ZANCO Journal of Pure and Applied Sciences

TYPE Original article
DOI http://dx.doi.org/10.21271/ZJPAS.38.3.15

OPEN ACCESS

*Corresponding author

Shaymaa A. Hantoosh
shymmaakram35@mtu.edu.iq

RECEIVED :09 /09 /2025
ACCEPTED :8/12/ 2025
PUBLISHED :30/06/2026

KEYWORDS:

Out-of- Distribution (OoD), Artificial Intelligent (AI), Convolutional Neural Network (CNN), Out-of-Distribution data (OoD), Maximum Softmax Probability (MSP), Cosine similarity (Cos-Sim), Mahalanobis Distance (Mah-d), Confidence Interval (CL).

Selective Classification Gates to Handle Out-of-Distribution OoD in CNN

Shaymaa A. Hantoosh

Middle Technical University, Continuous Education Center, Baghdad, Iraq

ABSTRACT

Detecting out-of-distribution (OoD) data is a critical process in artificial intelligence models that specialize in critical systems that require accurate results, such as medical systems, air navigation systems, and financial transaction systems. Therefore, researchers in the field of neural network programming place great importance on OoD handling. OoD detection approaches aim to identify inputs that deviate from a model's training distribution, preventing the overconfident and false predictions that neural networks suffer from. In this paper using Post-hoc scoring approaches which as selective classification gates to handle OoD in a classification convolutional neural networks CNN model. That model is supposed to implemented in a vending machine, it's trained on a label dataset of the Iraqi currency, that model suffers from OoD because of the convergent probabilities of the mathematical model of the neural network. With this approach, test four types; four methods were chosen to prevent out-of-distribution for the CNN model, which are: Confidence, Entropy, Cosine similarity, and Mahalanobis distance. Mathematical calculations were performed for each method with the tested data for the model and found threshold value to make as gate to pass currency classes and reject non-currency classes. Then the results of these gates comparing and analysis, it was concluded that the most accurate method of handling OoD is the Mahalanobis distance, which handled the OoD in model on real data from 85.29% to

1.Introduction

For the image's classification challenge, convolution neural networks (CNN) perform the best images processing techniques. Nevertheless, the practical application of neural networks apps faces several real-world difficulties. These issues are mostly dataset-related, including sample imbalance, train-test leakage, and the training dataset's limited size. User conduct is the source of further issues. For instance, these user-generated data are of low quality and are irrelevant (Vareldzhan et al., 2021, Szyc et al., 2021). Identifying unrelated, or unknown data is crucial for constructing an appropriate training dataset and for eliminating irrelevant samples during inference. Various formal definitions are suggested for these issues, including anomaly detection, novelty discovery, outlier detection, and Out-of-distribution detection, this study employs the conventional term meanings (Bulusu et al., 2020). In three learning techniques; supervised, semi-supervised, and unsupervised learning, OoD detection methodologies can be used to handle incorrect predication (Bulusu et al., 2020). In supervised learning, in-distribution and out-of-distribution samples with appropriate labeling supervised. Consequently, OoD detection constitutes a binary classification problem. Semi-supervised learning utilizes solely the in-distribution samples. Unsupervised learning employs no data annotation. Numerous techniques exist for out-of-distribution identification in vector datasets, including the local outlier factor, Mahalanobis distance, isolation forest, one-class support vector machine, autoencoder, variational autoencoder, and Gaussian mixture model for variational autoencoder, among others. The majority are executed in open-source libraries (Masaki et al., 2021, Sun et al., 2022). Neural networks models have been widely applied for real-world detection and its need to recognize inputs that do not belong to the data samples. While deep networks are vulnerable to (OoD) Predictive uncertainty is essential in safety-critical real-world scenarios such as autonomous driving, medical, financial, etc. When encountering some examples that the deep model has not been exposed to during training (Zhu et al., 2022, Szyc et al., 2021). In some researches that achieve high accuracy of classification process on the benchmark datasets (such as the latest models by Pham et al. (Pham et al., 2022)).

When the main goal is classification. So, Out-of-distribution (OoD) data identify and then eliminate from the predictions data, such samples would result in a cost process. In out-of-distribution data was present in the input samples, this problem considers selective classification. While, Selective Classification in the Out-of-Distribution data (SCOD) to prioritize the downstream classifier as the primary target (Xia and Bouganis, 2022, Liu et al., 2020) . Most deep image classification models are trained in the closed-world setting, the out-of-distribution (OoD) issue arises and deteriorates customer experience when the models are deployed in production, facing inputs coming from the open world (Wang et al., 2022). The main challenge to ensure the safety of DNNs is to estimate if a given input sample comes from the same data distribution for which a DNN was trained. This is very hard to estimate because the network usually extrapolates its decision while receiving new image samples. Another cause of the incorrect recognition of outlying samples can be the distributional shift of input data over time ; such as the time-dependent variations of an object’s appearance (Olber et al., 2023).

In this paper, a Selective classification gates to handle OoD in CNN is presented. The remainder of the paper is organized as follows: The rest of the paper is set up like this: Section II covers Out-of-Distribution data while OoD techniques present in Section III; Section IV present Selective classification method and Section IIV present Analysis of Approaches & Discussion. Finally, theses sections presented as this follow; Conflict of Interest, Acknowledgment, Author’s Contribution, and Conclusion.

2.Out-of-Distribution (OoD) Data

Lastly, Many algorithm available with techniques to isolate inputs samples that are’t belong to dataset by estimating class conditional posterior probabilities. (Xia and Bouganis, 2022),

Alternative methods attempt to recognize OoD samples by quantifying outlierness “d (x;Xi)” of input x in known training data Xi related to the class ci maximizing the posterior probability as show in equation1:

$$ci = \arg\max c2C\ P(cjx) \quad \ldots\ldots\ldots 1$$

While, the score d is estimated as distance, density, or some measure of outlierness, such as the Local Outlierness Factor (LOF) (Szyc et al., 2021).

