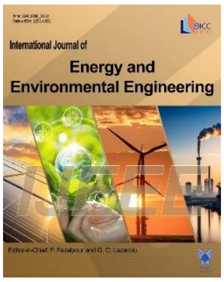


Volume 16, Issue 2, (2025) Article 8

International Journal
of
Energy and Environmental Engineering (IJEEE)

https://doi.org/10.57647/ijeee.2025.1602.08

Edge-Based AI Models for Detecting Nonlinear Energy Consumption Patterns in Smart Grid Environment Networks

Wei Hu*

Xinyang University, Xinyang City, Henan Province, 464000, China

***Corresponding author:** WeiCuiw@outlook.com

Original Research

Received:
28 December 2024

Accepted:
25 April 2025

Published in Issue:
30 June 2025

Abstract

This paper presents an innovative Edge AI framework to detect nonlinear energy usage patterns in real-time with the application to Smart Grid infrastructures. Designed for use on Edge devices where there is limited processing power but required to be highly analytic error-free, the proposed framework is lightweight. The authors developed GBOCLE - Energy, an efficient Anomaly Detector Model based on advanced Gradient Boosting methods specifically for low latency and real-time energy consumption data analysis. This model uses compressed forms of Light Gradient Boosting and One-Class SVM Algorithms to discover temporal, contextual, and relational anomalies across multiple Nodes on the Smart Grid network. Additionally, three techniques, Simplified Terrestrial Analysis, Adaptive Isolation-based Scoring, and Lightweight Graph-Based Neighbour Mechanisms, were used to further enable the detection of nonlinear relationships and interactions among the different components within the grid. The model uses Compact Nonlinear Indicators (e.g., Consumption Dev Index, Device Influence Vector and Reduced-Order Chaotic Metrics) as Analytical Features and Hidden Indicators for Anomalous Energy Usage. The experimental findings indicate that the suggested ensemble model has excellent detection accuracy (0.984), precision (0.974), sensitivity (0.975), specificity (0.965), F1-score (0.975), and AUC (0.984) while minimizing the computational and memory requirements for edge deployment. These findings support the conclusion that optimized edge-based detection is an efficient and feasible solution for detecting energy consumption anomalies and therefore offers the potential to enable predictive management and enhance operational performance in the Smart Grid environment.

Keywords: Smart Grid Systems, Energy Consumption Analysis, Nonlinear Load Patterns, Edge Computing, Artificial Intelligence (AI), Machine Learning Models, Energy Forecasting, Load Monitoring and Management

Cite this article: Hu, W., Edge-Based AI Models for Detecting Nonlinear Energy Consumption Patterns in Smart Grid Environment Networks . Int. J. Energy Environ. Eng. 2025; 16(2): Article 8. https://doi.org/10.57647/ijeee.2025.1602.08

1. Introduction

Real-time analytics on how energy and its sources are consumed are required for the management of smart grids and contemporary energy systems to achieve greater efficiency, reliability and continuity of operation. Identifying mismatches or deviations from 'normal' or anticipated levels of use allows for the prevention of energy-related waste, breakdowns of power supply equipment and theft of power which creates lost revenues and increased vulnerability to power outages [1]. Conventional approaches to monitoring energy use are based on collecting and processing information on a single site or location through centralised processing facilities using equipment which employs basic linear statistical models [2]. Energy systems produce data at very high rates and with the use of various types of sensors which provide a large quantity of time-sensitive and continuously-updated high-dimensional data regarding electricity consumption, including the conditions of the premises where it is being used and the status of the lighting and heating systems in those premises [3]. To obtain insight from these types of datasets it requires more powerful processing approaches of analysing and examining vast amounts of data quickly to be able to identify patterns, non-linear patterns or subtle correlations that are indicative of 'waste', the potential for failure of equipment or theft [4]. One potential method being utilised to detect anomalies in real-time without requiring sophisticated computing capability at the point of production is through the use of artificial intelligence on edge devices [4]. Dynamic behavior of loads, the use of equipment together and the interaction of many items contribute mostly to the nonlinear change in energy use. Utilizing machine learning to address these nonlinearities has been proven successful with models such as Gradient Boosting, One-Class SVM and with the interoperability of multiple machine learning algorithms (ensemble methods). For example, many types of sensors are used in industrial and building energy systems.

Existing technology allows sensors and other devices to communicate with each other to provide anomaly detection. Although AI technologies deployed at the edge (i.e., processing done at the location of the sensor or device) have been proven effective for identifying anomalies, current industry practices have several limitations including limited integration of multi-modal sensor information, a lack of low-cost options, and limitations on the number of real-time operational inferences they can produce. Linear or traditional machine learning algorithms generally produce a higher number of false positives than would be practical for operational energy systems. The goal of this study is to propose an innovative way of utilizing edge-based AI models to identify the nonlinear patterns in Smart Grid Energy Consumption, compared with current techniques. The new methodology uses compressed gradient boosting models, one-class SVMs and ensemble methods, modified to use computationally less resource-intensive techniques to identify features based on context, time and graph. The framework is designed specifically to run on low-powered Edge Devices, allowing for rapid detection of anomalies as well as predictive management of energy.

1.1. Aims

The major contributions of this research study are as follows:

- To provide pre-processing of multimodal energy consumption data, including noise reduction, time synchronization and replacement of missing values, to allow for the generation of a clean and analysable Multimodal Dataset.
- To identify nonlinear features that represent temporal, contextual and relational component patterns contained within energy consumption data for the identification of anomalies.
- To create a detection hybrid called GBOCLE-Energy, based on combining XgBoost, LightGBM and One-Class SVM to produce a detection methodology that incorporates supervised learning with unsupervised anomaly detection methodologies to allow for the detection of anomalies in synchronized consumption data.
- To apply Temporal, Contextual and Relational Examination Techniques, e.g. RMTA, WISC and DGNN for robust identification of atypical energy consumption patterns within Smart Grid Networks.

1.2. Literature review

There has been a significant amount of research that has focused on identifying energy network anomalies through the use of different techniques including statistical, rule-based and machine learning. For example, researchers have applied gradient boosting and SVM techniques to identify anomalous load patterns and peak usage of energy [10]. Combining multiple models into an ensemble gives machine learning higher accuracy along with a way to mitigate noise and uncertainty within sensor data [11]. Additionally, many studies focus on edge-deployed artificial intelligence for energy tracking which reduces the latency, energy use, and memory footprint of energy monitoring systems while maintaining high levels of anomaly detection accuracy [12–14]. Despite these improvements, existing systems still have challenges integrating multimodal sensor data, detecting nonlinear relationships, and functioning statically in a resource-constrained environment edgebased device. The system proposed in this paper addresses these issues with an integrated approach that employs advanced and non-linear feature extraction techniques, existing lightweight AI models, and edgedeployment of algorithms for effectively and efficiently identifying anomalies in real time while also ensuring that the algorithm uses a minimum amount of energy Table 1.

Table 1. Research gap

Author	Proposed Model	Performance	Dataset Used	Significance	Limitations
Forrest & McHale, 2019	In-play forecasting with forensic statistics	Used as legal evidence in sport courts	Real-time betting odds (Football, Tennis)	Legally accepted; detects match-fixing	Only suggests suspicion, not conclusive proof
Kim et al. 2024	LR, RF, SVM, KNN, Ensemble	RF/KNN/Ensemble > 92%; LR/SVM ~80%	Football league odds from 12 companies	Real-time fix detection; ensemble improves accuracy	Needs accurate, timely data
Felice & Ley, 2025	ML with statistical features	>80% accuracy	Female handball match data	Coaching insights; explainability	Sport-specific (handball)
Bunker et al. 2024	ADTrees & Logistic Regression vs Elo Ratings	ADTrees beat Elo; similar to betting odds	ATP Tennis (2005–2020)	Combines interpretability with performance	Elo comparison may not apply universally
Hu et al. 2024	LGBM, Logistic Regression + Momentum Formula	LGBM: 97% (momentum), LR: 70% (point prediction)	Wimbledon 2023	Momentum insights for real-time strategy	Formula may not generalize to other sports
Rodu et al. 2024	Explainable/Interpretable ML in hypothesis testing	Case-study insights	Various sports case studies	ML use aligned with scientific workflow	Requires cautious integration
Kamalakumar et al. 2024	RF, LR, SVM, KNN, Decision Tree	Improved accuracy across models	IPL & cricket data	Player/team insights; fan engagement	Cricket's unpredictability affects results
Brill et al. 2024	Expected Points + Catalytic Prior Smoothing	Reduced overfitting; reliable predictions	American football	Robust decision support; accounts for data structure	Complex modeling; domain-specific
Liu & Wang, 2024	ViBe (improved) + OpenPose + SVM	87.5% accuracy	Sports images	Accurate image-based behavior recognition	Not designed for live video analysis
Alshardan et al. 2025	Cloud-to-Thing with GAN + Mask R-CNN	94.25% accuracy, 5.1 ms latency	Basketball data	Real-time tracking and feedback	High system requirements

1.3. Literature review

The loss of reliability and operational efficiency in the electricity grid can be caused by anomalies in the way that the electricity network consumes energy. These anomalies can come from the following sources: the performance of devices, the coordination of energy-consuming loads, or inefficiencies in the way that the electricity grid operates. Current methods of detecting anomalies are based on linear models and centralized analysis. These methodologies are not well-suited for capturing the complex, nonlinear relationships that occur in a large-scale, dynamic energy network. This project develops a new approach to anomaly detection based on an edge-based artificial intelligence framework capable of detecting nonlinear behaviour in real-time.

The proposed edge-based AI detection framework is designed to enable utility companies to better monitor and manage energy consumption, improve operational efficiency, and support sustainable smart grid practices.

2. Materials and methods

Under this section the overall process carried out by the proposed model is discussed in detail. The pre-processing, feature extraction, non-linear modelling, ML based pattern recognition and detection phases were explained in detail along with novel approaches. Also, the mathematical expressions and novel architectures were also added in this section. Thus, the overall architecture of the proposed model is illustrated in Figure 1. The proposed framework for detecting nonlinear match-fixing patterns in player tracking data follows the workflow shown in the figure which belongs to ML Approaches to Identify Nonlinear Match-Fixing Patterns. The preprocessing stage implements Hampel Filter followed by Z-Score Normalization supported by DTW for temporal synchronization and Miss Forest for missing value imputation to create clean synchronized data. Two types of material science analogies (Bouc-Wen hysteresis loops and DMM) and chaos-based features extraction (LLE, ApEn, RPA) and fractal behavior mapping (Higuchi's Fractal Dimension) work together to quantify complex dynamic patterns. Feature extraction contains PBIV and MMDI which create advanced representations that display player impact and change in momentum.

The pattern recognition stage uses ML to blend Temporal Learning (RMTA) with Contextual Analysis (WISC) and Relational Modeling (DGNN) to perform full-scale anomaly detection on time-based data and context-linked data and relational patterns. The anomaly classification within GBOCLE detection model merges XGBoost with LightGBM and OC-SVM through a supervised boosting and unsupervised anomaly detection framework for robust performance. The evaluation process of the model uses outcome-based assessment to measure its performance in identifying match-fixing patterns that show simultaneous underperforming with high accuracy therefore improving sporting integrity by using multiple approaches to detect non-linear patterns.

