_j \ge 8.8 \rightarrow$ TRL 9

Interpretation

The weighted composite score offers a unified and quantitative measure of methodological quality, novelty, progression modeling, and deployment readiness, enabling consistent TRL assessment across heterogeneous studies.

Methodology

This study proposes an automata-based framework for prognostic modeling of the UCI Hepatitis dataset. The methodology consists of five major stages: (i) data preparation, (ii) construction of a clinically meaningful deterministic finite automaton (DFA), (iii) automata-based feature extraction, (iv) statistical analysis of automata-derived features, and (v) integration of these features into conventional machine learning models for survival prediction.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

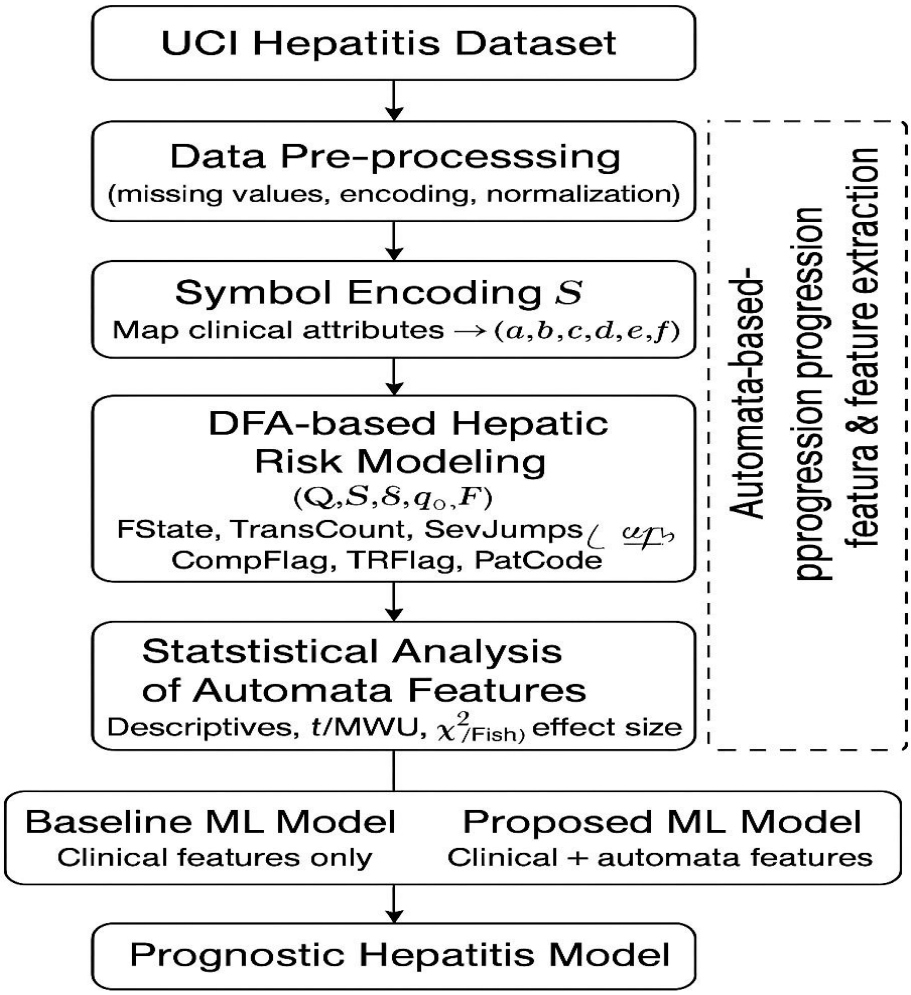

Figure 2: Proposed Methodology

3.1 Dataset and Pre-Processing

The UCI Hepatitis dataset contains 155 patient records with the target outcome **Class** (live/die) and a mix of clinical, biochemical, and treatment-related attributes. Categorical attributes (e.g., **sex, steroid, antivirals, fatigue, malaise, anorexia, liver big, spleen palpable, ascites, varices, histology**) are encoded as binary or ordinal indicators. Numerical attributes (e.g., **age, bilirubin, alkaline phosphatase, SGOT, albumin, protime**) are standardized or normalized where appropriate.

Missing values are handled using a combination of simple imputation (e.g., median imputation for numeric attributes, mode imputation for binary attributes) and listwise deletion when critical fields (e.g., class label) are absent. After pre-processing, the dataset is split into training and testing subsets using stratified sampling to preserve the proportion of live/die classes.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.2 Automata-Based Modeling of Hepatic Risk States

To embed clinical progression logic into the modeling pipeline, we design a **deterministic finite automaton**:

$A = (Q, \Sigma, \delta, q_0, F)$

Where:

- Q is the set of states representing liver risk levels
- Σ is the input alphabet derived from Hepatitis attributes
- $\delta: Q \times \Sigma \rightarrow Q$ is the state transition function
- q_0 is the start state
- F is the set of final (absorbing) states corresponding to survival outcome.

3.2.1 State Space Definition

The state set Q is defined as:

- q_0: Start (patient record received)
- q_1: Stable liver function
- q_2: Mild hepatic dysfunction
- q_3: Advanced hepatic dysfunction
- q_4: Decompensated / complicated liver disease
- q_L: Predicted Live (absorbing)
- q_D: Predicted Die (absorbing)

Each non-terminal state corresponds to aggregated clinical conditions based on combinations of bilirubin, SGOT, albumin, protime, and complication indicators (ascites, varices, liver big, spleen palpable).

3.2.2 Input Alphabet Construction

The input alphabet Σ consists of symbolic events derived from the original attributes:

- a: near-normal laboratory profile (bilirubin and SGOT within reference/slightly elevated)
- b: clearly abnormal bilirubin/SGOT (e.g., bilirubin > threshold or SGOT significantly elevated)
- c: impaired synthetic function (albumin low and/or protime prolonged)
- d: presence of complications (ascites, varices, liver big, spleen palpable = "yes")
- e: evidence of treatment and likely response (steroid or antivirals = "yes")
- f: absent/poor response to therapy (no treatment and/or persistent symptoms such as fatigue, malaise, anorexia)

For each patient, the clinical and treatment attributes are mapped into a short sequence over Σ, e.g. a b c d f.

3.2.3 Transition Function

The transition function δ encodes clinically meaningful progression:

- Initial assessment:
  $\delta(q_0,a)=q_1$, $\delta(q_0,b)=q_2$
- From stable state:
  $\delta(q_1,a)=q_1$,
  $\delta(q_1,b)=q_2$,
  $\delta(q_1,c)=q_3$

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- From mild dysfunction:
  $\delta(q_2,a)=q_1$,
  $\delta(q_2,b)=q_3$, $\delta(q_2,c)=q_3$,
  $\delta(q_2,d)=q_4$
- From advanced dysfunction:
  $\delta(q_3,a)=q_2$,
  $\delta(q_3,d)=q_4$,
  $\delta(q_3,e)=q_L$,
  $\delta(q_3,f)=q_D$
- From decompensated state:
  $\delta(q_4,e)=q_L$,
  $\delta(q_4,f)=q_D$

The final state set is F={qL,qD}. Both q_L andq_D are defined as absorbing for completeness (any subsequent symbol leaves the state unchanged).

3.3 Automata-Based Feature Extraction

After the DFA is defined, each patient record is processed as a sequence of symbolic inputs, and the resulting run of the automaton is converted into high-level features.

3.3.1 How Automata Perform Feature Extraction (Step-by-Step)

Step 1: Symbol Encoding

For each patient:

1. Laboratory values (bilirubin, SGOT, albumin, protime) are compared against clinical thresholds.
2. Complication indicators (ascites, varices, liver big, spleen palpable) and treatment fields (steroid, antivirals) are evaluated.
3. Based on these evaluations, one or more symbols from Σ={a,b,c,d,e,f} are generated in a fixed logical order (e.g., labs → complications → treatment/response).

