SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Robust Learning Under Distribution Shifts for Non-Stationary Data Environments

J. Jayaudhaya[1], Shanthi Vairavan[2], Sunil MP[3], Jitendra Kumar Katariya[4], Swapnil Maheshkumar Parikh[5], T. Shanthi[6], Shanthi R[7]

[1] Department of Electronics and Communication Engineering, R.M.D. Engineering College, Chennai, India. Email: jjn.eee@rmd.ac.in
[2] Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India
[3] Assistant Professor, Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, Karnataka, India. Email: mp.sunil@jainuniversity.ac.in, ORCID: 0000-0002-7737-4145
[4] Assistant Professor, Department of Computer Science & Application, Vivekananda Global University, Jaipur, India. Email: jitendra.kumar.katariya@vgu.ac.in, ORCID: 0009-0000-9940-9022
[5] Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India. Email: swapnil.parikh17761@paruluniversity.ac.in, ORCID: 0000-0001-7831-6927
[6] Associate Professor, Department of Electronics and Communication Engineering, Sona College of Technology, India. Email: shanthi@sonatech.ac.in, ORCID: 0000-0001-8962-5311
[7] Assistant Professor, Department of Mathematics, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India

Abstract

Machine learning (ML) systems in non-stationary environments, where data distributions vary with time, require robust learning. Conventional deep learning models make the assumption of stationary data, leading to poor performance in the face of concept drift and domain variability. This research suggests a coherent powerful learning model to overcome the change in the distribution of financial transactions. Financial fraud detection dataset, then scaling input data into a uniform range using min-max normalization. The architecture combines adaptive deep learning and distributed learning through an Intelligent Deep Neural Network (AHO-InDNN) that is an Archerfish Hunting Optimizer. It dynamically balances exploration and exploitation while adjusting to evolving fraud patterns. Various types of financial fraud drift, such as abrupt, gradual, and repeat changes, are modeled and identified with a lightweight drift detection module. Incremental learning strategies and online strategies allow real-time adaptation with resource constraints. Moreover, a mechanism of parameter evaluation based on Large Deviation Principle (LDP) is presented to minimize uncertainty and enhance robustness. Experimental results show that the proposed model achieves 98.74% accuracy, 98.42% precision, 98.52% recall, and 98.37% F1-score, outperforming conventional methods. The suggested framework is more stable, generalized, and resilient, offering a practical solution to fraud detection in the dynamic financial non-stationary setting.

Keyword: Robust Learning, Distribution Shift, Non-Stationary Data, Concept Drift, Adaptive Learning Systems, Deep Learning

Introduction

Contemporary ML systems are dynamic and constantly produce data that defies the notion of fixed distribution [1, 2]. The data properties in non-stationary environments evolve due to the user behavior, the system dynamics, or external forces, and thus learning becomes more complicated [3]. The most important problems are the distribution shift whereby training and testing data is not the same [4]. It is concept drift, where there is a shift in input-output relations, and domain shift, where there is a shift in input distributions, but the same task is being done [5, 6].

These models are primarily trained in a stationary form, without a continuous adaptation process [7, 8]. Thus, the performance is worse when the distributions are varied, which leads to a loss of accuracy in practice and inaccuracy in generalization [9]. This weakness offers the necessity to possess flexible and strong learning systems that can deal with the shifting flows of information. The best system must be able to recognize the patterns of data change, continuously update itself, and be capable of maintaining its performance at all times without full retraining [10, 11].

To alleviate these issues, the proposed presents a potent learning model in non-stationary environments with AHO-InDNN to balance exploration and exploitation. It is a combination of the drift detection and online incremental learning to adapt dynamically. An LDP-based optimization strategy increases stability and minimises uncertainty. The key contributions can be summed up as follows:

- Developed an AHO-InDNN adaptive learning framework integrated with drift detection, online learning, and incremental updating for non-stationary environments.
- Introduced an LDP-based parameter optimization strategy to improve robustness, stability, and reliable decision-making under dynamic uncertainty.
- Performed comprehensive evaluation demonstrating superior performance under sudden, gradual, and recurrent distribution shift scenarios compared with baseline models.

