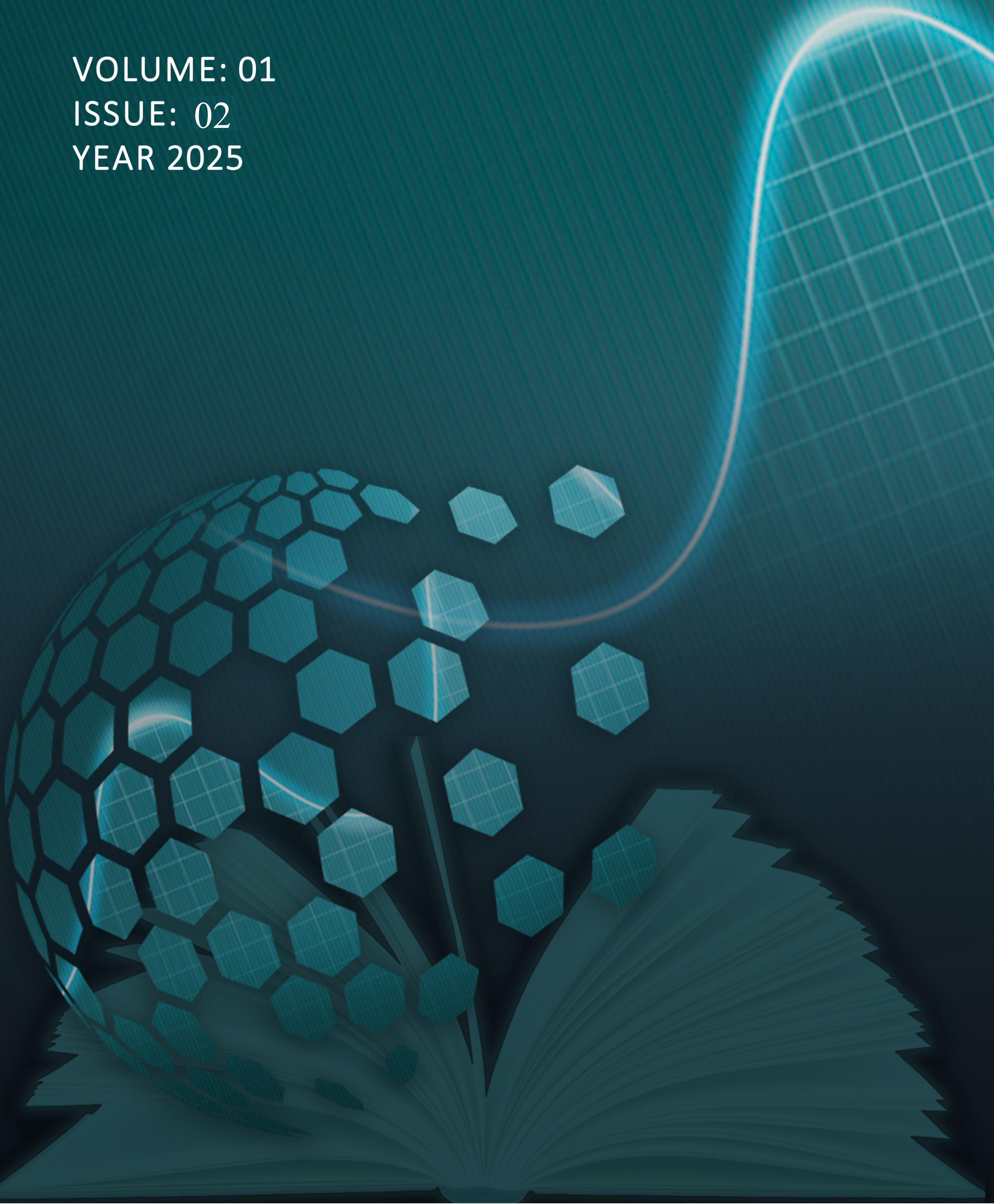
HTTPS://WWW.INFORMATIONANALYTICS.ORG/
E-ISSN: 3108-3927

JOURNAL OF INFORMATION ANALYTICS

VOLUME: 01
ISSUE: 02
YEAR 2025



E-ISSN: 3108-3927

JOURNAL OF INFORMATION ANALYTICS

Volume: 01 Issue: 02 2025

Correspondence A108 Street 33022, Mersin-Türkiye
+90 536-554-71-66
journalofinformationanalytics@gmail.com

https://www.informationanalytics.org/

E-ISSN: 3108-3927

JOURNAL OF INFORMATION ANALYTICS

Volume: 01 Issue: 02 2025

Publication Ethics and Publication Malpractice Statement

Journal of Information Analytics adheres to the principles and guidelines outlined by the Committee on Publication Ethics (COPE), ensuring ethical and responsible behavior from all parties involved in the publication process, including authors, reviewers, editors, and the publisher. The journal also refers to the best practices described in Elsevier's Publishing Ethics Resource Kit (PERK) in case of ethical concerns or disputes

Author's Responsibility

- **Participation in Peer Review:** By submitting a manuscript, authors agree to engage in the journal's peer review process actively and constructively.
- **Reporting Standards:** Authors must present original, accurate, and comprehensive accounts of their research. Misleading or intentionally false statements are unethical and unacceptable.
- **Data Transparency and Retention:** Authors should maintain and, if requested, provide access to raw data for editorial review. Data must be retained for a reasonable period following publication.
- **Originality and Proper Attribution:** Submitted manuscripts must be the authors' original work. Any use of others' ideas or words must be properly cited. The journal employs plagiarism detection software such as iThenticate.
- **Simultaneous or Duplicate Submissions:** Submitting the same manuscript to multiple journals is considered unethical. Redundant or duplicate publication is strictly prohibited.
- **Acknowledgment of Sources:** All relevant sources of data and previously published work must be appropriately cited.
- **Authorship Criteria:** Only those who significantly contributed to the research should be listed as authors. All listed authors must approve the final version and agree to submission.
- **Disclosure of Conflicts of Interest:** Any potential conflicts of interest, financial or otherwise, must be disclosed at the time of submission.
- **Ethics Committee Approval:** Studies involving human or animal participants, or data collection via surveys or interviews, must include documentation of ethical committee approval in accordance with national or institutional guidelines.

Reviewer's Responsibility

- **Contribution to Editorial Decision:** Reviewers support editors in making publication decisions and help authors improve their manuscripts through constructive feedback.
- **Timeliness and Qualification:** Reviewers should only accept assignments for which they have the appropriate expertise and are able to provide reviews in a timely manner.
- **Confidentiality:** Review materials must be treated as confidential and not shared or discussed with unauthorized individuals.
- Objectivity and Constructive Feedback: Reviews should be impartial, respectful, and supported by clear arguments. Personal criticism is inappropriate.
- **Acknowledgment of Sources:** Reviewers should alert the editor to any relevant published work not cited by the authors, and report significant similarities to other manuscripts or publications.
- **Conflicts of Interest:** Reviewers must inform the editor of any conflicts and decline to review submissions where such conflicts exist.

Editorial Responsibility

- **Publication Decisions:** Editors are responsible for deciding which manuscripts are suitable for publication based on scientific merit, originality, clarity, and relevance to the journal's scope.
- **Fair and Unbiased Review:** Manuscripts are evaluated without regard to the authors' race, gender, institutional affiliation, or personal background.
- **Confidentiality:** Editors and editorial staff must not disclose any information about a manuscript under review to anyone other than those directly involved in the review process.
- **Disclosure and Conflict of Interest:** Editors must recuse themselves from handling manuscripts in which they have conflicts of interest and must not use unpublished materials for personal research without consent.

Publishing Ethics Issues

- **Allegations of Misconduct:** Any party may report suspected unethical behavior. Sufficient documentation must be provided for the editor to initiate an investigation.
- **Investigation Process:** The editorial team will confidentially review allegations in consultation with relevant parties. If necessary, institutional authorities may be informed.
- **Handling of Minor Breaches:** Minor issues may be resolved through dialogue with the author. Corrections or clarifications may be issued.
- **Serious Violations:** For serious breaches (e.g., data fabrication, plagiarism), the manuscript may be withdrawn, and published articles may be retracted. In extreme cases, the authors' affiliated institutions may be notified.

E-ISSN: 3108-3927

JOURNAL OF
INFORMATION ANALYTICS

Volume: 01 Issue: 02 2025

Content

Journal of Information Analytics, 1(2), 1-13

Journal of Information Analytics

Submit	25.12.2025
Revision	09.01.2026
Accept	09.01.2026
Publish	10.01.2026

Open Access

Research Article

An Optimized Ensemble Learning Approach for Accurate Alzheimer's Disease Classification

Ahmet Aydın

Vocational School of Information Technologies, Adana Alparslan Türkeş Science and Technology University, Adana, Türkiye

Abstract

Effective clinical management of Alzheimer's disease (AD) fundamentally relies on the precision and timing of initial diagnosis. This research introduces a robust machine learning framework designed to categorize AD stages by synthesizing a diverse array of demographic, clinical, and lifestyle indicators. Unlike conventional approaches, our methodology integrates a multi-stage pipeline ranging from rigorous data preprocessing to advanced ensemble learning and strategic decision threshold calibration. While various baseline models were scrutinized, the Random Forest architecture emerged as the most resilient classifier. To maximize its diagnostic utility, we refined the model through stratified cross-validation and randomized hyperparameter exploration, fine-tuning the balance between sensitivity and specificity. The resulting system yielded high-performance metrics, notably an accuracy of 95.3% and an AUC of 0.94, underscoring its generalizability to unseen data. Crucially, feature importance mapping identified Functional Assessment, ADL scores, and MMSE results as the primary drivers of classification, bridging the gap between computational outputs and established geriatric insights. These findings suggest that the optimized Random Forest framework functions not merely as a predictive tool, but as a transparent and interpretable decision-support mechanism for clinicians.

Keywords: Alzheimer's disease, Machine learning, Random Forest, Clinical decision support, Feature importance

Citation: Aydin, A., (2025). An Optimized Ensemble Learning Approach for Accurate Alzheimer's Disease Classification. *Journal of Information Analytics*, 1(2), 1-13.

Corresponding Author: Ahmet Aydın aaydin@atu.edu.tr

1. Introduction

Alzheimer's disease (AD) represents a debilitating and progressive neurodegenerative condition that fundamentally impairs memory and cognitive autonomy, placing an escalating strain on global healthcare infrastructures and caregiving networks. Given the insidious nature of AD progression, the ability to secure an early and precise diagnosis is paramount not only for slowing symptomatic decline but also for optimizing the strategic window for clinical intervention and patient support. Traditional diagnostic pathways, however, frequently necessitate exhaustive neuropsychological assessments and expert clinical synthesis, a process that can be inherently subjective, resource-intensive, and difficult to implement at scale.