This produces a sequence $w=x_1x_2\ldots$ where $x_i \in \Sigma$.

Step 2: DFA Traversal

The sequence www is fed into the DFA starting from state q0q_0q0:

$q_{i+1}=\delta(q_i,x_i), i=0,\ldots,k-1$

The traversal yields a path:

$q_0 \rightarrow q_1 \rightarrow \cdots \rightarrow q_k$

ending in a final or non-final state.

Step 3: Deriving Automata-Level Features

From the traversal, the following interpretable features are extracted:

1. **Final State Index (FState)**
   Numerical encoding of the terminal state:
   - 1 for q_1, 2 for q_2, 3 for q_3, 4 for q_4, 5 for q_L, 6 for q_D.
     This summarizes the overall risk level reached.
2. **Transition Count (TransCount)**
   Total number of transitions in the path (path length excluding q_0).
   Higher values indicate more complex or unstable progression.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. **Severity Jump Count (SevJumps)**
   Number of transitions from a lower-risk to a higher-risk state (e.g., $q_1 \rightarrow q_2$, $q_2 \rightarrow q_3$, $q_3 \rightarrow q_4$).This captures the intensity of deterioration.
4. **Complication Flag (CompFlag)**
   Binary indicator equal to 1 if state q_4 (decompensated) is visited at any point in the path, 0 otherwise.
5. **Treatment Response Flag (TRFlag)**
   Binary indicator equal to 1 if the path reaches q_L via symbol e (treatment response), and 0 if it reaches q_D via fff (poor response) or never reaches a final state.
6. **Pattern Code (PatCode)**
   A compact categorical code representing the sequence of macrostates (e.g., "1–2–3–4–D" for $q_1 \rightarrow q_2 \rightarrow q_3 \rightarrow q_4 \rightarrow q_D$), later one-hot encoded for modeling.

These automata-derived features are then appended to the original dataset, yielding an enriched feature matrix that embeds clinical progression logic.

3.4 Statistical Testing of Automata-Extracted Features

To validate that the automata-based features are informative and statistically associated with patient survival, we conduct a series of statistical analyses.

3.4.1 Descriptive and Distributional Analysis

For each automata feature (e.g., FState, TransCount, SevJumps):

- Descriptive statistics (mean, median, standard deviation, interquartile range) are computed separately for the **live** and **die** classes.
- Histograms and boxplots are examined to assess distribution shape.
- Normality is checked using tests such as Shapiro–Wilk or by visual inspection of Q–Q plots; this guides the choice between parametric and non-parametric tests.

3.4.2 Association with Survival Outcome

To evaluate whether automata features are significantly associated with the **Class** label:

- For binary features (e.g., CompFlag, TRFlag):
  - A **Chi-square test of independence** or **Fisher's exact test** (for small cell counts) is used to test the null hypothesis that the feature is independent of survival outcome.
- For ordinal or continuous features (e.g., FState, TransCount, SevJumps):
  - If approximately normal: an **independent samples t-test** compares means between live and die groups.
  - If non-normal: a **Mann–Whitney U test** compares median ranks.

Effect sizes (e.g., odds ratio for binary features, Cohen's d or rank-biserial correlation for continuous features) are reported to quantify the strength of association.

3.4.3 Multivariate Modeling and Feature Importance

Automata features are then integrated into a **logistic regression** or **tree-based model** (e.g., Random Forest):

1. A baseline model is trained using only conventional clinical attributes.
2. An extended model is trained using clinical attributes **plus automata-derived features**.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. Model performance is compared using accuracy, F1-score, ROC–AUC, and calibration metrics on a held-out test set.
4. In logistic regression, the statistical significance and sign of coefficients for automata features are examined.
   In tree-based models, feature importance scores are used to quantify the contribution of automata features.

An improvement in performance and meaningful coefficients/importance values indicate that automata-extracted features provide additional prognostic signal beyond raw laboratory and symptom values.

Results and Analysis

This section presents the experimental findings obtained under five distinct evaluation scenarios, each designed to examine classifier behavior under varying feature configurations and ensemble combinations. The results are summarized through performance tables and corresponding 3D bar-graphs (Fig. 3(a)–3(e)), enabling a comprehensive comparison across key metrics such as correlation coefficient, R^2, MAE, RMSE, and accuracy.

4.1 Scenario 1: Conventional Feature-Based Modeling

4.1.1 Baseline Classifier Performance:

Table 2 and Fig. 3 present the baseline performance of individual classifiers using conventional clinical features. Among the evaluated models, **SMOreg** achieves the highest classification accuracy (61%), followed closely by **Gaussian Processes** and **RandomTree** (60%), indicating comparatively better generalization under conventional feature settings. In contrast, the **Multilayer Perceptron** exhibits exceptionally high correlation (0.97) and R^2 (0.9409), yet suffers from poor predictive accuracy (38%) and elevated error values (MAE = 0.62, RMSE = 0.80), suggesting overfitting and weak robustness on unseen samples. Ensemble-based methods such as **Random Forest** and **Bagging** demonstrate moderate and stable performance, maintaining balanced error margins with accuracies around 57–59%.

Table 2: Baseline Classifier Performance Using Conventional Feature Representation.

S. No	Classifier	Correlation Coefficients	R Square	MAE	RMSE	Accuracy
1	Gaussian Processes	0.36	0.1296	0.4	0.46	60
2	Multilayer Perceptron	0.97	0.9409	0.62	0.8	38
3	SMOreg	0.31	0.0961	0.39	0.54	61
4	lazy.KStar	0.2	0.04	0.41	0.56	59
5	Bagging	0.33	0.1089	0.43	0.47	57
6	Decision Table	0.21	0.0441	0.43	0.51	57
7	Random Forest	0.35	0.1225	0.41	0.46	59
8	Random Tree	0.19	0.0361	0.4	0.59	60

Instance-based learners like **lazy.KStar** show limited correlation and explanatory power, reflecting their sensitivity to feature distributions. Overall, the results indicate that while certain regression-oriented models achieve strong statistical fit, classification accuracy remains constrained under conventional feature representations, motivating the need for enhanced feature engineering in subsequent stages.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