The remaining structure of this research was as follows: Section 2 reviews the literature in robust learning in distribution shifts. Section 3 develops the suggested non-stationary data. Section 4 describes the implementation and design of the model. Section 5 evaluates results of performance. The section 6 ends with conclusion and future research directions.

Literature review

The concept drift research has been extensively studied in the fields of forecasting, streaming analytics, healthcare, and financial systems. Research [12] enhances Photovoltaic (PV) prediction based on adaptive federated learning with dual drift detection and selective retraining with a smaller Root Mean Square Error (RMSE) and faster adaptation, but was restricted by threshold tuning and poor real-world diversity. Likewise, the research [13] compares several adaptive classifiers in the concept drift context based on the scikit-multiflow framework, demonstrating the adaptability of models in a variety of algorithms though lacks real-life validation and more rigorous analysis of optimization.

Deep learning-based drift detection was explored in [14] using a DNN combined with an autoencoder (DNN+AE-DD), where reconstruction error and the 3σ rule are used for drift identification. Despite its sensitivity, it does not have a good cross-domain generalization. In network systems, [15] has used adaptive windowing with H2M networks, which are better with respect to latency and response, but it is based on pre-established traffic assumptions. In the same vein, [16] trained an ensemble variational autoencoder with Kolmogorov-Smirnoff testing to detect cloud drift, with high F-scores but high computational and scalability cost. In additional, [17] used variational autoencoders with KL-divergence testing to identify abnormal behavior in older adults, with over 91% F1-score, but was limited to small sample size and generalizability.

Research Gap

Although concept drift and non-stationary data management have advanced, current approaches are still disjointed, focusing on drift detection, adaptive learning, or optimization individually. Various methods are based on threshold-driven mechanisms, minimizing resilience to changing distributions. Some of the models are also less adaptable and have poor generalization in the real world. Moreover, model uncertainty and robustness optimization are under-explored, and certain methods are costly to compute or may need dataset-specific optimization. To address these issues, this research proposes a robust framework integrating lightweight drift detection with AHO-InDNN for online and incremental learning, enhanced with uncertainty-aware optimization and a LDP-based robustness evaluation to improve stability and adaptability.

Problem formulation

Training and testing data in ML were assumed to share the same underlying distribution, represented as: $Q_{Train}(Y, X) = Q_{Test}(Y, X)$. However, in the dynamic and real-world scenarios the assumption is not always true since data distributions evolve with time as customers adjust their behavior, new fraud strategies are invented, seasonal spending trends and dynamics of online payments. To give an example, the tendencies of fraud that could be traced in the previous weeks may be ineffective now as the fraudsters are coming up with other types of transactions or disguises. This breakdown can be formally stated as: $Q_{Train}(Y, X) \neq Q_{Test}(Y, X)$. Progressive model performance deterioration is caused by distributional mismatch. To be more specific, concept drift is change in the conditional distribution between input and output over time: $Q(Y|X)_s \neq Q(Y|X)_{s+1}$. Let $C_s = \{(y_j, x_j)\}_{j=1}^{m}$ denote data at time s, and e_θ represent the model with parameters θ_s. The goal is to minimize the expected loss in moderating distributions: $K_s(\theta) = \mathbb{E}_{(Y,X)\sim Q_s}[\ell(e_\theta(Y), X)]$. To handle this dynamic environment, the model is incrementally updated at each time step: $\theta_{s+1} = \theta_s + \Delta\theta_s$. Where $\Delta\theta_s$ is adaptively learned using the drift-directed uncertainty estimation-based optimization strategy based on the AHO approach.

Methodology

The methodology suggests a solid adaptive learning model of non-stationary data settings, as demonstrated in Figure 1. It is created to operate on continuous data streams in which distributions can vary because of abrupt, gradual or periodic drift. In contrast to the static models, the framework constantly checks incoming data and updates the model to maintain prediction accuracy in the changing conditions.