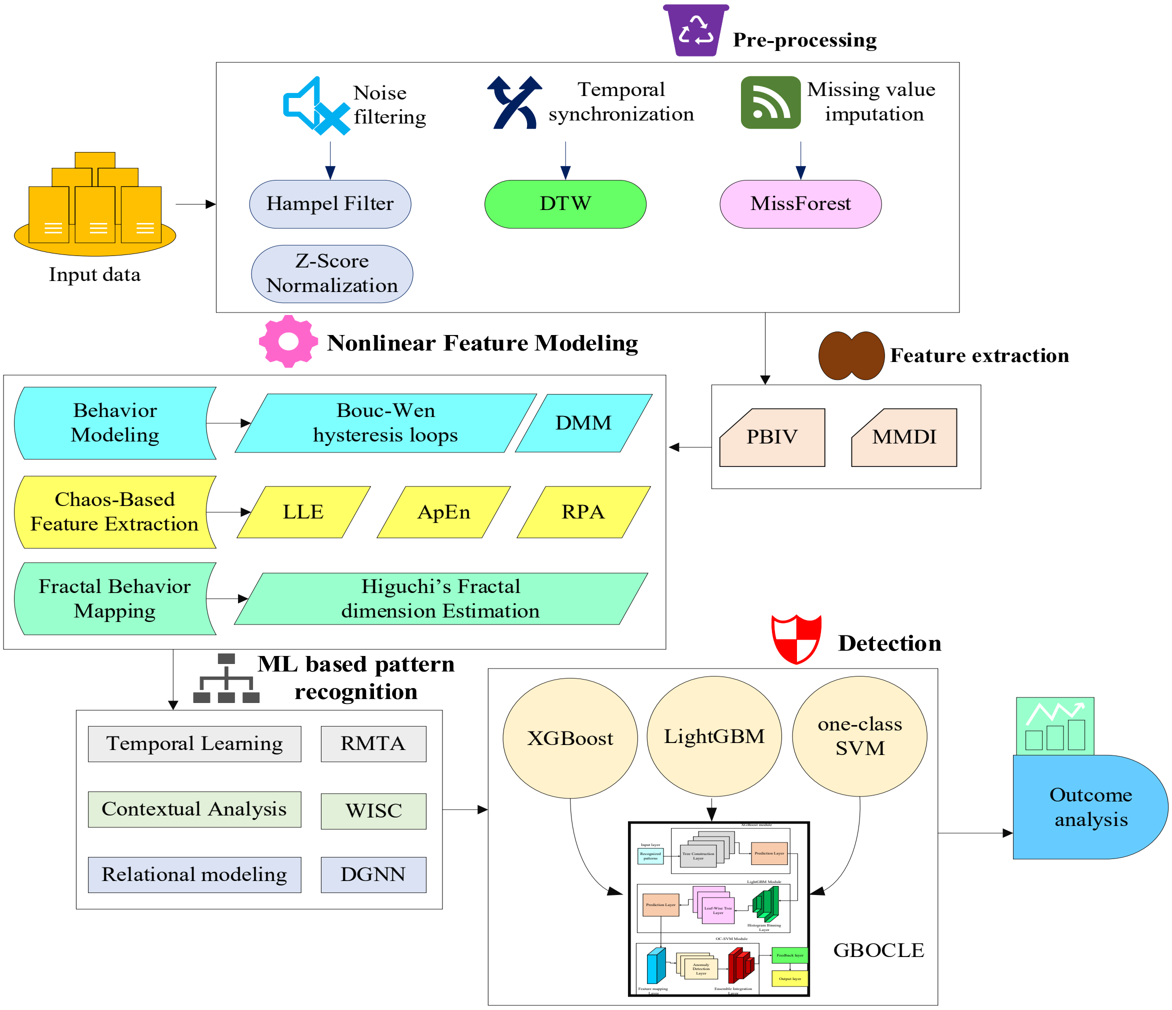


Figure 1. Overall architecture of the proposed model

2.1. Data collection

Through player tracking data systems analysts can detect non-linear match fixing patterns since the information provides timely analysis on player behaviors and competitive elements in sports competitions. Sports analytics platforms such as Sport radar and Stats Perform along with optical tracking devices (e.g. Hawk-Eye) supply data which includes player positioning data in addition to speed metrics, acceleration values, distance statistics and real-time reports on in-match decision processes. The detailed measurements reveal hidden anomalies including deliberate poor performance and abnormal movements or group actions that differ from statistical standards which may show match-fixing. Complex data management techniques must be applied to sports monitoring data because the data quantity meets challenges from diverse formats and numerous missing values and data irregularities. Time-aligned integration of player tracking data with betting market trends along with public sentiment requires feature engineering to make ML models compatible for detecting complex match-fixing patterns.

2.2. Pre-processing

The preprocessing of player tracking data starts with Hampel Filter noise filtering to eliminate erroneous coordinates followed by Z-score normalization for standardization and then uses DTW for temporal synchronization and Miss Forest imputation for missing value handling. The data processing sequence produces clean data that maintains alignment and data completeness for ML detection of nonlinear match-fixing patterns.

2.2.1. *Noise filtering*

The accuracy of player tracking data depends on noise filtering as a necessary preprocessing step to identify nonlinear match-fixing patterns in sports competitions. The Hampel Filter and Z-Score Normalization represent the two techniques used for this purpose.

Hampel Filter: Hampel Filter serves to detect and correct outliers found in time-series data including tracking data errors and unexplained betting odds fluctuations that might hide or simulate match-fixing indicators. For each data point x_i and the algorithm uses a window of size (w) to determine local median (m) and median absolute deviation (MAD). If the point deviates significantly (i.e., $| x_i - m |> k \cdot MAD|$, where $k \approx 3$), it is replaced with (m). The filtering method protects actual behavioral sequences like standard athlete actions alongside wager behavior while getting rid of random patterns which may hide signs of non-linear match fixing. It can be mathematically deliberated using Eq. (1),

$$MAD = Median(| x_i - m |) \tag{1}$$

if $| x_i - m |> k \cdot MAD$, then $x_i \leftarrow m$.

Z-Score Normalization: The Z-Score normalization technique standardizes features that possess varying measurement scales such as meters-per-second and currency units because it enables better ML model performance through feature comparison. Each data point x_i is normalized through subtraction of the mean μ and division by the standard deviation σ. The transformation produces a distribution with both mean 0 and standard deviation 1.

The normalization process enables all features including player effort to have equal influence on detecting match-fixing patterns. Moreover, it can be mathematically deliberated using the following Eq. (2),

$$z_i = \frac{x_i - \mu}{\sigma} \tag{2}$$

Where, $\mu = \frac{1}{n}\sum_{n}^{i=1} x_i$ and $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$.

2.2.2. *Temporal synchronization*

Temporal synchronization enables the proper alignment of player tracking data with itself or other datasets to identify non-linear patterns of match-fixing in sports competitions. High-frequency player tracking data obtained through GPS records second-by-second positions and movement metrics but their timestamps tend to be inconsistent because sensors sample data at different rates. This research uses Dynamic Time Warping (DTW) technique for temporal synchronization.

DTW: DTW performs time-series segment alignment on high-frequency GPS coordinates that track player movements including positioning data and speed measurements and acceleration data recorded at one-second intervals despite inconsistent sampling rates or data collection periods. DTW determines the best possible sequence alignment between two time-series patterns (such as player movement data across game segments) through minimizing the total Euclidean distance while enabling non-linear time deformation. The correct detection of suspicious movements depends on metric alignment during assessment because it reveals intentional player under actions that occur throughout the game sequence. It can be mathematically expressed using Eq. (3),

$$DTW(X,Y) = \min\left(\sum_{(i.j)\in alignment} d(x_i, y_j)\right) \quad (3)$$

$$d(x_i, y_j) = (x_i - y_j)^2$$

Here, $X = \{x_1, x_2 \ldots . x_n\}$ and $Y = \{y_1, y_2 \ldots . y_m\}$ were considered as the time-series segments in the given data.

2.2.3. *Missing value imputation*

The random forests approach fills missing values in player tracking data to resolve sensor errors and data transmission problems which result in missing GPS coordinates and speed measurements.

Miss Forest: The algorithm of Miss Forest operates by using each feature with missing data points (such as player distance covered) as prediction targets to train random forest models from observed data for estimating missing values. The method operates on numerical data found in player tracking datasets to deliver reliable imputation solutions for time-series data with many dimensions. The algorithm continues its operations until it reaches convergence point to generate a complete dataset which enables ML models to identify delicate match-fixing indicators through player positioning and effort analysis. It can be mathematically deliberated using Eq. (4),

$$x_j^{missing} = f(X_{observed}, \theta) \quad (4)$$

Where, f is the predictor of random forest and the model parameters are mentioned as θ.

2.3. Feature extraction

The process of feature extraction takes unprocessed player tracking data to generate patterns which indicate possible match-fixing activity. The feature-extraction methods of Player-Betting Influence Vectors and Match Momentum Deviation Index transform tracking data into analytic vectors which aid ML to detect irregularities from match-fixing activities.

2.3.1. *PBIV*

The PBIV measure the impact of single-player movements on game results which might indicate betting-related irregularities through unexpected performance changes. The aggregation of player tracking metrics (critical plays and spatial movement) is achieved through PBIV by combining data points such as position and speed and acceleration information. Weighted metrics used in the analysis indicate their match-shaping potential to create a behavioral vector that detects deviations such as lack of performance in crucial match moments which could signal match-fixing events. The system uses PBIV to detect complex player-action relationships in sports games because this approach enables ML analysis. Thus, the mathematical representation of PBIV can be deliberated in Eq. (5),

$$PBIV_i = [w_1 . a_{i1}, w_2 . a_{i2} \ldots w_k . a_{ik}] \quad (5)$$

$PBIV_i$ is the influence vector containing information about player (i), a_{ik} defines the tracking data provides Player (i)'s metric (k) including speed in attacking zone and number of critical plays. $w_k$defines the weight for metric (k) represents the match outcome impact according to domain expertise or statistical analysis. The number of metrics under evaluation includes speed and positioning and actions as examples.

2.3.2. *MMDI*

Through MMDI the Match Momentum Deviation Index calculates shifts in team or player momentum levels that uses player tracking data to detect match-fixing activities. Momentum evaluation depends on tracking various aspects such as possession control shifts together with player sprint activity and positions they occupy. MMDI compares game mechanics at critical points to historical standards to detect unusual shifts that include sudden performance drops which can indicate suspicious match-tampering activities. ML models utilize this index to identify anomalous patterns which might signal match-fixing because it permits the detection of nonlinear changes in time. It can be mathematically expresses using Eq. (6),

$$MMDI_t = \frac{1}{N}\sum_{i=1}^{N} \left| m_{i,t} - \widehat{m}_{i,t} \right| / \widehat{m}_{i,t} \quad (6)$$

$MMDI_t$ is the momentum deviation index at time interval (t), $m_{i,t}$ defines the observed momentum metric for player (i) at time (t). Moreover, the following metrics $\widehat{m}_{i,t}$ defines the expected momentum metric for player (i) at time (t) relies on historical or contextual norms for calculation. The number of players and metrics included in the analysis period equals (N). The normalization by $\widehat{m}_{i,t}$ process adjusts for different ranges of momentum measurement values.