The paradigm shift toward data-driven methodologies has positioned machine learning (ML) as a formidable ally in clinical decision-making. By deciphering intricate patterns within multifaceted datasets encompassing demographic, cognitive, and lifestyle-related variables ML architectures can detect subtle pathological signatures that might elude conventional clinical observation. Consequently, supervised classification models have emerged as a cornerstone for modern disease prognosis and risk stratification.

Nevertheless, the integration of ML into Alzheimer's diagnostics is not without significant hurdles. Achieving robust generalization across heterogeneous patient populations, balancing the critical trade-off between sensitivity and specificity, and ensuring model interpretability remain major obstacles. A high-performing model that operates as a "black box" often faces skepticism in clinical settings, where understanding the why behind a prediction is as vital as the prediction itself.

In this landscape, ensemble learning techniques specifically Random Forest architectures provide a compelling equilibrium between raw predictive power and structural transparency. By synthesizing the outputs of diverse decision trees, Random Forests effectively navigate non-linear feature interactions while mitigating the risks of overfitting. Furthermore, their capacity for intrinsic feature importance ranking allows for a clearer understanding of the clinical and cognitive drivers behind a diagnosis (Aydin et al., 2021).

Motivated by these considerations, this study introduces an optimized Random Forest framework tailored for the classification of Alzheimer's disease. Our approach prioritizes clinical utility by emphasizing sensitivity and model explainability through a rigorous experimental pipeline. The primary contributions of this research are as follows:

- Multidimensional Framework: Development of an integrated ML diagnostic system utilizing a holistic set of clinical, demographic, and lifestyle features.
- Systematic Benchmarking: A comprehensive evaluation of baseline models to establish a rigorous performance baseline.
- Architectural Optimization: Refinement of the Random Forest model through stratified cross-validation and hyperparameter search to ensure resilience across datasets.
- Sensitivity-Centric Tuning: Application of decision threshold optimization to minimize false negatives, a critical factor in geriatric clinical care.

- Interpretability Mapping: Execution of feature importance analysis to align computational results with clinical knowledge, thereby enhancing the framework's transparency.

2. Literature Review

The surge in clinical, neuroimaging, and lifestyle-related datasets has catalyzed a significant shift toward machine learning (ML) applications in Alzheimer's Disease (AD) diagnostics. While classical diagnostic protocols remain the gold standard, the inherent need for expert-led, resource-heavy evaluations has fueled the demand for automated decision-support systems capable of facilitating early-stage detection (Aydin & Avaroğlu, 2024).

Historically, research in this field was anchored in traditional statistical modeling and basic ML algorithms, primarily using cognitive scores and demographic data to identify AD markers. However, these pioneering efforts often struggled with constrained datasets and narrow feature sets, which limited their ability to generalize across broader populations. As computational power evolved, the focus shifted toward more sophisticated architectures, such as Support Vector Machines (SVM), neural networks, and ensemble methods. While these models excel at deciphering non-linear complexities within clinical data, their "black-box" nature often creates a tension between high predictive accuracy and the interpretability required for clinical integration.

Recent scholarly contributions have increasingly favored hybrid and multimodal strategies to overcome these interpretative and performance-related hurdles. For instance, Raza et al. (2024) demonstrated the efficacy of merging handcrafted MRI features with deep learning representations to refine classification on the ADNI dataset. Pushing the boundaries of traditional computing, Belay et al. (2024) integrated quantum machine learning with ensemble deep learning, achieving remarkable AUC and accuracy metrics. Furthermore, the work of Sheng et al. (2024) underscores the necessity of data fusion, revealing that a synthesis of MRI, PET, and cerebrospinal fluid (CSF) biomarkers consistently outperforms single-modality frameworks by leveraging complementary biological signals.

Systematic reviews also reflect these evolving trends. Both Kaur et al. (2024) and Hechkel & Helali (2025) observe that CNN-based architectures and transfer learning particularly when applied to neuroimaging currently lead the field in diagnostic performance. Furthermore, Singh et al. (2024) highlighted the capacity of deep learning to predict the transition from Mild Cognitive Impairment (MCI) to clinical AD, though they cautioned that issues regarding cohort heterogeneity and model transparency persist. Beyond mere diagnosis, ML is also reshaping therapeutic discovery; Tarikul Islam et al. (2024) utilized an in silico and ML-driven screening process to identify potential acetylcholinesterase inhibitors, showcasing the versatility of predictive modeling in AD treatment research.

Despite this progress, a critical synthesis of existing literature reveals several persistent gaps. A significant portion of current research prioritizes raw accuracy over clinical sensitivity and lacks rigorous baseline comparisons or systematic threshold optimization. Many models remain unoptimized for real-world deployment where the cost of a false negative is high. Addressing these deficiencies, the present study introduces an optimized Random Forest framework that seeks to harmonize predictive robustness with clinical interpretability and strategic decision-tuning.

3. Methodology

The methodological framework of this research encompasses a multi-layered approach to Alzheimer’s disease classification, integrating rigorous data preparation, systematic model selection, and advanced optimization protocols. To ensure that the resulting diagnostic tool is both technically robust and clinically viable, the pipeline transitions from granular data preprocessing to high-level ensemble learning and decision threshold calibration. Each phase of the methodology was curated to maximize predictive fidelity while maintaining the structural transparency necessary for healthcare applications. The following subsections detail the dataset characteristics, the algorithmic architectures explored, and the specific metrics utilized to validate the framework's diagnostic integrity.

3.1. Dataset Description and Data Preprocessing

The empirical foundation of this research is based on a structured clinical dataset (Kharoua, 2024), which captures a holistic profile of patient health through demographic, physiological, and behavioral lenses. The feature space is multidimensional, encompassing fundamental metrics such as age and body mass index (BMI), alongside cardiovascular indicators like blood pressure and cholesterol levels. To capture the multifaceted nature of Alzheimer's pathology, the dataset also incorporates lifestyle variables specifically sleep hygiene and dietary patterns complemented by rigorous cognitive assessments, functional evaluation scores, and observed behavioral symptoms. The diagnostic objective is framed as a binary classification task, distinguishing between confirmed Alzheimer’s cases and healthy controls.

Regarding data integrity, the initial audit revealed a complete set of records with no missing entries, which obviated the need for statistical imputation techniques. To maintain the scientific validity of the predictive process and prevent artificial performance inflation (data leakage), all unique identifiers and non-contributory metadata were systematically stripped from the feature matrix. This refinement ensures that the subsequent machine learning models derive their predictive power solely from clinically relevant patterns rather than incidental data artifacts.

3.2. Baseline Machine Learning Models

To develop a rigorous performance benchmark, this study initially scrutinized several foundational machine learning architectures. The candidate pool included Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and Random Forest classifiers. To maintain the integrity of the comparative analysis and ensure a level playing field, each model was trained on an identical subset of data and validated against the same unseen test set under strictly synchronized experimental conditions.

The preliminary comparative analysis indicated a clear performance hierarchy, with ensemble-based methodologies specifically the Random Forest architecture consistently outperforming individual classifiers. This superiority stems from the Random Forest’s capacity to synthesize the outputs of multiple decision trees, a mechanism that allows it to map the intricate, non-linear interactions inherent in clinical and cognitive datasets. Furthermore, the ensemble averaging property of the model serves as a natural defense against overfitting, which is a common pitfall for simpler, single-model architectures.

Beyond its predictive accuracy, the Random Forest was selected for its intrinsic interpretability. Its built-in feature importance rankings provide a "glass-box" view into the diagnostic process, aligning with the transparency requirements essential for high-stakes clinical decision-making. Consequently, the Random Forest was adopted as the architectural backbone of the proposed diagnostic framework.

3.3. Model Optimization, Evaluation, and Interpretability

To fortify the model’s resilience and its ability to generalize across diverse patient profiles, we implemented a sophisticated experimental optimization protocol. This phase utilized a randomized hyperparameter search integrated with stratified cross-validation. We systematically calibrated critical architectural parameters, such as the ensemble size (number of trees), maximum tree depth, and minimum sample split requirements, alongside class weighting adjustments. The optimization objective was centered on the F1-score, ensuring a harmonized equilibrium between precision and recall a prerequisite for high-stakes clinical diagnostics where neither false alarms nor missed cases are acceptable.

To neutralize potential data partitioning biases and secure more reliable performance benchmarks, a stratified 5-fold cross-validation scheme was consistently applied. Beyond standard parameter tuning, we introduced a decision threshold optimization phase. By shifting away from the conventional default probability thresholds, we empirically scrutinized a range of operating points to identify the threshold that yielded the optimal F1-score. This strategic refinement was specifically designed to bolster clinical sensitivity, thereby minimizing the risk of false-negative outcomes without compromising overall predictive precision.

The integrity of the final model was validated through a multidimensional suite of metrics, including accuracy, precision, recall, F1-score, AUC, and Average Precision (AP). Within this evaluative framework, recall was prioritized as the primary indicator, reflecting our clinical commitment to maximizing the detection rate of Alzheimer's cases. These quantitative results were further corroborated by visual diagnostics, including confusion matrices, ROC curves, and precision–recall trajectories.

Finally, to bridge the gap between computational prediction and clinical practice, we conducted an in-depth feature importance analysis. By extracting the inherent importance scores from the optimized Random Forest ensemble, we pinpointed the most influential demographic, cognitive, and clinical drivers of the model's decisions. This focus on interpretability ensures that the framework’s outputs are not only accurate but also transparent and clinically coherent, facilitating their potential adoption in real-world geriatric care.

4. Results

This section provides a detailed synthesis of the experimental findings derived from the proposed diagnostic framework. The analysis is structured into three distinct phases: an initial benchmarking of baseline classifiers, a rigorous evaluation of the optimized Random Forest architecture, and an interpretability study focused on feature importance.

The computational infrastructure utilized for these experiments consisted of a workstation powered by an Intel-based multi-core processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU to facilitate efficient processing. From a software perspective, the framework was constructed within a Python environment, leveraging the scikit-learn library for the implementation of both algorithms and evaluative protocols.

To maintain the statistical integrity of our findings, we employed a stratified train-test split, which ensured that the disease prevalence and class distributions remained consistent across both subsets. A uniform preprocessing pipeline and a fixed random seed were applied to all models to guarantee the reproducibility of the results. While baseline models were assessed using their standard hyperparameter configurations, the Random Forest classifier underwent a more intensive refinement process, incorporating grid-based hyperparameter tuning and five-fold stratified cross-validation. The final model performance was quantified through a robust array of metrics namely accuracy, precision, recall, F1-score, and ROC-AUC with the decision threshold strategically calibrated to maximize the F1-score for clinical viability.