The core of an OoD detector is a scoring function Ɵ that maps an input feature x to a scalar in ℝ, which indicate to an extending of the sample, thus it makes likely to be OOD. In testing, a threshold ℸ is decided, ensuring that the validation set retains at least a given true-positive rate (TPR), e.g. the typical value of 0.95. The input example is regarded as OOD if Ɵ (x) > ℸ and as ID (i.e., in-distribution) otherwise. In cases where a score indicating the ID-ness is convenient, we can mentally use the negative of OOD score as the ID score (Wang et al., 2022, Galil et al., 2023).

Out-of-distribution (OoD) detection aims to flag inputs that lie outside a model’s training distribution. Simple post-hoc score baselines include Confidence via maximum SoftMax probability (MSP), which treats lower max-probability predictions as more likely OoD, and Entropy of the SoftMax, which uses higher predictive entropy as an uncertainty OoD cue (both easy to compute from a classifier’s logits). Training-time strategies such as Outlier Exposure learn models to down-weight anomalies, while energy-based scoring often used to outperform raw SoftMax-based scores. In practice, systems mix these and calibrate a threshold for isolate a samples that aren’t belong to the training data and represent incorrect prediction “out-of-distribution (OoD)” in the model (Chan et al., 2021, Cheng and Vasconcelos, 2022, Olber et al., 2023).

3.Out-of-Distribution (OoD) Techniques

Handling out-of-distribution (OOD) data consider a main problem in deploying neural networks for real-world applications, as models frequently encounter inputs that differ from their training distribution (Averly and Chao, 2023). This challenge can be addressed by a variety of methods that have been proposed to improve robustness and reliability (Azeez and Hamadameen, 2024). Broadly, these methods can be categorized into uncertainty-based approaches, which use SoftMax confidence, Bayesian inference, or ensembles to flag anomalous inputs; representation-based approaches, which leverage embedding spaces, distance metrics such as Mahalanobis distance, or contrastive learning to separate in-distribution and OOD data; and generative approaches, which model the training distribution explicitly with density estimation or energy-based methods.(Raghuram et al., 2021) (Techapanurak et al., 2020)

Recent advancements also include post-hoc calibration and hybrid methods that combine multiple paradigms. These strategies enable neural networks to detect and adapt to unseen data, reducing overconfident misclassifications and enhancing safety in domains such as autonomous driving, natural language processing, and medical diagnosis (Karunanayake et al., 2025). Despite progress, challenges remain in scaling to high-dimensional settings, generalizing across domains, and integrating methods seamlessly into end-to-end learning frameworks. In this paper use the standard types of selective classification to handle Out-of-Distribution, these types are:

3.1 Confidence (Max Softmax Prob)

The Maximal Softmax Probability (MSP) serves as a straightforward, model-agnostic confidence metric commonly employed as a benchmark for (OOD) detection. For logits \(z(x) \in \mathbb(Bulusu et al., 2020)^K \) corresponding to input (x), class posteriors are computed using the softmax function, and the maximum value is utilized as a confidence score; out-of-distribution (OOD) inputs are identified when this score is below a specified threshold (Liu et al., 2020) (Wang et al., 2022). MSP is effective because, on average, in-distribution samples provide greater maximum than misclassified or out-of-distribution data, facilitating efficient separation using a singular scalar statistic (Alsinayi et al., 2025).

Nonetheless, MSP is susceptible to overconfidence—contemporary deep networks can allocate elevated softmax probabilities to inputs that deviate significantly from the training distribution (and are frequently miscalibrated), so

constraining MSP's dependability without further processing. Notwithstanding these constraints, MSP continues to serve as a robust, parameter-free benchmark and a prevalent reference point for OoD detectors in vision, natural language processing, and speech tasks (Liu et al., 2020) (Hsu et al., 2020).

In this work used Con method to separate the output predication of the CNN model by compute the confidence for a set of samples that model train it and find threshold value T to be the threshold to separate the two clusters (currency & non currency) as seen in Figure 1 , from this figure noted that threshold value equal to " 0.995" and this value can be separate the samples of classes that belong to a set of currency from the all other non-currency.

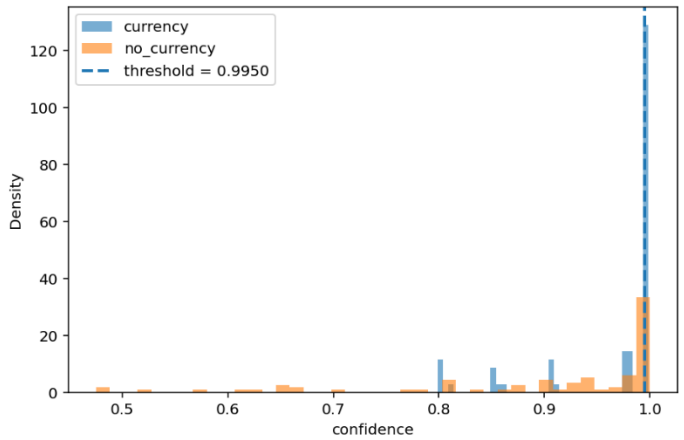


Figure 1: Confidence (MSP) and Threshold value

3.2 Entropy (Softmax Uncertainty)

Handling The entropy of the softmax serves as a straightforward and efficient uncertainty metric for OoD detection. Given class posteriors in equation 2 (Chan et al., 2021)

$$pk(x) = softmax(zk) \qquad \ldots\ldots\ldots 2$$

While the predictive entropy given in e equation 3, optionally normalized by log K

$$\mathrm{H(p)} = -\sum_k p_k \log p_k \qquad \ldots\ldots\ldots 3$$

The predictive entropy is elevated when the network exhibits uncertainty and diminished when it demonstrates confidence; thus, applying a threshold on H(p) can identify probable out-of-distribution inputs (Hassen and Abdlrazaq, 2024). This strongly aligns with the traditional maximum softmax probability (MSP) baseline, which demonstrated that out-of-distribution (OOD) and misclassified instances typically exhibit fewer "peaked" softmax outputs compared to in-distribution examples. Nevertheless, contemporary networks frequently exhibit miscalibration and may display excessive confidence for out-of-distribution data (Chan et al., 2021).

