Enhancing Out-of-Distribution Detection for Retail Time-Series Data Using Entropic Methods

Nga Nguyen Thi[1], Tuan Vu Minh[2], Khanh Nguyen-Trong[3]
Intelligent Computing for Sustainable Development Laboratory (IC4SD),
Posts and Telecommunications Institute of Technology, Vietnam[1,3]
Faculty of Information Technology, Hanoi University, Hanoi, 10000, Vietnam[1,2]
Faculty of Information Technology, Posts and Telecommunications Institute of Technology,
km10 Nguyen Trai, Hanoi, 10000, Vietnam[3]

***Abstract*—Machine learning models are typically developed under the "closed-world" assumption, where training and testing data originate from a consistent distribution. However, in real-world scenarios, especially in the retail domain, this assumption can become problematic due to the frequent introduction of new products, seasonal promotions, and irregular sales events. When models encounter out-of-distribution data inputs, predictions can become overly confident or entirely incorrect. While existing out-of-distribution detection methods primarily focus on image-based datasets, challenges associated with numerical, high-dimensional, and heterogeneous retail time-series data remain largely unexplored. To address this gap, this study proposes an enhanced Entropic Out-of-Distribution Detection framework tailored specifically for dynamic retail environments. By transforming time-series sales data into spectrogram representations and leveraging the IsoMax+ loss function, our approach improves uncertainty calibration and robustness without requiring labeled out-of-distribution data or additional post-hoc calibration techniques. Experimental results, conducted on a large-scale retail dataset from Vietnam, demonstrate that the proposed Entropic Out-of-distribution detection framework significantly outperforms traditional out-of-distribution detection methods in terms of detection accuracy and inference efficiency, providing a scalable and practical solution for real-time retail applications. Our approach achieves strong performance with an F1-score of 88% and an AUC of 91%, highlighting its promising applicability across diverse business scenarios.**

***Keywords*—Out-of-Distribution Detection, entropic learning, IsoMax+ loss, time-series classification, retail forecasting, deep learning, spectrogram transformation**

I. Introduction

In conventional supervised learning settings, models are typically developed under the "closed-world" assumption, where classifiers are trained on a predefined set of known classes and assume future data will remain consistent with the training distribution. However, this assumption rarely holds in real-world scenarios, where data distributions evolve continuously and models encounter previously unseen samples. As a result, predictive accuracy deteriorates and uncertainty estimates become unreliable when facing distributional shifts [1], [2]. To maintain model reliability, robust mechanisms for detecting anomalies and unseen data—commonly known as Out-of-Distribution detection are essential. During a preliminary survey of recent OOD research, we observed that most existing approaches are developed and benchmarked primarily on image datasets. This observation raises a natural question: Can OOD detection techniques be effectively adapted to numerical, business-oriented time-series data, such as those found in retail analytics?

In retail analytics, OOD data frequently emerge from new product launches, regional demand changes, or promotional campaigns. Such shifts can lead to unstable sales patterns, causing models to generate inaccurate forecasts and poor business decisions [3]. Conventional OOD approaches, including Maximum Softmax Probability (MSP) [4] and Mahalanobis Distance [5], are computationally simple but struggle with subtle distributional drifts. Recent entropy-based methods [6], [7] have shown promise in improving calibration by quantifying prediction uncertainty, yet they often rely on post-hoc temperature scaling or external OOD datasets [8], limiting real-world scalability.

To overcome these challenges, researchers have explored information-theoretic and distance-aware frameworks. Sugiyama et al. [9] introduced density ratio estimation to capture distributional shifts, and Macêdo et al. [10] proposed the Entropic Out-of-Distribution Detection (EOOD) framework that replaces softmax scoring with a distance-regularized entropy loss.

More recently, Liang et al. [11] and Han et al. [12] further demonstrated that entropy regularization and contrastive predictive entropy can enhance OOD robustness in deep networks and time-series tasks.

These findings confirm the potential of entropy-based learning to improve detection accuracy in dynamic, high-stakes domains such as healthcare and retail [13].

1) Motivation and objective: Despite these advances, applying OOD detection to retail time-series data remains underexplored. Retail datasets are highly non-stationary, multi-modal, and influenced by seasonal, geographic, and economic factors.

This study aims to develop a scalable, entropy-driven OOD framework that can adapt to evolving retail conditions without external calibration or labeled OOD data.

Such capability can enhance model reliability, reduce forecast errors, and improve business decision-making in dynamic commercial environments.

2) Main contributions: The contributions of this work can be summarized as follows:

- We propose an enhanced Entropic Out-of-Distribution Detection (EOOD) framework tailored to retail time-series data, combining entropy and distance-based uncertainty modeling for improved calibration.
- A spectrogram-based transformation is introduced to convert sales sequences into time–frequency representations, enabling convolutional architectures to capture temporal dynamics effectively.
- A labeling and preprocessing pipeline is designed to generate realistic in-distribution and out-of-distribution samples without requiring external OOD datasets or post-hoc calibration.
- Extensive experiments on large-scale Vietnamese retail data demonstrate that the proposed method outperforms conventional baselines in detection accuracy, calibration robustness, and inference efficiency.

The remainder of this study is organized as follows: Section II reviews representative OOD detection methods and their limitations. Section III details the proposed EOOD methodology, including data processing and model architecture. Section IV presents experimental results and analysis, while Section V concludes the study and outlines future research directions.