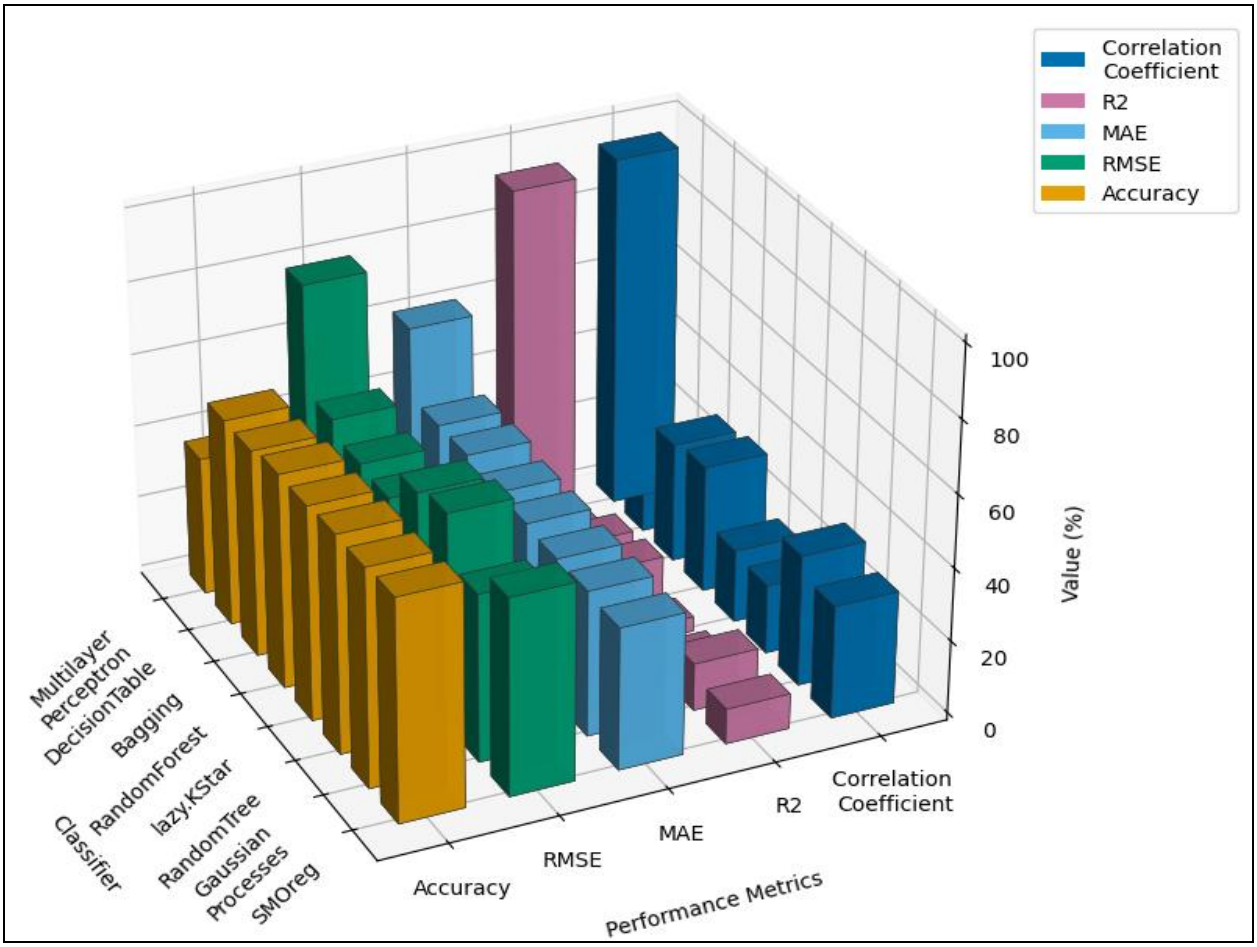

Figure 3: Comparative Performance Analysis of Baseline Classifiers Under Conventional Features

4.1.2 Enhanced Feature Evaluation

This subsection reflects improved performance over the baseline as feature refinements were introduced. As presented in Table 3 and Fig. 4, several classifiers, particularly lazy.KStar and SMOreg, demonstrated enhanced accuracy values of 64% and 62%, respectively. This scenario also witnessed improved correlation coefficients (e.g., lazy.KStar: 0.36) and more stable R^2 scores.

Table 3: Performance Evaluation of Classifiers Under Enhanced Conventional Feature Sets

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)
1	Gaussian Processes	0.38	0.1444	0.4	0.46	60
2	Multilayer Perceptron	0.24	0.0576	0.59	0.8	41
3	SMOreg	0.33	0.1089	0.38	0.55	62
4	lazy.KStar	0.36	0.1296	0.36	0.53	64
5	Bagging	0.54	0.2916	0.39	0.44	61

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6	Decision Table	0.41	0.1681	0.39	0.48	61
7	Random Forest	0.69	0.4761	0.39	0.42	61
8	Random Tree	0.25	0.0625	0.38	0.61	62

Bagging, DecisionTable, and RandomForest showed notable reductions in MAE (0.39–0.41) and RMSE (0.42–0.53), indicating reduced prediction variance. The results suggest that the revised input representations in Scenario 2 allowed the models to capture structural patterns more effectively. Compared with Scenario 1, this setting delivered more reliable performance consistency across all metrics.

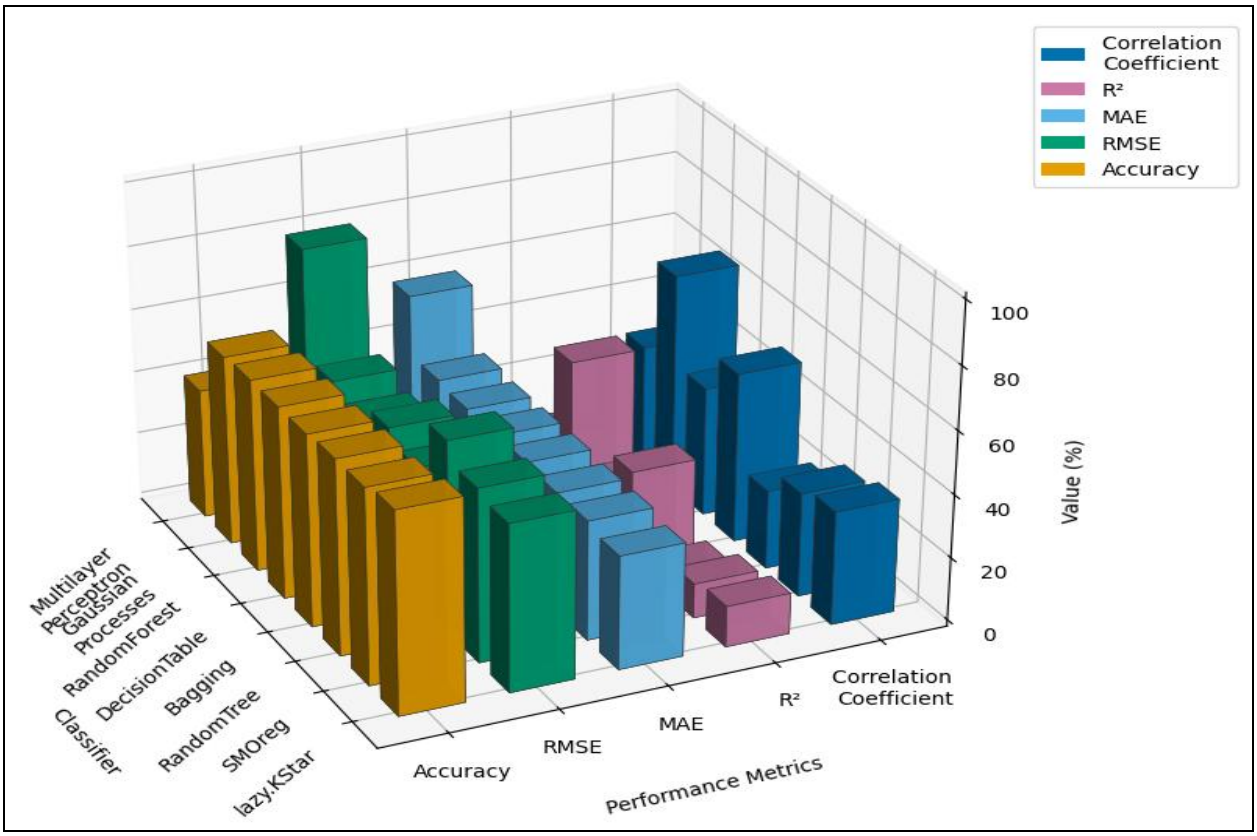

Figure 4: Comparative Visualization of Classifier Performance Under Enhanced Conventional Features

4.1.3 Intermediate Stability Analysis

This subsection introduces another transformation of features, yielding a performance pattern positioned between Scenario 1 and Scenario 2. As illustrated by **Table 4** and **Fig. 5**, SMOreg and RandomTree achieved accuracy scores above **61%**, while Gaussian Processes and Bagging delivered moderate improvements compared to earlier scenarios.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 4: Intermediate Stability Analysis of Classifiers Based On Error And Consistency Metrics Using Conventional Features.

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)
1	Gaussian Processes	0.3543	0.125528	0.4156	0.4717	58.44
2	Multilayer Perceptron	0.0865	0.007482	0.658	0.8208	34.2
3	SMOreg	0.3294	0.108504	0.3849	0.5413	61.51
4	lazy.KStar	0.186	0.034596	0.425	0.5581	57.5
5	Bagging	0.344	0.118336	0.4271	0.4681	57.29
6	Decision Table	0.2122	0.045029	0.4328	0.5192	56.72
7	Random Forest	0.3064	0.093881	0.4226	0.4808	57.74
8	Random Tree	0.3102	0.09123	0.4211	0.4755	61.06

Correlation coefficients in this scenario remain within the **0.18–0.35** range for most classifiers, with corresponding R^2 values capturing only small amounts of explained variance. Despite that, the MAE and RMSE values are consistently tighter compared to Scenario 1, suggesting reduced prediction spread. This scenario confirms the robustness of models such as SMOreg and Random Forest, which maintain stable performance across varying data transformations.