In this work, Entropy separate the result of predication that CNN model predicate it, and then a threshold found value T to separate the two clusters (currency & non currency) as seen in Figure 2 , from this figure noted that threshold value equal to " 0.036" and this value suppose it separated the samples of classes that belong to a set of currency from the other non-currency.

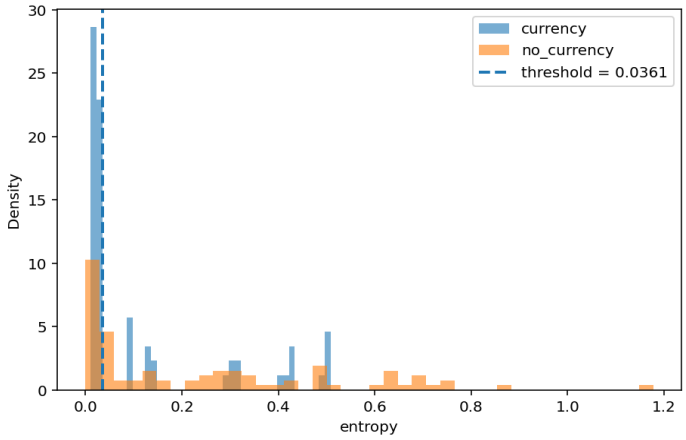


Figure 2: Entropy (Softmax Uncertainty) and Threshold value

3.3 Cosine similarity

Cosine similarity (Cos-Sim) is a metric utilized to measure the similarity of two non-zero vectors, regardless of their magnitude. Cosine similarity measures the angle between vectors in a high-dimensional space, rather than evaluating their Euclidean distance. This renders it especially advantageous in neural models, where embeddings (vector representations) encapsulate semantic or hidden attributes. Cos-Sim focuses on direction, not magnitude, making it robust to varying feature norms and it give better separability for example in many embedding spaces, cosine similarity correlates better with

semantic similarity than Euclidean distance. To defined Cosine Similarity, suppose two vectors x and y as in equation 4: (Techapanurak et al., 2020).

$$cosine\_sim(x,y) = \frac{x \cdot y}{\|x\|\|y\|} \quad \ldots\ldots\ldots 4$$

Where the dot product is in equation 5

$$x \cdot y = \sum_{i=1}^{n} x_i y_i \quad \ldots\ldots\ldots 5$$

The range of cosine similarity is [1;0;-1], 1 is perfectly similar (same direction), while 0 is an orthogonal (no similarity), and -1 represent diametrically opposed (opposite direction).

In this work (Cos-Sim) method separate the clusters in the CNN model by compute the maximum similarity between samples and determine each class by find an appropriate threshold value T to separate the two clusters (currency & non currency) as seen in Figure 3 , from this figure noted that threshold value equal to " '0.995" and this value can be separate the samples of classes that belong to a set of currency from the all other non-currency.

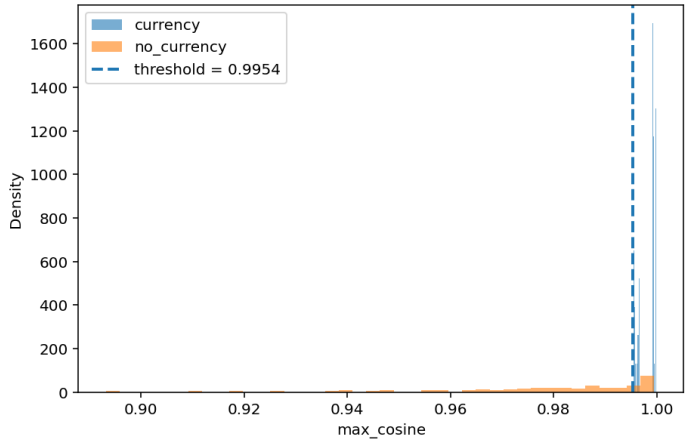


Figure 3: Cosine similarity (Cos-Sim) and Threshold value

3.4 Mahalanobis Distance

Mahalanobis distance (Mah-d) can be defined as a statistical measure of the distance between a point or a way to determine the closeness among a set of distribution belonging to a specific class. It treats all variables equally and accounts for correlations between variables and scales them according to variance. This makes it especially useful for multivariate anomaly or OoD detection, classification, and separating overlapping distributions (Podolskiy et al., 2021, Anthony and Kamnitsas, 2023).

Mahalanobis distance is defined as in Equation 6 to handle OoD if we have set of samples x belong to a vector Ψ(x):

$$D_m(x) = min_c \sqrt{(\Psi(x) - \mu_c)^T \sum^{-1} (\Psi(x) - \mu_c)} \quad \ldots. 6$$

In this equation define Ψ(x) represent as observation vector, the mean vector of the distribution represents as μc, while Σ^{-1} is an inverse covariance matrix. If Σ is the identity matrix, Mahalanobis distance reduces to be an Euclidean distance (Podolskiy et al., 2021, Anthony and Kamnitsas, 2023).

