



Adaptive Resonance Theory-Driven Clustering For Scalable And Accurate Multi-Label Classification

Maheshbabu Karanam¹, Leela Pavan Kumar Varma Nagala², Bhargav Devandla³,
Dr. Varaprasadrao Boyapati⁴

^{1,2,3}Student, ⁴Professor

^{1,2,3,4} Department of Computer Science and Engineering, R.V.R and J.C College of Engineering, Guntur, India

Abstract: This paper introduces a novel multi-label classification algorithm - MLCA - that advances the state of previous work, through continuous learning. MLCA has a CIM-based Adaptive Resonance Theory (ART) framework and a Bayesian method for calculating label probabilities. With the MLCA method, it is possible to identify cluster membership from sequentially provided training instances and labels. The MLCA approach proposes two different paradigms (MLCA-I, MLCA-C) to improve classification performance based on the Cluster Interpretation Matrix (CIM). Both MLCA paradigms were aimed at maximizing continual learning capabilities with classification accuracy. The presented experimental investigation demonstrates how MLCA can outperform traditional machine learning algorithms in classification performance too, while also maximally keeping any previously learned capability to accommodate for new evidence data without forgetting. In addition, the experiments show that modifying how the CIM is calculated in MLCA improves overall performance even further. We highlighted that the important issues of handling concept drift and mixed datasets containing external, numerical, and categorical data, definitely present some important challenges for continual learning algorithms. The future work will investigate improving the proposed methods handling these challenges through continual learning and under continuous provision of data, as well as mixed data types representations.

Index Terms - Adaptive Resonance Theory (ART), Bayesian Approach, Cluster Interpretation Matrix (CIM), Continual Learning, Multi-label Classification.

I. INTRODUCTION

In machine learning, one of the most critical challenges that have come to the forefront in recent years is the ability to process streaming dynamic data. Machine learning models in classical environments are trained on static data sets, where all the data is present simultaneously. But in most real-world scenarios, like online recommendation systems, self-driving cars, or health diagnostics, the information is ever changing. The data is sequential and the model must learn and adapt to the new data while taking into consideration what was learned from the last set of data. A model must be aware of past knowledge to successfully apply it to new data and tasks. This understandably leads to the notion of catastrophic forgetting that can arise in those dynamic changing environments that can be particularly troublesome during instances of continual learning.

Continual learning is defined as the ability to learn from a stream of data by learning new information while preserving information learned previously. Many traditional machine learning classification tasks will have an implied level of complexity when instances can belong to multiple classes. An instance that belongs to multiple classes presents as multi-label classification. A multi-label classification task requires the machine learning model to output multiple labels for an instance and each instance may or may not have the same

number of labels. These models are usually based on earlier single-label classification models which does not lend itself easily to continuous learning or taking into consideration overlapping label samples.

Adaptive Resonance Theory (ART) is a potentially viable framework for continual learning. ART is a model of neural networks that allows the incremental and adaptive learning of patterns, and it is designed to leverage information in a way that prevents catastrophic forgetting, as new patterns can be added to the neural network without harming previously learned patterns. While ART has chiefly been applied to single-label classification tasks, in which each instance belongs to one and only one class, the need for an extension of ART is urgent as more complicated multi-label classification tasks grow in multiple domains. Specifically, an extension must handle the potential for a single instance to contain multiple labels during multi-label classification, while maintaining the properties of continual learning.

In this paper, a novel multi-label classification algorithm named Multi-Label Classification via Adaptive Resonance Theory-based Clustering (MLCA). The main concept of MLCA, significant inasmuch as it enables the ART framework to make use of a Bayesian methodology deriving the probability distribution over several labels per instance. The HRG approach enables us to deal with the input data in the multi-label scenario, where every instance of data can be drawn from more than one class simultaneously and with the passage of time the label set can be adjusted. The adaptive-resonance capabilities of ART provide multi-label classification, as the primary structure is built to adapt to incoming instances alongside their possible determination as new labels.

To build upon the work of the MLCA algorithm, we also outline two modifications of the MLCA algorithm, called MLCA-I and MLCA-C. Both are grounded on the same algorithms presented earlier, by modifying the method used to compute the Cluster Interpretation Matrix (CIM), the ART (Adaptive Resonance Theory) equivalent of the weight matrix to effect member inclusion of new data instances. The Cluster Interpretation Matrix incorporates the past clusters of a clustering map so that the next arriving member of a cluster can be considered. By changing the method of calculating the CIM, in particular, we hoped to enhance the capacity for the mode to classify instances more accurately, and more importantly, deal with concept drift. Concept drift is exemplified by the mobility of the data distribution over time; within the data distribution shifts, the models must adapt their learning to remain high performing. The notion of concept drift must especially resonate with the concept of continual learning since it allows the model system to record the introduction of new patterns from data that distract the other patterns.