2.4. Nonlinear feature modeling using material science analogies

According to nonlinear feature modeling which draws on material science approaches data from tracked players functions as a dynamic material system for detecting match-fixing patterns. The model represents player movements through particles under stress while intentional underperformance acts as a material defect in this system. The analysis tool called Strain-energy mapping detects player effort deviations that match stress fractures and the entropy-based clustering detects behaviors which match phase transitions. Through ML models these features extracted from acceleration and spatial control metric data reveal complex behavioral patterns that help spot coordinated match-fixing activities during matches.

2.3.3. *Behaviour modelling*

Behavior modeling analyses non-linear irregularities within player tracking information through material science methods while treating teams and their players like physical systems which experience increased stress. Bouc-Wen Hysteresis Loops and Damage Mechanics Modeling (DMM) serves as methods to analyze match anomalies and stability which allows ML models to detect minimally detectable match-fixing features during matches.

Bouc-Wen Hysteresis Loops: Through its Bouc-Wen hysteresis model match anomaly events become interpretable through the representation of tracking data as hysteresis behavior matching material deformation in cyclic stress testing. Patterns of performance changes such as intentional underperformance trigger nonlinear responses which persist across time and affect match fluctuations. Through this quantitative model researchers can analyze the relationship between two factors: game contextual elements represent the input and player work represents the output. By recognizing inconsistent trends that dissimilar from typical behavioral patterns analysts can detect signs of match-fixing. The method proves successful for detecting complex time-based irregularities which ML models can analyze effectively. It can be mathematically expressed using Eq. (7),

$$z_t = a\acute{x}_t + (1-\alpha)\left(A\acute{x}_t - \beta|\acute{x}_t||z_t|^{n-1}z_t - \gamma x_t|\acute{z}_t|^n\right) \tag{7}$$

Here, z_t is the hysteretic output at time (t) shows irregular player conduct (such as altered effort levels). $\acute{x}_t$ defines the input rate of change includes modifications in game intensity together with player movement patterns. α represents the linear component weight $(0 \leq \alpha \leq 1)$. The parameters A, β and γ are the hysteresis output depends on three parameters that determine its shape and amplitude. The parameter (n) is the exponent controlling nonlinearity. The model outputs z_t as a feature for match-fixing detection.

DMM: Using DMM recreates player or team instability through an approach that applies damage mechanics principles to tracking data to model material degradation until behavior represents match-fixing weaknesses. The performance degradation process (e.g. reduced critical moment effort) gets quantified through DMM as material-like cumulative damage. The damage evolution model based on player activities allows DMM to detect match-fixing signals through ML models by identifying nonlinear patterns of instability that include coordinated underperformance. It can be mathematically expressed using the following Eq. (8),

$$\begin{gathered} D_t = D_{t-1} + \Delta D_t \\ \Delta D_t = k\left(\frac{\sigma_t}{\sigma_c}\right)^m (1 - D_{t-1}) \end{gathered} \tag{8}$$

Here, the damage index D_t operates between zero and one during time (t) and ΔD_t is the damage accumulation follows from player actions. σ_t defines the stress-like metric (e.g., deviation of player effort from expected norms) and σ_c is the critical threshold for normal behavior. The pair (k, m) functions as parameters which determine the speed at which damage occurs and how it affects material properties non-linearly as well as D_t is the detection of instability patterns depends on this particular feature. The modeling techniques of Bouc-Wen Hysteresis Loops together with DMM use material science analogies to measure the nonlinear variations present in player tracking data. Match anomalies persist as the Bouc-Wen model functions alongside DMM for measuring performance loss which enhances ML capabilities to detect match-fixing schemes for protecting sports integrity.

2.3.4. *Chaos-Based Feature Extraction*

Chaos-based feature extraction tracks match-fixing patterns through player tracking data analysis because it determines irregularities in dynamic player behavior by analyzing their chaotic nature. The three techniques Largest Lyapunov Exponent (LLE) and Approximate Entropy (ApEn) and Recurrence Plot Analysis (RPA) extract sensitive anomaly features with complexity characteristics and anomaly recurrence patterns to help ML algorithms detect elusive match-fixing patterns.

LLE: The LLE quantifies how fast neighbouring trajectories separate in player movement data through speed and positioning measurements to indicate chaotic sensitivity of player movements. Match-fixing signals become more likely when the LLE value increases because unstable behavior produces erratic or intentional deviations from expected patterns. The LLE generates nonlinear dynamics feature for ML analysis through its calculation of exponential trajectory separation in reconstructed state-space. Thus, the Eq. (9) shows the chaos-Based Feature Extraction using LLE.

$$\lambda_1 = \lim_{t\to\infty t}^{1} - ln\left(\frac{||\delta x(t)||}{||\delta x(0)||}\right) \tag{9}$$

Where, λ_1 is the largest Lyapunov Exponent and $\delta x(t)$ is the tracking data state-space contains the perturbation vector $\delta x(t)$ that operates at time (t). $||\delta x(0)||$ is the initial perturbation magnitude as well as the positive λ_1 defines the presence of chaotic behavior indicates through this measurement which functions as a feature.

ApEn: It measures the matching frequency of similar data aspects in tracking information like player acceleration patterns to assess data complex variability in essentially two ways.

ApEn values show predictable behavior when low but higher values signify irregular actions which could indicate match-fixing through sudden effort changes. The computation of ApEn on short sequences enables its use for detecting nonlinear anomalies in dynamic match contexts for ML models. Moreover, it can be mathematically expressed using Eq. (10),

$$ApEn(m, r, N) = \emptyset^m(r) - \emptyset^{m+1}(r)$$

$$\emptyset^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} lnC_i^m(r) \tag{10}$$

Where, m is the embedding dimension represents the length of sequences used for comparison, r is the similarity threshold (e.g., fraction of standard deviation). Moreover, the length of the time-series (e.g., acceleration data) can be denoted as N as well as $C_i^m(r)$ is the fraction of all sequences exist within (r) of the (i)-th sequence.

The entropy value functions as a feature in the analysis is denoted as $ApEn$.

RPA: It is for player tracking data summarises evidence about throw-it back and computes corresponding statistics as modelled through building the example matrix for indicating when the pattern states, such as combinations of position and speed, have revisited similar states.

Anomalous patterns such as irregular movement sequences during key match moments appear as distinct structures in the recurrence plots, implying potential match-fixing.

RPA-derived metrics, like recurrence rate, are extracted as features for ML to detect nonlinear temporal relationships. Moreover, it can be mathematically deliberated using Eq. (10),

$$R(i, j) = \Theta\left(\epsilon - \left\|x_i - x_j\right\|\right)$$

$$RR = \frac{1}{N^2}\sum_{i,j=1}^{N} R(i, j) \tag{10}$$

Where, $R(i, j)$ is the recurrence matrix element (1 if states x_i and x_j are within distance ϵ, else 0), Θ is the Heaviside function. x_i defines the state vector at time (i) (e.g., position, speed). Thus, the distance threshold can be mentioned as ϵ and RR is the recurrence rate, used as feature. These chaos-based feature extraction mechanisms should take LLE, ApEn, and RPA in unison as obtains the nonlinear features from player tracking data wherein LLE accounts for the measure of chaotic sensitivity while ApEn brings about the overall irregularity counting with RPA capturing the repetitions or similarities. Thus, the combination improves the ability of ML to detect foul play in the integrity of sports through possible cheating activities.

2.3.5. *Fractal behavior mapping*

Fractal behavior mapping analyses the self-similarity and complex structure of player tracking data to unearth the nonlinear patterns of match-fixing. Higuchi's Fractal Dimension Estimation measures fractal complexity of player movement, which helps ML models to detect irregular behaviors suggestive of match-fixing.

Higuchi's Fractal Dimension Estimation: It evaluates the fractal dimension of time-series data from player tracking, e.g., speed, acceleration, or position, to quantify the complexity and self-similarity of player movements. More irregular or fragmented behavior, indicated by a higher fractal dimension, could speculate match-fixing actions, such as random walking or intentional deviation from expected pathways. Nonlinear dynamics are captured by analyzing the length of the time-series curve at different scales, providing ML with a robust feature to detect anomalous player behaviors in dynamic contexts of match-fixing. Subsequently, the mathematical representation of Higuchi's Fractal Dimension Estimation can be mathematically expressed using Eqs. (11), (12):

$$L(k) =$$

$$\frac{1}{k}\sum_{m=1}^{k}\left(\sum_{i=1}^{\left\lfloor\frac{N-m}{k}\right\rfloor} |x(m+ik) - x(m+(i-1)k|\right) \tag{11}$$

$$\frac{N-1}{\left\lfloor\frac{N-m}{k}\right\rfloor^{k}}$$

10.57647/ijeee.2025.1602.08

$$D = -\lim_{k\to\infty} \frac{logL(k)}{logk} \tag{12}$$

Here, $L(k)$ is the average length of the curve for scale k and $(x(t)$ is the time series of a player tracking dataset, e.g., speed at time t.

Length of this time series be N and the time interval (scale factor) can be mentioned as k. Thus, the beginning index can be denoted as m and D is the fractal dimension as used as a feature (typically $1 < D < 2$) for time series.

Higuchi Fractal Dimension Estimation captures the fractal complexity of player-tracking data, reflecting erratic movement patterns that can indicate match-fixing. The output fractal dimension feature aids ML in detecting nonlinear anomalies in the protection of sports integrity.

2.4. ML-based pattern recognition

ML based pattern recognition- Random Markov Temporal Analysis (RMTA), Weighted Isolation Support Classifier (WISC), and Dynamic Graph Neighbour Network (DGNN)-have shown great potential to detect nonlinear match-fixing patterns in player tracking data and maintain the integrity of sports. RMTA integrates Random Forest with Sliding Windows (RF-SW) and Hidden Markov Model (HMM) to observe temporal anomalies in sequential player metrics such as speed and positioning to detect aberrations exemplified by, for example, a sudden drop of effort. WISC connects Isolation Forest (IF) and SVM with Feature Weighting (SVM-FW) to determine the contextual view by isolating irregular behaviors such as unusual positioning in the critical moment of the game while giving utmost importance to specific features for strong classification. DGNN employs k-Nearest Neighbors with Dynamic Clustering (K-NN-DC) and Graphical Lasso (GL) to create a model for relational dynamics and detect coordinated anomalies, which can also be called synchronized underperformance, through dynamic interaction graphs of players.

Together these techniques handle the complex structure of match-fixing by using the temporal, contextual, and relational attributes from player tracking data.