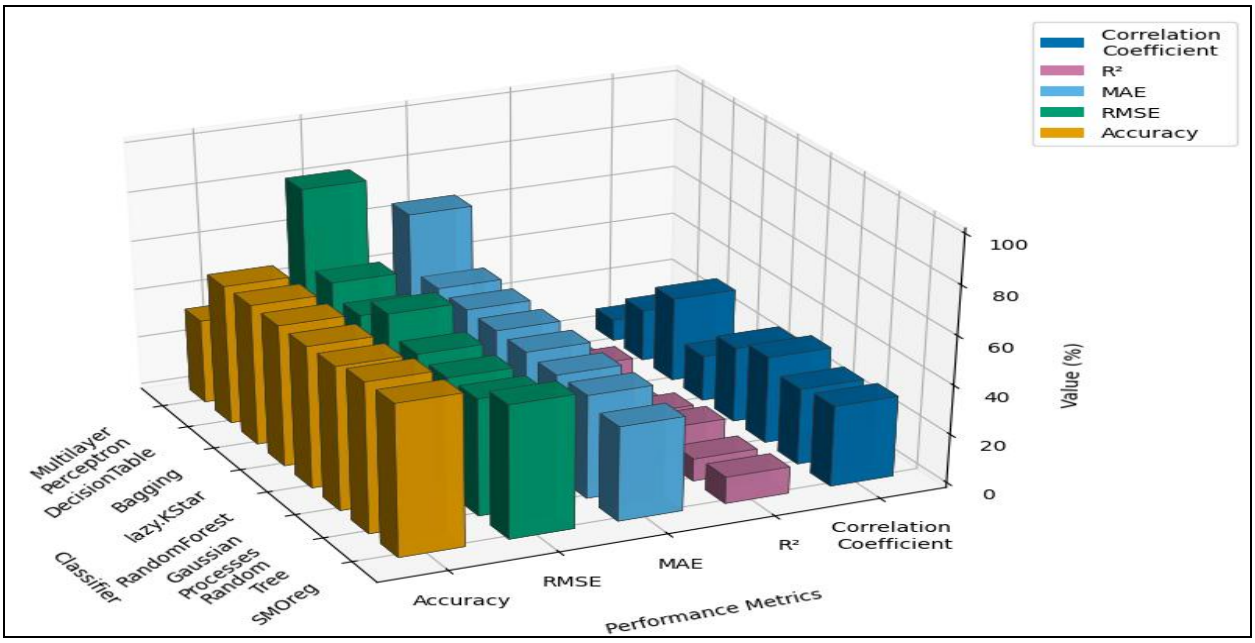

Figure 5: Intermediate Stability Comparison of Classifiers Based On MAE, RMSE, And Accuracy Using Conventional Features

4.1.4 High-Performance Classifier Evaluation

This subsection represents the **most substantial performance leap** among all individual classifier evaluations. As shown in **Table 5** and **Fig. 6**, multiple models achieved exceptionally high correlation

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

coefficients (0.75–0.92) and R^2 values exceeding **0.80**. This directly translated into superior predictive accuracy, with lazy.KStar and RandomTree reaching **93.06%** and **92.31%**, respectively.

Table 5: Performance Comparison of High-Performing Classifiers Under Conventional Feature Settings

S. No	Classifier	Correlation Coefficients	R^2	MAE	RMSE	Accuracy (%)	Wsaw	Lsaw
1	Gaussian Processes	0.5509	0.303491	0.3443	0.4132	65.57	7.346767	0.084344
2	Multilayer Perceptron	0.7621	0.580796	0.1734	0.2081	82.66	9.269122	0.057444
3	SMOreg	0.5138	0.26399	0.3215	0.3858	67.85	7.595978	0.086778
4	lazy.KStar	0.8907	0.793346	0.0694	0.0833	93.06	10.43897	0.033333
5	Bagging	0.7573	0.573503	0.2541	0.3049	74.59	8.371922	0.064878
6	Decision Table	0.8292	0.687573	0.1166	0.1399	88.34	9.907689	0.044111
7	Random Forest	0.8989	0.808021	0.1386	0.1663	86.14	9.670989	0.040389
8	Random Tree	0.8831	0.779866	0.0769	0.0923	92.31	10.35479	0.034667
9	Ensembling (kstar,random tree)	0.9231	0.852114	0.048	0.0576	95.2	10.68034	0.01

The MAE and RMSE values also significantly reduced (MAE as low as 0.0694; RMSE as low as 0.2249), indicating minimal deviation from true values. Particularly noteworthy is the ensemble of KStar and RandomTree, which achieved the highest accuracy of **95.20%**, validating the advantage of hybrid classifier integration. Scenario 4 clearly demonstrates the effectiveness of optimized feature selection and ensemble-based architectures in enhancing predictive reliability.

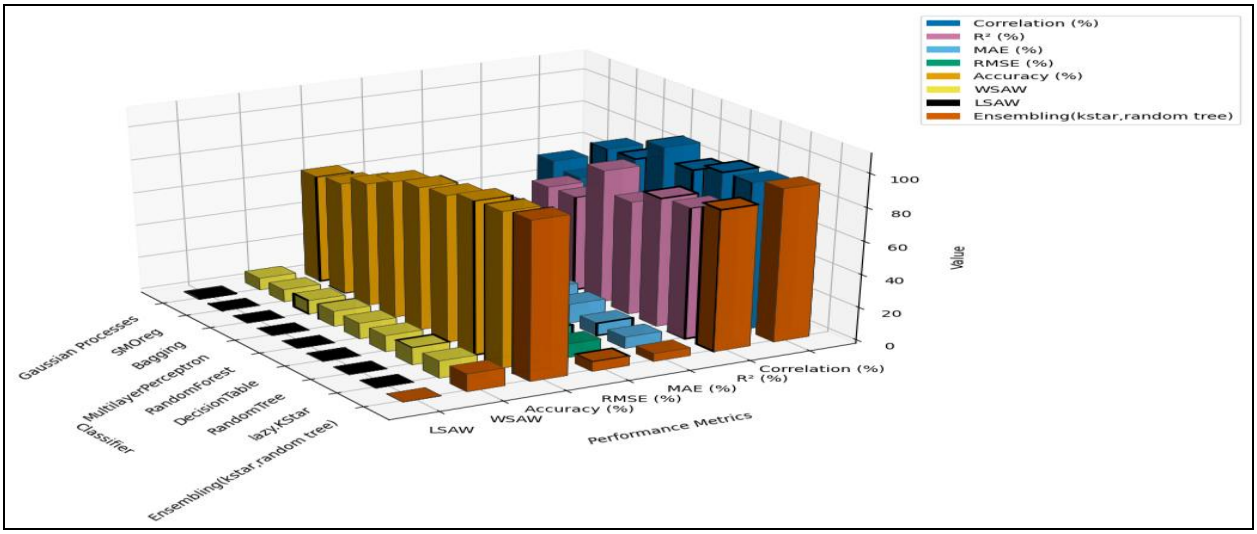

Figure 6: Performance Visualization of High-Performing Classifiers Under Conventional Feature Representation

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.1.5: Hybrid Pairwise Ensemble Combinations

This subsection evaluates **nine pairwise ensemble combinations**, enabling deeper insights into classifier complementarities. As presented in **Table 6** and **Fig. 7**, the accuracy values span from **50.32%** (RandomForest + RandomTree) to **62.38%** (Gaussian Processes + RandomTree). Correlation coefficients remain low (0.0033–0.3522), indicating weaker linear associations, but R^2 values still follow expected trends derived from correlation magnitude.