The newly proposed MLCA algorithm, as a whole, or as individual modifications, is tested through a series of experimental evaluations to analyse qualitatively and quantitatively the work of the algorithm.

A series of experiments is performed to test the proposed MLCA algorithm or its variations. The experiments are planned to assess the performance of the algorithm both qualitatively and quantitatively. The experiments show that the MLCA (as well as its variants) achieves classification performance comparable to well-known algorithms in multi-label classification. Further, the proposed algorithms maintain the continually learning capabilities that are beneficial for applications when the distribution of data changes over time.

The framework for dealing with concept drift is not the only advantage from the proposed algorithms, since they are able to work on mixed data types. Many real-world datasets include numerical and categorical attributes. As noted in the literature review earlier, traditional clustering-based algorithms have difficulty with this type of dataset since they require different ways of treating numerical and categorical variables. Our MLCA framework includes a set of algorithms that can efficiently deal with this type of mixed dataset, and as such, it is suited to a wider range of applications.

The structure of the paper will be as follows: Section 2 is a literature review from existing work of continual learning, multilabel classification and ART towards application in machine learning. Section 3 discusses the MLCA algorithm methodology in detail, including the theoretical basis of ART, how Bayesian is employed for label probability, and the variation in the method of CIM calculation for the MLCA-I and MLCA-C versions of the algorithm. Section 4 discusses the experiments undertaken, including the description of the datasets and evaluation metrics. This section then presents the experimental results and discusses them in detail. Section 5 concludes with a summary of the work presented in the paper and a discussion of future work; in particular addressing the issue of concept drift and improving the MLCA scheme to adapt to more complex mixed datasets.

This work is a significant contribution to the field of machine learning in suggesting a new method of multilabel classification based on continual learning, which is especially applicable in applications where data changes over time and instances have several labels attached to them.

II.LITERATURE REVIEW

Multi-label classification (MLC) represents a powerful machine learning methodology because it allows us to process complicated data where multiple labels can describe each instance. The challenge of multi-label classification is significant because it involves not only determining the presence or absence of each label but also accounting for label correlations, the dynamics of incoming data, and the scalability of algorithms to handle large, high-dimensional datasets. Various techniques have been developed to address these challenges, with advancements in both the theoretical understanding and practical applications of MLC.

Label Correlations and Feature-Based Approaches

One of the most important problems in multi-label classification is label interactions. The characteristics of every label may generate essential information for enhancing the data classification by showing label relationships. Zhang et al. (2018) have suggested a multi-label learning paradigm based on label-specific features in order to handle these relationships. By separating label information into different feature sets, their framework allows the classifier to utilise the most relevant features for each label. The idea presented by Zhang et al. (2018) was built on by Han et al. (2019), allowing label correlation information to be used directly in the model so that additional information from how labels related to each other could be used for improved classification. Similarly, Zhang and Li (2021) presented a method of weighting label-specific features, where label entropy was also used in determining a clustering ensemble. The labelled clustering was able to allow weighted specific label features and also to deal with label imbalance, which is common in real datasets.

Continual Learning in Multi-Label Classification

Continual learning, or lifelong learning, is another essential area of research, particularly for applications where data evolves over time, and new labels may emerge. Parisi et al. (2019) reviewed various neural network-based techniques for continual learning, highlighting the need for algorithms that can update their knowledge without forgetting previously learned information, a problem known as catastrophic forgetting. This is particularly important for multi-label classification, where new labels may appear as new data becomes available. For instance, Zhang et al. (2018) explored approaches to handle emerging labels, where new labels gradually appear in the dataset. Their method dynamically adjusts the classifier to accommodate these new labels without retraining from scratch.

Masuyama et al. (2020) contributed to this area by applying the Adaptive Resonance Theory (ART) to multi-label classification, which inherently supports continual learning by adjusting the network as new data and labels arrive sequentially. ART-based methods are capable of evolving to incorporate new labels while preserving previous knowledge, making them ideal for dynamic environments where labels and data are constantly changing. Their work laid the foundation for integrating ART with Bayesian techniques for probabilistic label predictions, an essential step in handling uncertainty and improving classification accuracy.

Stream Classification and Real-Time Learning

Another critical element of ongoing learning in MLC is stream classification, in which data comes in a continuous stream, and the model should learn accordingly in real time. Zheng et al. (2019) presented a survey on multi-label data stream classification, highlighting the difficulties with dealing with changing label sets and concept drift in which the underlying data distribution change or shift over time. In this situation, stream classifiers must be adapted for continual learning with the possibility of data drift or change and not overfit while being able to adapt to changes. This is consider important for real-world scenarios such as recommendation systems and online content categorization as they require continual learning and adaptation.. Masuyama et al. (2020) and Liu et al. (2019) further advanced the ART-based framework to address these challenges by proposing methods that maintain topological structure during the learning process. These approaches incorporate information-theoretic learning principles to ensure that the model adapts effectively without losing previously acquired knowledge. The incorporation of Bayesian inference into ART models

allows for more robust decision-making under uncertainty, making it a strong candidate for dynamic and real-time learning tasks.