In this work used Mah-d method to separate the output predication of the CNN model by compute the minimum Mah-d for a set of samples that model train it and find threshold value T to be the threshold to separate the two clusters (currency & non currency) as seen in Figure 4 , from this figure noted that threshold value equal to "'87.054" and this value can be separate the samples of classes that belong to a set of currency from the all other non-currency.

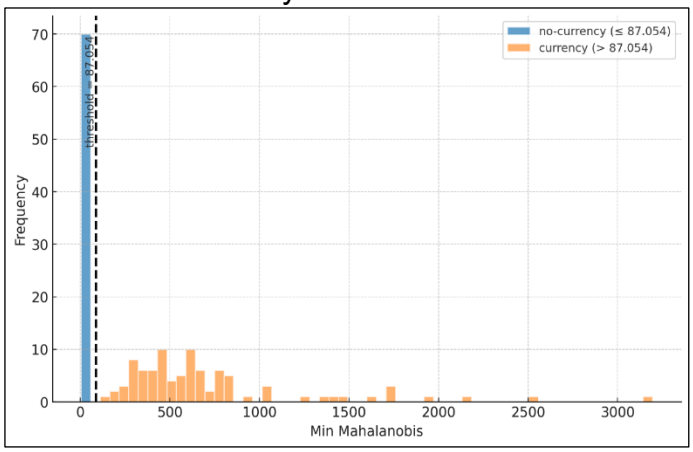


Figure 4: Mahalanobis distance and Threshold value

4.Selective Classification

Selective Classification can be defined the classifier as a pair of functions, the classifier f(x) (in our case given by equation (7) that produces a prediction y, and a binary rejection function

$$g(x;\ t) = \begin{cases} 0\ (reject\ prediction);\ if\ S(x) < t \\ 1\ (accept\ prediction);\ if\ S(x)\ \ t\ , \end{cases} \ldots 7$$

where an operating threshold represents t while S considers a scoring function which is typically a measure of predictive confidence or S measures uncertainty. Lastly, a selective classifier chooses to reject if it is uncertain about a prediction (Xia and Bouganis, 2022).

5.Analysis of Approaches & Discussion

In this paper CNN model has been designed and trained on a synthesis dataset of a Iraqi currencies, this dataset contains 800 samples: divided as “currency “ set which contain 700 samples of seven Iraqi backbone (250,500, 1000, 5000, 10000, 25000,50000) IQD , and “no currency” set which contain 100 different images of other countries backbones and variety images. CNN architecture is composed of three convolution layers (32, 64, 128) and take input layer (128,128,3ch) with average pooling to these layers and used AdamW as optimizer. This model train with activation function ReLU and learning rate “0.001”, output layer contains SoftMax activation function to predicate the correct class for the label seven class. This model train on currency dataset which divides into 70% of the samples data and 30% of the test models, the model validation on 20% from training data, while “no currency” dataset used to set the probabilities in validation process. The model epoch continuous until the model is stable, thus this model stabilized by prediction after about 28 epoch, by get the maximum value in validation accuracy “99.51%” and the most minimum validation loss “0.05441”, Precision “1.00”, Recall “1.00” F1-score “1.00” the values of accuracy and loss function ratios of the proposed CNN model during the training the proposed CNN model can be seen in Figure 5. In test process the performance of the proposed CNN model which check by the test samples achieved accuracy about 99.98, and this percentage consider perfect performance for the proposed model.