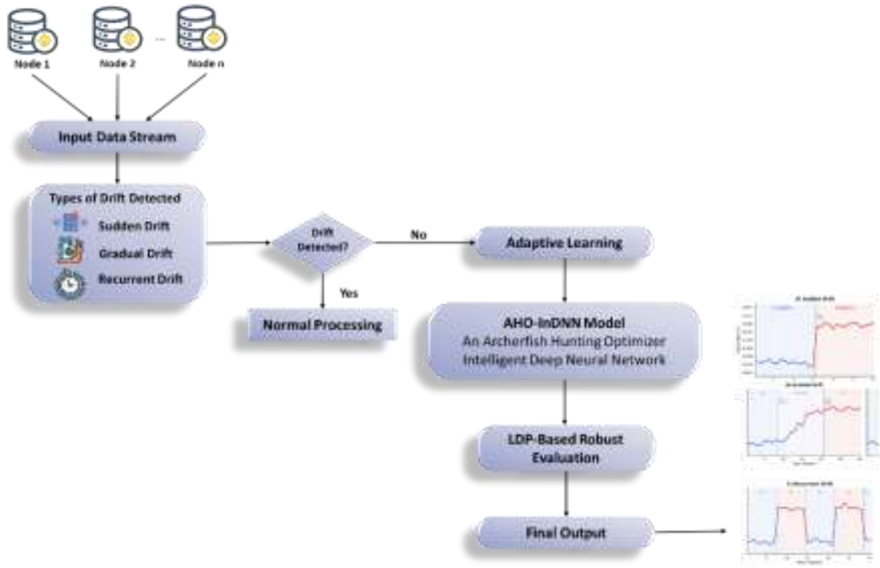

Figure 1: Proposed Robust Adaptive Learning Framework for Non-Stationary Environments

The model consists of four modules; data input, drift detection, adaptive learning and output prediction. Streaming data are preprocessed and analyzed to detect concept drift through statistical and feature changes. After the drift, the AHO-InDNN model re-estimates the parameters to learn effectively. Incremental and online learning allows constant adaptation without re-training. LDP-based mechanism minimizes the uncertainty and stabilizes updates, and distributed learning increases the scalability and computational efficiency of multiple nodes.

Dataset

The research utilizes PaySim artificial mobile money transaction data in Kaggle (https://www.kaggle.com/datasets/sriharshaeedala/financial-fraud-detection-dataset), which was developed to replicate actual financial transactions without compromising their privacy. It contains 30 days of sequential transaction records (743 time steps) with types including CASH-IN, CASH-OUT, PAYMENT, DEBIT, and TRANSFER. Features include transaction amount, time step, anonymized user IDs, and fraud label. It has a time-series structure that can be used in concept drift analysis, adaptive learning, and fraud detection in the changing transaction patterns.

Min-Max Normalization for Scaling Input Data into a Uniform Range

To ensure stable learning and faster convergence, Min–Max normalization is applied to numerical features, scaling values into the range [0,1]. This reduces magnitude variation and improves robustness under distribution shifts, as expressed in Equation (1).

$$y' = \frac{y - y_{min}}{y_{max} - y_{min}} \tag{1}$$

Where y is dynamic data patterns as raw transaction feature; y_{min} observed value as a way to maintain lower-bound stability; y_{max} is observed value as a way to maintain scale variation; y' prime normalized feature as a way to allow robust and drift-resilient learning.

Drift Detection Module for Identifying Concept Drift

The drift detection module of the proposed framework tracks transaction streams to detect changes in the data distribution. It initiates a change in the model when a drastic change is detected to reduce the overall cost of the computation and to retain the high detection levels. The model includes three types of concept drift, like sudden (immediate adaptation), gradual (patterns change gradually as learning progresses slowly), and recurrent (patterns repeat with the use of historical knowledge). These drifts are due to the rapid change of fraud strategies (Table 1), changing user spending behavior, and periodic attack cycles in the fraud detection systems, and adaptive learning is a solution to maintain strong performance.

Table 1. Types of Concept Drift and Examples

Drift Type	Description	Real-World Example
Sudden Drift	Abrupt change in data distribution	New fraud attack pattern
Gradual Drift	Slow, continuous changes over time	Slowly changing user spending behavior
Recurrent Drift	Periodic, repeating patterns	Seasonal fraud campaigns

Hybrid statistical divergence and mean shift enhance the drift detection by being more sensitive to the changes of the distribution and less prone to false positives and stable.