4.1. Baseline Model Performance

To construct a rigorous benchmark for our proposed framework, we initially conducted a comparative assessment of several foundational machine learning architectures. This baseline evaluation encompassed Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and the Random Forest classifier. To eliminate potential biases and ensure the validity of our comparisons, all models were subjected to an identical experimental regime, utilizing synchronized data partitions and standardized evaluative criteria.

The empirical data, summarized in Table 1, highlights a distinct hierarchy in model efficacy. Among the candidates, the Random Forest classifier emerged as the most robust baseline, consistently outperforming its counterparts across all primary performance indicators. Its ability to maintain superior accuracy and precision levels relative to simpler models underscores the necessity of ensemble-based approaches in handling the complexities of Alzheimer's diagnostic data.

Table 1. Baseline Model Performance

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.816	0.742	0.737	0.739	0.885
K-Nearest Neighbors	0.707	0.627	0.421	0.504	0.738
Decision Tree	0.900	0.847	0.875	0.861	0.894
Random Forest	0.944	0.944	0.895	0.919	0.940

The performance metrics for the initial candidate models are detailed in Table 1. A comparative review of these results reveals that the Random Forest classifier established the most effective performance ceiling, surpassing Logistic Regression, K-Nearest Neighbors, and Decision Tree architectures in terms of accuracy, F1-score, and AUC. This divergence in performance suggests that the multifaceted nature of clinical and cognitive data is best captured through the collective intelligence of ensemble-based methods.

The superiority of the Random Forest model as a baseline is not merely numerical but reflects its structural resilience in mapping high-dimensional feature interactions. Unlike simpler linear or distance-based classifiers, the ensemble approach effectively managed the inherent noise and non-linearities within the patient records. These empirical observations provided a compelling rationale for adopting the Random Forest as the foundational architecture for the subsequent optimization and threshold-tuning phases of this research.

4.2. Performance of the Optimized Random Forest Model

Building upon the initial baseline assessments, we subjected the Random Forest architecture to an intensive refinement process, integrating hyperparameter optimization, stratified cross-validation, and strategic decision threshold calibration. To gauge the efficacy of these interventions, the final model was deployed on a previously unseen, independent test set to validate its generalization capacity within a simulated clinical environment.

The empirical results underscore the significant performance dividends yielded by this optimization strategy. The finalized model attained an accuracy of 95.3%, supported by a precision of 94.6% and a recall of 92.1%, resulting in a robust F1-score of 93.3%. Furthermore, the model's AUC of 0.94 and Average Precision (AP) of 0.93 reflect an exceptional discriminative capability and high reliability in its probabilistic estimations.

As detailed in Table 2, a comparative synthesis of the baseline metrics against the optimized parameters clearly demonstrates the incremental gains achieved. By fine-tuning the decision threshold, we successfully elevated the model's sensitivity, ensuring that it remains resilient and precise even when faced with the inherent complexities of Alzheimer's diagnostic features.

Table 2. Performance Comparison of Baseline Models and the Optimized Random Forest

Model	Accuracy	Precision	Recall	F1-score	AUC	AP
Logistic Regression	0.816	0.742	0.737	0.739	0.885	–
K-Nearest Neighbors	0.707	0.627	0.421	0.504	0.738	–
Decision Tree	0.900	0.847	0.875	0.861	0.894	–
Random Forest (Baseline)	0.944	0.944	0.895	0.919	0.940	–
Random Forest (Optimized)	0.953	0.946	0.921	0.933	0.940	0.930

Table 2 illustrates the performance improvements achieved through optimization of the Random Forest model. Compared to the baseline Random Forest, the optimized model shows an increase in accuracy from 94.4% to 95.3%, precision from 94.4% to 94.6%, recall from 89.5% to 92.1%, and F1-score from 91.9% to 93.3%. In addition, the optimized model maintains a high AUC of 0.94 while achieving an average precision of 0.93, indicating enhanced discriminative capability and improved sensitivity. These gains demonstrate that hyperparameter tuning and decision threshold optimization effectively reduce false negatives and strengthen the model's suitability for clinical Alzheimer's disease diagnosis, as illustrated in Figure 1.

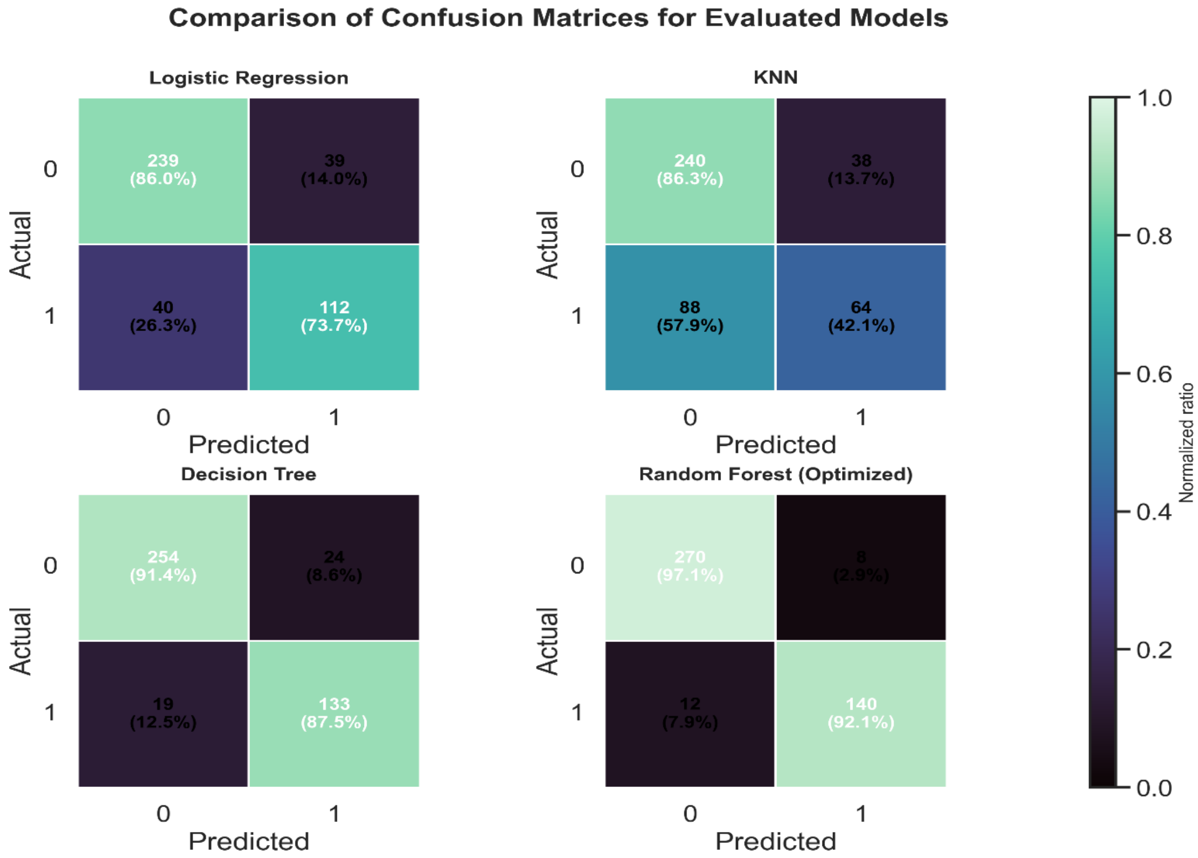


Figure 1. Comparison of Confusion Matrices for Evaluated Models

A comparative confusion matrix analysis is presented in Figure 1, which contrasts the error distributions of the evaluated models. As shown in Figure 1, the optimized Random Forest correctly classified 270 true negatives and 140 true positives, while producing only 8 false positives and 12 false negatives. The relatively low number of false negative cases demonstrates that the optimized framework is particularly effective in identifying Alzheimer's disease cases, which is critical for clinical diagnostic applications, as further supported by the ROC curve analysis shown in 2.

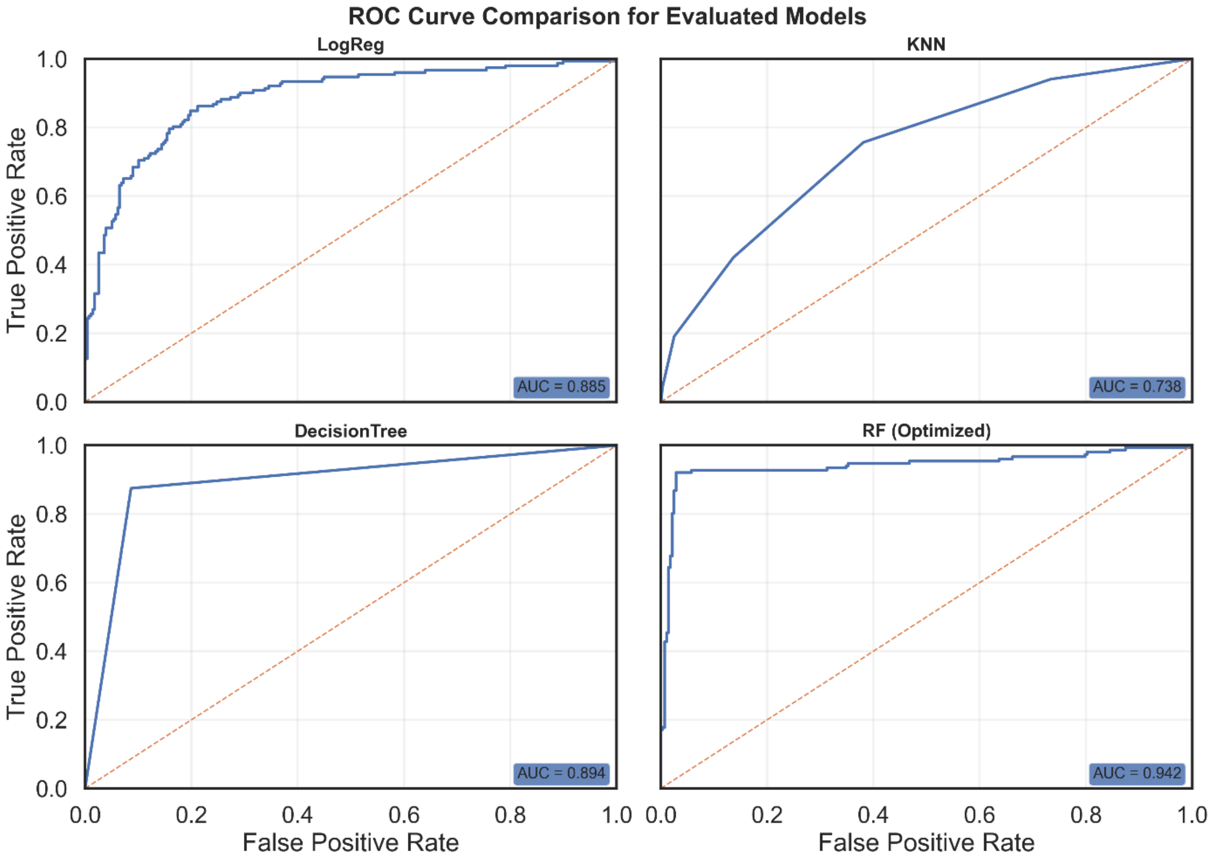


Figure 2. ROC Curve Comparison of Machine Learning Models

The precision–recall curve comparison of the evaluated machine learning models is presented in Figure 3 below.