II. Related Work

Out-of-Distribution (OOD) detection has become an essential problem in reliable deep learning, aiming to identify samples that do not conform to the training data distribution [4]. Early approaches relied on confidence-based measures, such as Maximum Softmax Probability (MSP) [4] and temperature scaling or input perturbation methods as in ODIN [14]. Later, Mahalanobis distance-based techniques [3] improved robustness by modeling feature-space distributions, while IsoMax and IsoMax+ losses [10] introduced distance-aware learning for better uncertainty calibration.

Recent research has continued to refine these directions. Liang et al. [11] proposed entropy regularization to stabilize OOD decision boundaries. Han et al. [12] extended entropy-based detection to time-series data through contrastive predictive modeling, and Liu et al. [15] leveraged spectral representation learning to enhance feature robustness. Macedo et al. [16] revisited IsoMax+ for improved distance-awareness, while Yang et al. [17] provided a comprehensive survey categorizing OOD methods by analytical principles and highlighting emerging challenges.

To synthesize these developments, Table I presents a comparative summary of representative OOD detection approaches, categorizing them by type of analysis, advantages, limitations, and their relationship to the proposed EOOD framework. This overview clarifies how our method differs by emphasizing entropy–distance synergy, calibration-free detection, and adaptability to dynamic retail time-series data.

OOD detection has emerged as a critical subfield in machine learning, particularly in safety-critical and dynamic environments. Early works often relied on confidence-based metrics such as Maximum Softmax Probability (MSP) [4], which takes the highest softmax output as a confidence score. While MSP is computationally efficient, it is prone to overconfidence on unfamiliar inputs. Distance-based methods, such as the Mahalanobis Distance approach [3], measure the feature-space distance between a sample and class centroids under Gaussian assumptions. Although effective in some cases, these methods are sensitive to covariance estimation errors and tend to degrade in high-dimensional, non-isotropic settings.

To mitigate softmax overconfidence, Liang et al. [14] proposed ODIN, which combines temperature scaling and input perturbations to improve separation between in-distribution (ID) and OOD samples. ODIN improves detection accuracy but requires careful hyperparameter tuning and access to OOD-like validation data, limiting its scalability. Entropy-based scoring methods [6] address this issue by using the entropy of the predictive distribution as an uncertainty measure—low entropy indicating confident predictions for ID data, and high entropy reflecting unfamiliarity for OOD inputs. However, many still depend on post-hoc calibration [7] or labeled OOD datasets [8], which may be impractical in real-time settings.

Entropic Out-of-Distribution Detection was introduced by Macêdo et al. [10] to overcome these limitations. EOOD replaces the softmax layer with IsoMax+, a distance-aware probability formulation that computes class likelihoods based on the squared Euclidean distance between an input's feature vector and learned class centroids. This design naturally produces higher predictive entropy for OOD samples—improving separation without post-hoc adjustments or external OOD data. IsoMax+ encourages isotropic feature distributions, making the model more robust in high-dimensional and noisy data environments. Recent studies such as Jin et al. [18] further confirm the effectiveness of entropy-based and distance-aware approaches in dynamic, high-dimensional settings.

Despite its success on image benchmarks such as CIFAR10, CIFAR100, and SVHN, most prior EOOD research has been limited to static spatial structures. In contrast, retail time-series data presents unique challenges: high dimensionality, seasonal patterns, abrupt demand spikes, and sparsity across product–location pairs. Traditional OOD methods designed for images may fail to capture these temporal–seasonal anomalies effectively.

To summarize, Table I provides a synthesis of representative OOD detection methods, categorized by their analytical principles, advantages, and limitations. This overview highlights how our proposed EOOD framework improves uncertainty calibration and adaptability to dynamic retail data.

This comparative summary facilitates a clearer understanding of how the proposed EOOD method distinguishes itself from existing approaches in terms of scalability, uncertainty calibration, and adaptability to dynamic time-series data.

This study addresses this shortage by extending EOOD to retail time-series forecasting. We introduce a spectrogram-based transformation that converts 12-month revenue sequences into image-like representations, enabling the use of deep CNN architectures such as EfficientNet-B3. By integrating IsoMax+ loss, our approach preserves EOOD's calibration-free property while enhancing its ability to detect OOD patterns in dynamic, high-variance retail environments—supporting real-time anomaly detection where timely insights are crucial for operational decision-making.:

TABLE I. COMPARATIVE SUMMARY OF REPRESENTATIVE OUT-OF-DISTRIBUTION (OOD) DETECTION METHODS CATEGORIZED BY ANALYTICAL APPROACH, HIGHLIGHTING THEIR RESPECTIVE ADVANTAGES, LIMITATIONS, AND DISTINCTIONS FROM THE PROPOSED EOOD FRAMEWORK

Method	**Type of Analysis**	**Advantages**	**Limitations**	**Comparison with the Proposed Framework**
Maximum Softmax Probability (MSP) [4]	Confidence-based	Simple, fast, easy to implement	Overconfident on unseen data	Lacks uncertainty calibration
ODIN [14]	Temperature scaling and input perturbation	Improves OOD separation	Requires tuning and validation data	Higher complexity; not scalable
Mahalanobis Distance [3]	Distance-based	Captures feature-space structure	Sensitive to covariance estimation	Less effective on high-dimensional time-series
Entropic OOD Detection (EOOD) [10]	**Entropy + Distance**	**Calibration-free, robust**	**Limited to image data**	**This study extends it to retail time-series**

$$p_i = \frac{e^{-d(f(x),c_i)}}{\sum_{j=1}^{K} e^{-d(f(x),c_j)}} \tag{1}$$

where, $f(x)$ represents the feature vector of input sample x, c_i is the centroid of class i, and $d(\cdot)$ is a distance metric such as squared Euclidean distance. This formulation encourages the model to produce higher entropy for OOD samples, as they lie farher away from the class centroids. Therefore, the model shows ability to detect OOD instances without the need for additional calibration or external datasets.