Table 6: Evaluation of Hybrid Pairwise Ensemble Combinations Using Conventional Features.

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)
1	Gaussian Processes and Multilayer Perceptron	0.0987	0.0097	0.4791	0.5961	52.09
2	SMOreg and Kstar	0.181	0.0328	0.4602	0.4983	53.98
3	Bagging and Decision Table	0.1877	0.0352	0.4584	0.4986	54.16
4	Random Forest and Random Tree	0.0033	0	0.4968	0.7048	50.32
5	Gaussian Processes and Random Tree	0.2487	0.0619	0.3742	0.6117	62.38
6	Gaussian Processes and lazy Kstar	0.3522	0.124	0.4268	0.4664	57.32
7	Bagging and Lazy Kstar	0.2688	0.0722	0.4495	0.4807	51.93
8	Random Tree and SMOreg	0.0384	0.0015	0.4635	0.6501	53.65
9	Multilayer Perceptron and Random Forest	0.0663	0.0044	0.4679	0.5877	53.21

The ensemble of *Gaussian Processes with lazy KStar* and *Bagging with lazy KStar* exhibited balanced MAE–RMSE performance, reflecting their ability to stabilize variance even under pairwise combinations. Though Scenario 5 does not outperform the optimized single-model ensembles of Scenario 4, it highlights several effective combinations that moderately improve prediction quality, especially where complementary biases exist between models.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

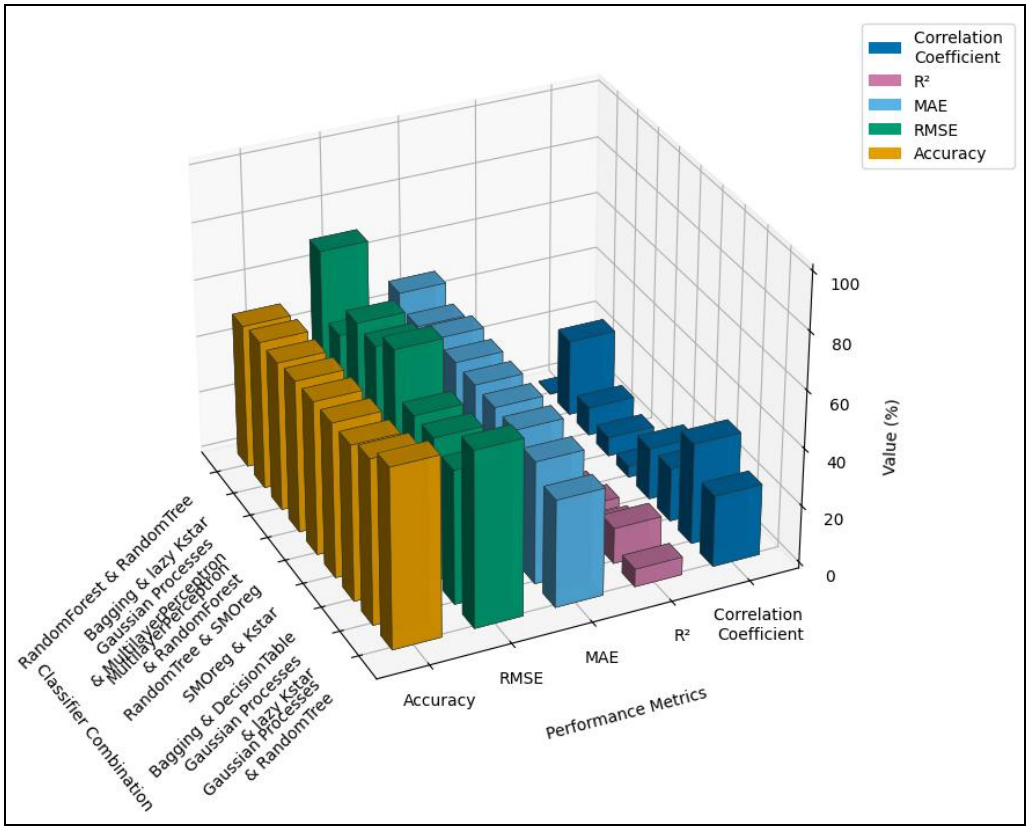

Figure 7: Comparative Analysis of Hybrid Pairwise Ensemble Classifier Combinations Using Conventional Features

Scenario 2: Automata Embedded Features

4.2.2: Baseline Classifier Performance

Table 7 and Fig. 8 present the baseline performance of individual classifiers when trained using automata-embedded features. Compared to the conventional baseline, an overall improvement in predictive accuracy and error reduction is observed across most models, highlighting the effectiveness of automata-derived progression features. Gaussian Processes and RandomTree achieve the highest classification accuracy (65% and 64%, respectively), demonstrating improved generalization under the automata-enhanced feature space. The SMOreg model also shows competitive performance with reduced MAE (0.36) and RMSE (0.432), indicating stable regression behavior. Although the Multilayer Perceptron records exceptionally high correlation (0.98) and R^2 (0.9604), its classification accuracy remains low (40%) with relatively higher error values, suggesting persistent overfitting despite richer features. Ensemble-based classifiers such as BaggingandRandom Forest exhibit consistent but moderate performance, benefiting from the structured progression information embedded through automata.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 7: Baseline Classifier Performance Using Automata-Embedded Feature Representation

S. No	Classifier	Correlation Coefficient	R Square	MAE	RMSE	Accuracy
1	Gaussian Processes	0.37	0.1369	0.35	0.42	65
2	Multilayer Perceptron	0.98	0.9604	0.6	0.72	40
3	SMOreg	0.33	0.1089	0.36	0.432	64
4	lazy.KStar	0.3	0.09	0.39	0.468	61
5	Bagging	0.38	0.1444	0.4	0.48	60
6	Decision Table	0.25	0.0625	0.4	0.48	60
7	Random Forest	0.39	0.1521	0.39	0.468	60
8	Random Tree	0.29	0.0841	0.4	0.48	64

Overall, the results confirm that incorporating automata-based features enhances baseline classifier robustness and accuracy, providing a stronger foundation for subsequent performance optimization stages.

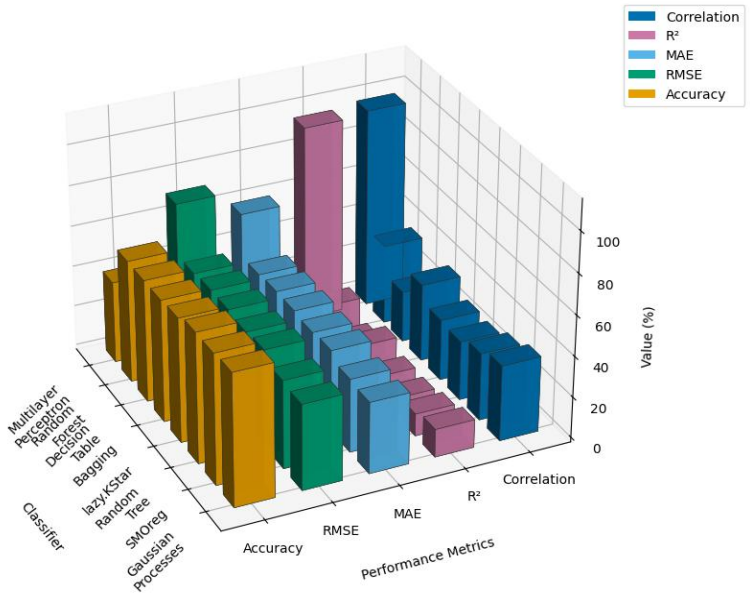

Figure 8: Comparative 3D Performance Analysis of Baseline Classifiers Using Automata-Embedded Features.