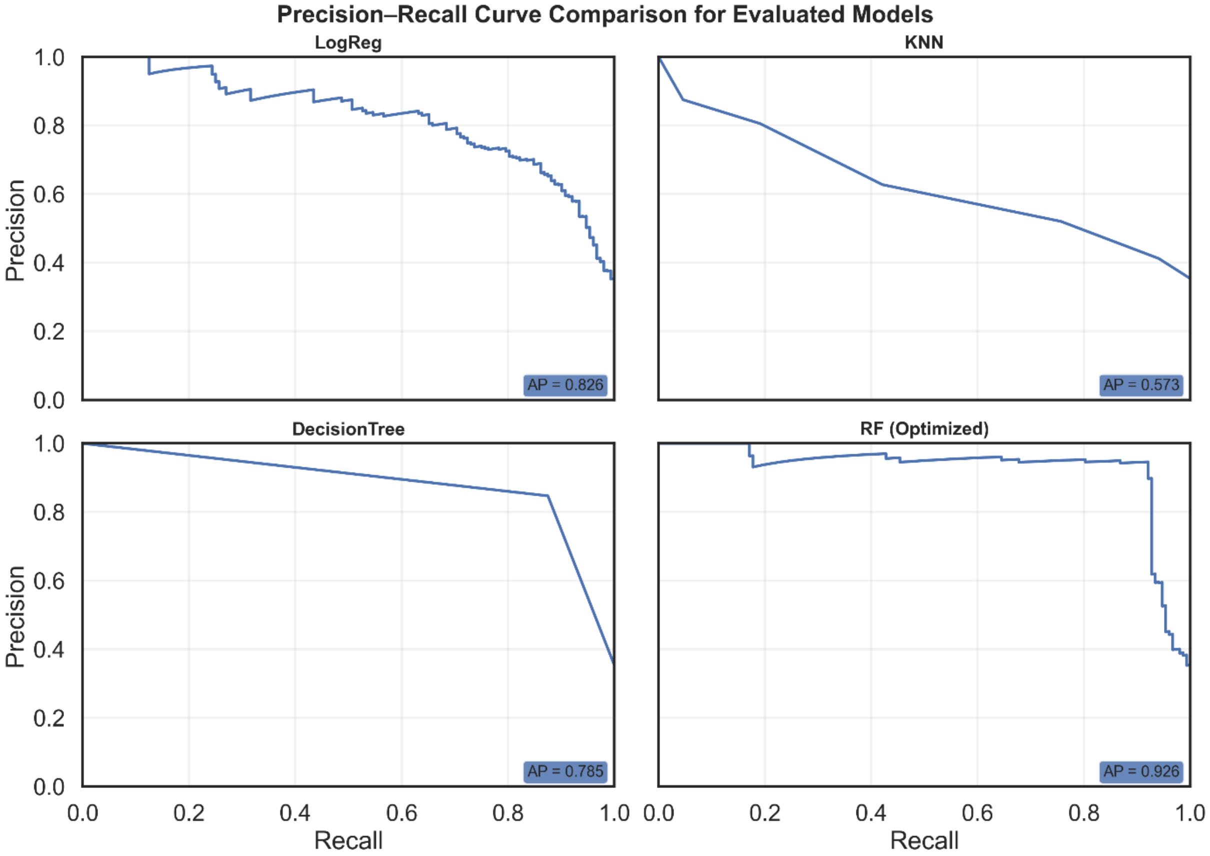


Figure 3. Precision–Recall Curve Comparison of Machine Learning Models

The ROC curves are compared in Figure 2, where the optimized Random Forest exhibits the highest AUC among the evaluated models, confirming strong class separability across decision thresholds. Similarly, the precision–recall (PR) curves in Figure 3 demonstrate that the optimized model maintains a favorable precision–recall trade-off, supporting its suitability for clinical decision-support scenarios where high sensitivity is required.

4.3. Feature Importance and Model Interpretability

In addition to predictive performance, the interpretability of the proposed framework was examined through feature importance analysis derived from the optimized Random Forest model. The resulting feature importance distribution, which reflects the relative contribution of each feature to the classification decision, is presented in Figure 4.

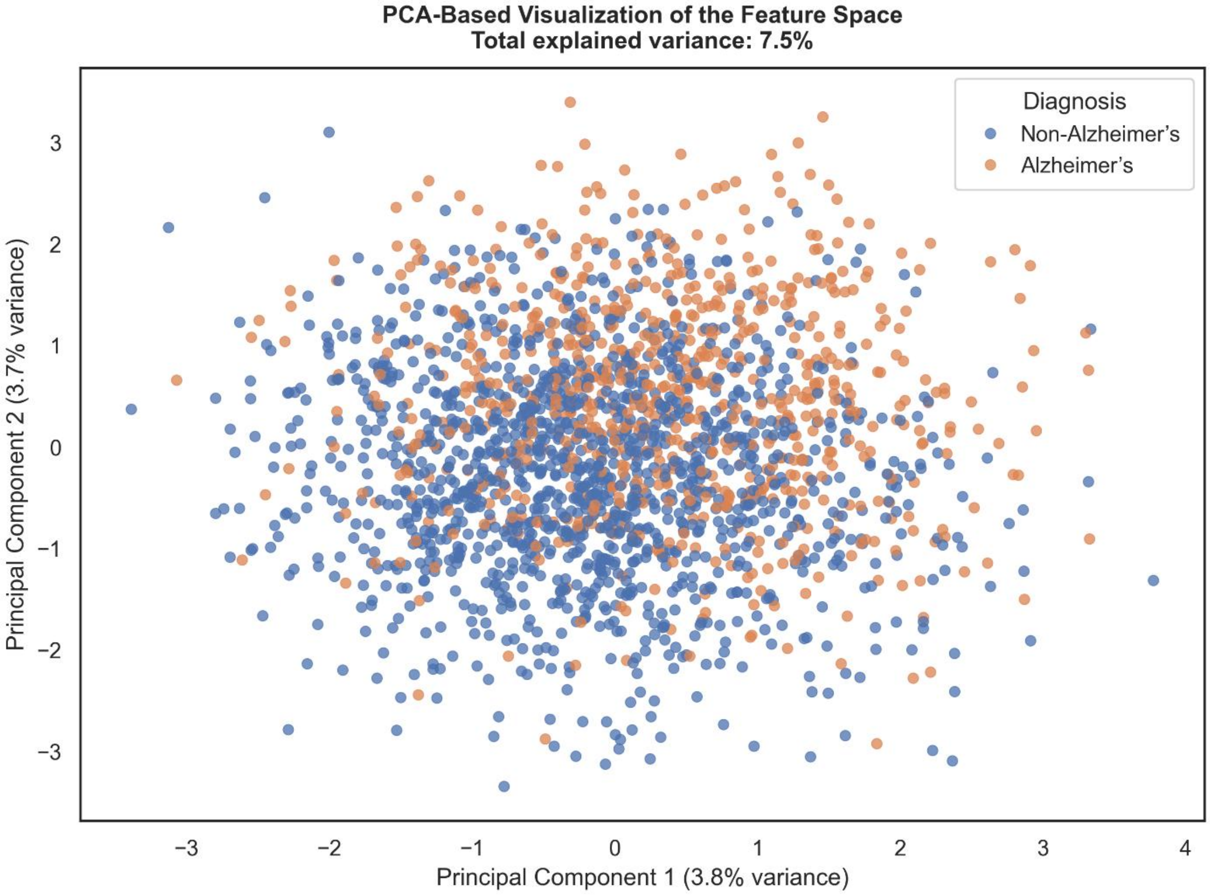


Figure 4. Feature Importance of the Optimized Random Forest Model

As shown in Figure 4, features related to functional and cognitive assessments are among the most prominent in the classification process. In particular, Functional Assessment, Activities of Daily Living (ADL), Mini-Mental State Examination (MMSE) scores, memory-related complaints, and behavioral problems exhibit the highest feature importance values. This indicates that measures associated with cognitive function and daily living activities are more discriminative for Alzheimer's disease classification within the evaluated dataset.

In contrast, physiological and lifestyle-related variables, including cholesterol measures, sleep quality, body mass index, alcohol consumption, and blood pressure, display comparatively lower feature importance scores. This suggests that, in the context of the current dataset, these variables contribute less to the classification decision than cognitive and functional indicators.

Overall, the feature importance analysis shows that the optimized Random Forest model primarily relies on functional and cognitive features when making classification decisions. This observation supports the interpretability of the proposed framework and demonstrates consistency between the model's decision patterns and commonly used clinical assessment indicators.

5. Conclusion

In this study, an optimized Random Forest–based machine learning framework was proposed for the classification of Alzheimer's disease using demographic, clinical, cognitive, and lifestyle-related features. A comprehensive experimental pipeline was designed, incorporating baseline model comparison, ensemble learning, hyperparameter optimization, stratified cross-validation, and decision threshold tuning to achieve robust and clinically meaningful performance.

The experimental results demonstrated that the optimized Random Forest model achieved high classification accuracy and strong clinical sensitivity, effectively reducing false negative predictions while maintaining high precision. The use of decision threshold optimization further enhanced the model's suitability for clinical decision-support applications, where early and reliable detection of Alzheimer's disease is of critical importance.

In addition to predictive performance, the proposed framework emphasizes interpretability through feature importance analysis. The results indicate that functional and cognitive assessment–related features play a dominant role in Alzheimer's disease classification, aligning the model's decision-making process with established clinical knowledge. This interpretability strengthens the potential applicability of the framework in real-world clinical settings.