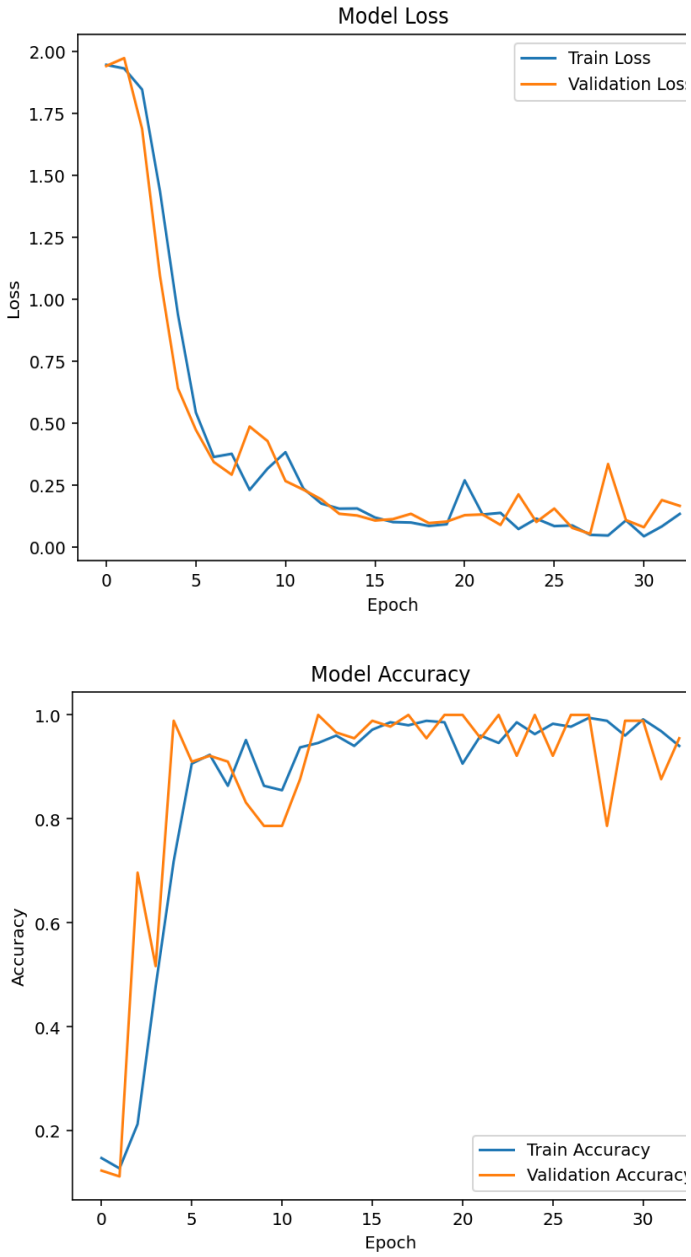


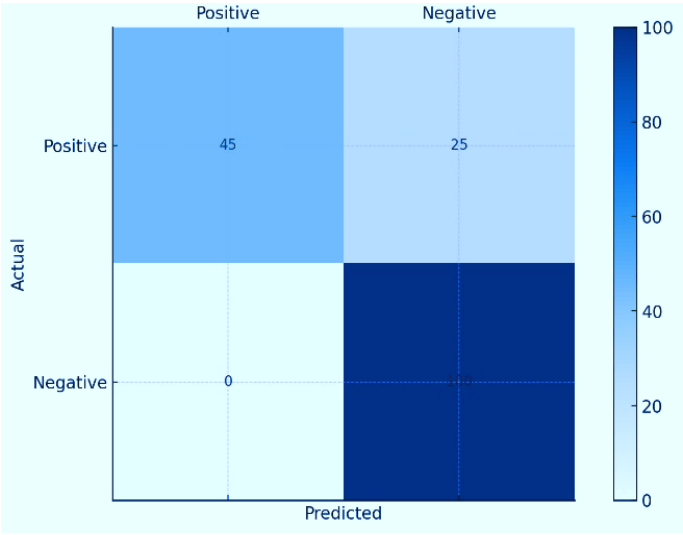
Figure 5: The accuracy and loss function ratios of the proposed CNN model during the training and validation processes

After that the model test on un seen data from the real world that gate from multi-sites which contain set of samples contain 170 samples divided as “70 currency label and 100 no currency label “ , the result of this test process get the following result : the accuracy is decrease to be “ 85.294 “, while precision “1.00”, Recall “ 0.64”, F1-score “0.78”. In this type of supervised model can be used selective classification gate to handle Out-of- Distribution, thus the overfitting of model can be correct. This model implemented by

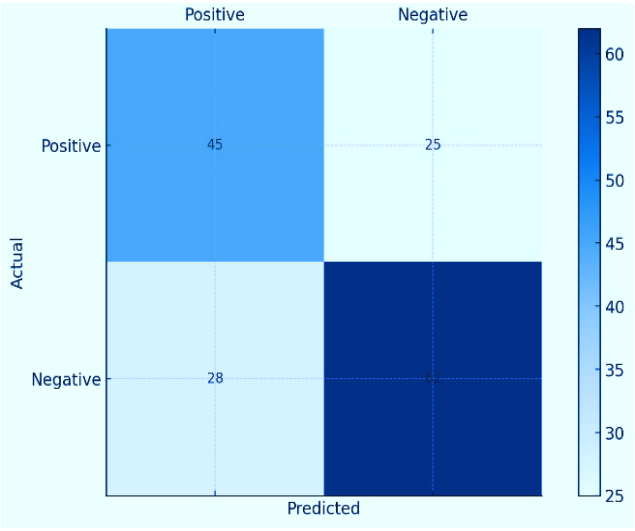
Python | spyder version 6.0.5, this program run on an ASUS computer model: "ROG Zephyrus M16 GU604VI_GU604VI".

In this model, four types of classification selection gates were proposed to handle external distributions. The classification selection gates used to split the probabilities that predicate as output class, the standard gates used which are: Confidence, Entropy, Cosine similarity, and Mahalanobis distance. The model was tested with unseen label data (currency and non-currency) and at each gate calculated values of the test samples and last found suitable threshold to separate the samples from all probabilities to make the correct prediction of the class. Thus, this method helps to get the correct decision of the prediction. The confusion matrix of the predication of proposed model and the predications of the proposed model with each gate (Confidence, Entropy, Cosine similarity, and Mahalanobis distance) can be created from the result of these methods as shown in Figure 6.

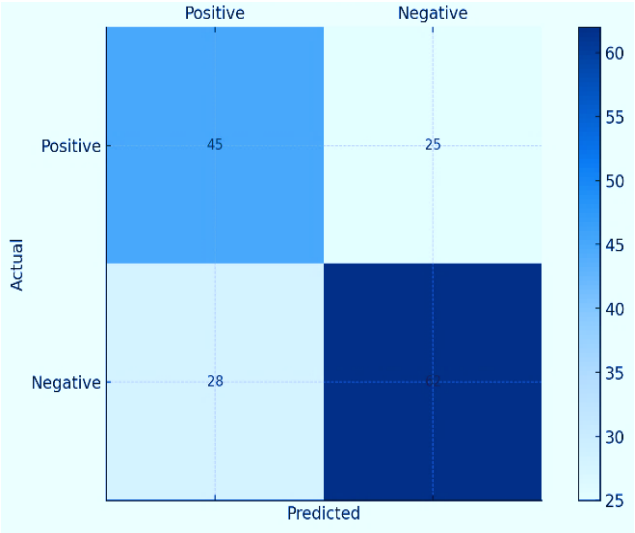
These matrices show the numbers of the correct prediction which represent by (TP,TN) and the number of incorrect predications which are (FP, FN) that get from the unseen data samples that shown in Table 1.



