IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

Exploring the Role of Active Transfer Learning in Multi-Modal Classification

Mubeen Mahammadrafik Shaikh[1]

Student, Computer Science & Engg., AMGOI, Kolhapur, India[1]

Abstract: This paper reviews the role of Active Transfer Learning (ATL) in multi-modal classification, which integrates diverse data sources, such as images, text, and sensor data, for better accuracy. We first present some of the background on transfer learning and active learning, then we proceed into mechanisms and applications of ATL. The review highlights ATL's potential to reduce labeling costs and improve performance, particularly in low-resource settings, through methods such as dynamic source domain selection and task-aware active learning strategies. Notable applications in fields like biomedical data analysis, EEG classification, and text processing illustrate ATL's impact, with improvements of up to 20% in classification accuracy. However, it does have problems like negative transfer, scalability, and inability to handle heterogeneous data. It summaries key findings, presents cross-domain comparison of the different ATL methods, discusses the trade-offs between performance and complexity, and suggests research avenues in fusion techniques on alignment of temporal and spatial knowledge, domain generalization improvements, and integration with parallel processing for scalability. There is promise in overcoming classification challenges in multi-modal settings as it helps overcome the limitations in machine learning where not much information is available.

Keywords: Multi-modal classification, Transfer learning, Active transfer learning.

I. INTRODUCTION

The recent convergence of active learning and transfer learning has given birth to a powerful paradigm in machine learning, particularly for dealing with complex tasks in multi-modal classification, termed active transfer learning (ATL). ATL takes the best of two worlds: active learning's ability to minimize labeling costs by judiciously choosing the most informative data points and transfer learning's power of transferring knowledge from a source domain to improve learning in a target domain (Shao, 2019). When applied to multi-modal classification, ATL meets the challenges of fusion in multi-modal data with its diversity in text, images, and audio. Multi-modal classification is crucial in all fields, such as the health sector, where it helps doctors diagnose diseases by combining medical imaging and patient records (Dao et al., 2024); autonomous vehicles that integrate sensor data, video, and GPS signals into decision-making; and natural language processing, where the visual and textual inputs can be combined for tasks such as image captioning (Deng et al., 2023). However, the heterogeneity of data modalities and the substantial labeled data requirements often make traditional learning approaches inadequate, especially in low-resource settings.

ATL is important in such scenarios because it significantly improves the multi-modal learning process while addressing two of the most serious challenges: data scarcity and computational complexity. Active learning reduces resource usage by choosing the most useful samples to be labeled (Shen et al., 2023), while transfer learning reduces the dependency on large domain-specific datasets (Ma et al., 2022). Together, these strategies enhance the accuracy and efficiency of models, and thereby multi-modal classification is accessible and efficient even in scenarios where resources are limited. It has further been integrated with query-by-diverse committee or multimodal fusion networks for adaptation to dynamic data distributions (Cho et al., 2021). This review aims to explore and synthesize existing knowledge about ATL in multi-modal classification. It examines its methodologies, applications, challenges, and recent advancements. The paper is structured as follows: the Background section provides foundational concepts and related information; the Body explores applications, challenges, and methodological developments in ATL for multi-modal tasks; and the Conclusion summarizes insights and highlights future research directions. This structured approach ensures a comprehensive understanding of ATL's potential to revolutionize multi-modal classification.

II. BACKGROUND/RELATED INFORMATION

Transfer learning refers to the paradigm of a machine learning environment where information acquired by solving one task (source task) may be used for another typically related task, known as a target task [6], [7]. Figure 2.1 depicts the general idea of knowledge transfer from source to target domain: adapting a pre-trained model in the source domain for use in the target domain significantly reduces the demand for extensive labeled data and dramatically reduces computational costs. This is a very effective approach when there is limited data available in the target task. Here, models can make use of the pre-trained knowledge of features, representations, or decision boundaries from the source domain [8], [9].