Overall, the findings of this study suggest that optimized ensemble learning approaches can provide effective, interpretable, and reliable tools for Alzheimer's disease diagnosis. Future work may focus on validating the proposed framework on external datasets, extending the approach to multi-stage disease classification, and incorporating longitudinal data to further enhance early detection and disease progression analysis.

References

Aydin, A., & Avaroğlu, E. (2024). Contact classification for human–robot interaction with densely connected convolutional neural network and convolutional block attention module. Signal, Image and Video Processing, 18(5), 4363-4374. https://doi.org/10.1007/s11760-024-03078-4

Aydin, A., Salur, M. U., & Aydin, İ. (2021). Fine-tuning convolutional neural network based railway damage detection. IEEE EUROCON 2021-19th International Conference on Smart Technologies, 216-221. https://ieeexplore.ieee.org/abstract/document/9535585/

Hechkel, W., & Helali, A. (2025). Unveiling Alzheimer's disease early: A comprehensive review of machine learning and imaging techniques. Archives of Computational Methods in Engineering, 32(1), 471-484. https://doi.org/10.1007/s11831-024-10179-3

Jenber Belay, A., Walle, Y. M., & Haile, M. B. (2024). Deep Ensemble learning and quantum machine learning approach for Alzheimer's disease detection. Scientific Reports, 14(1), 14196. https://doi.org/10.1038/s41598-024-61452-1

Kaur, A., Mittal, M., Bhatti, J. S., Thareja, S., & Singh, S. (2024). A systematic literature review on the significance of deep learning and machine learning in predicting Alzheimer's disease. Artificial Intelligence in Medicine, 154, 102928. https://doi.org/10.1016/j.artmed.2024.102928

Rabie El Kharoua. (2024). Alzheimer's Disease Dataset. https://doi.org/10.34740/KAGGLE/DSV/8668279

Raza, H. A., Ansari, S. U., Javed, K., Hanif, M., Mian Qaisar, S., Haider, U., Pławiak, P., & Maab, I. (2024). A proficient approach for the classification of Alzheimer's disease using a hybridization of machine learning and deep learning. Scientific Reports, 14(1), 30925. https://doi.org/10.1038/s41598-024-81563-z

Sheng, J., Zhang, Q., Zhang, Q., Wang, L., Yang, Z., Xin, Y., & Wang, B. (2024). A hybrid multimodal machine learning model for Detecting Alzheimer's disease. Computers in Biology and Medicine, 170, 108035. https://doi.org/10.1016/j.compbiomed.2024.108035

Singh, S. G., Das, D., Barman, U., Saikia, M. J., Singh, S. G., Das, D., Barman, U., & Saikia, M. J. (2024). Early Alzheimer's disease detection: a review of machine learning techniques for forecasting transition from mild cognitive impairment. Diagnostics, 14(16). https://doi.org/10.3390/diagnostics14161759

Tarikul Islam, M., Aktaruzzaman, M., Saif, A., Riyad Hasan, A., Hasan Sourov, M. M., Sikdar, B., Rehman, S., Tabassum, A., Abeed-Ul-Haque, S., Hasan Sakib, M., Alam Muhib, M. M., Ahasan Setu, M. A., Tasnim, F., Rayhan, R., M. Abdel-Daim, M., & Obayed Raihan, M. (2024). Identification of acetylcholinesterase inhibitors from traditional medicinal plants for Alzheimer's disease using in silico and machine learning approaches. RSC Advances, 14(47), 34620-34636. https://doi.org/10.1039/D4RA05073H

Journal of Information Analytics, 1(2), 14-20

Journal of Information Analytics

Submit	25.12.2025
Revision	09.01.2026
Accept	09.01.2026
Publish	10.01.2026

Open Access

Research Article

Data-Driven Soccer Ball Trajectory Prediction Using Recurrent Neural Networks

Hanan Abu Kwaider, Faruk Serin*

Department of Computer Engineering, Mersin University, Mersin, Türkiye

Abstract

Accurate prediction of soccer ball trajectories enables advanced tactical analysis and real-time sports insights. This study proposes a deep learning approach for forecasting future ball positions from real match footage. Using the SAM2 segmentation model, frame-by-frame ball coordinates were extracted and refined through frame-reduction methods. Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU) were trained with ball velocity as an additional feature. GRU achieved the lowest one-step-ahead Euclidean error while maintaining a consistent MAPE of 0.03 across both methods. Keyframe extraction yielded greater stability over longer horizons (1–10 steps). The approach demonstrates superior accuracy while operating directly on real-world footage, underscoring its potential for real-time soccer analytics and AI-assisted coaching systems.

Keywords: Soccer ball trajectory prediction, BiLSTM, GRU, Motion-based frame reduction, Keyframe extraction, Sports analytics

Citation: Abu Kwaider, H. and Serin, F., (2025). Data-Driven Soccer Ball Trajectory Prediction Using Recurrent Neural Networks. *Journal of Information Analytics*, 1(2), 14-20.

*Corresponding Author: Hanan Abu Kwaider hanan.abuquader@gmail.com

A version of this work was presented in abstract form at *The Fibonacci International Conference on Engineering, Technology, and Mathematics*.

1. Introduction

The analysis of ball movements in sports has become an important research area, driven by advances in computer vision and deep learning. In soccer, predicting future ball positions provides valuable insights for tactical analysis, strategic planning, and the development of AI-assisted coaching (Naik et al., 2022). Recent progress in object detection, segmentation, and tracking has enabled the transformation of raw match footage into structured data, allowing for more accurate modeling of ball trajectories and game dynamics (Juhai, 2024).

Despite these advances, ball trajectory prediction remains challenging due to the high variability of motion—such as rapid speed changes, sudden direction shifts, and frequent occlusions. These dynamics require models capable of capturing complex temporal dependencies across sequential frames, especially under real-world match conditions.

This study introduces a deep learning-based pipeline for forecasting soccer ball trajectories from real match footage. The proposed pipeline integrates segmentation-based tracking using SAM2 (Ravi et al., 2024) with recurrent neural networks—Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU) architectures (Chung et al., 2014), (Schuster & Paliwal, 1997)—to learn temporal motion patterns from sequential ball positions. To improve data quality and reduce redundancy, two frame-reduction methods are explored: keyframe extraction using PySceneDetect AdaptiveDetector (Detection Algorithms — PySceneDetect 0.6.6 Documentation, n.d.), and motion-based thresholding that filters frames with minimal displacement. Additionally, ball velocity is calculated and incorporated as an input feature to enhance motion representation.

This approach contributes to advancing sports analytics by offering a reliable and scalable method for short-term trajectory forecasting, with potential applications in tactical evaluation and AI-driven coaching systems.

2. Literature Review

With the growing adoption of computer vision and machine learning in sports analytics, numerous studies have addressed ball tracking and trajectory prediction across different sports. These works have established solid foundations for analyzing match footage and forecasting in-game events, but often face limitations such as reduced accuracy over long distances, challenges with real-world data, or a narrow focus on specific aspects of play.

Yutaro et al (2022) developed a model for predicting soccer passes by combining a 3D CNN, LSTM, and Transformer encoder to learn spatial relations between players and the ball. While effective for short-distance passes, it struggled with longer-range predictions.

Anar & Hande (2022) used optical tracking data and a neural network to estimate the ball's 2D position based on player behavior, achieving R^2 scores of 0.79 (x-axis) and 0.92 (y-axis), though performance declined when possession changed over large distances.

Samriddha (2021) proposed a dual-model framework—the Opponent Proximity and Pass Region models—to predict pass recipients from abstract top-view visualizations, achieving strong performance on a custom dataset.

Yang et al (2023) designed a stereo vision-based system integrating an ANN for detection and 3D trajectory estimation, yielding mean errors of 29.6 cm (x), 7.2 cm (y), and 11.7 cm (z) in simulation.

Kim et al (2023) applied Set Transformers with BiLSTM networks to infer ball motion from player trajectories, obtaining less than 3.7 m mean position error and 64.7% player-ball possession accuracy.

Overall, prior approaches demonstrate valuable progress but typically rely on player tracking or simulated data. This study addresses these limitations by predicting ball trajectories directly from real match footage using segmentation-based tracking and recurrent neural networks.

3. Methodology

This study presents a deep learning-based approach for predicting future soccer ball positions from real match footage. The workflow includes data collection, segmentation, and tracking using the SAM2 model, data cleaning and feature engineering, and model training with BiLSTM and GRU architectures. Each model forecasts the ball's motion over short temporal windows and is evaluated across multiple prediction horizons using standard error metrics.

3.1. Data Collection

Match videos were obtained from publicly available YouTube sources. Separate datasets were prepared for training and testing to ensure model generalization. Each video was converted into individual frames and processed using the SAM2 segmentation and tracking model. Prior to running SAM2, players and the ball were manually labeled to improve segmentation accuracy. SAM2 then generated JSON outputs containing frame-level information such as object labels, bounding boxes, and centroid coordinates. For this study, only ball-related data were extracted for subsequent preprocessing and modeling.

3.2. Data Preparation and Preprocessing

While SAM2 effectively tracked the ball, occasional failures produced missing coordinates ($x, y = 0, 0$). These values were corrected using linear interpolation to preserve temporal continuity. Rare visual misidentifications, typically limited to two or three frames per video, had a negligible impact on data quality.

To minimize redundancy and enhance efficiency, two frame-reduction strategies were applied. The first used PySceneDetect's AdaptiveDetector to identify scene changes between consecutive frames. A threshold of 2.0 was empirically selected to eliminate redundant frames while preserving motion continuity. The second strategy employed motion-based thresholding, discarding frames with ball displacement below 8 pixels to remove near-static intervals. Threshold values were determined experimentally to balance information retention and compactness.

Following frame reduction, positional data were Min-Max normalized across training and testing sets. A velocity feature, derived from frame-to-frame displacement, was added to the normalized x and y values, forming a three-dimensional feature vector per time step. Finally, data were organized into 10-frame input sequences to predict ball positions for n future steps (1–10). Predictions were expressed in frames rather than seconds due to variability in ball speed.

3.3. Modeling

Recurrent neural networks (RNNs) are well-suited for sequential data modeling, making them appropriate for forecasting soccer ball trajectories (Wang et al., 2017). In this study, two architectures were implemented: Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU).