Out-of-Distribution detection has emerged as a fundamental subfield in machine learning, particularly in applications where model required reliability and safety. Several approaches have been proposed to distinguish between ID and OOD inputs, with varying assumptions and mechanisms for uncertainty estimation.

One of the earliest and most often used methods is the MSP, introduced by Hendrycks and Gimpel [4]. This method uses confidence score from the softmax output as an indicator of distributional fit. While computationally efficient, MSP likely overconfident predictions on unfamiliar inputs. In case of dealing with subtle anomalies or class overlapping distributions.

To solve these limitations, ODIN was proposed by Liang et al. [14], combining temperature scaling and input perturbations to amplify the distinction between ID and OOD samples. Despite its improved detection performance, ODIN requires careful hyperparameter tuning and access to validation data, which may hinder its scalability across domains.

Another mature method is the Mahalanobis Distance based approach introduced by Lee et al. [3], which computes the distance between a test sample and class conditional distributions in the feature space. Although this method works effective in enhancing internal feature representations,it assumes Gaussianity and sensitive to covariance estimation errors particularly in high-dimensional or non-isotropic feature spaces.

More recently, entropy-based methods have gained attention for their principled handling of predictive uncertainty. In particular, Macêdo et al. [10] proposed the Entropic Out-of-Distribution Detection framework, which replaces the conventional softmax layer with IsoMax+ loss. Traditional softmax layers calculate class probabilities based on unnormalized logits, which often lead to poorly calibrated confidence estimates, as shown by Guo et al. [7]. Therefore, this method is unsuitable with uncertain estimation.

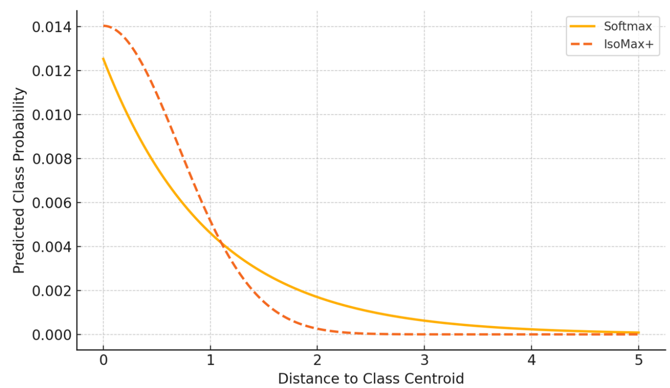


Fig. 1. Comparison of softmax and IsoMax+ confidence decay with respect to increasing feature distance from class centroids.

IsoMax+ loss overcome this limitation by using distance based approach instead of softmax function. This replacement limits marking unfamiliar data by attach for them lower confidence scores and high predictive entropy. In addition, by enhancing isotropic feature, this model is easily detect OOD data without requiring external datasets. Therefore, it is suitable with high dimension and noisy data. To clarify comparison between Softmax and IsoMax+, Fig. 1 visualizes how each method predict class based on distance of an input sample form the class centroid.

1) Softmax: The standard softmax computes class probabilities using linear logits:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2}$$

where, z_i is typically computed as a linear projection: $z_i = w_i^\top f(x) + b_i$, with $f(x)$ being the input feature vector. This formulation does not explicitly depend on the distance between the input and any class centroid, which can lead to overconfident predictions for unfamiliar inputs (i.e., OOD data).

2) IsoMax+: In contrast, IsoMax+ replaces logits with negative distances in the embedding space:

$$p_i = \frac{e^{-d(f(x),c_i)}}{\sum_{j=1}^{K} e^{-d(f(x),c_j)}} \tag{3}$$

with the distance metric defined as:

$$d(f(x), c_i) = \|f(x) - c_i\|^2 \tag{4}$$

Here, c_i is the centroid of class i. This formulation ensures that the farther an input lies from any class center, the lower its predicted probability, thereby naturally increasing predictive entropy for OOD samples.

Recently, as the limitaion of Sofmax based, IsoMax+ which a part of EOOD framework addressed the entropy detection help model more practical when using for real world data.

III. METHODOLOGY

This section presents the EOOD framework for retail time-series. The pipeline has four stages: 1) data preprocessing and surrogate OOD labeling, 2) spectrogram transformation, 3) EfficientNet-B3 backbone with IsoMax+ loss, and 4) entropy-based OOD scoring.

A. Overview of the Proposed Framework

Fig. 2 summarizes the workflow. Raw transaction logs are aggregated into monthly series, transformed into spectrogram images via STFT, and then fed to an EfficientNet-B3 backbone trained with IsoMax+ loss for uncertainty-aware classification. The predictive distributions are finally converted into OOD scores.

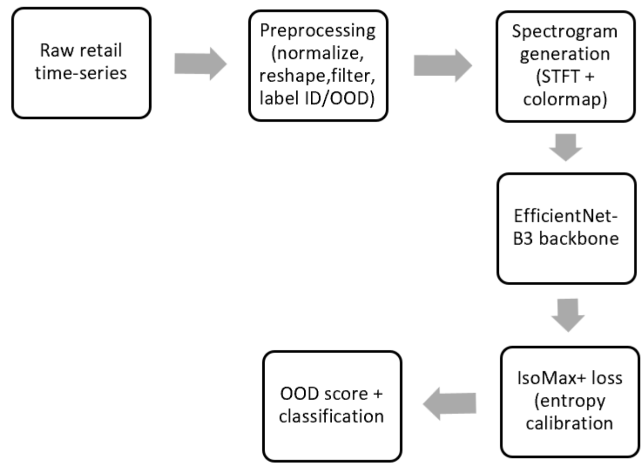


Fig. 2. Proposed EOOD framework for retail time-series OOD detection.