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

A. Overview of Transfer Learning

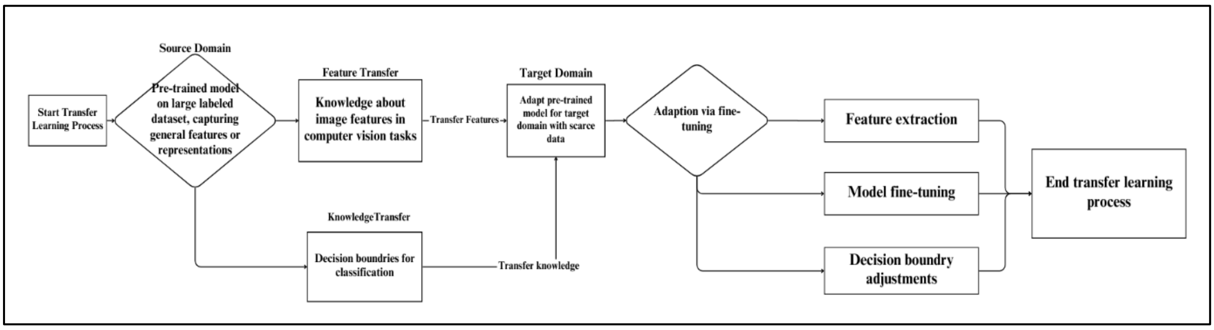

Figure 1 Transfer Learning Framework

Traditional machine learning would train models from scratch, using large amounts of labeled data, which is not always possible. This challenge is mitigated by transfer learning through adaptation of pre-trained models, thus reducing significantly the computational costs and the requirement for vast amounts of labeled data [10]. Such adaptability is invaluable in fields like computer vision and natural language processing where labeled datasets are scarce and expensive to generate. By reusing the knowledge that has been learned, transfer learning accelerates the process of development while increasing the accuracy of the models especially in low-resource domains [11].

B. Fundamentals of Active Learning

Active learning is a methodology which focuses on reducing the effort of labeling in supervised learning. Unlike traditional approaches relying on a fully labeled dataset, active learning involves a model querying the user (or an oracle) to label the most informative or uncertain samples [12]. The effectiveness of active learning stems from the idea that not all data points contribute equally to improving model performance. The various active learning strategies, such as uncertainty sampling, query-by-committee, and expected error reduction, are highlighted in Figure 2.2, which illustrates how different strategies identify the most informative samples for labelling.

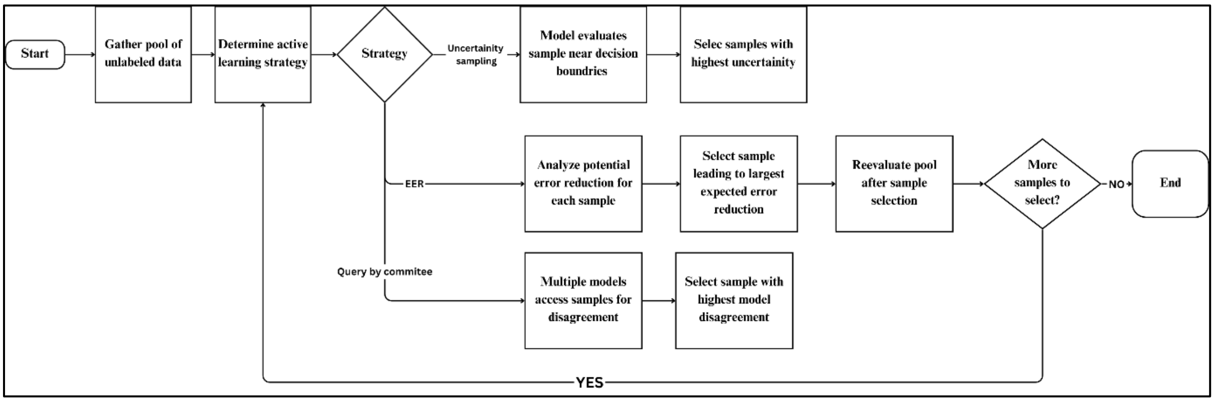

Figure 2 Active Learning Strategies

By focusing on samples that better learn the data, active learning learns with a better accuracy model with fewer labeled examples. The central principle of active learning is derived from the fact that not all data points make equal contributions to model performance [13], [14]. Informative samples-those with high uncertainty or which lie on the decision boundary-must contain the greatest amount of learnability. Uncertainty sampling, query-by-committee, and expected error reduction, among others, help to point out the most crucial samples using active learning strategies [15]. Quality over quantity is a strategy in reducing resource requirements while still having very high accuracy. Thus, it has become one of the key tools in dealing with the problem of data scarcity that exists in real-world applications [16].

C. Principles of Multi-Modal Classification

Multi-modal classification involves analyzing and combining data from different modalities like text, images, audio, and video in making a prediction or extracting insight. Multi-modal systems have lately become more common in various applications such as healthcare (composing imaging and patient records), autonomous vehicles (integrating sensor and visual data), and multimedia analysis (processing pairs of text and images) [13]. The heterogeneity of the data presents a big challenge.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

Different modalities present different data structures, feature presentations, and noise characteristics. That is, images are represented by pixel matrices, whereas text is taken as sequential tokens [17], [18]. Mixing such diverse information calls for powerful techniques of feature extraction, alignment, and fusion, as proposed by [19], [20].