BiLSTM captures temporal dependencies from both past and future contexts (Xu et al., 2024), while GRU offers a simpler, more computationally efficient alternative capable of modeling nonlinear motion patterns. Comparing both allows examination of the trade-off between accuracy and model complexity. Both models followed the same structure:

- Input layer
- Recurrent layer with 128 units (bidirectional for BiLSTM, unidirectional for GRU),
- Dropout layer (rate = 0.15),
- Second recurrent layer with 64 units,
- Another dropout,
- Dense layer with 32 ReLU-activated neurons,
- Final dropout layer.
- The output layer produced 2 × n values representing the x and y coordinates of the ball across n future frames.

Training used the Adam optimizer (learning rate = 0.0001) with the Huber loss function for 150 epochs, a batch size of 16, and a 10% validation split.

4. Results and Discussions

The accuracy of the models was evaluated using 1-step-ahead trajectory predictions under both preprocessing strategies—keyframe extraction and motion-based thresholding, and metrics including Mean Absolute Percentage Error (MAPE), R^2 for x and y, and the average Euclidean distance error.

The BiLSTM and GRU models were evaluated across prediction horizons ranging from 1 to 10 frames using two preprocessing strategies: keyframe extraction and motion-based thresholding. Performance was assessed through average Euclidean distance error, R^2 scores for the x and y coordinates, combined R^2, and MAPE. Tables 1 and 2 present representative results for 1-, 5-, and 10-step-ahead predictions, illustrating short-, mid-, and long-term forecasting performance for threshold-based and keyframe-based, respectively.

Table 1. Prediction Performance for Trajectories Obtained Using Threshold-Based Frame Reduction

n_future	Model	Avg Euclidean Error (px)	R^2 X	R^2 Y	Combined R^2	MAPE
1	BiLSTM	20.95	0.9829	0.9841	0.9975	0.03
1	GRU	26.18	0.9775	0.9571	0.9958	0.03
5	BiLSTM	65.35	0.7724	0.9041	0.9721	0.05
5	GRU	67.22	0.8071	0.8709	0.9742	0.06
10	BiLSTM	90.27	0.5451	0.8331	0.9459	0.07
10	GRU	111.86	0.3357	0.7944	0.9227	0.07

Table 2. Prediction Performance for Keyframe-Extracted Trajectories

n_future	*Model*	*Avg Euclidean Error (px)*	$R^2 X$	$R^2 Y$	*Combined* R^2	MAPE
1	BiLSTM	47.04	0.8342	0.9862	0.9893	0.02
1	GRU	37.46	0.884	0.9726	0.9917	0.03
5	BiLSTM	50.15	0.7723	0.9498	0.984	0.03
5	GRU	58.59	0.7076	0.9404	0.9798	0.04
10	BiLSTM	86.57	0.259	0.9159	0.952	0.05
10	GRU	78.58	0.4138	0.9061	0.9607	0.04

Both models achieved high short-term accuracy, with error gradually increasing as prediction horizons extended. Under the keyframe extraction approach,

- GRU exhibited superior short-term performance (steps 1–4),
- while BiLSTM maintained higher stability across medium-range horizons (steps 5–9).
- At the longest horizon (n = 10), GRU slightly regained advantage, suggesting better long-term generalization.

Using threshold-based reduction, GRU consistently achieved lower Euclidean errors and higher R^2 scores than BiLSTM across most horizons, particularly for short-term forecasts. The performance gap narrowed mid-range but widened again at extended horizons, indicating that GRU benefits more from compact, motion-focused inputs. These findings emphasize that frame-reduction methods influence not only model accuracy but also their comparative strengths—GRU excelling in short-range prediction and BiLSTM showing better temporal consistency.

Figures 1–2 visualize 1-step-ahead predictions for both models under keyframe and threshold-based preprocessing, confirming GRU's advantage in capturing immediate motion trends.

Figure 1 - 1-step-ahead prediction with threshold preprocessing

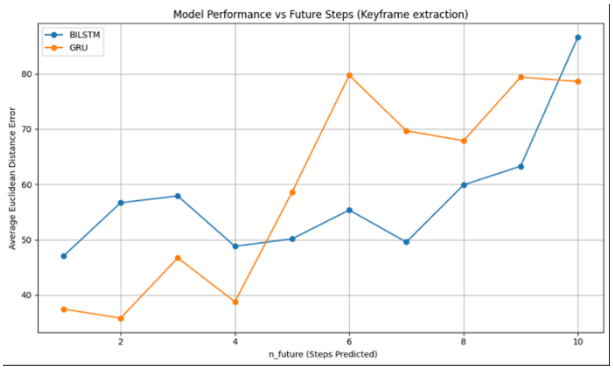


Figure 2- 1-step-ahead prediction with keyframe preprocessing

In practical terms, achieving sub-30 px average errors for immediate steps and below 90 px for 10-step forecasts demonstrates strong predictive capability for broadcast-resolution data. This precision supports reliable trajectory overlays and tactical visualization in live analytics environments.

5. Conclusion

This study introduced a deep learning approach for forecasting soccer ball trajectories directly from real match footage. The proposed pipeline integrates the SAM2 segmentation model with BiLSTM and GRU architectures to learn motion dynamics from ball trajectories obtained from videos with two frame-reduction strategies. The approach is lightweight, independent of player tracking, and suitable for real-world sports analytics applications.

Experimental results demonstrate that the proposed approach achieves strong predictive accuracy while operating directly on unstructured video data, highlighting its practical potential for real-time analytics and coaching applications.

Potential extensions of this work involve improving segmentation robustness under occlusions and camera motion, expanding the dataset, and incorporating player trajectory prediction for more comprehensive tactical modeling.

References

Anar, A., & Hande, A. (2022). Prediction of the ball location on the 2D plane in football using optical tracking data. Academic Platform Journal of Engineering and Smart Systems. https://doi.org/10.21541/apjess.1060725

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling (No. arXiv:1412.3555). arXiv. https://doi.org/10.48550/arXiv.1412.3555

Detection Algorithms—PySceneDetect 0.6.6 documentation. (n.d.). Retrieved May 5, 2025, from https://www.scenedetect.com/docs/latest/api/detectors.html#scenedetect.detectors.adaptive_detector.AdaptiveDetector

Juhai, W. (2024). Analyzing players' tactical positions and movement trajectories in soccer matches using machine vision algorithms. Journal of Electrical Systems. https://doi.org/10.52783/jes.3126

Kim, H., Choi, H.-J., Kim, C. J., Yoon, J., & Ko, S.-K. (2023). Ball trajectory inference from multi-agent sports contexts using set transformer and hierarchical Bi-LSTM. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 4296–4307. https://doi.org/10.1145/3580305.3599779

Naik, B. T., Hashmi, M. F., & Bokde, N. D. (2022). A comprehensive review of computer vision in sports: Open issues, future trends and research directions. Applied Sciences, 12(9), Article 9. https://doi.org/10.3390/app12094429

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024). SAM 2: segment anything in images and videos (No. arXiv:2408.00714). arXiv. https://doi.org/10.48550/arXiv.2408.00714

Samriddha, S. (2021). Who will receive the ball? Predicting pass recipient in soccer videos. Journal of Visual Communication and Image Representation. https://doi.org/10.1016/j.jvcir.2021.103190

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681. https://doi.org/10.1109/78.650093

Wang, L., z, l, & Yi, Z. (2017). Trajectory predictor by using recurrent neural networks in visual tracking. IEEE Transactions on Cybernetics, PP, 1–12. https://doi.org/10.1109/TCYB.2017.2705345

Xu, Y., Pan, Q., Wang, Z., & Hu, B. (2024). A novel trajectory prediction method based on CNN, BiLSTM, and multi-head attention mechanism. Aerospace, 11(10), Article 10. https://doi.org/10.3390/aerospace11100822

Yang, Y., Kim, D., & Choi, D. (2023). Ball tracking and trajectory prediction system for tennis robots. Journal of Computational Design and Engineering, 10(3), 1176–1184. https://doi.org/10.1093/jcde/qwad054

Yutaro, H., Rei, K., Ryota, Y., Kenta, K., & T., N. (2022). Pass Receiver prediction in soccer using video and players' trajectories. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Wo Rkshops (CVPRW). https://doi.org/10.1109/CVPRW56347.2022.00394

Journal of Information Analytics, 1(2), 21-28

Submit 25.12.2025
Revision 06.01.2026
Accept 07.01.2026
Publish 08.01.2026

Journal of Information Analytics

Research Article

Open Access

Comparative Evaluation of Video-Based Repetitive Movement Analysis Using Multiple Methods

Yasin Badur[1,], Faruk Serin[2]*

[1] Web Design and Coding Program, Vocational School of Technical Sciences, Harran University, Şanlıurfa, Türkiye

[2]Department of Computer Engineering, Faculty of Engineering, Mersin University, Mersin, Türkiye

Abstract

This study aims to analyze and compare repetitive human movements using skeleton-based keypoints extracted from videos. Using keypoints detected with the YOLOv11 algorithm, we apply the Discrete Fourier Transform (DFT), Principal Component Analysis (PCA), and Matrix Profile. The DFT provides frequency-domain characterizations of motion, PCA attenuates noise to yield more salient patterns, and the Matrix Profile enables precise detection of recurring motifs in the time series. Empirical findings indicate that Fourier-based analyses better capture global structure and provide more discriminative similarity measures, while the Matrix Profile complements them by detecting repeated motifs and onset/offset boundaries. This study aims to contribute to the development of more accurate and reliable approaches in the field of human motion analysis.

Keywords: Motion Analysis, Fourier Transform, Matrix Profile, Dynamic Time Warping,

Repetitive Movements, Skeleton-Based Features

Citation: Badur, Y., & Serin, F. (2025). Comparative Evaluation of Video-Based Repetitive Movement Analysis Using Multiple Methods, Journal of Information Analytics,1(2), 21-28.

*Corresponding Author: Yasin Badur yasin.badur.nk@gmail.com

A version of this work was presented in abstract form at *The Fibonacci International Conference on Engineering, Technology, and Mathematics*.

1. Introduction

Video-based human pose estimation is an active research topic in computer vision and AI, yet robust generalization remains challenging due to complex spatiotemporal structure, occlusions, viewpoint changes, heterogeneous capture conditions, and limited or imbalanced datasets. Although state-of-the-art models achieve high per-frame keypoint accuracy (e.g., Chen et al., 2023), the detection and comparison of recurring motion patterns and the measurement of cross-video similarity remain underexplored, reducing discriminative power and making reliable evaluation more difficult (e.g., Usman & Zhong, 2022).