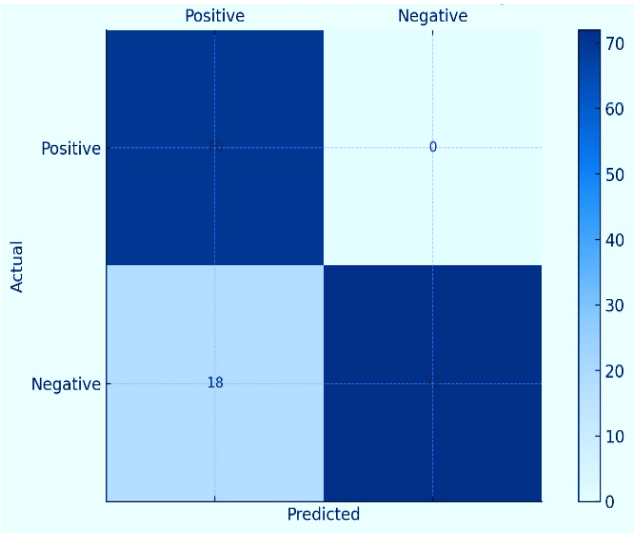
a) confusion matrix of the predication of proposed model



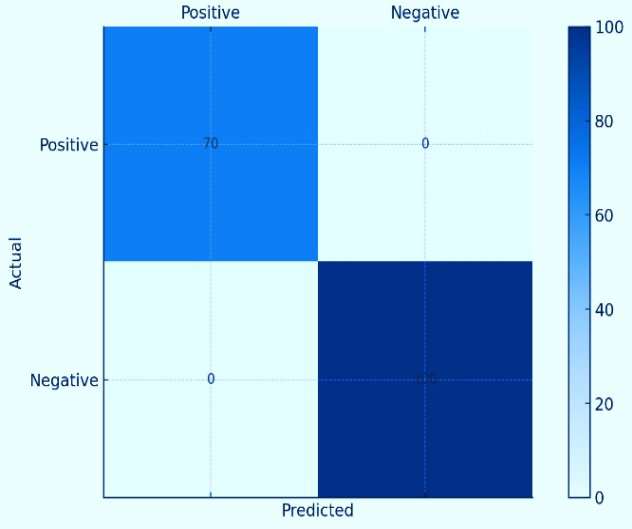
b) confusion matrix of the predication of Confidence



c) confusion matrix of the predication of Entropy



d) confusion matrix of the predication of Cosine Similarity



e) confusion matrix of the predication of Mahalanobis distance

Figure 6: Confusion Matrices of the Predication of Proposed Model and The Predications of Proposed Model with Each Method (Confidence, Entropy, Cosine Similarity, Mahalanobis Distance).

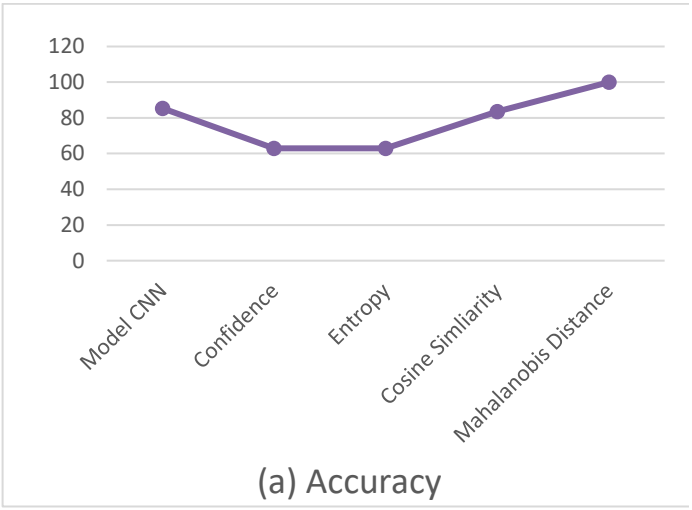
Table 1: The numbers of the correct prediction which represent by (TP,TN) and the number of incorrect predications which are (FP, FN) of the unseen data samples

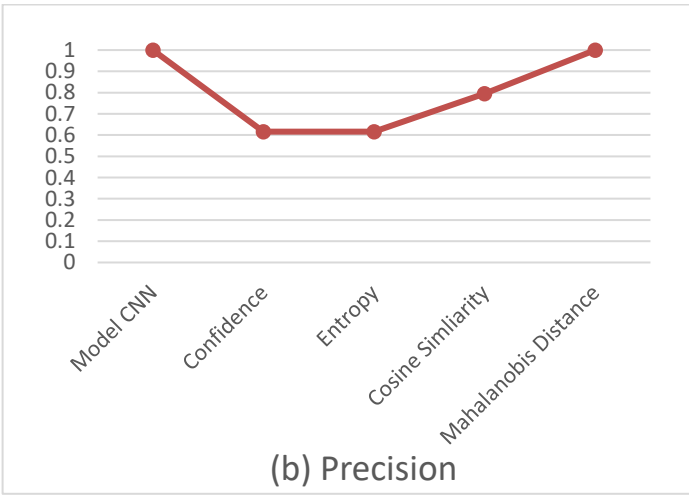
Metrics	Model Predictio-n	Confi-dence	Entr-opy	Cosine Simliarit y	Mahal-anobis
TRUE POSITIVE (TP)	45	45	45	70	70
FALSE NAGTIVE (FP)	25	25	25	0	0
FALSE POSITVE (FP)	0	28	28	18	0
TRUE NEGTIVE (TN)	100	62	62	72	100
SUM POSITIVE	70	70	70	70	70
SUM NEGTIVE	100	100	100	100	100
SUM ALL	170	170	170	170	170