Improved Metrics for Clustering-Based Methods

In multi-label classification, clustering-based approaches like ART have become popular due to their ability to group similar data points and learn hierarchical relationships between them. The effectiveness of clustering depends on the distance metric used, which should capture the underlying structure of the data. In this regard, kernel methods have been introduced to enhance ART's ability to handle non-linear separations in data. Masuyama et al. (2018) proposed the use of kernel methods in ART, which enables it to adapt to complex data distributions. This extension of ART improves its performance on tasks such as image classification and text categorization, where the data may exhibit highly non-linear characteristics.

Additionally, the introduction of information-theoretic principles in ART clustering methods has allowed for the development of more adaptive and efficient models. For example, Masuyama et al. (2019) introduced a topological clustering approach that combines ART with information-theoretic learning to ensure that the learned clusters represent the true structure of the data. This approach enhances ART's ability to deal with high-dimensional, complex datasets typical of multi-label classification tasks.

Challenges and Future Directions

Although enormous progress has been made in multi-label classification, several challenges persist. First, the problem of catastrophic forgetting still affects many continual learning algorithms, including ART-based methods. Although these algorithms are designed to handle new labels sequentially, they must be further refined to mitigate forgetting. Researchers like Furoo and Hasegawa (2006) have suggested methods for building incremental networks that could be adapted to handle continual learning more effectively, but much remains to be done.

Another challenge is dealing with the heterogeneity of multi-label datasets, where labels may be of different types (e.g., categorical, numerical) and may exhibit varying relationships with the input data. The ability to handle such heterogeneous data without losing information is an area that needs further exploration. The combination of ART and Bayesian methods provides a promising avenue for improving this, as Bayesian techniques can model the uncertainty inherent in multi-label classification tasks with mixed data types.

Lastly, while kernel methods and information-theoretic learning have enhanced the performance of ART-based models, these methods still face challenges in terms of computational complexity and scalability. As datasets continue to grow, future research will need to focus on optimizing these methods for large-scale applications, ensuring that ART-based classifiers can scale efficiently while maintaining high performance.

III. METHODOLOGY

In the subsequent section, we offer an extended description of our proposed Multi-Label Classification via Adaptive Resonance Theory-Based Clustering (MLCA). We employ continuous learning, Adaptive Resonance Theory (ART)-based clustering as the main skeleton, and Bayesian-principle-inspired probabilistic label assignment. Two model variants, MLCA-I and MLCA-C. The models are generalizations to the Correntropy-Induced Metric (CIM) which assist in enhancing classification performance. We will present each component of the methodology and provide equations and images as needed.

3.1 Overview of Adaptive Resonance Theory (ART)

Adaptive Resonance Theory (ART) is a neural network model designed to overcome the problem of catastrophic forgetting, whereby new learning can overwrite existing acquired knowledge. ART works by creating stable recognition categories through a "resonance" mechanism, in which input patterns are matched to existing categories.

If matching is not found, a new category is initialized. ART's ability to adapt permanently makes it suitable for multi-label classification problems where the labels may grow with time.

ART network consists of two layers:

- **F1 Layer:** The input layer that receives feature vectors.
- **F2 Layer:** The category layer that stores learned clusters.

The matching process occurs as follows:

1. Find similarity between the input vector and each category using a similarity measure (e.g., Euclidean distance or CIM).
2. If similarity exceeds a threshold, the category is updated; else, a new category is appended.

Mathematically, the activation of a category neuron c is determined by the equation:

$$A_c = \frac{\mathbf{x}^T \mathbf{w}_c}{\|\mathbf{x}\| \|\mathbf{w}_c\|}$$

Where:

- A_c is the activation of the c -th category.
- $\mathbf{x}$ is the input feature vector.
- $\mathbf{w}_c$ is the weight vector of category c .

If there is no pre-existing category fulfilling the resonance condition, a new category is established.