We address this gap by modeling pose trajectories as multivariate time series and conducting pattern-based similarity analysis using dynamic time warping (DTW), Fourier analysis, principal component analysis (PCA), and the Matrix Profile to capture temporal regularities beyond visual matching. Our aims are to characterize the strengths and limitations of existing techniques, identify core gaps, and provide practical guidance—covering datasets, metrics, and protocols—for fair and reproducible assessment. The anticipated impact spans sports and rehabilitation (fine-grained performance/impairment analysis), security (recurrent/anomalous behavior detection), and human–machine interaction/AR.

The contributions: (1) we reframe the task of detecting and comparing recurring motions in skeleton-based video by specifying explicit assumptions and evaluation criteria; (2) we present a gap analysis that surfaces limitations of current methods and distills practitioner-oriented design principles; and (3) we propose actionable solution directions and experimental blueprints—covering dataset choice, metrics, and protocols—to enable fair and reproducible comparisons.

2. Proposed Method

2.1. Analyzed Pose-Estimation Algorithms

OpenPose provides open-source, multi-person pose estimation with optional hand/face tracking, though its computational load can limit real-time deployment (Cao et al., 2017). MediaPipe is lightweight and fast, runs in real time on mobile/low-power devices, and is widely used in healthcare, fitness, and AR (Lugaresi et al., 2019). PoseNet targets web/mobile via TensorFlow.js/TFLite and remains robust at low resolution, but is comparatively constrained in multi-person scenes (Papandreou et al., 2018). The YOLO family offers a strong speed–accuracy trade-off for detection and pose, with pose-optimized variants adopted in sports, rehabilitation, and entertainment.

We applied YOLOv11 for per-frame keypoint extraction owing to its real-time throughput and stability, high keypoint localization accuracy, and straightforward cross-platform deployment; it is well suited to repetitive-movement analysis. Our evaluation targets repetitive actions—rope skipping, hands-clapping, and mixed exercise sequences.

2.2. Video Data

Seven videos were selected to span movement types and conditions. Criteria included: (i) repetitive actions (jumping, running, warm-ups, therapy), (ii) sufficient resolution/quality for reliable keypoint detection, and (iii) varied lighting/backgrounds. Preprocessing trimmed relevant segments, standardized frame rates, and normalized formats/resolutions. Each video was processed with YOLOv11 to obtain per-frame keypoints; repetition frequency and accuracy were derived downstream from time-series analyses (Fourier, Matrix Profile, DTW).

2.3. Repetitive-Movement Analysis Methods

Per-frame poses are modeled as multivariate time series over 17 keypoints × (x, y) (34 dimensions). We apply complementary analyses to capture periodicity, alignment, and structural similarity:

- PCA projects the 34-D frame representation onto principal components (typically PC1), denoising and exposing dominant motion patterns.
- Fourier analysis quantifies periodic structure using (i) separate x/y channels and (ii) complex signals $x + iy$ to capture planar dynamics; dominant spectral peaks indicate repetition rates
- DTW aligns sequences under variable execution speeds and yields sequence-level similarity (Sakoe & Chiba, 2003).
- Matrix Profile discovers motifs and local anomalies via nearest-neighbor subsequence distances; we apply it to the PCA-reduced 1D series to recover repetitions and onset/offset boundaries.
- Similarity metrics include Euclidean distance (spatial proximity), cosine similarity (angular consistency of limb/trajectory vectors), and Pearson correlation (linear association for intra-class consistency and inter-class separability).

DTW addresses temporal alignment; Fourier captures periodicity; Matrix Profile localizes motifs and segment boundaries; geometric/angle/correlation metrics assess spatial and directional coherence.

3. Experimental Results

3.1. Time Series Analysis of Keypoint Coordinates

Across Rope Skipping 1–2 (Figure 1), the skeleton keypoint trajectories exhibit a clear and stable periodicity. Cranial keypoints show low-amplitude, near-parallel fluctuations, whereas distal joints (wrists, ankles, knees) display higher-amplitude oscillations that carry most of the discriminative signal. In Rope Skipping 1, brief transients—consistent with crossed-hands execution and occasional occlusion—appear predominantly in the wrist channels; these remain local and do not disrupt the global cycle structure. Rope Skipping 2 presents more regular, phase-locked cycles.

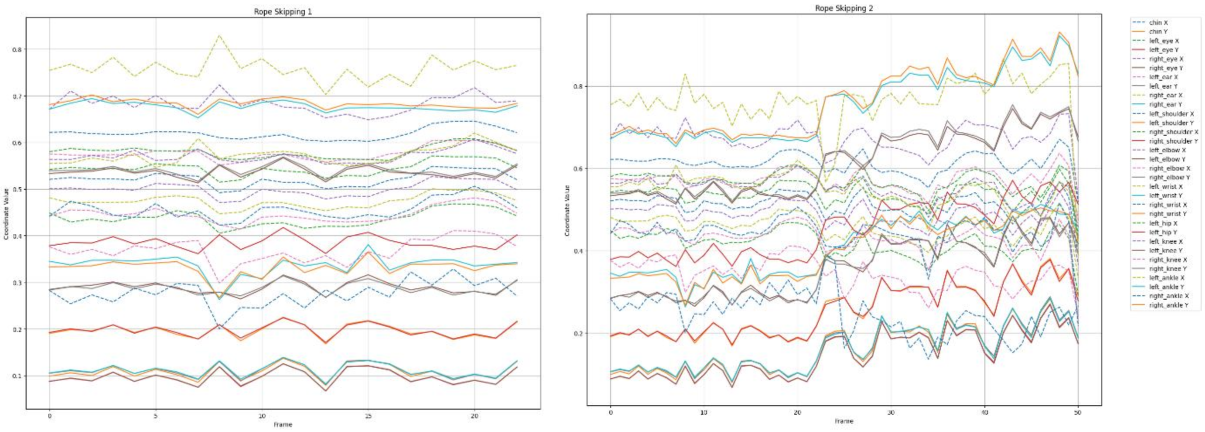


Figure 1. *Keypoint analysis of rope skipping movements*

For "Jumping with hands clapping" (Figure 2), onset/offset phases are demarcated by pronounced plateaus and transition regions. The first clip shows a short initial quiescence followed by highamplitude rhythmic oscillations; the second clip runs at a lower cadence, yielding longer periods and more salient phase transitions. Transient drops in wrist/shoulder signals are consistent with occlusion or temporary out-of-frame motion and do not alter the repetition structure.

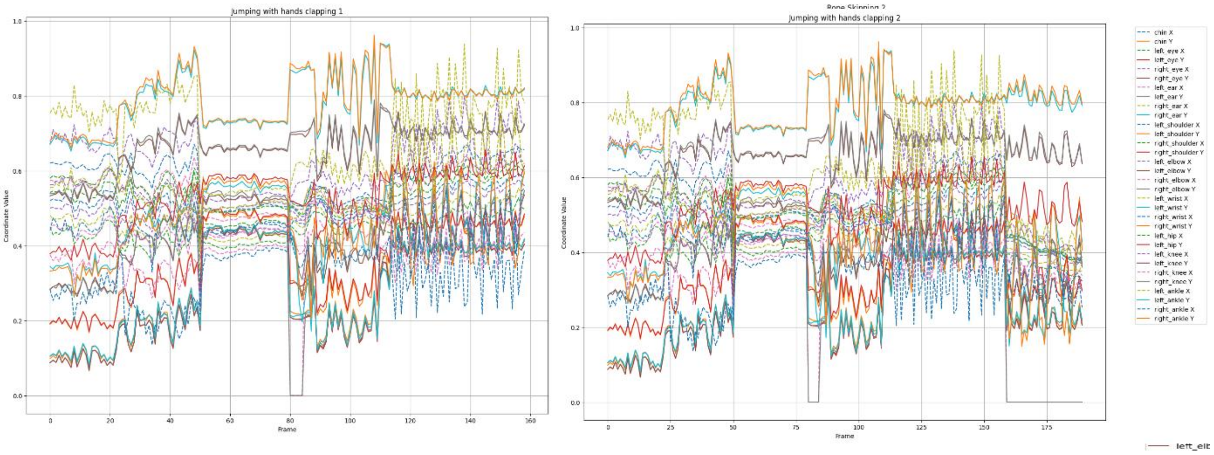


***Figure 2**. Keypoint analysis of jumping with hands clapping*

The figures show that (i) head-region channels serve as a stable reference, (ii) distal joints constitute high-variance, information-rich channels, and (iii) despite local variability from execution style and cadence, the global periodic structure is preserved. These findings are consistent with—and reinforce—subsequent PCA, Fourier, and Matrix Profile analyses, indicating that the recordings furnish reliable, comparable time series for cross-video evaluation.

3.2. Analysis of Time Series of All Movement Coordinates Using PCA

We applied PCA to the x–y coordinates of 17 keypoints per motion, extracting a single principal component to obtain a one-dimensional, interpretable time series. Compared with the raw multi-channel traces (Figure 1–2), the PCA projection (Figure 3) makes periodicity, repetition counts, and segment boundaries more salient. Even in the clip containing successive, heterogeneous actions, recurring patterns were clearly exposed. Overall, PCA reduced noise and dimensionality, improved cross-record comparability, and provided a compact, information-dense representation for downstream analyses (Fourier analysis, DTW, Matrix Profile).