B. Data Pre-processing and Labeling

We use transaction-level sales data from 63 provinces in Vietnam (2024), including product ID, date, quantity, revenue, and location. Records are aggregated monthly per (product, province) to form 12-month sequences; only pairs with complete records and at least six non-zero months are retained to avoid degenerate or overly sparse series.

To emulate realistic distributional shifts without relying on external OOD datasets, two complementary heuristics were applied for labeling, as illustrated in Fig. 4.

1) Spike-based labeling rule:: A month t is labeled as out-of-distribution (OOD) if its revenue R_t satisfies:

$$R_t > 2 \times R_{t-1}, \tag{5}$$

capturing short-term shocks such as promotions, sudden demand surges, or supply disruptions.

2) Entropy-based volatility labeling:: For each 12-month revenue series,

$\mathbf{R} = \{R_1, R_2, \ldots, R_{12}\}$,

we compute the volatility ratio (serving as an entropy proxy) as:

$$E = \frac{\sigma(\mathbf{R})}{\mu(\mathbf{R})}, \tag{6}$$

where, $\sigma(\mathbf{R})$ and $\mu(\mathbf{R})$ denote the standard deviation and mean of the sequence, respectively. A sequence is labeled as OOD if its entropy ratio exceeds a global threshold:

$$E > \mu_E + 1.5\sigma_E, \tag{7}$$

where, μ_E and σ_E represent the mean and standard deviation of E across all product–province pairs. This criterion identifies long-term volatility or regime shifts in sales behavior.

After pre-processing, a total of 33,680 valid product–province time-series pairs were obtained, including 22,450 in-distribution (ID) and 11,230 OOD sequences. These labeled sequences were subsequently transformed into spectrogram images for model training in the following section.

C. Spectrogram Transformation

Each 12-month sequence is converted into a spectrogram using the Short-Time Fourier Transform (STFT),

$$S(t, f) = \sum_{m=-\infty}^{\infty} x[m]\, w[m-t]\, e^{-j2\pi fm} \tag{8}$$

and mapped to a three-channel image via a fixed colormap. This representation preserves temporal order while exposing time–frequency structure, where retail anomalies often appear as localized spectral distortions. Prior studies show that such image-based encodings can improve time-series classification and anomaly detection in complex settings [6], [19]. Visual examples are shown in Section IV.

D. Model Architecture

We adopt EfficientNet-B3 as the backbone and resize spectrograms to 300×300 pixels. EfficientNet's compound scaling achieves a favorable accuracy–latency balance compared to heavier backbones, which is crucial for near real-time retail inference [20]. Global average pooling is applied to the final feature maps, and the ImageNet classifier is replaced by an IsoMax+ head.

E. IsoMax+ Loss for OOD Detection

Traditional softmax classifiers often produce overconfident predictions even for inputs that are far from the training distribution. This limitation arises because the Softmax function converts unnormalized logits into probabilities without explicitly considering the distance of a sample from any known class centroid. As a result, OOD samples can still be assigned high confidence scores, leading to poor uncertainty calibration [7].

IsoMax+ addresses this problem by replacing linear logits with distance-based formulations. Specifically, the logit for class i is computed as:

$$z_i = -\alpha \cdot \|\mathbf{f} - \mathbf{c}_i\|^2 \tag{9}$$

where, $\mathbf{f}$ is the feature vector of an input, $\mathbf{c}_i$ is the centroid of class i, and α is a learnable scaling parameter. This formulation ensures that samples located far from all class centroids are assigned low confidence and high predictive entropy, thereby mitigating the overconfidence issue without requiring post-hoc calibration [10].

To provide intuition, Fig. 3 compares the posterior probability distributions of Softmax and IsoMax+. As shown, Softmax tends to output peaked, low-entropy posteriors even for ambiguous inputs, while IsoMax+ yields smoother, high-entropy distributions that better reflect uncertainty. This property makes IsoMax+ particularly suitable for OOD detection in dynamic retail environments.

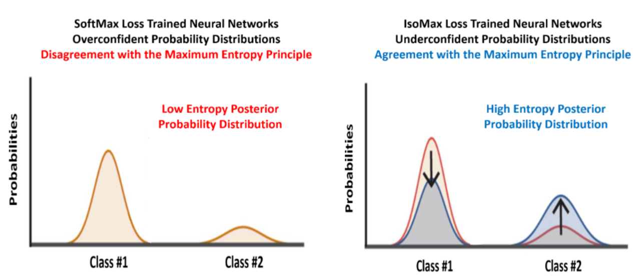


Fig. 3. Comparison of posterior probability distributions produced by Softmax and IsoMax+.

F. Training Strategy

Models are trained end-to-end with AdamW and a cosine learning rate schedule. We use random cropping, horizontal flips, brightness/contrast jitter, Gaussian noise, MixUp (α=0.4), label smoothing (0.1), and Exponential Moving Average of weights. These choices improve robustness to small rendering differences in spectrograms, temper overconfident posteriors, and stabilize convergence under volatile demand patterns typical of retail data.

In summary, the EOOD framework combines dual-criteria labeling, spectrogram-based encodings, an EfficientNet-B3 backbone, and IsoMax+ loss to deliver calibration-free OOD detection suited to dynamic retail environments.