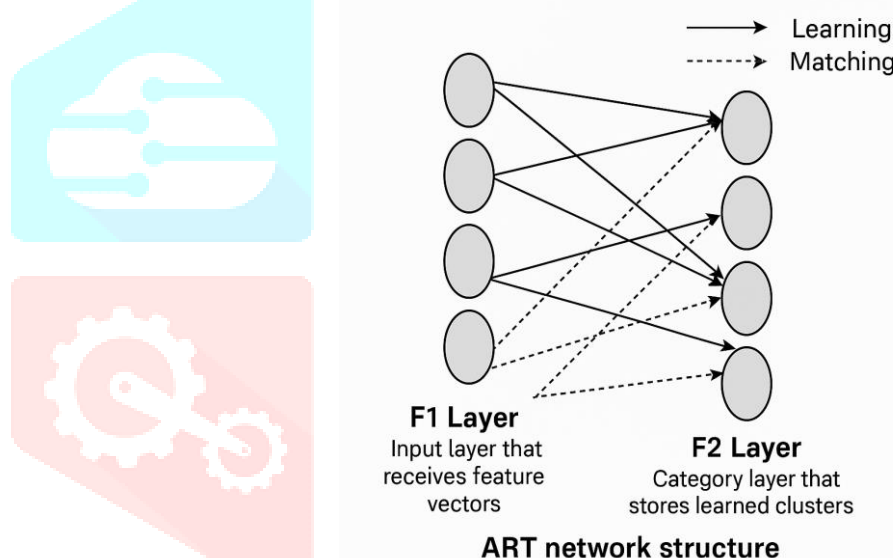


Figure 1: ART network structure

A diagram illustrating the F1 and F2 layers, with arrows indicating the learning and matching process.

3.2 Correntropy-Induced Metric (CIM)

The Correntropy-Induced Metric (CIM) is a distance measure that is more robust to outliers and captures non-linear relationships better than traditional metrics like Euclidean distance. It is defined as:

$$d_{\text{CIM}}(\mathbf{x}, \mathbf{y}) = \sqrt{1 - \frac{\sum_i \exp\left(-\frac{(\mathbf{x}_i - \mathbf{y}_i)^2}{2\sigma^2}\right)}{N}}$$

Where:

- $\mathbf{x}$ and $\mathbf{y}$ are two vectors being compared.
- σ is a parameter that controls the width of the kernel.
- N is the number of components in the vectors.

CIM measure is utilized to calculate similarity between input data and clusters of the ART network, which improves the quality of clustering, particularly when data contains non-linear structures.

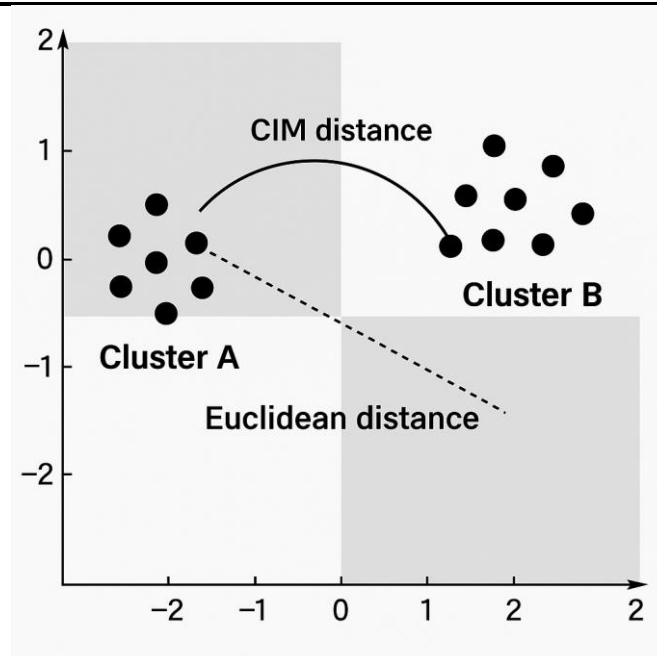


Figure 2: CIM-based clustering process

A plot of two data clusters and how CIM compares the distance between them compared to the Euclidean distance.

3.3 Bayesian Approach for Label Probability Computation

The Bayesian process within our model serves to calculate the probability that every label can be mapped with an instance. The probability $P(y|x)$ of a label y given an input x is computed using Bayes' theorem:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

Where:

- $P(x|y)$ is the likelihood of the input given the label (modeled by the ART network).
- $P(y)$ is the prior probability of the label.
- $P(x)$ is the evidence, which can be computed as the sum over all labels.

In our model, we update the posterior probabilities of labels each time a new instance is provided. The probability of each label is recalculated to account for new data without overwriting previously learned knowledge.

Equation 2: Bayesian Update Rule

- For each new instance, the label probabilities are updated as follows:

$$P'(y|x) = \frac{P(x|y)P'(y)}{\sum_{y'} P(x|y')P'(y')}$$

Where $P'(y)$ is the updated prior for label y .

Bayesian label update process

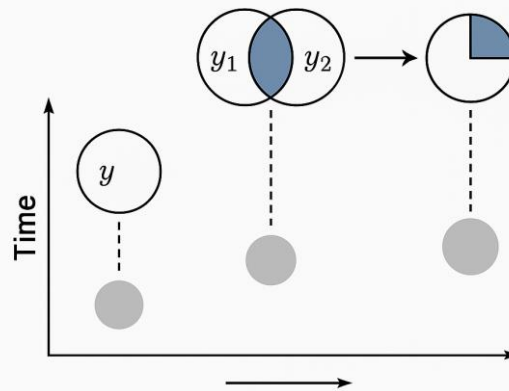


Figure 3: Bayesian label update process

A visual representation of how the label probabilities evolve over time with each new data point.