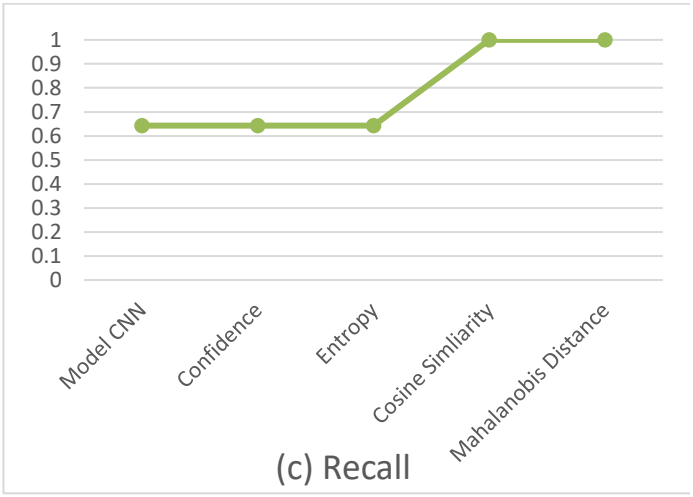
So, the metrics: Accuracy, Precision, Recall, F1-score calculated from the prediction result of each method as shown in Table 2 which display values of these metrices. The metrics (accuracy, precision, Recall, and F1-score) that calculate from the proposed model with each gate can be represented in plots shown in Figure 7.

Table 2: Metrics of Performance of for models and each selective classification gate

Type of Model	ACCURACY	PRECISION	RECALL	F1-SCORE
Original Model CNN	85.294	1.000	0.643	0.783
Confidence	62.941	0.616	0.643	0.629
Entropy	62.941	0.616	0.643	0.629
Cosine Simliarity	83.529	0.795	1.000	0.886
Mahalanob-is Distance	100.000	1.000	1.000	1.000







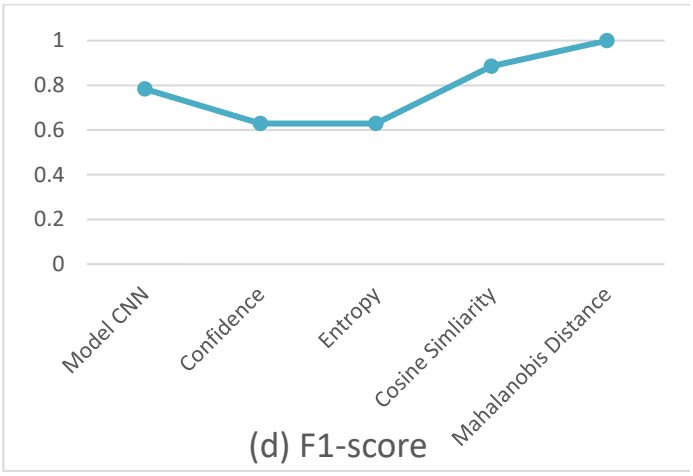


Figure 7: The metrics: Accuracy, Precision, Recall, F1-score calculated from the prediction result of model and model with each method (Confidence, Entropy, Cosine Similarity, Mahalanobis Distance)

In financial Transactions and small dataset its important to measure the Confidence Interval (CL) of the model with each selective gate, to find the confidence intervals of accuracy which can be calculated by use “Confidence-T” function that calculate the “margin of error” of the “mean of accuracy” for the proposed model and the model with each gate, a small value of alpha “ 0.001” using in function to exceed the higher financial risk in vending machine, by this alpha value achieve confidence level 99.9%. in this percentage use “z-value” about “3.29”, z-value noted how many stander deviations away from the mean. the lower (CL) and the upper (CL) can be get for the model with each gate as shown in

Table 3.

Table 3: The Confidence Interval (CL) of the model with each selective gate t

Type of Model	ACCURACY	CL 99.9% [lower, upper]
Original Model CNN	85.294	[76.36 94,23]
Confidence	62.941	[50.75, 75.13]
Entropy	62.941	[50.75, 75.13]
Cosine Simliarity	83.529	[74.17, 92.89]
Mahalanobis-Distance	100.000	[100, 100]

From these metrics notes that the accuracy of the confidence and entropy methods very worse and they weak” moderate” separability because they inherit the overconfidence problem of softmax function in CNN and their upper bounds don’t reach the stronger models means, while CNN looks good under Cosine Similarity and strong separability because it was trained for that exact task, and from CL noted the entropy method and Cosine Similarity have a wide interval, higher uncertainty. While there’s no statistically significant difference between Cosine Similarity and the original model. But for OoD detection, feature-distance metrics like Mahalanobis succeed and achieved high result with near-perfect because they use representation-level signals that better separate in- and out-of-distribution data, separability and it performs perfectly (no uncertainly). Mahalanobis distance leverage internal learned representations, which tend to generalize better and highlight OoD differences. The use of selective classification gates concluded that Mahalanobis distance achieved high prediction of the proposed model compared to other gates with this model , this gate can increase the accuracy of the model successfully and also can get high value of precision, Recall, F1-score, thus Mahalanobis distance considered the most suitable to prevent overfitting of the proposed model in real world.

Conclusions

In this paper attempt to handle Out-of-distribution OoD in a CNN model to make it more general for real data. This model specializes in detecting and

classifying currency papers, this model trains and validates with high accuracy but when used on unseen data, accuracy decreases to “85.29”. So, this requires finding an appropriate method to handle OoD in the model. This method focuses on using selective classification gates; four standard types of gates: “Confidence, Entropy, Cosine similarity, and Mahalanobis distance” and a threshold is found to be a gate of separating the correct prediction of currency classes from other non-currency. From the result of each gate, it has been concluded that Confidence and Entropy do not improve the accuracy of the model; it achieves a low separation by accuracy of 62.941. Also, Cosine similarity gives moderate separation by achieving an accuracy of “83.529”, while Mahalanobis distance achieves a high accuracy of 100.000 and can improve the separation of samples. This prompted us to say that it gave ideal results for processing external distributions of this type of model.