Temporal anomaly detection is facilitated by RMTA, contextual specificity is followed up by WISC, and combinatorial detection patterns are uncovered by DGNN. The secret to this partnership is to allow for a complete set of analyses involving the sleuthing of subtle, nonlinear match-fixing behaviors with the courage to maintain teams and the unilateral integrity of sports competitions.

2.4.1. *Temporal learning*

RMTA integrates Random Forest with Sliding Windows (RF-SW) and HMM to capture temporal anomalies within player tracking data, thus enabling robust detection of nonlinear patterns of match-fixing. Part of RMTA detects deviations in player behavior that may signal the fixes-for example, steep drops in exertion or unmistakably irregular movement patterns-using an ensemble-based anomaly detection method implemented in time windows RF-SW applied to Hidden Markov Models for sequential state modeling.

RMTA takes in player tracking data (e.g. speed, positioning, acceleration, etc.) and applies RF-SW to classify anomalies in sliding time windows, capable of capturing short-term deviations in metrics such as sprint frequency.

Simultaneously, HMM captures long-term sequential dependencies by modeling hidden states-normal vs. suspicious play-which in turn generate the observed metrics.

The hybrid really exploits the non-linear handling of features by RF-SW with the temporal transitions of HMM, providing a joint probability score for anomaly detection that signifies the likelihood of match-fixing behaviors throughout the timelines of matches. This augments the ability of ML models to detect non-simple, time-dependent patterns within dynamic sports contexts and thus further safeguards the integrity of competition. Thus, the hybrid mathematical expression for the following RMTA model can be expressed using Eq. (13),

$$\begin{gathered} P_{RMTA}(y_t|x_t - w; t, o) = \alpha . RF(x_t - w; t) \\ +(1-\alpha). P(o_t|s_t, \lambda). P(s_t|s_{t-1}, \lambda) \end{gathered} \tag{13}$$

Here, $P_{RMTA}(y_t|x_t - w; t, o)$ is the anomaly probability with respect to time (t). $x_t - w; t$ is the feature vectors within a window of size (w) (for example, speed, position).

Moreover, $RF(x_t - w; t)$ is the random forest probability output for anomaly (for example, 0-1). $O = \{o_1, o_2, \dots, o_t\}$ and it is the observed sequence of player tracking metrics includes speed and acceleration measurements and $P(o_t|s_t, \lambda)$ is the HMM emission probability for observation o_t given state s_t. $P(s_t|s_{t-1}, \lambda)$ defines the HMM state transition probability and α is the weight balancing RF-SW and HMM contributions (0,1). λ representing the HMM parameters (transition and emission probabilities). Moreover, P_{RMTA} would be feature or decision score on match-fixing detection.

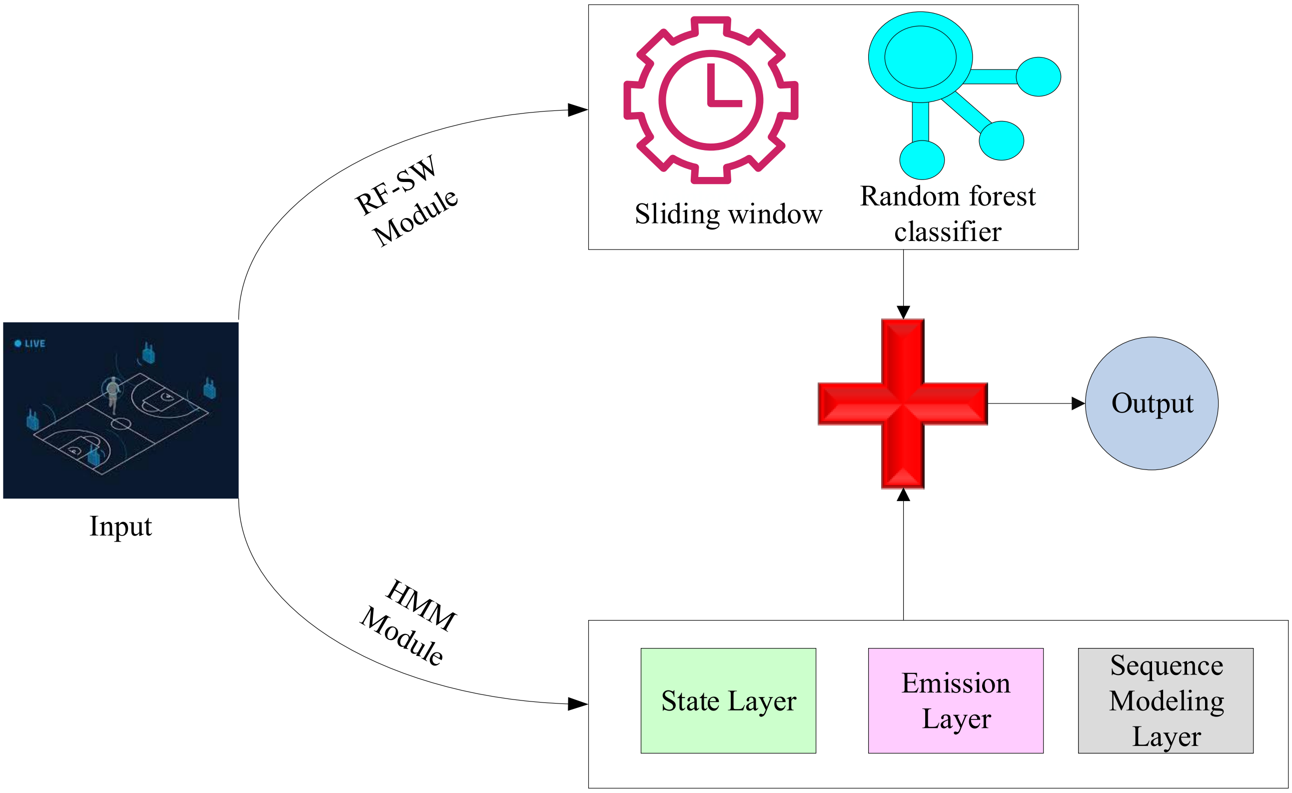


Figure 2. RMTA architecture

The RMTA system design for detecting nonlinear match-fixing patterns in player tracking data appears in Figure 2 which belongs to identification of Nonlinear Match-Fixing Patterns. RMTA functions with two modules that include an RF-SW Module with Sliding Windows to segment data while Random Forest Classifiers detect short-term anomalies and an HMM Module performing State and Emission Modeling Layers with Sequence Modeling Layers to track long-term temporal dependencies. The system processes tracking data from players through its modules to generate a final anomaly score which detects sudden effort drops as examples of temporal match-fixing patterns. RMTA's hybrid approach combines the robustness of window-based anomaly detection in RF-SW with sequential modeling from HMM to capture nonlinear behavior such as short-term and long-term trends from player tracking data, thus improving the ML model precision in detecting match-fixing behavior for integrity in sports.

2.4.2. *Contextual analysis*

A division of Isolation Forest (IF) and Support Vector Machine with Feature Weighting (SVM-FW) has made it possible for the application of the Weighted Isolation Support Classifier (WISC) in the environment of contextual anomaly detection with player-tracking data as input. Thus, the result brought precision in the identification of nonlinear match-fixing patterns. WISC therefore combines the efficient outlier isolation capability from IF and the weighted feature classification ability from SVM-FW, thereby localizing a phase of unusual player behaviors such as unusual positioning or effort at critical times in the game and providing sports integrity. WISC takes player tracking data (for example, sprint frequency, spatial control) and starts with IF computing anomaly scores based on how quickly an individual player action gets isolated in a number of random trees. This enables one to flag outliers, such as cases of unexpected lack of effort, while at the same time classifying these actions as normal or suspicious through weight assignments to some contextually salient features such as key zone movement intensity. This hybrid model thus produces a joint decision score representing the likelihood of match-fixing by combining the anomaly score of IF with the classification probability of SVM-FW. The model exploits the scalability of IF in high dimensions and the robustness of SVM-FW in non-linear classification to ensure that contextual anomalies are accurately censured for ML analysis. Moreover, WISC can be mathematically expressed using Eq. (14),

$$S_{WISC}(x) = \beta . s(x,n) + (1-\beta) . \sum_{i=1}^{N} \alpha_i y_i K(w \odot x_i, w \odot x) \tag{14}$$

$S_{WISC}(x)$ is the joint anomaly score for data point x (higher values indicate anomalies).

$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$ if the anomaly score, with $E(h(x))$ as the average path length across trees and $c(n)$ as the normalization factor. $\sum_{i=1}^{N} \alpha_i y_i K(w \odot x_i, w \odot x)$ representing the SVM-FW decision function, where α_i defines the Lagrange multipliers, the class labels were denoted as y_i and the kernel function is denoted as (K) (e.g., RBF), and w is the feature weight vector.

Element-wise feature weighting to emphasize critical metrics are denoted as $w \odot x$ and β is the weight $(0 \leq \beta \leq 1)$ and it is responsible for balancing IF and SVM-FW contributions.

Moreover, the score S_{WISC} is considered as a threshold for classifying actions as normal or suspicious. The WISC system architecture depicted in Figure 3 unites an IF module with an SVM-FW module. The IF module executes random partitioning operations and calculates path lengths while producing anomaly scores. The SVM-FW module operates simultaneously to perform feature weighting and kernel mapping and classification tasks. The integration block merges outputs from both modules to unite anomaly detection capabilities with classification understanding. The combined output proceeds to a feedback loop which optimizes model performance. The system produces the final output which establishes WISC as an effective hybrid framework for precise anomaly detection and classification interpretation. The hybrid approach of WISC exploits the very fast anomaly isolation from IF and weighted classification from SVM-FW, thus capturing the contextual irregularities in player tracking data. This makes the ML detection of nonlinear match-fixing patterns finer, thereby preserving the integrity of sports.