Online & Incremental Learning for Continuous Model Adaptation

The framework online incremental learning module to revise the model as it notices changes in distribution. The model is retrained using the new information received, and is not retrained entirely, which is effective in adapting and preserving previously learnt information. This is the selective means of learning that allows the system to optimize its parameters continuously by new streams of data and uncertainty data. This is particularly essential in fraud detection, as transaction patterns tend to fluctuate as the attackers change their methods of fraud, as their behavior changes, and new patterns of attack emerge. The suggested approach allows constant adaptation, minimizes the computational cost, and is stable in non-stationary conditions. It enhances the reliability of prediction and a robust performance in dynamic and imbalanced transaction flows, which are effective in performing real-time fraud detection.

InDNN for Robust Learning under Distribution Shifts

InDNN is designed to deal with non-stationary data situations that are affected by concept drift and domain variability. It has a three-layer architecture, with input, multiple hidden layers, and output. Unlike standard DNNs, InDNN employs dynamic depth and neuron structure to be robust. It has 64 input neurons, as many as seven hidden layers, and 64 output neurons to learn high-dimensional features. A parameter tuning strategy controls the hidden neurons in a dynamic manner to prevent overfitting and underfitting. The neuron output is defined as (3):

$$x_p^{m+1} = \sigma(w) = \sigma\left(\sum_{j=1}^{n} \omega_{jp}^{m} + \omega_p^{n+1}\right) \tag{2}$$

where $\sigma(w)$ is the activation function, x_p^{m+1} denotes the output of the p^{th} neuron in layer $m+1$, ω_{jp}^{m} represents weights, and ω_p^{n+1} denotes bias. The ReLU activation function is employed to ensure efficient gradient propagation and adaptability under shifting data distributions. To explicitly handle uncertainty caused by distribution shifts, a regularized loss function $F(\theta)$ is used (3):

$$F(\theta) = -\frac{1}{M}\sum_m \sum_p s_{mp} \log x_{mp} \tag{3}$$

where $\sum$ denotes aggregation over all samples and classes; s_{mp} is the true label, $\log x_{mp}$ is the predicted output, $\theta = \{\omega, b\}$, and $F(\theta)$ is a robustness regularization term that mitigates overfitting under drift conditions. To optimize the parameters efficiently in dynamic environments, an adaptive moment-based gradient optimization is employed (4):

$$\theta_s = \theta_{s-1} - \propto \frac{\beta_1 m_{s-1} + (1-\beta_1)\,\nabla_\theta F(\theta_{s-1})}{\sqrt{\beta_2 v_{s-1} + (1-\beta_2)\left(\nabla_\theta F(\theta_{s-1})\right)^2 + \epsilon}} \tag{4}$$

Where θ_s updated parameters, θ_{s-1} previous state, $\propto$ learning rate, β_1, β_2 decay factors, m_{s-1}, v_{s-1} gradient moments, $\nabla_\theta F(\theta_{s-1})$ current gradient, and ϵ ensures stability under distribution shifts. Dropout improves generalization under unseen shifts, while drift-aware updates enable incremental learning when significant distribution changes are detected.

AHO for Optimal Parameter Tuning

The AHO method is used in this research to improve the AHO-InDNN model's parameter setting and adaptive learning in non-stationary situations. AHO balances exploration and exploitation to handle distribution shifts such as concept drift and domain variability, inspired by the hunting behavior of archerfish, where each agent represents a candidate solution of network parameters, hyperparameters, and adaptation strategies.

- **Population Initialization:** Initially, the population of archerfish is randomly generated within the defined search space. The initial position of the j^{th} archerfish is given as (5):

$$q(j,0) = \left[\left(q_1^K + \sigma_1(q_1^V - q_1^K)\right), \cdots, \left(q_c^K + \sigma_c(q_c^V - q_c^K)\right)\right] \tag{5}$$

Where $q(j,0)$ denotes the initial candidate solution; j is the archerfish index; c is dimensionality; q_c^K and q_c^V represent lower (K) and upper (V) bounds of parameters; $\sigma_c \in [0,1]$ ensures randomness for diversity and robustness under distribution shifts.