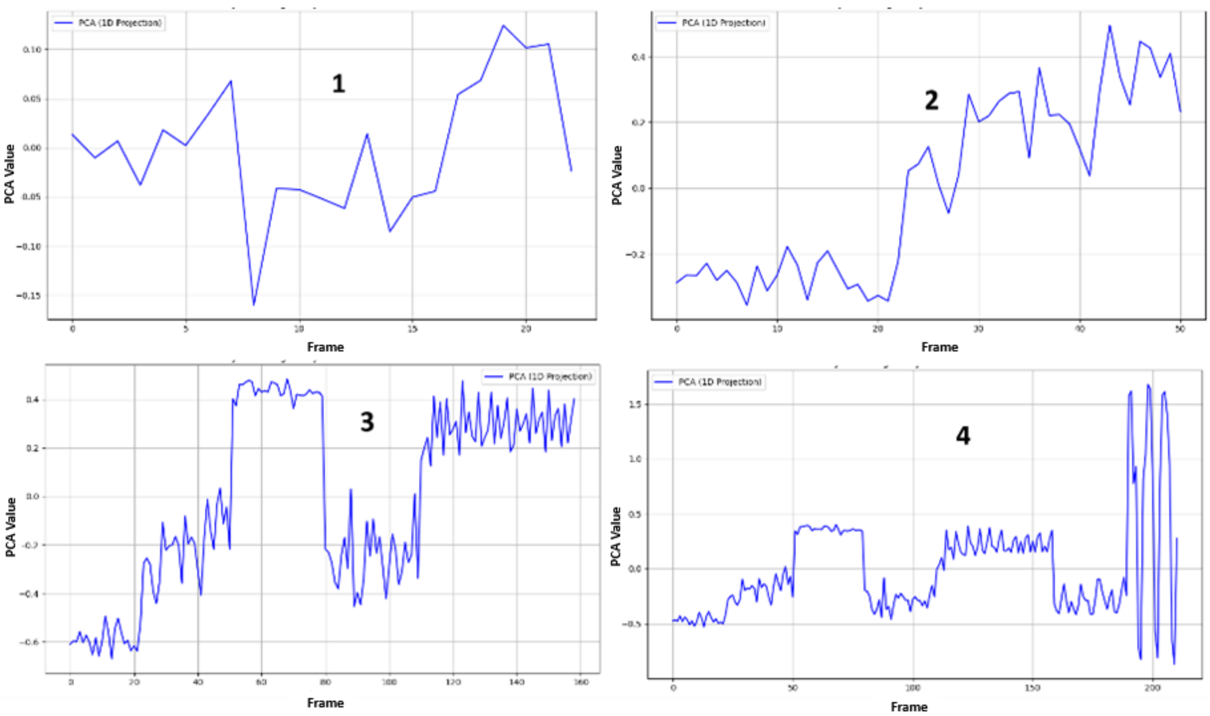


Figure 3. *PCA-based analysis of all movement coordinates (1: Rope Skipping-1, 2: Rope Skipping-2, 3: Jumping with Hands Clapping-1, 4: Mixed Exercises).*

3.3. Pattern Analysis of Keypoints Using Fourier Transform (With and Without PCA)

Across all datasets, we computed Fourier spectra separately for the PCA-reduced 1D series and for the original multivariate series. While Fourier analysis—especially after PCA—provides a stable basis for between-sequence similarity, it is insufficient on its own for accurate repetition counting; accordingly, we use Matrix Profile to recover repetition structure and boundaries, and Fourier (often after PCA) for discriminative inter-sequence similarity. As shown in Table 1, Fourier-based period estimates provide an approximate repetition count..

Table 1. Pattern Frame Count and Approximate Repetition Number of Videos (Fourier)

Video Name	Total Frames	Actual Repetitions	Approx. Period (frames)	Approx. Repetitions (Without PCA / With PCA)
rope skipping-1	380	14	190	2/2
rope skipping-2	620	19	207	3/24
rope skipping-3	283	9	28	10/10
rope skipping-4	305	12	17	18/18
jumping with hands clapping-1	276	9	276	1/9
jumping with hands clapping-2	208	4	26	2/4
different exercises	327	3	47	7/3

For pattern similarity, Table 2 (without PCA) indicates generally high cosine similarity—reflecting directional proximity—alongside weak Pearson correlations and middling Euclidean/DTW distances. In contrast, Table 3 (with PCA) shows marked reductions in Euclidean and DTW distances and more coherent association patterns: similar pairs exhibit high positive correlation, whereas out-of-phase pairs show strong negative correlation. Notably, “Jumping with Hands Clapping 1–2” and “Rope Skipping 1–

4” emerge as strongly similar (low distances, high correlation), while “Rope Skipping 2–3” displays a pronounced negative correlation indicative of phase/misalignment effects.

Table 2. Pattern Similarity Ratios of Keypoint Averages of Repetitive Movements in Videos Obtained by Fourier Transform (Without PCA)

Motion Pair	Euclidean Distance	Pearson Correlation	Cosine Similarity	DTW Distance
rope skipping 1 - 2	20.17	0	0.88	9.65
rope skipping 1 - 3	9.35	-0.01	0.82	4.71
rope skipping 1 - 4	5.80	-0.04	0.88	3.03
rope skipping 2 - 3	8.76	-0.01	0.84	4.68
rope skipping 2 - 4	5.18	-0.02	0.90	2.75
rope skipping 3 - 4	6.40	0.02	0.85	3.38
hands clapping 1 - 2	7.47	-0.06	0.89	4.32
hands clapping 1 - different exercises	8.46	-0.01	0.93	4.81
rope skipping 2 - different exercises	9.28	-0.01	0.90	4.64

Table 2 shows that, in the analyses conducted without PCA, cosine similarity scores were generally high, whereas Pearson correlation remained low; nevertheless, the motions were found to be similar in terms of directionality.

Table 3. Pattern Similarity Ratios of Keypoint Averages of Repetitive Movements in Videos Obtained by Fourier Transform with PCA

Motion Pair	Euclidean Distance	Pearson Correlation	Cosine Similarity	DTW Distance
rope skipping 1 - 2	2.12	-0.63	-0.63	1.64
rope skipping 1 - 3	2.07	0.54	0.54	0.96
rope skipping 1 - 4	1.00	0.83	0.82	0.86
rope skipping 2 - 3	2.65	-0.98	-0.98	2.14
rope skipping 2 - 4	0,8	-0.33	-0.33	0.47
rope skipping 3 - 4	1.70	0.31	0.31	1.21
hands clapping 1 - 2	0.58	0.84	0.84	0.29
hands clapping 1 - different exercises	3.35	0.35	0.35	2.98
rope skipping 2 - different exercises	3.36	-0.40	-0.39	3.04

In sum, Fourier analysis effectively reveals the periodic nature of the motions and—especially when combined with PCA—provides a reliable and consistent basis for similarity assessment. However, it is not sufficient on its own for accurate repetition counting; complementary approaches such as Matrix Profile are recommended for that purpose.

3.4. Keypoint Pattern Analysis with Stumpy (Matrix Profile) and PCA

As summarized in Table 4, Matrix Profile was applied to the PCA-reduced (1D) keypoint time series to detect motifs, onset/offset boundaries, and repetition structure. Matrix Profile was applied to the PCA-reduced (1D) keypoint time series to detect motifs, segment boundaries (onset/offset), and repetition structure. Across all datasets, the method reliably exposed periodicity and localized repeating segments, while also differentiating heterogeneous motion sequences. Similar pairs exhibited low DTW distances and high positive correlation, whereas out-of-phase pairs showed strong negative correlation indicative of phase effects. Euclidean distance tracked structural proximity, and cosine similarity remained consistent with correlation under z-scoring.

Table 4. Keypoint Pattern Analysis with Stumpy (Matrix Profile) and PCA

Video Pairs	Best Pattern Window Size 1	Estimated Repetition Count 1	Best Pattern Window Size 2	Estimated Repetition Count 2	Actual Repetition Count	DTW Distance	Correlation	Euclidean Distance	Cosine Similarity
rope skipping 1-2	20	18.33	20	12.40	4	0.36	-0.63	0.99	0.35
rope skipping 1-3	20	18.33	20	2.42	9	0.38	-0.90	0.99	0.38
rope skipping 1-4	20	18.33	20	5.87	9	0.36	0.01	0.99	0.29
rope skipping 2-3	20	12.40	20	2.42	2	0.09	0.56	0.99	0.07
rope skipping 2-4	20	12.40	20	5.87	9	0.09	0.69	0.99	0.07
rope skipping 3-4	20	2.42	20	5.87	5	0.17	-0.01	0.99	0.10
rope skipping 1-different exercises	20	18.33	20	3.37	4	0.69	0.73	0.99	0.68

Overall, Matrix Profile is well suited for repetition counting and segmentation and for revealing temporal motif topology, but it is less discriminative than Fourier-based analysis for fine-grained structural similarity. Accordingly, we recommend using Matrix Profile to recover repetition structure and boundaries, in conjunction with Fourier (and related metrics) for precise inter-motion similarity assessment.

4. Conclusion

This study evaluated skeleton-based analyses of repetitive human motion in video. A YOLOv11-based inference pipeline achieved high detection accuracy and throughput, underscoring its suitability for applied settings such as sports performance analysis, rehabilitation monitoring, and human–robot interaction. Among the downstream analyses, Fourier-based representations most effectively quantified inter-motion similarity, yielding lower dynamic time warping (DTW) distances and higher cosine-similarity scores, while PCA improved signal quality by attenuating noise. The Matrix Profile method excelled at recovering repetition structure—localizing motifs and delineating onset/offset boundaries—but was less precise than Fourier analysis for fine-grained structural similarity. Taken together, Fourier is preferred for discriminative similarity measurement, whereas the Matrix Profile is a complementary tool for robust motif and boundary detection; using both in tandem provides a comprehensive characterization of repetitive motion.

Generalizability is limited by the scope and diversity of the available datasets. Future research should employ larger, better-balanced corpora spanning varied motion types and environmental conditions (e.g., viewpoints, occlusions, illumination) and pursue real-time–oriented enhancements, including computational optimization, multi-view fusion, angle/topology-invariant features, and domain adaptation.

References

Agarwal, S., Gupta, M., Khandelwal, S., Jain, P., Aggarwal, A., Singh, D., & Mishra, V. K. (2021). FitMe: a fitness application for accurate pose estimation using deep learning. 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), 232–237.

Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3686–3693.

Chen, Y., Xie, X., Yin, W., Li, B., & Li, F. (2023). Structure guided network for human pose estimation. *Applied Intelligence*, 53(18), 21012–21026. https://doi.org/10.1007/S10489-023-04521-8/FIGURES/7

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).

Dantone, M., Gall, J., Leistner, C., & Van Gool, L. (2013). Human pose estimation using body parts dependent joint regressors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3041–3048.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*.

Papandreou, G., Zhu, T., Chen, L. C., Gidaris, S., Tompson, J., & Murphy, K. (2018). Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 269-286).

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2022). Video-based human action recognition using deep learning: A review. https://arxiv.org/abs/2208.03775v1

Sakoe, H., & Chiba, S. (2003). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, *26*(1), 43-49.

Samkari, E., Arif, M., Alghamdi, M., & Al Ghamdi, M. A. (2023). Human pose estimation using deep learning: a systematic literature review. *Machine Learning and Knowledge Extraction*, 5(4), 1612–1659.

Usman, M., & Zhong, J. (2022). Skeleton-based motion prediction: A survey. *Frontiers in Computational Neuroscience*, 16, 1051222. https://doi.org/10.3389/FNCOM.2022.1051222/BIBTEX

Wang, C., Zhang, F., & Ge, S. S. (2021). A comprehensive survey on 2D multi-person pose estimation methods. *Engineering Applications of Artificial Intelligence*, 102, 104260.

Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. *32nd AAAI Conference on Artificial Intelligence*, AAAI 2018, 7444–7452. https://arxiv.org/abs/1801.07455v2