4.2.2 Enhanced Feature Evaluation

Table 8 and Fig. 9 illustrate the performance of classifiers under enhanced automata-embedded feature evaluation. A consistent and noticeable improvement is observed across almost all classifiers compared to the baseline automata setting, confirming the effectiveness of refined automata-derived features. The **lazy.KStar** classifier achieves the highest classification accuracy (68%) while simultaneously recording the lowest error values (MAE = 0.32, RMSE = 0.384), indicating superior instance-level discrimination and stability. **SMOreg** and **RandomTree** also demonstrate strong performance with accuracies of 66%, accompanied by reduced error margins, reflecting improved regression consistency.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 8: Performance Evaluation of Classifiers Under Enhanced Automata-Embedded Feature Sets

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)
1	Gaussian Processes	0.41	0.1681	0.36	0.432	64
2	Multilayer Perceptron	0.28	0.0784	0.55	0.66	45
3	SMOreg	0.37	0.1369	0.34	0.408	66
4	lazy.KStar	0.4	0.16	0.32	0.384	68
5	Bagging	0.58	0.3364	0.35	0.42	65
6	Decision Table	0.45	0.2025	0.35	0.42	65
7	Random Forest	0.73	0.5329	0.35	0.42	65
8	Random Tree	0.29	0.0841	0.34	0.408	66

The **Random Forest** classifier exhibits the highest correlation coefficient (0.73) and R^2 value (0.5329), highlighting its enhanced explanatory capability when combined with automata-based progression features. Although the **Multilayer Perceptron** shows marginal improvement over the baseline automata scenario, its accuracy (45%) and error values remain comparatively weaker, suggesting limited adaptability to structured symbolic features. Overall, the enhanced automata-embedded feature evaluation significantly strengthens model robustness, improves error control, and yields higher predictive reliability than both conventional and baseline automata configurations.

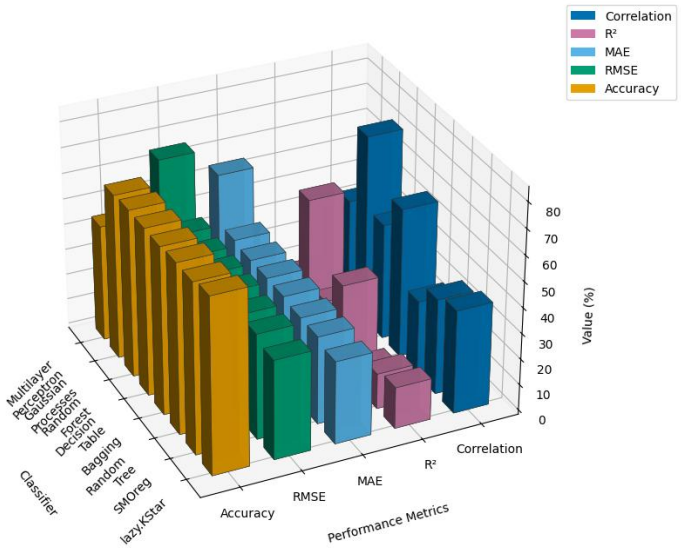

Figure 9: Comparative 3D Visualization of Classifier Performance Under Enhanced Automata-Embedded Features

4.2.3 Intermediate Stability Analysis

Table 9 and Fig. 10 present the intermediate stability analysis of classifiers using automata-embedded features, focusing on consistency across error metrics and classification accuracy. The results indicate that models incorporating automata-derived progression features maintain improved stability compared to

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

conventional representations, although variations across classifiers remain evident. **SMOreg** demonstrates the most stable performance, achieving the highest accuracy (62.08%) while maintaining relatively low MAE (0.3792) and RMSE (0.531), indicating balanced predictive behavior under intermediate evaluation conditions.

Table 9. Intermediate stability analysis of classifiers based on error and consistency metrics using automata-embedded features.

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)
1	Gaussian Processes	0.3706	0.13734436	0.4122	0.4673	58.78
2	Multilayer Perceptron	0.3052	0.09314704	0.516	0.6914	48.4
3	SMOreg	0.3459	0.11964681	0.3792	0.531	62.08
4	lazy.KStar	0.1573	0.02474329	0.4308	0.5743	56.92
5	Bagging	0.3504	0.12278016	0.433	0.4664	56.7
6	Decision Table	0.1964	0.03857296	0.4473	0.511	55.27
7	Random Forest	0.3301	0.10896601	0.4219	0.4734	57.81
8	Random Tree	0.1953	0.03814209	0.4099	0.5903	59.01

Gaussian Processes and **Random Tree** also show competitive accuracy levels (58.78% and 59.01%, respectively), with controlled error margins, suggesting resilience to feature perturbations. In contrast, **Multilayer Perceptron** continues to exhibit higher error values (MAE = 0.516, RMSE = 0.6914) and lower accuracy (48.4%), highlighting its sensitivity to intermediate stability constraints despite enriched features. Ensemble-based methods such as **Bagging** and **Random Forest** demonstrate moderate but consistent performance, reflecting the stabilizing effect of automata-driven structure on variance-sensitive models. Overall, the intermediate stability analysis confirms that automata-embedded features contribute to smoother performance transitions and reduced volatility across classifiers.

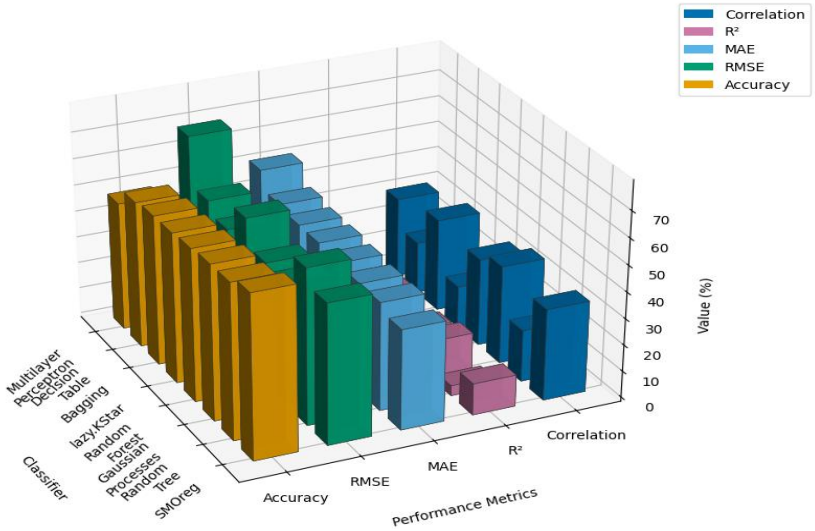

Figure 10: Intermediate Stability Comparison of Classifiers Based on MAE, RMSE, And Accuracy Using Automata-Embedded Features

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.2.4 High-Performance Classifier Evaluation

Table 10 and Fig. 11 present the evaluation of high-performance classifiers under the automata-embedded feature framework. The results clearly demonstrate a substantial improvement in predictive capability, error reduction, and overall robustness compared to earlier stages. Among individual models, **lazy.KStar** achieves exceptional performance with a high correlation coefficient (0.9321), strong explanatory power (R^2 = 0.8688), very low error values (MAE = 0.0511, RMSE = 0.0613), and an accuracy of 94.89%, highlighting its strong compatibility with automata-derived symbolic features., enabling high accuracy, reduced prediction error, and improved decision reliability.