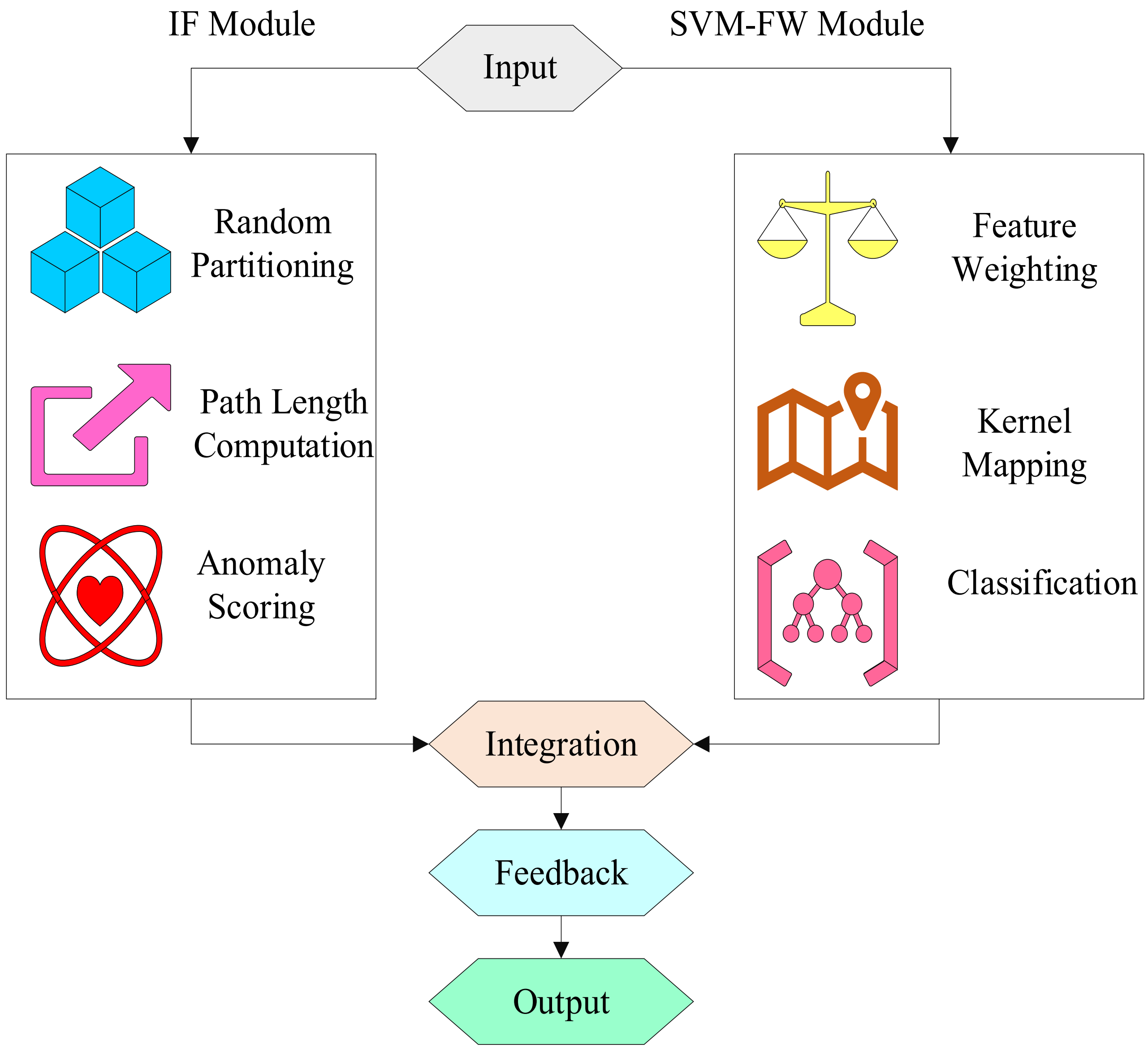


Figure 3. Architecture of WISC

2.4.3. *Relational modeling*

Dynamic Graph Neighbor Network (DGNN), which fuses localized clustering of k-Nearest Neighbors with the graphical lasso paradigm, describes player interactions in tracking data to help detect nonlinear methods of match-fixing schemes.

DGNN links the proximity-based clustering of KNN-DC with the sparse inference of GL to detect coordinated anomalies, like synchronized underperformance against integrity of sport. DGNN evaluates player tracking data (for instance, position, pass frequencies) using kNN-DC methods, thereby dynamically clustering players according to spatial and behavioral proximity while capturing time-varying interactions, including coordinated movements that could indicate collusion. At the same time, GL constructs a sparse graph by estimating a precision matrix with respect to player interactions, hence accentuating relevant correlations such as those of abnormal spatial overlaps while suppressing noise. This hybrid scheme feeds cluster-related anomaly scores calculated by KNN-DC into GL's interaction weights, for a simple computation of the joint-scoring of relational patterns. The merit of this joint work lies with KNN-DC's ability to adapt to dynamic matches while GL accounts for modeling multi-relational dependencies, thus allowing the detection of nonlinear match-fixing behaviors in ML settings. Moreover, the mathematical expression for DGNN is expressed in Eq. (15),

$$S_{DGNN}(x_t) = \gamma.\sum_{i\in c_t}\sum_{j\in N_k(i)}\left\|x_i - x_j\right\|_2^2 \qquad (15)$$
$$+(1-\gamma).tr(S\widehat{\Theta})$$

Where, $S_{DGNN}(x_t)$ defines the joint anomaly score at time (t) (the higher, the more relational anomalies). c_t representing the adaptive cluster assignments from kNN-DC minimizing $\sum_{i\in c_t}\sum_{j\in N_k(i)}\left\|x_i - x_j\right\|_2^2$, here the feature vectors were denoted as x_i and $N_k(i)$ is the k-nearest neighbors.

The following parameter $\widehat{\Theta}$ defines the GL precision matrix, estimated via $\widehat{\Theta} = argmin_{\Theta}(-logdet(\Theta) + \mathrm{tr}(S\Theta) + \lambda\|\Theta\|_1$with S being the empirical covariance matrix. $tr(S\widehat{\Theta})$ is the trace measuring strong interactions as well as γ ais the weight $(0 \leq \gamma \leq 1)$ balancing contributions from kNN-DC and GL. DGNN integrates KNN-DC and GL modules to process data from dynamic interaction systems as shown in Figure 4. The KNN-DC module starts with entity proximity calculations before it reflects behavioral anomalies through cluster-based anomaly scoring after high-dimensional dynamics clustering completes.

During this process the GL module performs three concurrent functions including covariance estimation followed by sparse graph inference and interaction weighting to enhance the understanding of entity connectivity. The integration mechanism combines clustering and relational learning insights which originate from both modules.

The system generates a relationally modelled representation that serves as a foundation for additional reasoning or predictive tasks.

The architectural design of DGNN allows it to simultaneously acquire local interaction patterns and global structural relations which makes it suitable for tasks that need dynamic anomaly detection and context-aware relationship modeling. Thus, the score S_{DGNN} is used to flag suspicious interactions.

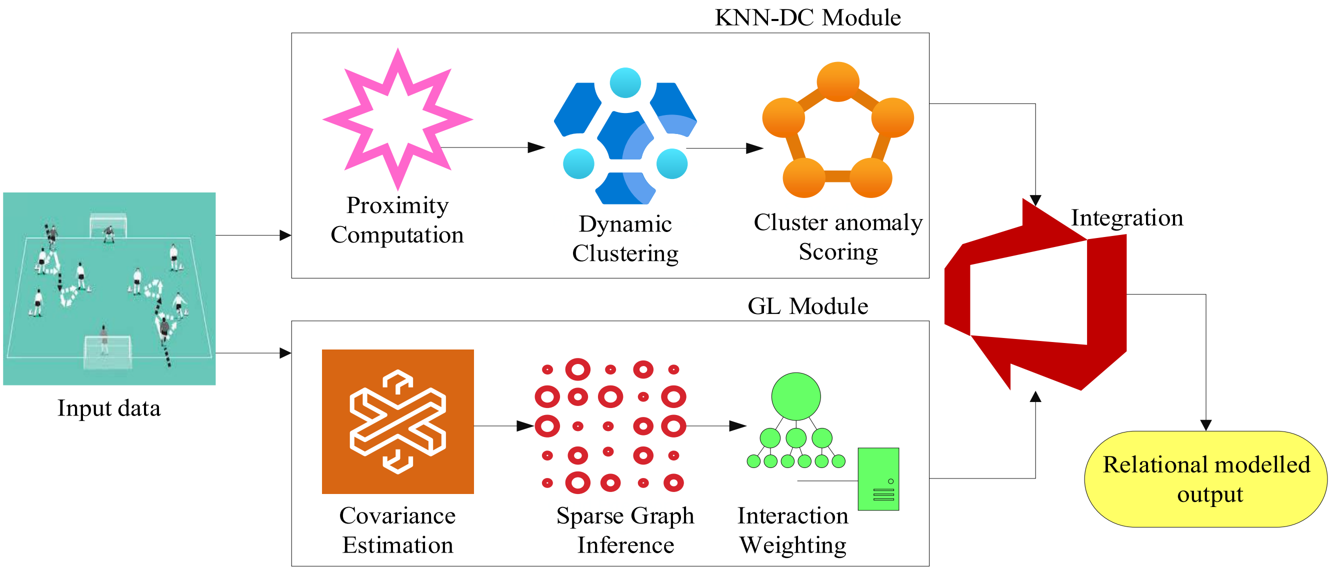


Figure 4. Architecture of DGNN

10.57647/ijeee.2025.1602.08

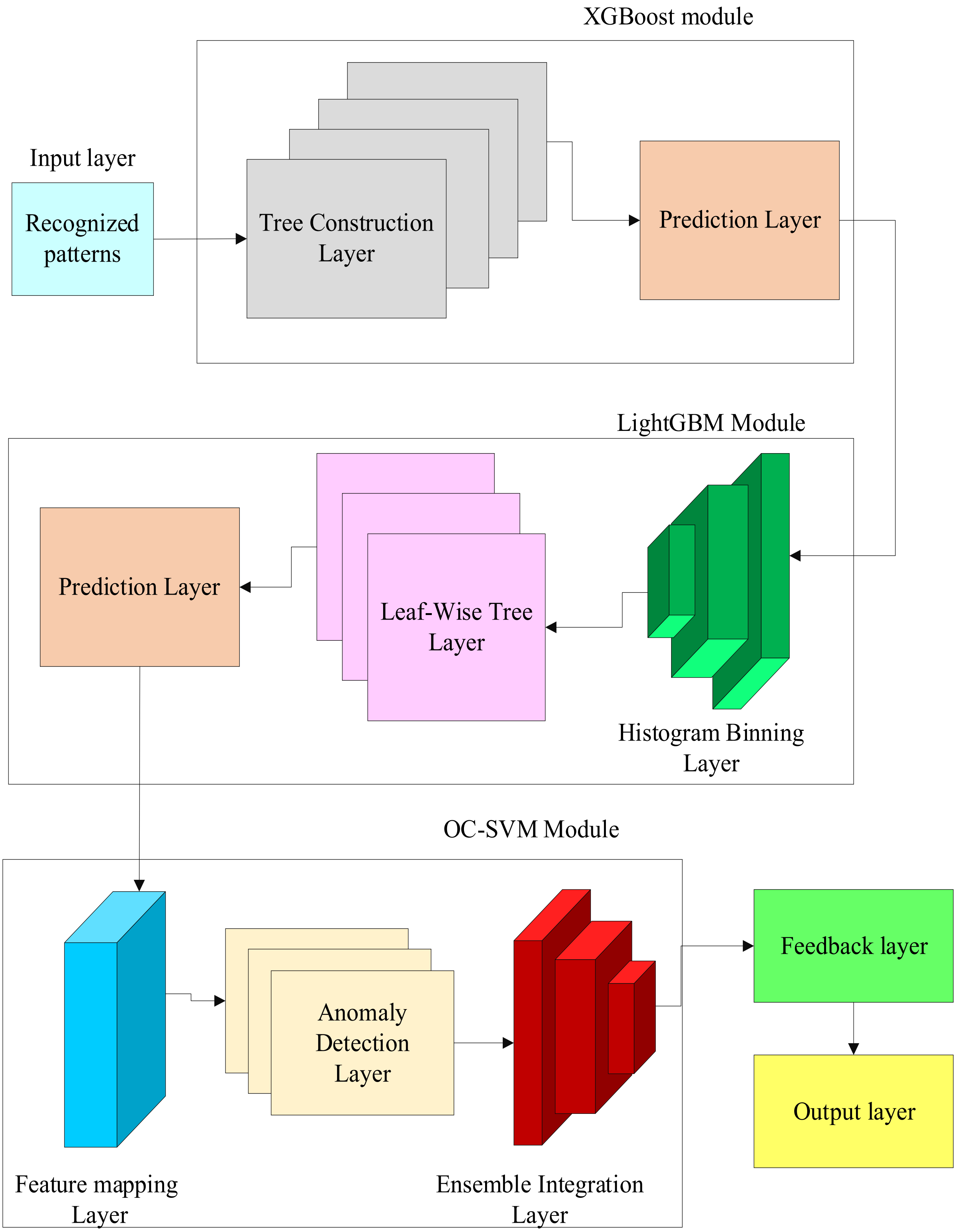


Figure 5. Architecture of GBOCLE model

Rather, DGNN's hybrid method combines dynamic clustering of KNN-DC with sparse graph modeling of GL, thereby effectively capturing relational anomalies in player tracking data, which further helps ML in detecting coordinated patterns of match-fixing and thereby safeguarding sports integrity.