- **Shooting Behavior (Exploration Phase):** This phase performs global search to adapt to sudden distribution changes. The position update is defined as (6):

$$q(j,s+1) = -\left(q(j,s) - q_{in}(j,s)\right) f^{-(\|q_{in}(j,s)-q(j,s)\|_2)^2} + q(j,s) \tag{6}$$

Where $q(j,s+1)$ denotes the updated proposed model parameters; $q(j,s)$ is the current parameter set of the j^{th} candidate model; $q_{in}(j,s)$ is the estimated optimal solution under distribution shift; j is candidate index; s is iteration; $\|\cdot\|^2$ measures adaptation distance, guiding robust learning.

$$q_{in}(j,s) = q(j,s) + \left(0, \cdots, \frac{u^2}{2\text{g}} \times \sin 2\emptyset, \ldots, 0\right) + \epsilon \tag{7}$$

In equation (7), $q_{in}(j,s)$ is the estimated optimum adaptive solution in the case of distribution shift; $q(j,s)$ represents the state of the current model; u^2 is the adaptive update strength; 2g is scaling stabilization; ϕ is the constant parameters; 0 is the uncertainty added to ensure robustness. In this case, ϵ is in the range of -0.5 to 0.5, which increases uncertainty and noise-resistance in non-stationary data streams.

- **Jumping Behavior (Exploitation Phase):** This phase refines solutions for gradual and recurrent shifts. The position update is expressed as (8):

$$q(j,s+1) = -\left(q(j,s) - q_{in}(j,s)\right) f^{-(\|q_{in}(j,s)-q(j,s)\|_2)^2} + K(j,s) \tag{8}$$

$q(j,s+1)$ denotes the updated candidate solution; $q(j,s)$ is the current solution; $q_{in}(j,s)$ represents the estimated optimal (target) solution; $(\|\cdot\|^2)$ is the Euclidean norm measuring distance; $f^{-(\|\cdot\|_2)^2}$ controls convergence; $K(j,s)$ is the local exploitation component refining parameters under distribution shifts.

$$q_{in}(j,s) = q(j,s) + \left(0, \cdots, \frac{u^2}{2\text{g}} \times \sin 2\emptyset, \ldots, \frac{u^2}{2\text{g}} \times \sin^2 \emptyset, \ldots, 0\right) + \epsilon \tag{9}$$

In equation (9), $q_{in}(j,s)$ is the estimated configuration of the target that should be followed in adaptation to the distribution shift; j is the index of the candidate; s is the iteration; $q(j,s)$ is the reference state in which to update; $\sin^2 \emptyset$ determines the magnitude of the adaptive step; u^2 determines the search intensity; $2g$ normalizes scaling; 0 modulates exploration exploitation; ϵ introduces stochastic perturbation for robustness.

- **Lévy Flight Reinitialization:** To avoid stagnation, Lévy-based random re-initialization is used (10):

$$q_{in}(j,s) = q(j,s) + \sigma\left[u_1^{\frac{q_1}{(1/\beta)}}, \cdots, u_c^{\frac{q_c}{(1/\beta)}}\right] \tag{10}$$

$q_{in}(j,s)$ represents the reinitialized solution for escaping stagnation; $q(j,s)$ is the previous state; σ controls perturbation strength; u_1 and u_c scales direction per dimension; $u_1^{\frac{q_1}{(1/\beta)}}$ and $u_c^{\frac{q_c}{(1/\beta)}}$ induces Lévy-flight-based step variation; β regulates step distribution, enabling long jumps for improved adaptation under dynamic distribution shifts, it represent (11):

$$\begin{cases} q_j \sim e_m(0,\ \gamma^2), \gamma = \left(\frac{\Gamma(\beta+1)\times\sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{\beta+1}{2}\right)\times 2^{\left(\frac{\beta-1}{2}\right)}\times\beta} \right)^{\frac{1}{\beta}} & , j\epsilon\{1,\dots,c\} \\ u_j \sim e_m(0,\ \acute{\gamma}^2), \acute{\gamma} = 1 & , j\epsilon\{1,\dots,c\} \end{cases} \tag{11}$$

where $q_j \sim e_m(0,\ \gamma^2)$ produces Levy-based directional components to search in a global scale; $u_j \sim e_m(0,\ \acute{\gamma}^2)$ provides stabilized stepwise control; $j \in \{1,\dots,c\}$ is the parameter index with c total dimensions; $e_m(\cdot)$ is the Gaussian sampling; γ and $\acute{\gamma}$ is calculated using the Gamma function $\Gamma(\cdot)$ and $\pi\beta$ to control heavy-tailed search; $\acute{\gamma} = 1$ ensures normalized variance. In general, the hybrid AHO-InDNN architecture is an integration of metaheuristic optimization with deep neural learning that enhances the parameter tuning, adaptive convergence, resistance to drift, and generalized prediction in changing conditions.