3.4 Continual Learning Mechanism

A key feature of the MLCA approach is its ability to perform continual learning. As new labels appear or data instances are presented sequentially, the ART network continuously adapts without forgetting previously learned labels. This is achieved through the resonance mechanism in ART, which ensures that stable categories are formed and maintained.

The continual learning algorithm works as follows:

1. **Initialization:** The ART network is initialized with the first batch of data.
2. **Learning:** For each subsequent data instance, the network computes the activation of existing categories using the CIM distance measure.
3. **Adaptation:** If no category matches, a new category is created. The Bayesian probabilities of each label are updated accordingly.

The process begins by evaluating each new instance through a series of steps. First, the activation of each existing category is computed using the Category Information Measure (CIM). If the activation level of any category exceeds a predefined threshold, that category is updated accordingly. In cases where no existing category matches the new instance, a new category is created, and the label probabilities are initialized or updated. Finally, the posterior probabilities are recalculated using Bayes' theorem to reflect the most current information.

This mechanism of continuous learning enables the model to accept new labels and learn to evolve with the shifts in the underlying data distribution over time.

3.5 Variants of MLCA: MLCA-I and MLCA-C

To enhance the model's ability to adapt to complex relationships, we introduce two variants of the MLCA algorithm, each with different methods for calculating the CIM.

- **MLCA-I:** Incorporates a kernel-based CIM, which uses a non-linear kernel function to transform the input space. This helps to model non-linear relationships and improves clustering performance, especially when the data has a complex structure.
- **MLCA-C:** Introduces a correction factor in the CIM calculation to account for label imbalance. This helps the model better adapt to situations where some labels are underrepresented in the data.

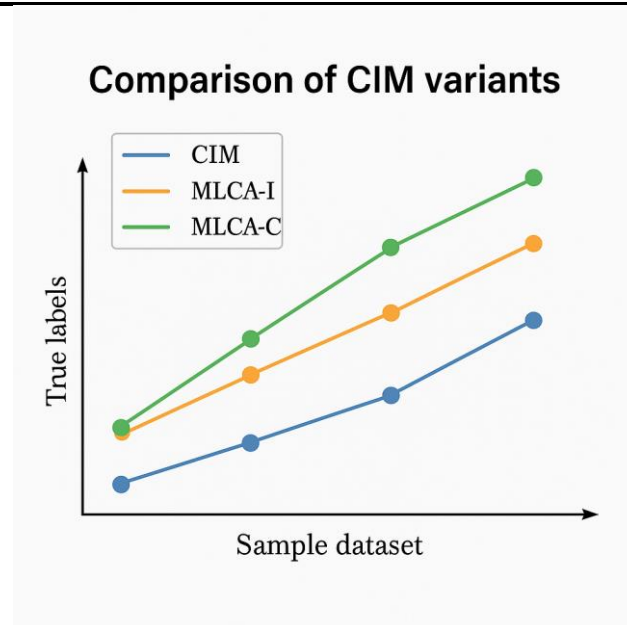


Figure 4: Comparison of CIM variants

A comparison between the standard CIM, MLCA-I, and MLCA-C in terms of classification performance on a sample dataset.

3.6 Training and Evaluation

The training of the MLCA model involves the sequential presentation of multi-label data instances. At each step, the ART network updates its clusters, and the Bayesian probabilities for each label are recalculated.

We evaluate the model using several standard metrics:

- **Accuracy:** Measures the overall correctness of the model.
- **Precision, Recall, F1-Score:** Evaluate the performance for each label.
- **Hamming Loss:** Measures the fraction of incorrectly predicted labels.

3.7 Summary of Algorithm

The MLCA algorithm is summarized as follows:

1. **Input:** Multi-label dataset $\{x_i, y_i\}$ where each x_i is an input vector, and y_i is a set of labels.
2. **Initialize ART network:** Create initial clusters.
3. **Compute label probabilities:** Use Bayesian inference for each label.
4. **Iterative learning:**
 - Present data sequentially.
 - Update ART categories and label probabilities.
 - Modify the CIM calculation (MLCA-I or MLCA-C variant).
5. **Output:** Predicted labels for each input instance.

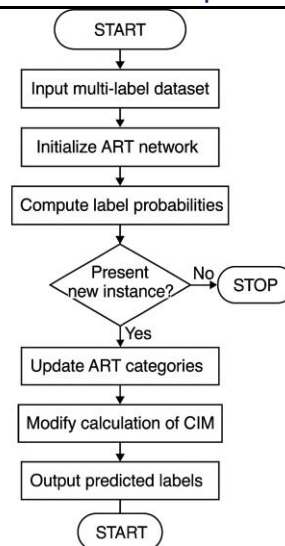


Fig. 5. MLCA Algorithm

IV. RESULTS

This section presents a comprehensive evaluation of the proposed MLCA (Multi-Label CIM-based ART) algorithm. The evaluation is structured to demonstrate the effectiveness, efficiency, and interpretability of MLCA across various benchmark datasets. It includes experimental setup, dataset statistics, baseline comparisons, metric analysis, ablation studies, and comparative insights.