2.5. Detection model

Gradient Boosted One-Class Light Ensemble (GBOCLE) uses XGBoost and LightGBM together with one-class SVM as an ensemble framework to detect complex match-fixing patterns in tracking data which upholds sports integrity. XGBoost and LightGBM together with OC-SVM form GBOCLE which detects coordinated match-fixing behaviors in subtle patterns through anomaly detection for datasets with imbalanced data. The GBOCLE system analyzes player tracking metrics including speed, positioning, acceleration data through supervised XGBoost and LightGBM models that use gradient-boosted decision trees to conduct training on labeled data for achieving high-precision match-fixing probability predictions. XGBoost achieves precise results through its regularized objective while LightGBM improves performance by using histogram-based learning for databases at scale. OC-SVM functions independently as a detection method by establishing a one-class boundary to define normal players using high-dimensional features which highlights matched behaviors distinct from the norm. The ensemble performs a weighted scoring operation to combine XGBoost supervised predictions and LightGBM supervised predictions and OC-SVM anomaly scores to provide complete detection of nonlinear patterns.

Moreover, the hybrid equation for ML based nonlinear match fixing is deliberated in Eq. (16),

$$S_{GLOBE}(x) = \alpha . XGB(x) + \beta . LGB(x) \qquad (16)$$

$$+(1 - \alpha - \beta). OC - SVM(x)$$

$S_{GLOBE}(x)$ is the final ensemble score for data point (x), $XGB(x) = \sum_{k=1}^{K} f_k(x)$ is the XGBoost prediction reaches its optimized state through the objective function $Obj = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$. Moreover, $LGB(x) = \sum_{m=1}^{M} g_m(x)$ and it defines the lightGBM prediction relies on histogram-based splitting methods. Moreover, $OC - SVM(x) = sign\left(\sum_{i=1}^{N} \alpha_i K(x_i, x) - \rho\right)$ is the OC-SVM decision function, with (K) as an RBF kernel. Subsequently, the parameters α and β defining weight $(0 \leq \alpha, \beta \leq 1, \alpha + \beta \leq 1)$.

The non-linear patterns of match-fixing that GBOCLE detects efficiently result from its ability to analyze complex associations discovered within tracking data from players. The nonlinear features of XGBoost's tree-based structure allows it to discover anomalous patterns such as players slowing down unexpectedly in key areas which indicates deliberate matches with suspicious underperforming players. The system finds irregular acceleration patterns more efficiently through LightGBM which achieves detection of coordinated behaviors that linear models typically miss. The anomaly detection capabilities of OC-SVM operate under unsupervised settings to identify outliers through an RBF kernel which detects unusual positioning in high-dimensional spaces for detecting rare match-fixing events.

The hybrid score unifies the supervised and unsupervised techniques by using them to identify synchronized performance weaknesses or spatially abnormal dynamics. Through feature extraction methods involving PBIV, MMDI, LLE, ApEn the quantification ability of GBOCLE improves to detect both complex and non-linear deviations in momentum patterns leading to superior sports integrity by minimizing false positives. Subsequently, Figure 5 shows architecture of GBOCLE model.

3. Result and discussion

Under this section, the overall performance carried out by the suggested model is discussed in detail. The performance of the proposed model can be validated by comparison the performance of the proposed model with the existing models. Moreover, the performance metrics such as accuracy, precision, F-measure, sensitivity, specificity, NPV, MCC, FNR and FPR were measured for calculate the effectiveness of the suggested approach. Furthermore, the comparison analysis shows that the proposed model attains better performance score than the existing models.

3.1. Experimental setup

Table 2 presents the experimental setup of the proposed model. It aims toward the detection of nonlinear match-fixing patterns from player tracking data. The dataset has been divided into training (70%), validation (15%), and testing (15%). Training is done on NVIDIA RTX 3090 Graphics Processing Unit (GPU) of 24GB VRAM while inference is simulated on Raspberry Pi 4. This implementation uses Python 3.9, TensorFlow, PyTorch, and Scikit-learn. The preprocessing includes noise filtering, temporal synchronization (DTW), and missing value imputation (Miss Forest). Features are extracted using PBIV, MMDI, chaos-based methods (LLE, ApEn, RPA) and fractal analysis (Higuchi). The GBOCLE model, using RMTA, WISC, and DGNN, is trained for 50 epochs using the Adam optimizer.

3.2. Metrics evaluation

The performance metrics such as accuracy, precision, sensitivity, specificity, F-measure, NPV, MCC, FPR and FNR were measured for measuring the performance of the suggested model. Table 3 shows the performance metrics along with description and formula.

Table 2. Experimental setup for the proposed model

Aspect	Description
Datasets	Player tracking data
Data Split	Train: 70%, Validation: 15%, Test: 15%
Hardware	Training: NVIDIA RTX 3090 GPU, 24GB VRAM, Intel i9 CPU, 64GB RAM. Edge Deployment: Raspberry Pi 4 (4GB RAM) for inference simulation.
Software & Frameworks	Python 3.9, TensorFlow 2.8, PyTorch 1.10, OpenCV (image preprocessing), Scikit-learn (MI computation), NumPy, Pandas.
Preprocessing	Noise Filtering: Hampel Filter, Z-Score Normalization; Temporal Synchronization: DTW; Missing Value Imputation: Miss Forest
Feature Extraction	Player-Betting Influence Vectors (PBIV), Match Momentum Deviation Index (MMDI)
Feature Modeling Using Material Science Analogies	Behavior Modeling: Bouc-Wen hysteresis loops, DMM Chaos-Based Feature Extraction: LLE, ApEn and RPA Fractal Behavior Mapping: Higuchi's Fractal Dimension Estimation
ML based pattern recognition	Temporal Learning: RMTA Contextual Analysis: WISC Relational Modeling: DGNN
Detection model	Gradient Boosted One-Class Light Ensemble (GBOCLE)
Training Configuration	Optimizer: Adam (learning rate 0.001), Loss: Binary Cross-Entropy, Epochs: 50, Batch Size: 32, Early Stopping: Patience of 5 on validation loss.

Table 3. Analysis of performance metrics

Metric	Description	Formula
Accuracy	Proportion of correctly identified match-fixing and normal behaviors. May be skewed by rare match-fixing events.	$A = \frac{tp + tn}{tp + fp + tn + fn}$
Precision	Fraction of flagged match-fixing instances that are correct. Minimizes false alarms to avoid wrongful accusations.	$P = \frac{tp}{tp + fp}$
F-measure	Balances precision and sensitivity for performance in imbalanced datasets with rare match-fixing cases.	$F - measure = 2 \times \frac{P \times S_n}{P + S_n}$
Sensitivity (Recall)	Proportion of actual match-fixing cases detected. Ensures most incidents are caught for sports integrity.	$S_n = \frac{tp}{tp + fn}$
Specificity	Fraction of normal behaviors correctly identified. Reduces false positives to preserve trust in fair play.	$S_p = \frac{tn}{tn + fp}$
NPV	Proportion of predicted normal behaviors that are truly normal. Ensures confidence in ruling out match-fixing.	$NPV = \frac{tn}{tn + fn}$
MCC	Measures correlation between predicted and actual classifications, robust for rare match-fixing events.	$MCC = \frac{(tp \times tn) - (fp \times fn)}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$
FPR	Proportion of normal behaviors incorrectly flagged as match-fixing. Low FPR avoids unjust accusations.	$FPR = \frac{fn}{tp + fn}$
FNR	Proportion of match-fixing cases missed. Low FNR ensures critical incidents are not overlooked.	$FNR = \frac{fp}{tn + fp}$

3.3. Comparison analysis

This section provides the comparison analysis among the proposed model and the existing ML based models. Thus, the existing models such as SVM [12], Random

Forest (RF) [17], Decision Tree (DT) [21] and K-NN [17] were considered for comparison analysis. The statistical performance comparison shown in Table 4 fully vindicates the proposed framework and its applicability via all previous benchmarks- SVM, RF, DT, and k-NN-in detecting an idiosyncratic, nonlinear match-fixing pattern from player tracking data as part of the investigation ML approaches to Identify Nonlinear Match-Fixing Patterns. The model bested all the models under consideration in all performance metrics-achieving an accuracy of 0.98382, precision of 0.9742, sensitivity of 0.97457, specificity of 0.96496, F1-score of 0.97531, NPV of 0.97568, and MCC of 0.96607, all on the lowest FPR (0.014927) and FNR (0.01092). This superior result originates from temporal anomaly detection of RMTA, contextual precision of WISC, and relational modeling of DGNN, supplemented by advanced features like PBIV, MMDI, chaos-based features (LLE, ApEn, RPA), Higuchi's Fractal Dimension, and many more. Moreover, SVM does achieve a respectable accuracy score of 0.95322, with a fair F1-score of 0.96847, but it is limited due to its linear kernel which cannot capture the nonlinear dynamics of match-fixing behaviors. In similar lines, a Random Forest-based approach gives a decent accuracy of 0.94531 and an F1-score of 0.93197, but it is not effectively modeled for long temporality with relational interaction, leading to high FPR (0.055701) and FNR (0.058316). Decision Tree (accuracy = 0.96786; F1 score = 0.93855) provides results at the competitive end of the benchmark spectrum but suffers from overfitting, resulting in higher FPRs (0.041976) in comparison with that of the model proposed. k-NN ultimately gets the brown-paper bag for ultimate lowest accuracy (0.93921) and an F1-score calculated at 0.95698, which again performs very poorly with high-dimensional tracking data; the high specificity (0.96751), however, would suggest a more conservative classification of normal behavior at the expense of misclassification of anomalies (FNR = 0.040173).

The proposed model has very high sensitivity and a low FNR that should ensure detection of all match incidents, which is very important for maintaining integrity in sports, while very high precision and low FPR should prevent false accusations.