LDP for Enhancing Model Stability and Robustness

LDP measures the probability of exceptional bursts in streaming data in financial systems. It differentiates between the normal variation, which can be a minor adjustment of a transaction, and the unusual and impactful variations (unauthorized transaction). LDP increases model responsiveness to important changes but is stable in ordinary behaviour by costing more for the rare patterns. Let $\{Xs\}_{s\geq 0}$ denote the incoming data stream. The LDP is defined as (12):

$$\lim_{s\to\infty} \frac{1}{s} \log P(X_s \,\epsilon\, A) = -\inf_{x\epsilon A} I(x) \tag{12}$$

Where s denotes time in the streaming financial data, X_s represents transaction stream, and A indicates abnormal conditions. $P(X_s \,\epsilon\, A)$ denotes the probability of such events. The term $\lim_{s\to\infty}$ captures long-term behavior, while log reflects the exponential decay rate. The function $I(x)$ rarity of fraud behavior, and $\inf_{x\epsilon A}$ represents the minimum deviation cost, where higher values indicate rarer and more significant distributional shifts. LDP detects rare distributional shifts related to fraud, including abrupt increases in total transactions or abnormal spending behavior, to successfully distinguish between changes in normal/unusual fraud.

Result

Results from experiments and an analysis of the suggested AHO-InDNN model's performance in non-stationary data environments using the Python tool. It evaluates the model's effectiveness in handling concept drift through comparative analysis with baseline methods using standard classification metrics.

Figure 2: Analysis of Data Distribution Shift in Fraud Detection (a) Sudden, (b) Gradual, and (c) Recurrent Concept Drift

Figure 2 illustrates how fraud rates change over time under three different concept drift scenarios. In (A) Sudden Drift, the fraud rate sharply jumps at a specific point, showing an immediate shift from a low baseline (Concept C_1) to a higher fraud pattern (Concept C_2). In (B) Gradual Drift, the change occurs gradually over time, the model behavior gradually changes to another concept, and then stabilizes at a higher rate of fraud. In (C) Recurrent Drift, the system switches between low and high fraud rate patterns cyclically, with changes being cyclical with past ideas reoccurring with time. Combined, these plots illustrate that patterns of real-world fraud can dynamically change, and adaptive learning is necessary

Performance Evaluation

Python and standard libraries were used to carry out the experiments. The model has been trained and tested on PaySim data, in a simulated non-stationary environment with added distribution shifts. A measure of performance was the metrics of classification. The AHO-InDNN was contrasted with K-Nearest Neighbors (KNN) [18], Synthetic Minority Over-Sampling Technique Boosting (SMOTEBoost) with cost-sensitive learning [19], and Multi-Head Deep Recurrent Neural Network (MH-DRNN) [20].

To assess the effectiveness of the proposed framework towards detection of strong and adaptable fraud detection (Table 2 and Figure 3) to distribution changes, the following measures are applied: Accuracy, the ratio of the number of transactions correctly classified, is a measure that is used to measure the stability of the model when the data distributions change. Precision means how many of the predicted cases of fraud are actually fraudulent, decreasing false alarms in dynamic financial environments. Recall is the test that analyzes the identification of actual cases of fraud. F1-score is a measure that aims to balance both recall and precision and thus ensures a good overall performance.

Table 2: Performance Comparison of Models under Dynamic Distribution Shifts

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN [18]	91.67	-	95	93.33
SMOTEBoost + Cost-sensitive [19]	-	78	85	81
MH-DRNN [20]	98.5	97	98	97
AHO-InDNN [Proposed]	98.74	98.42	98.52	98.37