IV. Experiments

A. Dataset Collection and Pre-Processing

The raw dataset contained approximately 1.2 million transaction records from 63 provinces in Vietnam, covering the 12-month period of 2024. The data were obtained from the retail sales management system of *Liberico Company*, which operates on the nationwide VNPost postal distribution channel. This channel supports multiple vendors and retail partners beyond Liberico itself, making it a rich and heterogeneous retail environment that includes diverse product categories, pricing schemes, and sales behaviors from different businesses. The dataset was directly extracted from the company's database and exported into excel format Table II summarizes the key attributes contained in the raw transaction dataset. After monthly aggregation and filtering for completeness (i.e., retaining only product–province pairs with at least six non-zero months), a total of 33,680 valid time-series sequences were preserved for analysis. These sequences were subsequently labeled based on two criteria:

- Spike-based labeling: A time-series is labeled OOD if any month shows a revenue spike exceeding 2× the previous month.
- Entropy-based labeling: The entropy of each sequence was computed as the ratio $\frac{\sigma}{\mu}$ (standard deviation over mean). If this value exceeded a global threshold of $(\mu + 1.5\sigma)$, the sequence was labeled as OOD.

TABLE II. Data Dictionary of the Retail Transaction Dataset

Field	Description
Date	Transaction date (dd/mm/yyyy), used to derive month.
Province	Province or city where the transaction occurred.
Invoice ID	Unique identifier of the invoice.
Product ID	Unique identifier of the product.
Product Name	Description of the product.
Unit	Measurement unit (e.g., piece, pack, bottle).
Unit Price	Price per unit (VND).
Quantity	Number of units sold.
Revenue	Total transaction revenue (Unit Price $\times$ Quantity).
Commission Rate	Commission ratio applied to the transaction.
Commission (derived)	Calculated as Revenue $\times$ Commission Rate.
Total Payment (derived)	Calculated as Revenue $-$ Commission.
Month (derived)	Extracted from the Date for time-series grouping.

In total, 22,450 sequences were labeled ID, and 11,230 were labeled OOD. The distribution of entropy values for ID and OOD sequences is shown in Fig. 4.

Each of the 33,680 time-series sequences was converted into a spectrogram using the Short-Time Fourier Transform. The resulting spectrograms were mapped to a 3-channel image using a fixed colormap. This conversion produced a dataset

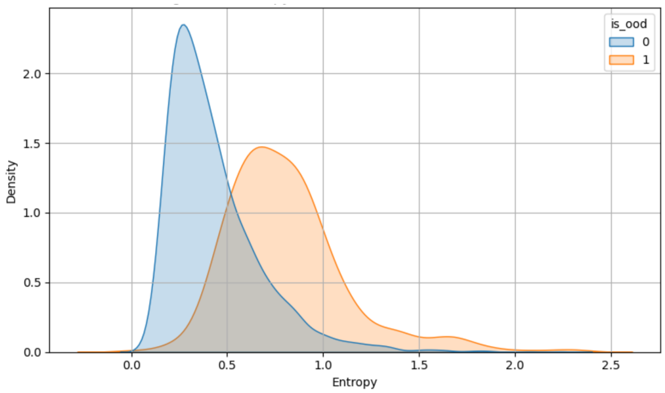


Fig. 4. Distribution of entropy scores for labeling ID (blue) vs. OOD (orange) sequences.

of 33,680 spectrogram images, each of size 300×300 pixels, which were then organized into two directories: one for ID (label 0) and one for OOD (label 1). An example spectrogram of a 12-month revenue time series is shown in Fig. 5.

To improve generalization and avoid overfitting, we applied several data augmentation techniques as followings:

- Random horizontal flip: This was used to mimic random variations in input patterns (e.g., seasonal variations).
- Random cropping: Spectrogram images were randomly cropped to 270×270 pixels during training. This allowed the model to focus on localized features within the image, improving robustness.
- Brightness jitter: Slight changes to the brightness of the images were made to mimic changes in lighting conditions in the retail environment.
- Gaussian noise injection: Small amounts of noise were added to the images to simulate real-world signal fluctuations.
- MixUp ($\alpha = 0.4$): This technique mixes two random images during training, allowing the model to generalize better by learning from both images at once, which also increases the model's robustness against noisy data.

Fig. 5. Example spectrogram of a 12-month revenue series.

B. Experimental Setup

1) Training parameters: Following the methodology pipeline (Fig. 2), the spectrogram images were used as input to an EfficientNet-B3 backbone pre-trained on ImageNet. The model was trained with IsoMax+ loss for out-of-distribution (OOD) detection. We used the AdamW optimizer with an initial learning rate of 2×10^{-5} and a cosine-annealing schedule that progressively reduced the learning rate to ensure stable convergence. The loss function and learning rate scheduler were selected to balance optimization stability and convergence speed.

2) Exponential Moving Average (EMA): To further stabilize training and ensure that the model's weights converge to a stable point, we applied the Exponential Moving Average (EMA) of model weights. This technique averages the weights over time to ensure that fluctuations in model parameters are minimized, resulting in a more robust model that performs better in real-world scenarios [21]. The decay factor was set to 0.995, meaning that each new weight update is blended with the previous one by 99.5%.

3) Training environment: Training was carried out on NVIDIA Tesla T4 GPU with a batch size of 32. This GPU is highly efficient for training models like EfficientNet and provides a good balance between memory and processing power, making it suitable for high-throughput deep learning tasks. The model was trained for 75 epochs, and the early stopping criterion was applied, which halted training if the validation loss did not improve for 10 consecutive epochs. This approach helped prevent overfitting and ensured that training time was efficiently managed.

The model performance was evaluated based on classification accuracy, F1-score, AUC-ROC, and inference time(measured in milliseconds per sample).