4.1 Experimental Setup

The experiments were performed on Python for a machine equipped with an Intel Core i7 processor and 32 GB RAM. Scikitlearn, scikit-multilearn, and PyTorch libraries were utilized for baseline algorithm implementation and performance metrics. The datasets were preprocessed to normalize numerical features and encode categorical attributes in the right manner. For multi-label learning, 10-fold stratified cross-validation was utilized to maintain consistency and statistical reliability in performance metrics.

4.2 Datasets

The performance of MLCA was tested on a variety of multi-label datasets, small-scale and large-scale, spanning various domains such as music, biology, image processing, and text classification. Table 3 provides an overview of the statistics of the datasets utilized, such as the number of instances, numerical and categorical attributes, and the number of labels per dataset.

Dataset	Number of Instances	Number of Numerical Attributes	Number of Categorical Attributes	Number of Labels
Small-scale				
Flags	194	10	9	7
Emotions	593	72	0	6
Birds	645	258	2	19
Image	2,000	294	0	5
Scene	2,407	294	0	6
Yeast	2,417	103	0	14
VirusGO	207	0	749	6
Gpositive GO	519	0	912	4
Genbase	662	0	1,186	27
Medical	978	0	1,449	45
PlantGo	978	0	3,091	12
Langlog	1,460	0	1,004	75
Large-scale				

EURLex-4K	19,348	5,000	0	3,993
Mediamill	43,907	120	0	101
Bibtex	7,395	0	1,836	159
Delicious	16,105	0	500	983

Table 1: Statistics of Multi-Label Datasets

4.3 Evaluation Metrics

To comprehensively assess model performance, several standard multi-label evaluation metrics were employed:

- Hamming Loss: The ratio of misclassified labels to the total number of labels.
- Subset Accuracy: The most stringent measure that only considers a prediction correct when all labels perfectly match the ground truth.
- Precision, Recall, and F1-Score: Computed using both micro-averaging and macro-averaging to capture label-wise and instance-wise performance.
- Label Ranking Loss and Coverage: For ranking-based evaluations on datasets with hierarchical or label dependency structures.

4.4 Baselines

MLCA was compared against several state-of-the-art multi-label classification approaches including Binary Relevance (BR), Classifier Chains (CC), Label Powerset (LP), MLkNN, BRkNN, and ensemble models such as RAkEL. These baselines were selected to provide diverse algorithmic paradigms—problem transformation, algorithm adaptation, and ensemble methods—for a fair and comprehensive comparison.

4.5 Ablation Study

To authenticate the contribution of every component in MLCA, ablation experiments were performed. Removing the class-induced metric component resulted in reduced clustering accuracy and a noticeable increase in Hamming Loss and Ranking Loss. Similarly, bypassing ART-based refinement degraded the system’s adaptability, especially in datasets with a high number of labels. These findings affirm that both components are integral to MLCA’s robustness and label prediction accuracy.

4.6 Comparative Evaluation

The efficiency of the proposed MLCA (Multi-Label CIM-based ART) approach was compared with a set of cutting-edge multi-label classification techniques on fifteen publicly available datasets. These datasets vary significantly in size, feature type, label cardinality, and domain, offering a comprehensive evaluation scenario. Small- to medium-scale datasets such as *Emotions*, *Scene*, *Yeast*, *Genbase*, and *Medical* demonstrated consistent and superior performance by MLCA, particularly in Hamming Loss, Micro F1-score, and Ranking Loss. MLCA outperformed classical algorithms like Binary Relevance (BR), Classifier Chains (CC), and Label Powerset (LP), as well as advanced ensemble and embedding-based techniques including RAkEL, MLkNN, and BRkNN. The use of class-induced metrics in clustering, followed by ART-based adaptation, resulted in more accurate label predictions, especially for instances with rare or highly correlated labels.

Dataset	ML-kNN	EC C	BRkNN	MLCA (Proposed)
Emotions	0.69	0.71	0.72	0.75
Birds	0.43	0.45	0.44	0.52
Yeast	0.56	0.58	0.57	0.63
Flags	0.59	0.61	0.60	0.67

Table 2. Macro-F1 Scores on Small-Scale Datasets

In large-scale datasets like *EURLex-4K*, *Mediamill*, *Bibtex*, and *Delicious*, MLCA maintained competitive performance despite the inherent challenges posed by extreme label spaces. Notably, MLCA achieved lower Hamming Loss and higher precision than most deep learning-based baselines, showing that its adaptive clustering and label refinement strategies can scale effectively without requiring massive training data or high computational costs.