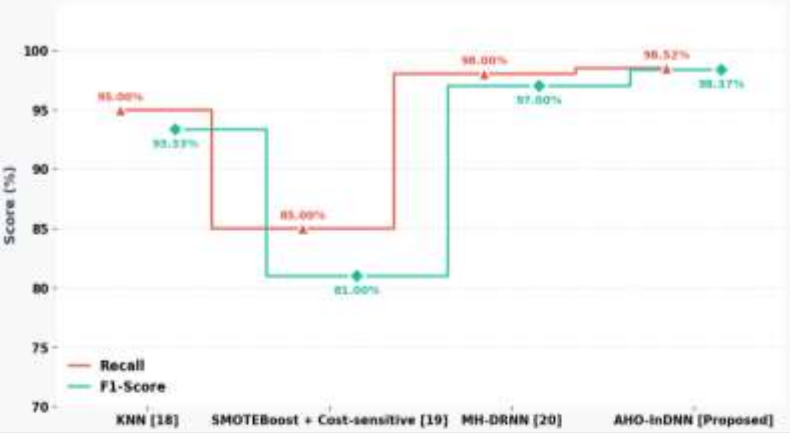

Figure 3: Recall and F1-score Comparison for Robust Fraud Detection

In this research, AHO-InDNN achieves 98.74%, outperforming KNN and MH-DRNN. The proposed model records 98.42%, exceeding SMOTEBoost and MH-DRNN. Overall, AHO-InDNN demonstrates strong adaptability, reliability, and high fraud detection capability in non-stationary environments.

Discussion

The suggested AHO-InDNN model identifies abrupt, gradual and repetitive patterns of fraud drift, guaranteeing stable, adaptive, and precise outcomes in dynamic and unbalanced financial conditions with constantly changing distributions of transactions. Current techniques like KNN [18] are not adaptable because they are based on fixed learning, which restricts them in aspects of drift. SMOTEBoost [19] can better deal with class imbalance, yet it cannot easily deal with continuous distribution changes. MH-DRNN [20] is also highly sequential, but lacks real-time adaptability and increased computational complexity in dynamic fraud situations. Conversely, the proposed AHO-InDNN model obtains better results. This enhancement is supported through drift-aware learning, online incremental, and uncertainty-informed optimization, which provides quick adaptation with a retention of learned knowledge. The framework is generally more robust, false detections are minimized, and generalization is better in the framework of constantly changing patterns of fraud.

Conclusion

In the case of real-life financial systems, detecting fraud is difficult because of the ever-evolving behaviors of the transactions and the shift in the attack techniques. This study introduced an AHO-InDNN-driven adaptive system to deal with sudden, gradual and recurrent concept drift in streaming data by combining drift detection, online incremental learning, uncertainty-sensitive adaptation, and LDP-based optimization. Experimental performance on PaySim data exhibits a high level of performance with 98.74% accuracy, 98.42% precision, 98.52% recall and 98.37% F1-score and a low number of classification errors in the confusion matrix. The model is always superior to the existing methods and it is very robust in non-stationary conditions. The suggested solution needs to be further elaborated to consider the intricacies of real-life financial systems. Future studies are needed to implement in real-time in a distributed environment, simplify lightweight architectures, reduce the cost of computation, scale up, and test the framework on actual financial data.