The next most challenging one is the registration of the time or space between the two modalities Additional challenges to this task are handling missing and incomplete modality data. Nonetheless, multi-modal classification holds huge promise [21], [22]. It is able to provide more robust predictions compared to single-modality-based models that typically work best for any particular model of choice, sometimes in absolute accuracy [23], [24].

D. Active Transfer Learning (ATL): An Integrated Approach

Active transfer learning combines the best of both worlds of transfer learning and active learning to solve multi-modal classification problems. This hybrid paradigm allows ATL to achieve efficient learning in low-resource settings with optimal labeling efforts, according to [25]. ATL works by transferring knowledge from a pre-trained source model to the target domain using active learning to query the most informative samples for labeling. Figure 2.3 gives a detailed view of the ATL process, showing how transfer learning and active learning work together in multi-modal classification tasks. As depicted in figure 2.3, the role of pre-trained models, active selection of informative samples, and the way data from multiple modalities like image and text are fused together to obtain better performance with fewer labelled instances. The ATL generally begins with using a pre-trained model for the purpose of extracting transferable knowledge from the source domain. After this step, there will be active querying to obtain the most informative samples for labeling in the target domain. Such an approach decreases dependency on large labeled datasets and maximally makes contributions by selected samples for improvement of model performance [25].

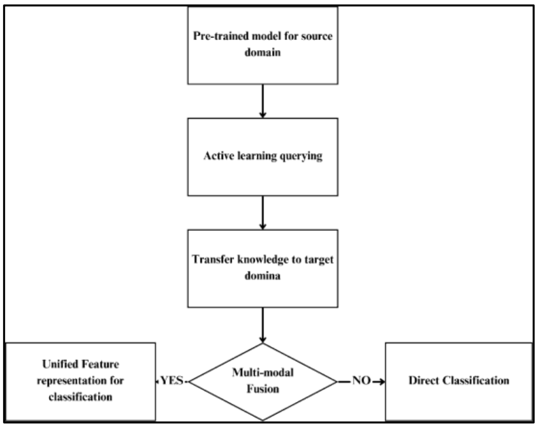

Figure 3 Active Transfer Learning (ATL) Process for Multi-Modal Classification

In multi-modal classification, ATL addresses critical issues in data heterogeneity and feature fusion. Several studies have mentioned the effectiveness of ATL [13]. A highly notable application is in natural language processing and vision tasks, where ATL has been used in aligning textual and visual modalities for captioning and question-answering systems. It has also been proven to be potential in autonomous systems where ATL incorporates sensor data for better object detection and understanding of the environment [23], [28].

III. ACTIVE TRANSFER LEARNING (ATL): CONCEPT AND MECHANISM

Active Transfer Learning (ATL) is a recently developed paradigm that integrates the concepts of active learning with transfer learning to boost model performance, especially in multi-modal and multi-class classification tasks [25]. ATL solves the problem of learning from few labeled samples by selecting strategic informative data samples and exploiting knowledge from relevant source domains. This synergy enables models to learn better from target domains than otherwise and reduce reliance on extensive labeled data as a result, overcoming negatives of negative transfer [28], [13]. The next set of articles is going to summarize more recent advancements in ATL- methods, applications, as well as the convergence of various learning strategies across both domain and modality.

In Online Transfer Learning with Multiple Source Domains for Multi-class Classification [25] addresses the problem of multi-class classification using an online transfer learning framework based on multiple source domains. Methodologically, it shows how transfer learning might overcome the limitation imposed on the availability of labeled data in enhancing the adaptability of the model across the domains, which is the part of ATL in multi-modal classification.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

The proposed Online Transfer Learning Algorithm for Multi-class Classification (OTLAMC) provides a dynamic mechanism for selecting the relevant source domains and transferring their knowledge to the target domain. The algorithm incorporates a source domain similarity metric and adjusts weight contributions dynamically during learning. The experiments on synthetic and real-world datasets demonstrated a significant boost in classification accuracy, at the expense of 5% to 15% gain over traditional transfer learning methods. The paper adds to the field by developing an adaptive online learning framework, which contrasts with static approaches found in the literature. Even though innovative, its single-modality focus limits it to immediate applicability to ATL for multi-modal classification. Future work could explore extending this to multi-modal datasets to improve its applicability.