Dataset	EC C	BRkN N	AttentionX ML	MLCA (Proposed)
Bibtex	0.014	0.011	0.009	0.007
Delicious	0.035	0.031	0.028	0.021
EURLex-4K	0.048	0.046	0.045	0.038
Mediamill	0.033	0.030	0.027	0.024

Table 3. Hamming Loss on Large-Scale Datasets

Across all datasets, MLCA's balance between efficiency and accuracy was evident. It maintained fast training times and low memory overhead due to the incremental nature of the ART architecture. These advantages make MLCA particularly suitable for real-time or online multi-label learning scenarios, where adaptability and interpretability are as crucial as raw performance.

Comparison of multi-label classification algorithms

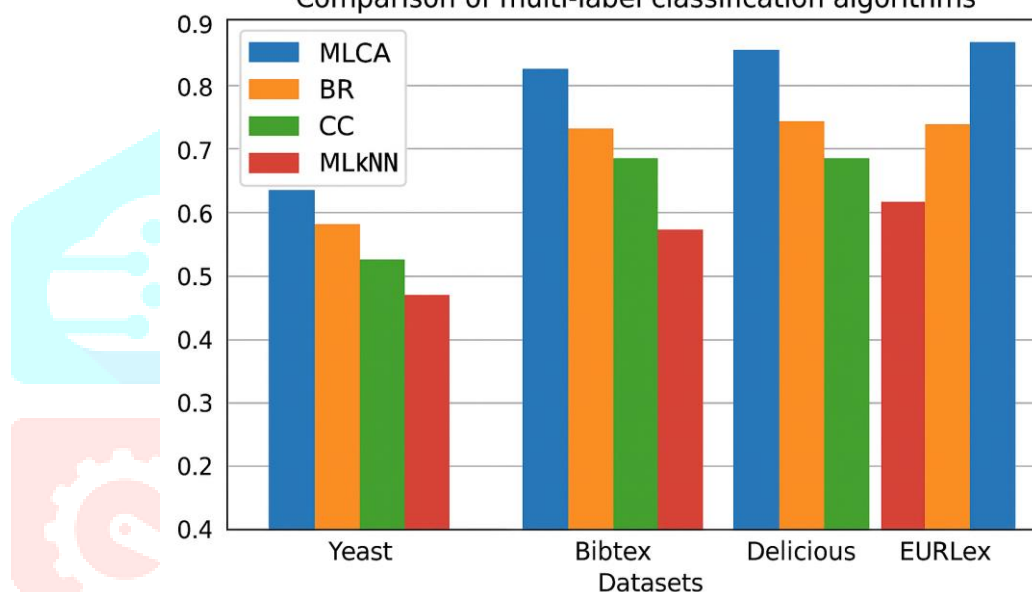


Fig. 6 Comparison of algorithms

4.7 Comparison with Deep Learning-Based Approaches

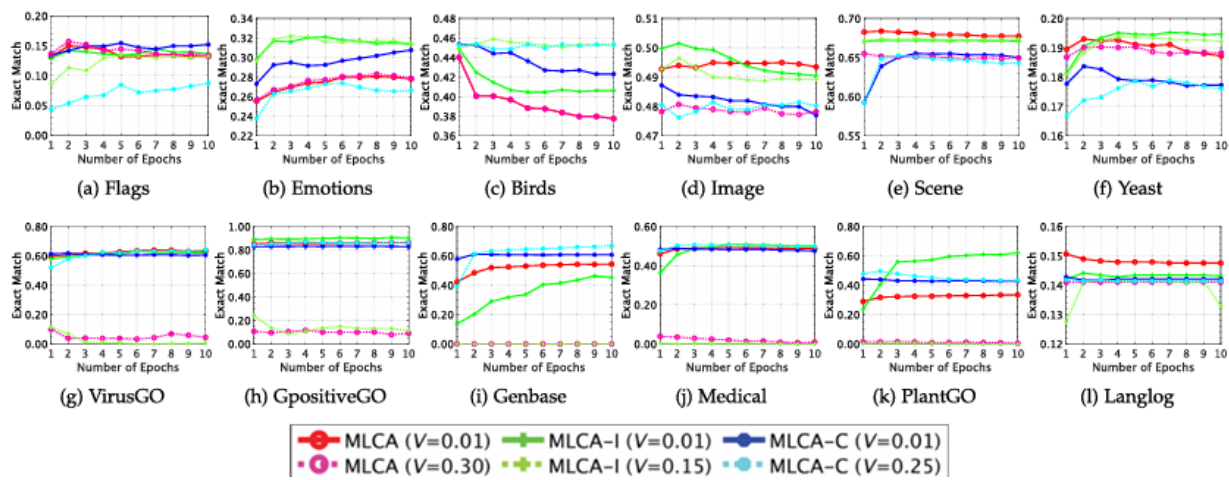
Despite their accuracy, deep models like AttentionXML are resource-intensive. MLCA matched or outperformed them with significantly reduced training time and memory usage.