References

1. Liu, Z., Lu, J., Xuan, J. and Zhang, G., 2024. Deep reinforcement learning in nonstationary environments with unknown change points. IEEE transactions on cybernetics, 54(9), pp.5191-5204. https://doi.org/10.1109/TCYB.2024.3356981
2. Wang, T., Li, M., Zheng, R., Cai, C., Lou, Y. and Zhu, S., 2023. Towards continual knowledge transfer in modeling manufacturing processes under non-stationary data streams: T. Wang et al. Applied Intelligence, 53(23), pp.29393-29413. https://doi.org/10.1007/s10489-023-05034-0
3. Reis, M.J.C.S., 2026. Deep-Learning-Driven Adaptive Filtering for Non-Stationary Signals: Theory and Simulation. Electronics, 15(2), p.381. https://doi.org/10.3390/electronics15020381
4. Benatia, M.A., Hafsi, M. and Ayed, S.B., 2025. A continual learning approach for failure prediction under non-stationary conditions: Application to condition monitoring data streams. Computers & Industrial Engineering, 204, p.111049. https://doi.org/10.1016/j.cie.2025.111049
5. Halstead, B., Koh, Y.S., Riddle, P., Pears, R., Pechenizkiy, M., Bifet, A., Olivares, G. and Coulson, G., 2022. Analyzing and repairing concept drift adaptation in data stream classification. Machine Learning, 111(10), pp.3489-3523. https://doi.org/10.1007/s10994-021-05993-w
6. Ma, S., Yuan, Z., Wu, Q., Huang, Y., Hu, X., Leung, C.H., Wang, D. and Huang, Z., 2023. Deep into the domain shift: Transfer learning through dependence regularization. IEEE Transactions on Neural Networks and Learning Systems, 35(10), pp.14409-14423. https://doi.org/10.1109/TNNLS.2023.3279099
7. Li, J., Yu, H., Zhang, Z., Luo, X. and Xie, S., 2024. Concept drift adaptation by exploiting drift type. ACM Transactions on Knowledge Discovery from Data, 18(4), pp.1-22. https://doi.org/10.1145/3638777
8. Wang, M., Yang, N., Gunasinghe, D.H. and Weng, N., 2023. On the robustness of ML-based network intrusion detection systems: An adversarial and distribution shift perspective. Computers, 12(10), p.209. https://doi.org/10.3390/computers12100209
9. Shaheen, K., Hanif, M.A., Hasan, O. and Shafique, M., 2022. Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks. Journal of Intelligent & Robotic Systems, 105(1), p.9. https://doi.org/10.1007/s10846-022-01603-6
10. Ghibi, O. and Weyns, D., 2024. Dealing with drift of adaptation spaces in learning-based self-adaptive systems using lifelong self-adaptation. ACM Transactions on Autonomous and Adaptive Systems, 19(1), pp.1-57. https://doi.org/10.1145/3636428
11. Cano, A. and Krawczyk, B., 2022. ROSE: robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams. Machine Learning, 111(7), pp.2561-2599. https://doi.org/10.1007/s10994-022-06168-x
12. H. Moayyed, H. Abdeltawab, L. Gomes and Z. Vale, "Adaptive Federated Learning for Robust PV Power Forecasting Under Weather-Induced Concept Drift," in IEEE Access, vol. 13, pp. 193356-193376, 2025. https://10.1109/ACCESS.2025.3630992
13. Pereira, E.V. and Da Silva, W.S., 2025. A comparison of approaches for handling concept drifts in data processed with machine learning. IEEE Access, 13, https://doi.org/10.1109/ACCESS.2025.3557229
14. Desale, K.S. and Shinde, S.V., 2023. Concept drift detection and adaption framework using optimized deep learning and adaptive sliding window approach. Expert Systems, 40(9), p.e13394. https://doi.org/10.3390/app15063056
15. Yu, X., Ruan, L., Evans, J.S. and Wong, E., 2025. Adaptive windowing-based concept drift detection and adaptation framework for human-to-machine applications over future communication networks. Journal of Optical Communications and Networking, 17(4), pp.338-351. https://doi.org/10.1364/JOCN.538964
16. Mehmood, T., Latif, S., Latif, R., Majeed, H. and Malik, A.W., 2024. DRIFTNET-EnVACK: adaptive drift detection in cloud data streams with ensemble variational auto-encoder featuring contextual network. IEEE Access, 12, pp.80020-80034. https://doi.org/10.1109/ACCESS.2024.3409433
17. Friedrich, B., Sawabe, T. and Hein, A., 2023. Unsupervised statistical concept drift detection for behaviour abnormality detection. Applied Intelligence, 53(3), pp.2527-2537. https://doi.org/10.1007/s10489-022-03611-3
18. Usman, A.U., Abdullahi, S.B., Liping, Y., Alghofaily, B., Almasoud, A.S. and Rehman, A., 2024. Financial fraud detection using value-at-risk with machine learning in skewed data. Ieee Access, 12, pp.64285-64299. https://doi.org/10.1109/ACCESS.2024.3393154

19. Al-Daoud, K.I. and Abu-AlSondos, I.A., 2025. Robust AI for financial fraud detection in the GCC: A hybrid framework for imbalance, drift, and adversarial threats. Journal of Theoretical and Applied Electronic Commerce Research, 20(2), p.121. https://doi.org/10.3390/jtaer20020121
20. Wang, L., 2026. MH-DRNN: An intelligent approach in financial fraud detection and prevention. Systems and Soft Computing, p.200483. https://doi.org/10.1016/j.sasc.2026.200483