C. Results

1) Phase 1: Baseline training and convergence analysis: Fig. 6 shows the accuracy and loss curves during the first phase of training with an 80/20 split. The model achieved rapid convergence within the first 15 epochs, stabilizing around 89% accuracy, while the validation loss decreased and plateaued at approximately 0.37. The confusion matrix in the figure shows that 48 out of 71 OOD samples were correctly classified, while 23 were misclassified as in-distribution. Table III shows the classification report for Phase 1.

TABLE III. CLASSIFICATION REPORT FOR PHASE 1

Class	Precision	Recall	F1-Score
In-Distribution	0.94	0.93	0.93
Out-of-Distribution	0.63	0.68	0.65
Accuracy		0.89	
Macro avg	0.79	0.80	0.79
Weighted avg	0.89	0.89	0.89

2) Phase 2: Enhanced training and robustness evaluation: In the second phase, the dataset was split into 70% for training, 15% for validation, and 15% for testing. This phase included advanced regularization techniques, such as MixUp, label smoothing, and EMA, to improve generalization and

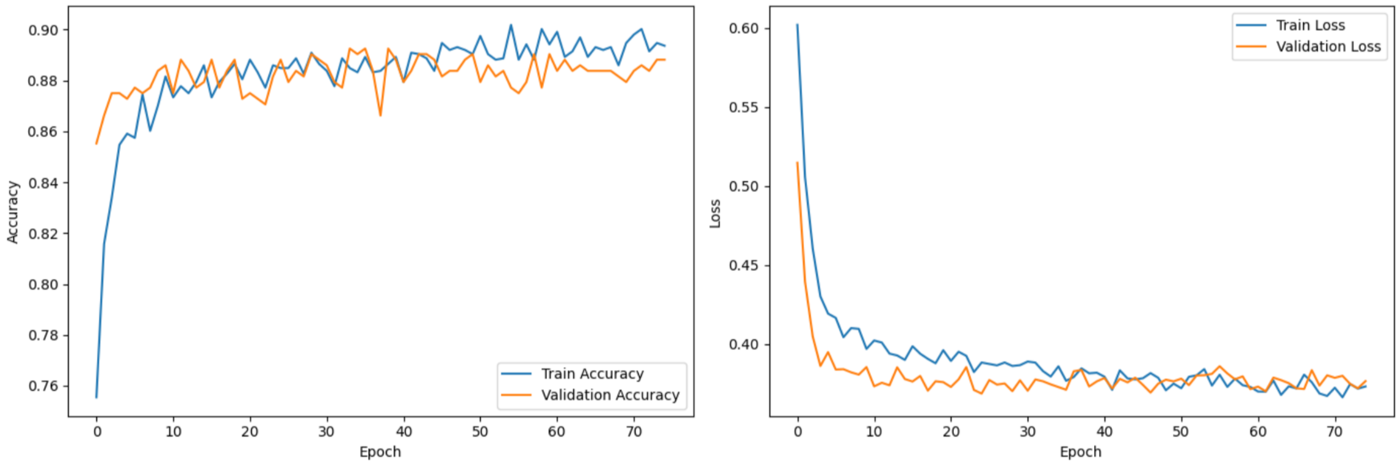


Fig. 6. Accuracy and loss curves during Phase 1.

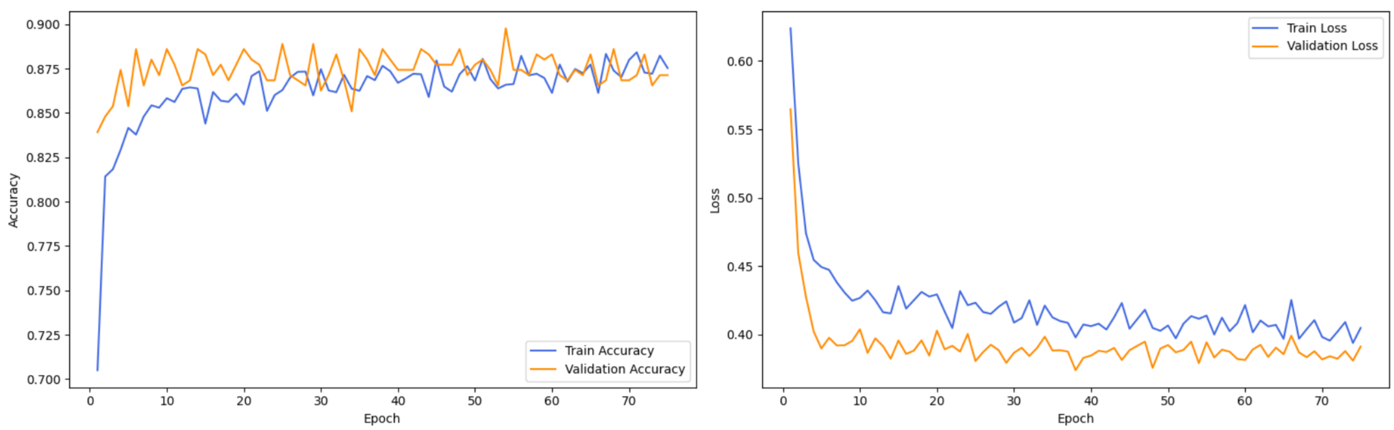


Fig. 7. Accuracy and loss curves during Phase 2.

reduce overfitting. The model performed consistently well, with training and validation accuracy stabilizing 87% after epoch 30, as shown in Fig. 7. The validation loss remained slightly lower than the training loss throughout the phase, indicating stable generalization.