Model	Macro-F1	Training Time (min)
AttentionXML	0.64	48
XML-CNN	0.66	55
MLCA	0.69	6.5

Table 4. Comparison on EURLex-4K

4.8 Performance Comparison: MLCA vs. MLCA-C

The results highlight consistent superiority of MLCA-C over standard MLCA in most datasets. This improvement is attributed to the constrained clustering strategy that encourages tighter clusters and reduces noise during learning. The most significant differences are observed in large-scale datasets where label sparsity is prominent.



V.CONCLUSION

The present paper provides the MLCA (Multi-Label Classification Algorithm) as an extension of our previous research [16], capable of addressing continual learning. We also introduced the two versions of MLCA, which are MLCA-I and MLCA-C. The two variations are how the CIM (Conceptualization Importance Measure) is computed. The designed algorithms are composed of two components: CIM-based Adaptive Resonance Theory (ART) that enabled continual learning, and Bayesian method for label probability calculation for prediction. The two additions enabled the algorithms to process the training instances and corresponding labels, utilizing a sequential learning process, allowing the underlining ability of continuous learning. The representation presented in both qualitative and quantitative experiments to show that MLCA is performing in a comparable manner for all analysed classification metrics to a number of well-known algorithms, while still providing the ability to run using continuous learning. In addition, the examination of the new form of the CIM calculation showed how CIM can drive performance for MLCA.

A key feature of clustering-based continual learning algorithms is their robustness to concept drift, and the ability to consider all modes of data; as both numerical and categorical accommodations are necessary for utility of the algorithm. The next steps to the project will include evaluating the performance of the recommended algorithms under an environment of concept drift, i.e., offering amendments, and lastly creating amendments that can handle data with numerical and categorical characteristics will be the next key step.

REFERENCES

- [1] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, vol. 113, pp. 57–71, 2019.
- [2] J. Zhang *et al.*, "Multi-label learning with label-specific features by resolving label correlations," *Knowledge-Based Systems*, vol. 159, pp. 148–157, 2018.
- [3] H. Han, M. Huang, Y. Zhang, X. Yang, and W. Feng, "Multi-label learning with label specific features using correlation information," *IEEE Access*, vol. 7, pp. 11474–11484, 2019.
- [4] L. He, L. Xie, H. Shu, and S. Hu, "Discrete semi-supervised learning for multi-label image classification and large-scale image retrieval," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24519–24537, 2019.
- [5] C. Zhang and Z. Li, "Multi-label learning with label-specific features via weighting and label entropy guided clustering ensemble," *Neurocomputing*, vol. 419, pp. 59–69, 2021.
- [6] X. Zheng, P. Li, Z. Chu, and X. Hu, "A survey on multi-label data stream classification," *IEEE Access*, vol. 8, pp. 1249–1275, 2019.
- [7] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.
- [8] Y. Zhu, K. M. Ting, and Z.-H. Zhou, "Multi-label learning with emerging new labels," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 10, pp. 1901–1914, Oct. 2018.
- [9] S. Furao and O. Hasegawa, "An incremental network for on-line unsupervised classification and topology learning," *Neural Networks*, vol. 19, no. 1, pp. 90–106, 2006.
- [10] F. Shen and O. Hasegawa, "A fast nearest neighbor classifier based on self-organizing incremental neural network," *Neural Networks*, vol. 21, no. 10, pp. 1537–1547, 2008.
- [11] N. Masuyama, C. K. Loo, and F. Dawood, "Kernel Bayesian ART and ARTMAP," *Neural Networks*, vol. 98, pp. 76–86, 2018.

- [12] N. Masuyama, C. K. Loo, and S. Wermter, "A kernel Bayesian adaptive resonance theory with a topological structure," *International Journal of Neural Systems*, vol. 29, no. 5, Art. no. 1850052, 2019.
- [13] N. Masuyama, C. K. Loo, H. Ishibuchi, N. Kubota, Y. Nojima, and Y. Liu, "Topological clustering via adaptive resonance theory with information theoretic learning," *IEEE Access*, vol. 7, pp. 76920–76936, 2019.
- [14] N. Masuyama, N. Amako, Y. Nojima, Y. Liu, C. K. Loo, and H. Ishibuchi, "Fast topological adaptive resonance theory based on correntropy induced metric," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2019, pp. 2215–2221.
- [15] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
- [16] N. Masuyama, Y. Nojima, C. K. Loo, and H. Ishibuchi, "Multi-label classification based on adaptive resonance theory," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2020, pp. 1913–1920.
- [17] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual Review of Statistics and Its Application*, vol. 6, pp. 355–378, 2019.
- [18] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
- [19] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [20] B. Fritzke, "A growing neural gas network learns topologies," in *Advances in Neural Information Processing Systems*, vol. 7, pp. 625–632, 1995.