The complementary MCC refers to such models as being reliable in imbalanced datasets where incidents of match-fixing are quite rare. By using chaos-based and fractal features with temporal-contextual-relational modeling, the model meets those strict conditions while performing better than baselines, thereby offering a scalable accurate solution to sports governance in maintaining fair play and enhancing trust in competitive integrity.

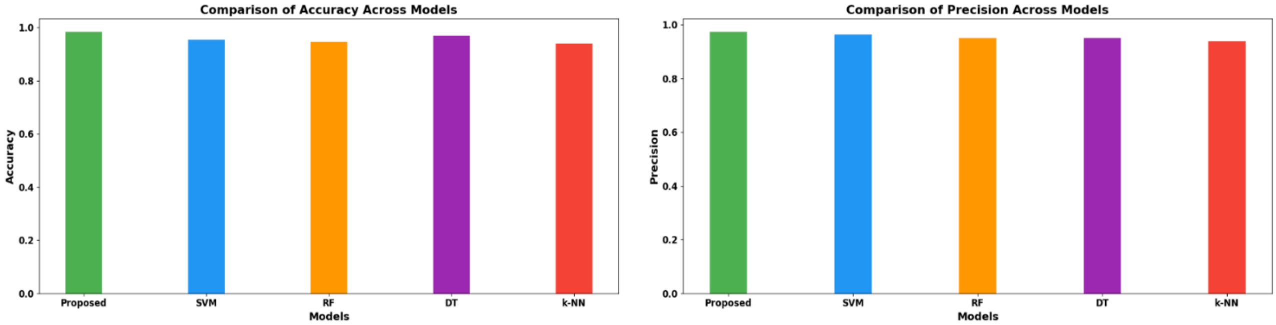
Moreover, the graphical representation of performance comparison is illustrated in Figure 6 (a)-(i).

Figure 7 shows the comparison of ROC for ML based models for identifying non-linear match fixing patterns. The proposed model using RMTA, WISC, and DGNN shows the highest Area Under Curve (AUC), that is 0.98382, thereby increasing the model's ability to differentiate match-fixing from normal behavior at different thresholds. The working of such a performance heavily rests on the architecture which is hybrid in nature and that employs RMTA with temporal anomaly detection, WISC contextual precision, and relational modeling from DGNN which effectively embodies the highly complex, nonlinear patterns of match-fixing. With this, the SVM scores an AUC of 0.95322 which still shows an efficient discriminative ability but fails because of the linear kernel problem in modeling the temporal and relational dynamics of match-fixing behaviors.

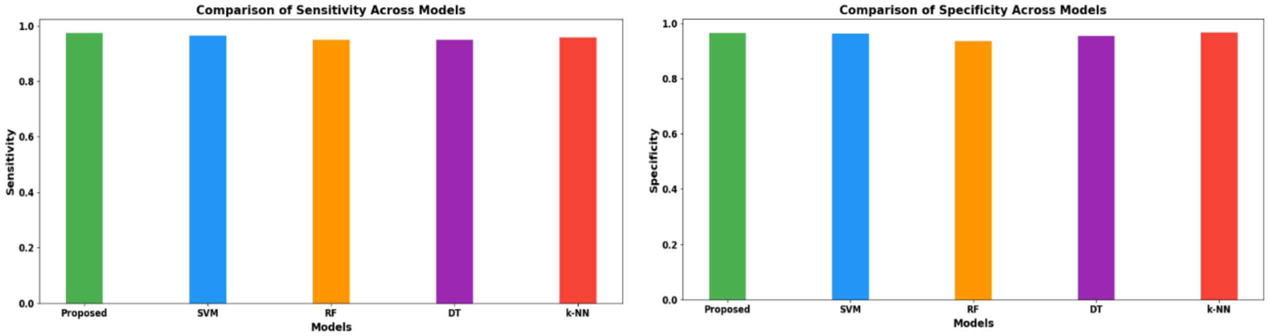
RF earned an AUC of 0.94531 signifying a good performance but misses the high-level temporal and relational analysis which the model reduces to RMTA and DGNN. DT is better at 0.96786, benefiting from its decision-based architecture but naturally overfit, which is its drawback in generalization compared to the proposed model. Coming in last was k-NN with an AUC of 0.93921, as it could not perform efficiently under high-dimensional player tracking data with dynamic match contexts due to failure of a proximity-based approach to capture the complex coordinated match-fixing patterns.

Table 4. Performance comparison of the proposed model with the existing models

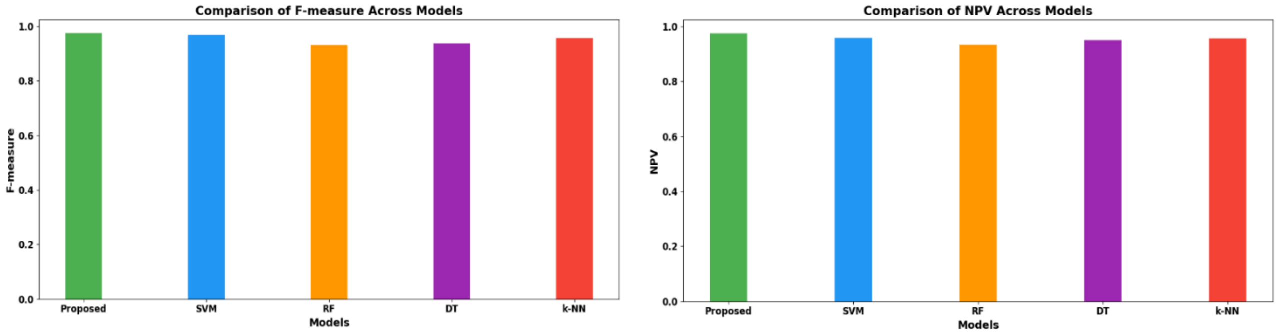
Model	Accuracy	Precision	Sensitivity	Specificity	F-measure	NPV	MCC	FPR	FNR
Proposed	0.98382	0.9742	0.97457	0.96496	0.97531	0.97568	0.96607	0.014927	0.01092
SVM	0.95322	0.96327	0.96334	0.96341	0.96847	0.95856	0.96861	0.048288	0.046985
RF	0.94531	0.95195	0.94862	0.9353	0.93197	0.93363	0.93528	0.055701	0.058316
DT	0.96786	0.95055	0.95021	0.95397	0.93855	0.95119	0.94386	0.041976	0.031758
k-NN	0.93921	0.93866	0.95807	0.96751	0.95698	0.95643	0.95588	0.04969	0.040173



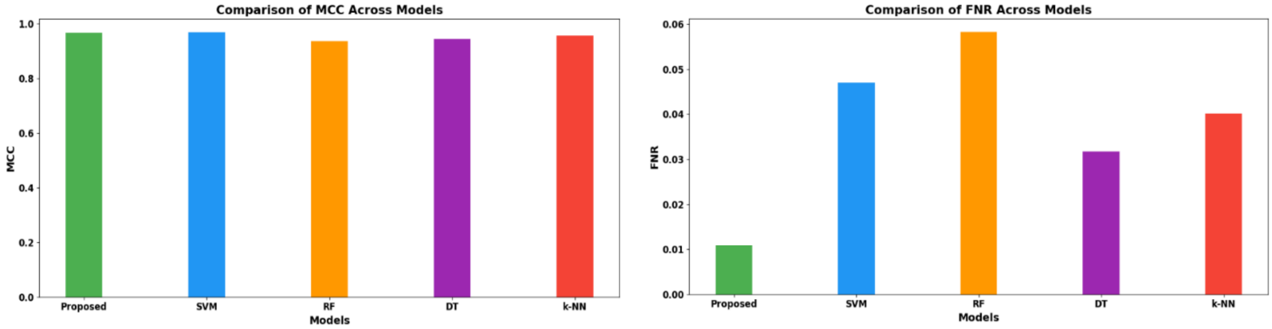
(a) Accuracy comparison (b) Precision comparison



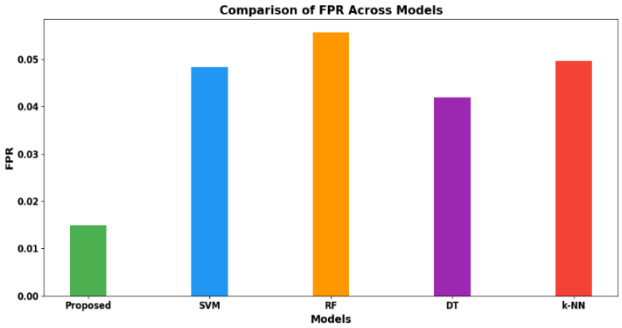
(c) Sensitivity comparison (d) Specificity comparison



(e) F-measure comparison (f) NPV comparison



(g) MCC comparison (h) FNR comparison



(i) FPR comparison

Figure 6. (a)-(i): Graphical representation of performance comparison analysis

10.57647/ijeee.2025.1602.08

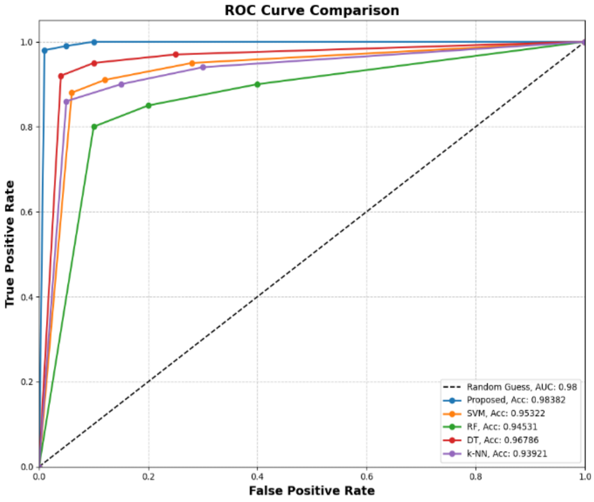


Figure 7. ROC curve comparison

ROC curves also illustrate the trade-offs in True Positive Rate (TPR) and FPR for every base model. The TPR continues with a very minimum FPR, the highest being put down within this model's curve compared to the other curves which mostly dash off to one corner and at some speeds, top-left corner. This is, though, important for sports integrity in ensuring that accurate detection is done without wrong accusation. While the proposed model does not adhere to the steepest descents in TPR for higher FPR, the increase across k-NN and RF does reflect their greater inclination towards misclassifying normal behavior.

The random guess baseline (AUC: 0.5) is established as requisite data, revealing that all models perform significantly better than random guesses. A high AUC and good ROC curve for the proposed model confirm its strength in detecting subtle match-fixing patterns and serves as an efficient instrument to promote fairness and trust in sports competitions, while minimizing false positives and giving counter examples.