Acknowledgment: The authors would like to thank the university that affiliated with it; “Middle Technical University (www.mtu.edu.iq),” for its support to complete and present this work.

Financial support: No financial support.

Potential conflicts of interest: The authors declare that they have no competing interests.

Author’s Contribution: This research contributes to finding a suitable method to handle Out-of-Distribution OOD in CNN model dedicated to classifying currency by using Selective Classification Gates. In this work tests many kinds of gates and compares results to find the appropriate gate for this type of model, this work gives a special thought to direct future research to run with the most appropriate gate.

References

Alsinayi, N. B. M., Alzuhary, M. H. O. & Gaznayee, H. a. A. Alsinayi, N. B. M., Alzuhary, M. H. O. & Gaznayee, H. a. A. 2025. A Comprehensive Remote Sensing Approach for Assessing Forest Degradation and Environmental Stress in Amedi District, Kurdistan Region of Iraq. *Zanco Journal of Pure and Applied Sciences,* 37**,** 46-69.

Anthony, H. & Kamnitsas, K. On the use of mahalanobis distance for out-of-distribution detection with neural networks for medical imaging. International workshop on uncertainty for safe utilization of machine learning in medical imaging, 2023. Springer, 136-146.

Averly, R. & Chao, W.-L. Unified out-of-distribution detection: A model-specific perspective. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023. 1453-1463.

Azeez, H. N. A. & Hamadameen, A. O. 2024. Solving Stochastic Transportation Programming Problems with Fuzzy Information on Probability Distribution Space Using a New Approach. *Zanco Journal of Pure and Applied Sciences,* 36**,** 61-84.

Bulusu, S., Kailkhura, B., Li, B., Varshney, P. K. & Song, D. 2020. Anomalous example detection in deep learning: A survey. *IEEE Access,* 8**,** 132330-132347.

Chan, R., Rottmann, M. & Gottschalk, H. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. Proceedings of the ieee/cvf international conference on computer vision, 2021. 5128-5137.

Cheng, J. & Vasconcelos, N. Calibrating deep neural networks by pairwise constraints. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022. 13709-13718.

Galil, I., Dabbah, M. & El-Yaniv, R. 2023. A framework for benchmarking class-out-of-distribution detection and its application to imagenet. *arXiv preprint arXiv:2302.11893*.

Hassen, S. & Abdlrazaq, A. 2024. Contextual Deep Semantic Feature Driven Multi-Types Network Intrusion Detection System for IoT-Edge Networks. *Zanco Journal of Pure and Applied Sciences,* 36**,** 132-147.

Hsu, Y.-C., Shen, Y., Jin, H. & Kira, Z. Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020. 10951-10960.

Karunanayake, N., Gunawardena, R., Seneviratne, S. & Chawla, S. 2025. Out-of-distribution data: an acquaintance of adversarial examples-a survey. *ACM Computing Surveys,* 57**,** 1-40.

Liu, W., Wang, X., Owens, J. & Li, Y. 2020. Energy-based out-of-distribution detection. *Advances in neural information processing systems,* 33**,** 21464-21475.

Masaki, A., Nagumo, K., Lamsal, B., Oiwa, K. & Nozawa, A. 2021. Anomaly detection in facial skin temperature using variational autoencoder. *Artificial Life and Robotics,* 26**,** 122-128.

Olber, B., Radlak, K., Popowicz, A., Szczepankiewicz, M. & Chachuła, K. Detection of out-of-distribution samples using binary neuron activation patterns. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023. 3378-3387.

Pham, H., Dai, Z., Xie, Q., Luong, M.-T. & Le, Q. V. 2022. Meta Pseudo Labels. 2020. *arXiv preprint arXiv:2003.10580*.

Podolskiy, A., Lipin, D., Bout, A., Artemova, E. & Piontkovskaya, I. Revisiting mahalanobis distance for transformer-based out-of-domain detection. Proceedings of the AAAI conference on artificial intelligence, 2021. 13675-13682.

Raghuram, J., Chandrasekaran, V., Jha, S. & Banerjee, S. A general framework for detecting anomalous inputs to dnn classifiers. International Conference on Machine Learning, 2021. PMLR, 8764-8775.

Sun, Y., Ming, Y., Zhu, X. & Li, Y. Out-of-distribution detection with deep nearest neighbors. International conference on machine learning, 2022. PMLR, 20827-20840.

Szyc, K., Walkowiak, T. & Maciejewski, H. 2021. Why Out-of-distribution Detection in CNNs Does Not Like Mahalanobis--and What to Use Instead. *arXiv preprint arXiv:2110.07043*.

Techapanurak, E., Suganuma, M. & Okatani, T. Hyperparameter-free out-of-distribution detection using cosine similarity. Proceedings of the Asian conference on computer vision, 2020.

Vareldzhan, G., Yurkov, K. & Ushenin, K. Anomaly detection in image datasets using convolutional neural networks, center loss, and mahalanobis distance. 2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), 2021. IEEE, 0387-0390.

Wang, H., Li, Z., Feng, L. & Zhang, W. Vim: Out-of-distribution with virtual-logit matching. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022. 4921-4930.

Xia, G. & Bouganis, C.-S. Augmenting softmax information for selective classification with out-of-distribution data. Proceedings of the Asian Conference on Computer Vision, 2022. 1995-2012.

Zhu, Y., Chen, Y., Xie, C., Li, X., Zhang, R., Xue, H., Tian, X. & Chen, Y. 2022. Boosting out-of-distribution detection with typical features. *Advances in Neural Information Processing Systems,* 35**,** 20758-20769.