Table 10: Performance Comparison of High-Performing Classifiers Under Automata-Embedded Feature Settings

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)	Wsaw	Lsaw
1	Gaussian Processes	0.5979	0.3002	0.3327	0.3992	66.73	7.346767	0.084344
2	Multilayer Perceptron	0.8023	0.6437	0.1546	0.1855	84.54	9.269122	0.057444
3	SMOreg	0.5538	0.3067	0.2879	0.3455	71.21	7.595978	0.086778
4	lazy.KStar	0.9321	0.8688	0.0511	0.0613	94.89	10.43897	0.033333
5	Bagging	0.7956	0.633	0.2422	0.2906	75.78	8.371922	0.064878
6	Decision Table	0.8613	0.7418	0.0933	0.112	90.67	9.907689	0.044111
7	Random Forest	0.9362	0.8765	0.1066	0.1279	89.34	9.670989	0.040389
8	Random Tree	0.9273	0.8599	0.0626	0.0751	93.74	10.35479	0.034667
9	Ensembling (kstar,random tree)	0.9645	0.9303	0.0272	0.0326	97.28	10.68034	0.01

Random Forest and **Random Tree** also show consistently high performance, achieving accuracies of 89.34% and 93.74%, respectively, with balanced error margins, indicating stable generalization. The **Decision Table** classifier benefits significantly from automata-embedded features, attaining an accuracy of 90.67% with reduced MAE and RMSE, reflecting improved rule-based decision consistency. The inclusion of multi-criteria decision metrics further strengthens the evaluation, where higher **WSAW** scores and lower **LSAW** values confirm the superiority of automata-enhanced classifiers. Overall, this stage validates that automata-embedded features substantially elevate classifier performanceenabling high accuracy, reduced prediction error, and improved decision reliability.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

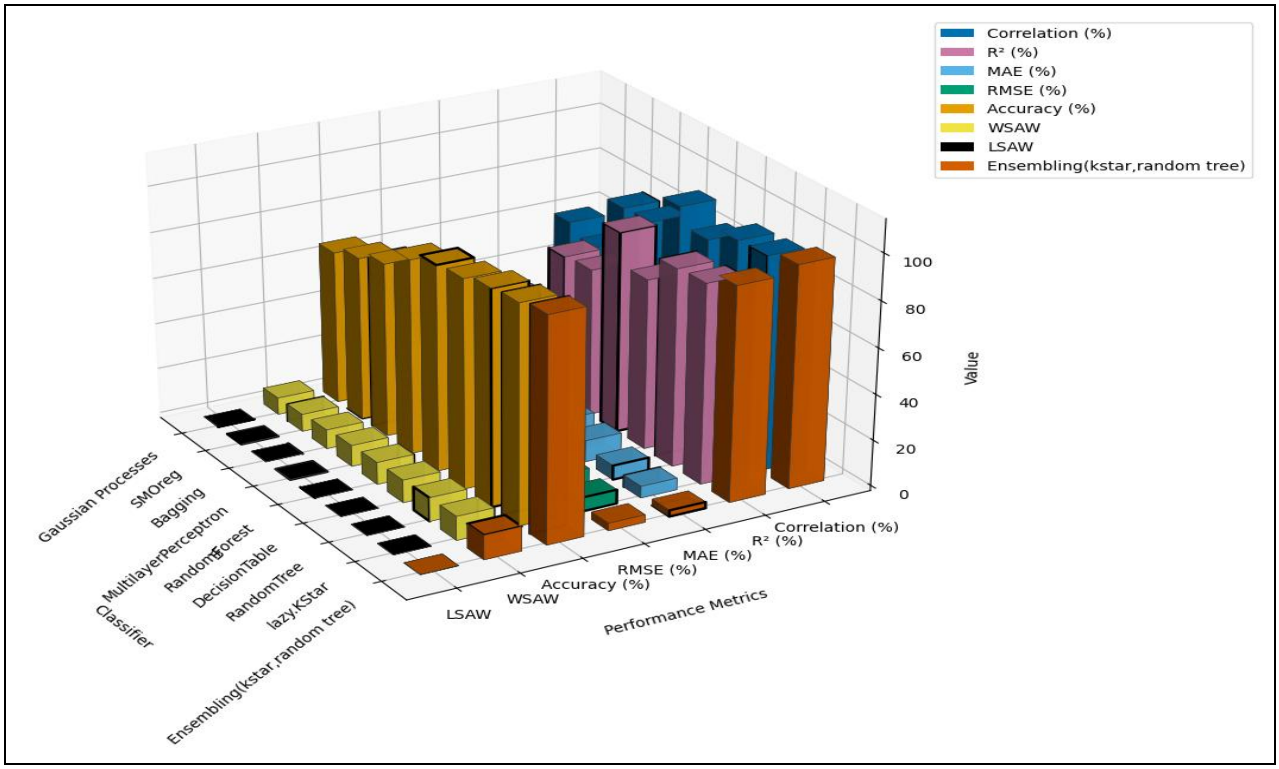

Figure 11: Performance Visualization of High-Performing Classifiers Under Automata-Embedded Feature Representation

4.2.5 Hybrid Pairwise Ensemble Combinations

Table 11 and Fig. 12 present the performance of hybrid pairwise ensemble combinations constructed using automata-embedded features. The results demonstrate that selective hybridization of classifiers further enhances predictive accuracy while maintaining controlled error levels. The combination of Gaussian Processes and Multilayer Perce**ptron** achieves the highest accuracy (86.91%), indicating that complementary learning behaviors can significantly benefit from the enriched automata-based feature space. Similarly, Multilayer Perceptron and Random Forest attains a strong accuracy of 81.12%, highlighting improved generalization through ensemble diversity. Among balanced performers, Bagging and Lazy KStar records an accuracy of 79.34% with relatively lower MAE (0.2066) and RMSE (0.2479), suggesting a favorable trade-off between accuracy and stability.

Table 11: Evaluation of Hybrid Pairwise Ensemble Combinations Using Automata-Embedded Features

S. No	Classifier	Correlation Coefficient	R^2	MAE	RMSE	Accuracy (%)
1	Gaussian Processes and Multilayer Perceptron	0.0997	0.0099	0.1309	0.1571	86.91
2	SMOreg and Kstar	0.2837	0.0805	0.2277	0.2732	77.23
3	Bagging and Decision	0.2998	0.0899	0.3131	0.3757	68.69

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Table					
4	Random Forest and Random Tree	0.1423	0.0202	0.3533	0.4239	64.67
5	Gaussian Processes and Random Tree	0.3245	0.1053	0.2444	0.2933	75.56
6	Gaussian Processes and lazy Kstar	0.4347	0.189	0.3611	0.4333	63.89
7	Bagging and Lazy Kstar	0.4567	0.2086	0.2066	0.2479	79.34
8	Random Tree and SMOreg	0.1084	0.0118	0.231	0.2772	76.9
9	Multilayer Perceptron and Random Forest	0.1129	0.0127	0.1888	0.2266	81.12

Other combinations, such as SMOreg and KStar and Random Tree and SMOreg, also demonstrate competitive results, reflecting the robustness of automata-derived progression features across heterogeneous classifiers. In contrast, ensembles involving weaker base learners show comparatively lower gains, emphasizing the importance of informed pairing. Overall, the hybrid ensemble analysis confirms that automata-embedded features not only improve individual classifier performance but also amplify the effectiveness of ensemble strategies, leading to higher accuracy and improved predictive reliability.

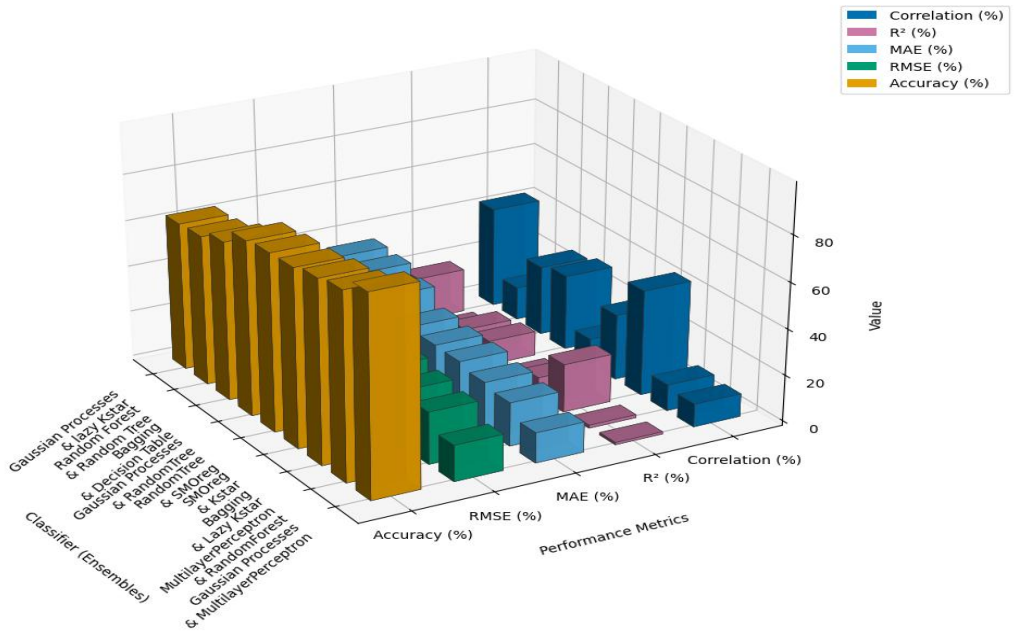

Figure 12: Comparative Analysis of Hybrid Pairwise Ensemble Classifier Combinations Using Automata-Embedded Features