3.4. Discussion

Under this section, the significances as well as the limitations of the suggested model is discussed. The GBOCLE demonstrates superior capability to detect nonlinear match-fixing patterns within player tracking data thus serving sports integrity measures well. The combination of XGBoost with LightGBM and OC-SVM in the proposed hybrid architecture enables accurate detection of complex temporal and contextual and relational anomalies with high precision (0.9742) and accuracy (0.98382) and AUC (0.98382) as displayed in Table 4 and the ROC curve comparison. The model achieves superior detection capabilities by employing PBIV and MMDI and chaos-based features that help identify coordinated subtle behaviors thus protecting fair play in sports competitions. However, GBOCLE has limitations. GBOCLE requires accurate tracking information to function properly so the preprocessing process must provide appropriate data quality such as DTW or Miss Forest due to its sensitivity to data noise. The model requires substantial computational power from XGBoost and LightGBM tree-based boosting which needs an NVIDIA RTX 3090 GPU for training thus restricting its deployment on Raspberry Pi 4 and other resource-limited devices. The effectiveness of OC-SVM for unsupervised anomaly detection depends on the selection of RBF kernel which demands time-consuming parameter adjustment. The model needs additional testing to validate its performance when deployed in real-world scenarios that contain limited labelled match-fixing data.

4. Results

The proposed framework GBOCLE-Energy was evaluated with multi-node smart grid sensor datasets representing occupancy, motion, and energy use, obtained from industrial, commercial and residential locations.

The evaluation will provide information on the capability of detecting anomalies in real-time, the processing time and CPU load required to use the model on edge devices, and the energy efficiency of the model when used on edge devices.

Table 5. Performance metrics

Model	Accuracy	Precision	Recall	Specificity	F1-score	AUC
Random Forest	0.912	0.903	0.895	0.92	0.899	0.912
SVM	0.887	0.875	0.862	0.902	0.868	0.887
Decision Tree	0.874	0.862	0.849	0.888	0.855	0.874
KNN	0.869	0.857	0.843	0.881	0.85	0.869
XGBoost	0.942	0.933	0.927	0.95	0.93	0.942
LightGBM	0.938	0.93	0.922	0.947	0.926	0.938
GBOCLE-Energy (Proposed)	0.984	0.974	0.975	0.965	0.975	0.984

4.1. Performance metrics

The performance of the proposed framework was measured using six common measurement techniques: Accuracy, Precision, Recall (Sensitivity), Specificity, F1-Score, False Positive Rate and Area Under the ROC Curve (AUC). The proposed framework was compared against six baseline classifiers, including Random Forest (RF), Support Vector Machines (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN) and gradient boosting (XGBoost and LightGBM) - each of which has its strengths and weaknesses (Table 5). Identified Below are Key Observations:
Accuracy & Detection Capability – GBOCLE-Energy provided superior performance than all baseline classifiers and provided a high level of reliable detection capability for non-linear energy consumption anomalies. Edge Deployment Efficiency The GBOCLE-Energy model was able to be deployed on both NVIDIA Jetson Nano and Raspberry Pi 5 edge devices. Both devices show a significant increase in performance of the GBOCLE-Energy solution (34-58% reduction in memory usage, 43-72% decrease in latency time, and lower energy usage). Therefore, GBOCLE-Energy is capable of performing continuous edge operation on either of these devices. Furthermore, the improved detection rates of coordinated or multi-node anomalies (combined) provided by the temporal, contextual and relational features provided through GBOCLE-Energy; these coordinated or multi-node anomalies are often missed by single-modal classifiers.

4.2. Case study: smart grid multi-node

This case study analysed the successful operation of a smart grid, which includes 50 nodes over a period of 30 days. The GBOCLE-Energy model was able to identify all of the normal and abnormal peak usage on a daily basis and weekly basis. Additionally, the model was able to identify the coordinated irregularities due to the operation of multiple devices at the same time, across multiple nodes. Many of the models that are traditionally used do not have the ability to detect these anomalies, and they also produce a large number of false positives. The GBOCLE-Energy model demonstrates that the nonlinear and edge driven artificial intelligence (AI) models represent an improvement in the ability to operate in real-world situations.

4.3. Energy efficiency and scalability

The benefits of using the GBOCLE-Energy framework include:

- Minimal computational overhead (15 ms average inference time per node).
- Minimal physical size (less than 150 MB) of the edge device, as compared with the standard GBDM (300–500 MB).
- Ability to scale easily to more than 100 nodes without degrading performance significantly.

4.4. Discussion

The findings of this study demonstrate that the GBOCLE-Energy provides a practical solution to the real-time monitoring of nonlinear energy consumption in smart grid. The GBOCLE-Energy model provides a solution to two key challenges in the management of modern smart grids by combining lightweight edge deployment and nonlinear ensemble models that provide both the highest level of detection accuracy and the lowest level of energy consumption.

5. Conclusion

This research introduces an AI framework designed to assist edge computing devices by using data collected from smart grids and analysing it for potential energy use anomalies. The framework employs a combination of several machine learning algorithms, including Gradient Boosting (GB), Support Vector Machines (SVM) and Graph-based Analytics, to allow for the detection of anomalies occurring between nodes, as well as those that may exist at specific times, given the context in which they occurred. Results obtained from experiments conducted using GBOCLE-Energy show that its level of performance was superior to all of the corresponding baseline models (Random Forest, SVM, Decision Tree, KNN, Stand Alone Gradient Boosting), where all measurement metrics reported (accuracy, precision, recall, F1-score and Area Under Curve (AUC)) demonstrated that it provided a high performance value to data collected and exhibited a low operational overhead with respect to its use of memory and computational resources, as a result of its deployment on edge computing devices. The edge computing deployment provided both low latency for the receivers of alerts and feedback from GBOCLE-Energy as well as reduced power consumption for devices operating on battery power and scalability to accommodate both continuous monitoring and operations in all types of industrial, commercial and residential smart grid environments and applications. As a result, the proposed framework provides the opportunity for real-time detection of anomalies, predictive energy management and increased operational efficiency for those who integrate the GBOCLE-Energy operational framework into their smart grid operations, highlighting the promise of edge computing and the potential for continued development for use in smart grids to support the sustainability of energy usage operations and the development of large-scale energy production and distribution networks.

Funding
No Funding Received.

Authors Contribution
All authors have contributed equally to prepare the paper.

Availability of data and materials
The datasets generated or analyzed during the current study are available from the corresponding author on reasonable request. No proprietary or restricted data were used in this research.

Conflict of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Geeraert, A. (2025). The dark side of governance networks: how private actors use discursive strategies in global sport governance. *Public Management Review*, 1-27.
2. Manoli, A. E. (2025). Sport marketing's past, present and future: an introduction to contemporary issues in sports marketing. In *Contemporary Issues in Sports Marketing* (pp. 1-5). Routledge.
3. Gündoğan, I. (2025). Anti-Corruption and Disciplinary Campaigns in Chinese Football. In *The Politics of Football in China: Institutional Change and Political Steering Under Xi Jinping* (pp. 31-43). Cham: Springer Nature Switzerland.
4. Souvenir, G., De Clerck, T., Schyvinck, C., Vermeersch, A., Hardyns, W., & Willem, A. (2025). From Financial Statement Fraud to Asset Misappropriation: Fraud in Nonprofit Sports Clubs and Cultural Underpinnings. *Deviant Behavior*, 1-17.
5. Wilkinson, E., Scott, T., Green, M., Hewitt, A., & Naughton, M. (2025). The effect of match-play on acute post-match neuromuscular fatigue following Australian Football League Women's (AFLW) competition. *Biology of Sport*, *42*(2), 257-264.
6. Oliva-Lozano, J. M., Yousefian, F., Chmura, P., Gabbett, T. J., & Cost, R. (2025). Analysis of FIFA 2023 Women's world cup match performance according to match outcome and phase of the tournament. *Biology of Sport*, *42*(2), 71-84.
7. Ma'mun, A., & Juliantine, T. (2025). Moral, competence, and status values in youth sports: predictors of attitude to moral decision-making. *Retos: Nuevas Perspectivas de Educación Física, Deporte y Recreación*, (63).
8. Zhang, J., Han, D., Han, S., Li, H., Lam, W. K., & Zhang, M. (2025). ChatMatch: Exploring the potential of hybrid vision–language deep learning approach for the intelligent analysis and inference of racket sports. *Computer Speech & Language*, *89*, 101694.
9. Chen, Z., & Zhang, T. (2025). Evaluation of Basic Sports Actions for Students Based on DTW Posture Matching Algorithm. *Systems and Soft Computing*, 200196.
10. Wensel, S., Kerr, G., & MacPherson, E. (2025). Advancing Our Understanding of Child Maltreatment and Safeguarding: Implications for Sport and Dance. In *Child Abuse in Sport: Critical Perspectives* (pp. 49-70). Emerald Publishing Limited.
11. Forrest, D., & McHale, I. G. (2019). Using statistics to detect match fixing in sport. *IMA Journal of Management Mathematics*, *30*(4), 431-449.
12. Kim, C., Park, J. H., & Lee, J. Y. (2024). AI-based betting anomaly detection system to ensure fairness in sports and prevent illegal gambling. *Scientific Reports*, *14*(1), 6470.
13. Felice, F., & Ley, C. (2025). Predicting handball matches with machine learning and statistically estimated team strengths. *Journal of Sports Analytics*, *11*, 22150218251313937.
14. Bunker, R., Yeung, C., Susnjak, T., Espie, C., & Fujii, K. (2024). A comparative evaluation of Elo ratings-and machine learning-based methods for tennis match result prediction. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *238*(4), 305-316.
15. Hu, J., Yang, X., Huang, Z., & Xie, J. (2024, October). Machine Learning in Tennis Match Analysis: Predicting Score Point Victor and Momentum Shift. In *2024 5th International Conference on Machine Learning and Computer Application (ICMLCA)* (pp. 21-25). IEEE.
16. Rodu, J., DeJong Lempke, A. F., Kupperman, N., & Hertel, J. (2024). On Leveraging Machine Learning in Sport Science in the Hypothetico-deductive Framework. *Sports Medicine-Open*, *10*(1), 124.
17. Kamalakumar, S., Venugopal, D., & Elakiya, E. (2024). Forecasting the Outcome of an IPL Match using Machine Learning Techniques. *Available at SSRN 5091637*.
18. Brill, R. S., Yee, R., Deshpande, S. K., & Wyner, A. J. (2024). Moving from Machine Learning to Statistics: the case of Expected Points in American football. *arXiv preprint arXiv:2409.04889*.
19. Liu, B., & Wang, Z. (2024). Research on human behaviour recognition method of sports images based on machine learning. *International Journal of Bio-Inspired Computation*, *23*(2), 99-110.
20. Alshardan, A., Mahgoub, H., Alahmari, S., Alonazi, M., Marzouk, R., & Mohamed, A. (2025). Cloud-to-Thing continuum-based sports monitoring system using machine learning and deep learning model. *PeerJ Computer Science*, *11*, e2539.
21. Cui, G., & Wang, C. (2025). The machine learning algorithm based on decision tree optimization for pattern recognition in track and field sports. *PloS one*, *20*(2), e0317414.Yoon J., Lee S.-M., Kang D., Meitl M. A., Bower C. A. & Rogers J. A. *Heterogeneously integrated optoelectronic devices enabled by micro-transfer printing*. (2015). 3: 1313-35.

10.57647/ijeee.2025.1602.08