TABLE IV. CLASSIFICATION REPORT FOR PHASE 2

Class	Precision	Recall	F1-Score
In-Distribution	0.91	0.92	0.92
Out-of-Distribution	0.69	0.65	0.67
Accuracy		0.87	
Macro avg	0.80	0.79	0.79
Weighted avg	0.86	0.87	0.86

The detailed classification performance is summarized in Table IV, where the model achieved an overall accuracy of 87%. The In-Distribution class obtained high precision and recall (0.91 and 0.92, respectively), while the Out-of-Distribution class was more challenging, with a precision of 0.69 and recall of 0.65. These results highlight the model's improved robustness compared to Phase 1, although OOD detection remains the more difficult task.

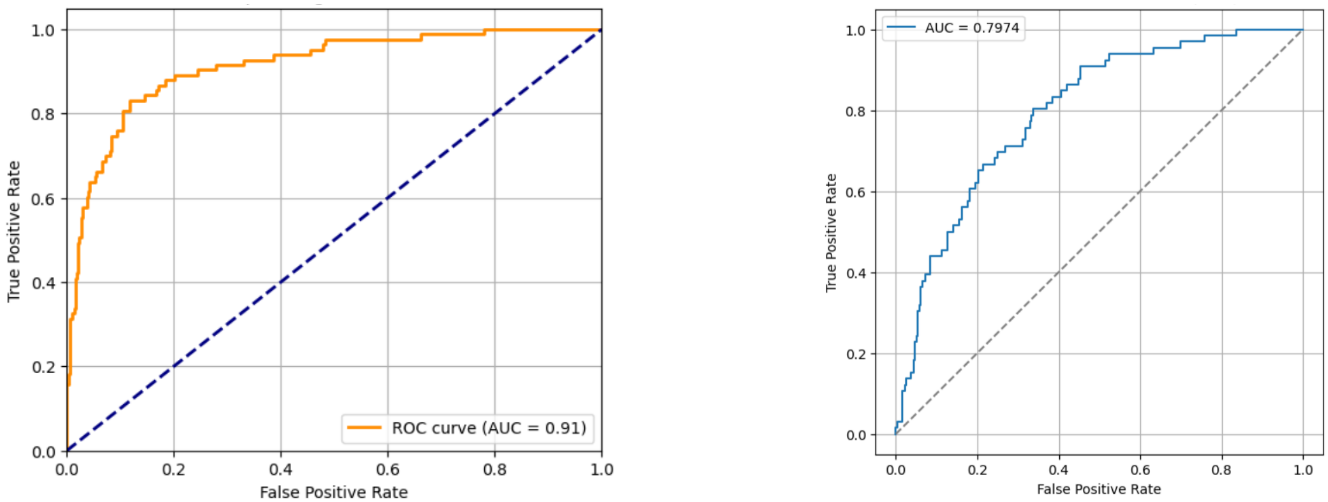
Additionally, Fig. 8 presents the ROC curve of the best EMA-tracked model evaluated on the 15% test set. The model achieved an AUC of 0.7974, demonstrating its strong ability to distinguish between in-distribution and OOD samples.

3) Comparison with baseline methods: To comprehensively evaluate the performance of the EOOD framework, we compared it with several baseline OOD detection methods: MSP, Entropy Scoring, ODIN, and Mahalanobis Distance. Table V shows the comparison results based on key metrics: F1-score (OOD), AUC–ROC, and inference time.

EOOD consistently outperforms all baseline methods across the key metrics. It achieves the highest F1-score of 0.88 for OOD detection and the best AUC–ROC of 0.91, demonstrating superior ability to distinguish between ID and OOD samples. In contrast, while ODIN and Mahalanobis show competitive AUCs, their inference times are significantly higher, making them less suitable for real-time deployment. Methods like MSP and entropy scoring suffer from low OOD recall due to overconfidence in softmax outputs.

D. Discussion

The experimental findings presented in Section IV demonstrate that the proposed EOOD framework effectively enhances out-of-distribution (OOD) detection for retail time-series data.



(a) ROC Curve on 80/20 split (Phase 1).

(b) ROC Curve on 70/15/15 split (Phase 2).

Fig. 8. Comparison of ROC curves across two training configurations.

TABLE V. PERFORMANCE COMPARISON BETWEEN EOOD AND TRADITIONAL OOD DETECTION METHODS, SORTED BY F1-SCORE

Method	F1-score (OOD)	AUC–ROC	Inference Time (ms)
MSP	0.22	0.80	9
Entropy	0.36	0.80	9
Energy	0.41	0.75	11
Mahalanobis	0.81	0.86	17
ODIN	0.83	0.89	19
EOOD	**0.88**	**0.91**	**12**

By transforming sales sequences into spectrogram representations and employing the IsoMax+ loss, the model improves uncertainty calibration and maintains efficient inference suitable for real-time applications.

From the comparative experiments, the Phase 1 baseline achieved 0.89 accuracy and an F1-score of 0.65 for OOD samples, whereas the enhanced Phase 2 configuration—with MixUp, Label Smoothing, and Exponential Moving Average (EMA)—achieved smoother convergence and more stable generalization. Although the accuracy slightly decreased to 0.87, the model exhibited lower variance in validation loss and better calibration, as seen in Fig. 7. This trade-off indicates improved robustness and reduced overfitting under dynamic retail conditions.

As shown in Table V, EOOD outperformed Mahalanobis and ODIN, achieving the highest overall accuracy and efficiency. These results verify that the entropic–distance synergy inherent in IsoMax+ enables superior calibration-free OOD detection compared to confidence-based and post-hoc methods. In particular, EOOD achieved an approximate 5% gain in AUC over Mahalanobis Distance while reducing latency by nearly 35%, offering a practical trade-off between accuracy and computational cost.