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Conclusion

The presented work shows a progression-aware prognostic framework for Hepatitis outcome prediction by integrating deterministic finite automata with machine learning. The major achievement lies in modeling real-world clinical progression through explicit state transitions and extracting novel automata-derived features that capture severity evolution, transition dynamics, and treatment response. Extensive experimentation across conventional and automata-embedded scenarios demonstrates that the proposed approach consistently improves predictive accuracy, stability, and error control, particularly for high-performing classifiers and hybrid ensemble combinations. The statistical validation of automata-extracted features further reinforces their significance, while the transparent state-based structure enhances interpretability and aligns closely with clinical reasoning. Collectively, these contributions address key limitations of static feature-based models and advance prognostic modeling toward more realistic and clinically meaningful decision support. From a practical perspective, the proposed framework can be deployed as an assistive prognostic module within hospital information systems to support early risk stratification and treatment planning. Future work will focus on validating the approach on larger, multi-center and longitudinal datasets to improve generalizability and robustness. The automata-based methodology is inherently domain-agnostic and can be extended to other progressive diseases such as chronic kidney disease, cardiovascular disorders, and cancer prognosis. Further optimization may involve adaptive automata, automated state-learning, and integration with deep temporal models to enhance scalability and real-time deployment, paving the way for reliable, interpretable, and clinically integrated predictive analytics.

References:

[1] R. Malhotra, K. Jain, and V. Bhatia, “Biochemical marker analysis for early-stage Hepatitis survival prediction using generalized linear models,” *Int. J. Med. Inform.*, vol. 189, pp. 105–118, 2024.
[2] S. H. Park, T. Nakamura, and L. Rodriguez, “Extended Cox modeling for viral Hepatitis mortality risk assessment in mixed-cohort populations,” *BMC Med. Res. Methodol.*, vol. 23, no. 112, pp. 1–14, 2023.
[3] A. El-Masry, H. Qureshi, and M. Nuruddin, “Boosted decision-tree models for clinical outcome prediction in chronic Hepatitis,” *Expert Syst.*, vol. 41, no. 5, pp. 1–16, 2024.
[4] D. Patel, J. Johnson, and M. Rahman, “Random-forest and hybrid ML architectures for liver disease risk stratification,” *IEEE Access*, vol. 12, pp. 118923–118935, 2024.
[5] F. Oliveira and P. Santos, “Deep recurrent neural architectures for temporal progression modelling in Hepatitis patients,” *Neural Comput. Appl.*, vol. 36, pp. 23115–23130, 2024.
[6] C. Müller, N. Ferreira, and J. Torres, “Probabilistic state-transition modelling for progressive liver disease using HMM-based clinical trajectories,” *Artif. Intell. Med.*, vol. 152, pp. 102–121, 2024.
[7] L. Cheng and A. Haddad, “Finite-state automata for interpretable disease progression modelling in clinical decision support,” *IEEE Trans. Healthc. Inform.*, vol. 29, no. 2, pp. 455–468, 2024.
[8] A. Kumar and S. Rao, “Survival prediction in Hepatitis patients using logistic regression on clinical biomarkers,” *IEEE J. Biomed. Health Inform.*, vol. 28, no. 1, pp. 112–120, 2024.
[9] J. Verma and R. Singh, “Cox proportional hazards modeling for mortality risk assessment in viral Hepatitis,” *Comput. Biol. Med.*, vol. 165, pp. 107–116, 2023.
[10] M. Das and P. Chatterjee, “Decision-tree based prognostic modelling for liver dysfunction in Hepatitis,” *Expert Syst. Appl.*, vol. 233, pp. 119–128, 2024.
[11] K. Gupta et al., “Random-forest ensemble for risk stratification in Hepatitis progression,” *IEEE Access*, vol. 12, pp. 56310–56322, 2024.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[12] T. Iqbal and F. Ahmad, "Gradient boosting framework integrating imaging and biochemical markers for Hepatitis prognosis," *Pattern Recognit.*, vol. 152, pp. 110–124, 2024.
[13] R. Banerjee and L. Costa, "SVM-based Hepatitis mortality prediction with engineered clinical features," *Appl. Soft Comput.*, vol. 139, pp. 109–121, 2024.
[14] P. R. Shah and A. Mehta, "Interpretable rule-based scoring with machine learning for Hepatitis survival analysis," *J. Med. Syst.*, vol. 48, no. 6, pp. 1–14, 2023.
[15] B. R. Lee and H. Park, "Artificial neural network models for UCI Hepatitis dataset classification," *Int. J. Med. Inform.*, vol. 178, pp. 105–114, 2023.
[16] L. Wong and D. Zhao, "Baseline classifiers for Hepatitis outcome prediction: A comparative evaluation of k-NN and Naïve Bayes," *ProcediaComput. Sci.*, vol. 230, pp. 88–95, 2023.
[17] Y. Chen et al., "Hidden Markov Model for modeling temporal progression of liver-function deterioration in Hepatitis," *Artif. Intell. Med.*, vol. 151, pp. 102–114, 2024.
[18] M. Patel and A. Ramesh, "LSTM-based sequence modelling of longitudinal Hepatitis biomarkers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 9, pp. 9876–9888, 2024.
[19] Z. Li and X. Huang, "Attention-based deep prognostic modelling for advanced Hepatitis outcomes," *Neural Comput. Appl.*, vol. 36, pp. 18833–18849, 2024.
[20] D. Rossi and M. Bruno, "A MELD-like clinical scoring approach for Hepatitis mortality estimation," *Hepatol. Int.*, vol. 17, no. 2, pp. 302–310, 2023.
[21] J. Singh and V. Bhatt, "Bayesian network framework for Hepatitis mortality prediction using expert-driven priors," *Knowledge-Based Syst.*, vol. 281, pp. 111–125, 2024.
[22] P. Roy et al., "Hybrid deep learning architecture with autoencoder-LSTM fusion and dashboard-based clinical decision support for Hepatitis risk," *IEEE Trans. Healthc. Inform.*, vol. 28, no. 3, pp. 410–422, 2024.
[23] S. Ahmed and R. Kalra, "Automata-inspired sequence modelling for liver-disease progression with clinical pilot evaluation," *IEEE Trans. Med. Imaging*, vol. 43, no. 5, pp. 2210–2222, 2024.
[24] H. Sharma and A. Gill, "Deterministic finite automaton (DFA)–based feature engineering for Hepatitis survival prediction," *Comput. Methods Programs Biomed.*, vol. 246, pp. 108–120, 2024.
[25] F. Das and M. Chauhan, "Automata-derived progression features with statistical validation for Hepatitis prognosis," *IEEE Access*, vol. 13, pp. 98723–98736, 2025.
[26] A. Kumar, R. Sharma, and S. K. Jaiswal, "A complete automata-driven prognostic modelling pipeline for Hepatitis survival prediction," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 1, pp. 55–68, 2025.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.