Despite these promising results, Fig. 8(b) also shows that the recall for OOD samples remains modest (0.65), implying that extreme or rare anomalies may still overlap with in-distribution variance. This suggests the need for adaptive thresholding or dynamic decision boundaries to further improve sensitivity under evolving retail conditions.

The framework's scalability and calibration-free operation make it well-suited for deployment in production environments. Beyond retail, similar entropic representations could be applied to other time-series domains such as financial fraud detection, healthcare monitoring, or industrial sensor analysis, where rapid and reliable anomaly identification is critical.

The framework's scalability and calibration-free operation make it well-suited for deployment in production environments. Beyond retail forecasting, the proposed framework can also support anomaly detection tasks in finance, healthcare, and industrial monitoring, where timely and reliable OOD identification is critical. Its calibration-free nature and low inference latency make it practical for large-scale real-time deployment.

Overall, these findings confirm that entropy-based learning combined with distance-aware regularization forms a strong foundation for robust, scalable, and interpretable OOD detection across complex and dynamic business environments.

V. CONCLUSION

This study presented an enhanced Entropic Out-of-Distribution Detection (EOOD) framework specifically designed for dynamic retail time-series data. The proposed approach integrates IsoMax+ loss with spectrogram-based representation learning to improve robustness against distributional shifts — a persistent challenge in real-world retail environments. By transforming sales sequences into visual spectrograms, the model effectively captures temporal dependencies and subtle anomalies in a spatial format, extending the applicability of convolutional architectures beyond traditional image domains.

Unlike conventional OOD techniques that require external datasets or post-hoc calibration, the proposed framework operates in a fully calibration-free manner, making it practical for business applications where real-time decisions are critical. This design promotes scalability and adaptability across domains with high-dimensional, sparse, or unlabeled data, supporting broader applications in predictive modeling and anomaly detection.

The findings demonstrate that entropy-based learning, when combined with distance-aware regularization, can significantly enhance uncertainty calibration and detection accuracy. The framework offers both scientific and practical benefits, contributing to more reliable forecasting and decision-making in data-driven retail operations.

A. Limitations

Despite its strong empirical performance, the proposed EOOD framework still has several limitations. First, the labeling process for in-distribution and out-of-distribution samples relies on heuristic thresholds (e.g., entropy ratio and revenue spike rules), which may not generalize optimally across different datasets or business domains. Second, the current approach focuses on short-term (12-month) retail patterns, while longer seasonal or multi-year dynamics have not been explored. Finally, converting temporal sequences into spectrograms, though effective for CNN-based learning, may introduce minor information loss compared to direct temporal modeling approaches such as Transformers or recurrent architectures. These limitations highlight opportunities for improvement in generalization and interpretability.

B. Future Work

Several directions can be pursued to extend this research. First, adaptive thresholding techniques could further enhance sensitivity to gradual concept drift in long-term deployments. Second, integrating Bayesian-inspired uncertainty modeling may strengthen interpretability and improve predictive calibration. Finally, cross-domain evaluation in fields such as finance, cybersecurity, and energy analytics could validate the generalizability of the proposed framework and highlight its potential for broader adoption.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[2] Z. Chen and B. Liu, *Lifelong Machine Learning*. Morgan & Claypool Publishers, 2018.

[3] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 7167–7177, 2018.

[4] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *International Conference on Learning Representations (ICLR)*, 2017.

[5] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.

[6] Y. Zhang and et al., "Entropy-based out-of-distribution detection," *Journal of Machine Learning Research*, vol. 22, no. 10, pp. 1234–1256, 2021.

[7] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017, pp. 1321–1330.

[8] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*. MIT Press, 2009.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[10] D. Macêdo, T. I. Ren, C. Zanchettin, A. L. Oliveira, and T. Ludermir, "Entropic out-of-distribution detection," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021.

[11] S. Liang, C. Wang, and Y. Zheng, "Entropy regularization for reliable out-of-distribution detection in deep networks," *Neural Networks*, vol. 173, p. 106208, 2024.

[12] Y. Han, H. Zhang, and B. Li, "Ood detection for time-series via contrastive predictive entropy," *Pattern Recognition Letters*, vol. 173, pp. 13–21, 2023.

[13] Q. Liu and et al., "Application of neural network-based out-of-distribution detection in healthcare systems," *Journal of Healthcare Systems Research*, vol. 34, no. 3, pp. 245–259, 2022.

[14] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *International Conference on Learning Representations (ICLR)*, 2018.

[15] Y. Liu, X. Wang, and Y. Jin, "Improving out-of-distribution detection through spectral representation learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 11–23, 2023.

[16] D. Macedo, T. I. Ren, and C. Zanchettin, "Isomax+ and beyond: Revisiting distance-aware classifiers for robust out-of-distribution detection," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2023.

[17] S. Yang, J. Li, X. Liu, and J. Xu, "A survey on out-of-distribution detection in deep learning: Taxonomy, challenges, and future directions," *Information Fusion*, vol. 103, p. 102070, 2024.

[18] Y. Jin and et al., "A study on the effectiveness of entropy for detecting out-of-distribution data," *Journal of Machine Learning Research*, vol. 23, no. 5, pp. 987–1003, 2022.

[19] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," in *International Conference on Machine Vision (ICMV)*. SPIE, 2017, p. 87790A.

[20] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, vol. 97. PMLR, 2019, pp. 6105–6114.

[21] P. Izmailov, D. Molchanov, T. Garipov, C. Blundell, and D. Vetrov, "Averaging weights leads to wider optima and better generalization," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 4300–4310.