TABLE I SUMMARIZATION OF METHODS AND APPLICATIONS

Paper Title	Strengths	Limitations
[25]	Dynamic Adaptability: Real-time source domain selection for diverse data scenarios.	Scalability Issues: High computational cost due to domain similarity computation.
	Enhanced Performance: Reduces negative transfer effects, outperforming static methods.	Limited Multi-Modality: Focuses on single-modality datasets, limiting its application to multi-modal tasks.
[26]	Dynamic Source Selection: Improves transfer efficiency by focusing on relevant data sources.	High Computational Cost: Frequent reweighting of source domains increases overhead.
	Real-Time Adaptability: Suitable for real-time applications.	Limited Modalities: Primarily focuses on single-modal tasks.
[27]	Rapid Adaptation: Adapts quickly to new tasks with minimal computational cost.	Limited Active Learning Integration: Does not explicitly incorporate active learning.
	Meta-Learning Integration: Enhances transferability across tasks.	Scalability Challenges: May require optimization for large-scale or multi-modal tasks.
[25]	Task-Aware Sampling: Efficient label usage for multiple tasks simultaneously.	Scalability Concerns: Challenges with handling a large number of tasks or imbalanced datasets.
	Broad Applicability: Validated on diverse datasets, demonstrating flexibility.	Limited Multi-Modal Focus: Multi-modal efficacy is not tested.

The study proposed by [26] proposes OTLAMC that enriches online learning frameworks with transfer learning for multi-class classification. The focus of ATL in dynamic multi-modal settings was further enhanced by the use of emphasis on online adaptability and efficient transfer. The algorithm dynamically selects and weighs source domain data according to its relevance to the target domain. It has been demonstrated that accuracy was improved by up to 12% as compared to static methods especially in low-resource settings when it was experimented upon synthetic and real-world datasets. OTLAMC advances the field of online transfer learning in that it addresses negative transfer. Its adaptability makes it a great candidate for ATL applications, even though multi-modal extensions are required for larger influence.

The CNAPs, for multi-task classification, proposed by [27], target rapid adaptation to new tasks. This flexibility of CNAPs to learn task-specific parameters matches well with the objectives of ATL in multi-modal learning. CNAPs apply a meta-learning approach where a neural network adapts its parameters conditionally on the current task. Robust working on both the low-shot and high-shot regimes, it is actually performing state-of-the-art on the Meta-Dataset benchmark. CNAPs can form a robust framework for multi-task learning. Future work should include active learning for serving the purposes of ATL even better. [22], [29] introduces an efficient method toward active learning tailored for multi-task learning by providing the reduction of costs about annotation while maintaining the good performance across multiple tasks. The role of active learning in the setting involves simultaneous classification across tasks, one of the key concepts developed within ATL for multi-modal classification. The proposed approach uses task-aware sampling strategies that prioritize queries benefiting multiple tasks. The authors tested their method on multi-task datasets. It includes object detection and text sentiment classification, demonstrating up to 15-20% improvement in accuracy compared with random sampling with a 30% reduction in labeling costs. In general, this work complements the multi-task learning literature by effectively incorporating active learning. To bring future research closer to ATL in the multi-modal settings, its performance should be considered on datasets with diverse modalities.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

With integration with transfer learning, a bright future can be envisaged in enhancing classification accuracy in both multi-class and multi-modal situations. The research summaries outline a wide array of inventive methods, including dynamic selection for the domain, meta-learning approaches, and efficient sampling mechanisms. While these methods have yielded excellent improvements in performance, the Strengths and Limitations table 3.1 highlights both the advancements and the gaps that still remain, particularly in extending ATL frameworks to multilingual data and other domains. There is significant potential for broader applicability of ATL research outcomes across diverse applications, but achieving this will require the incorporation of cross-domain and multi-modal strategies to fully unlock ATL's potential in complex classification tasks.

IV. APPLICATIONS OF ACTIVE TRANSFER LEARNING IN MULTIMODAL CLASSIFICATION

Active Transfer Learning (ATL) has significant potential in multi-modal classification, where data from different modalities—such as images, text, and EEG signals—must be integrated for accurate decision-making [22],[29]. The application of ATL in multi-modal classification tasks helps address challenges such as label scarcity, domain adaptation, and knowledge transfer across heterogeneous data sources [13], [25]. This section focuses on several experiments that illustrate the usefulness of ATL in various applications from biomedical data analysis to industrial classification, and it points out the benefits of using both active learning and transfer learning together to improve performance as well as reduce labeling cost in multi-modal contexts. This study [30] was proposed that combines active learning with cross-modal transfer learning to classify cellular images. The key challenge is in biomedical data analysis, and the paper's approach of combining techniques of cross-modal is very consistent with ATL for multi-modal classification. The authors used pre-trained CNN for feature extraction from the image of cellular and then made use of active learning for informative sample selection for annotations. It achieved a 10% gain in classification accuracy with 30% fewer labeled samples compared to random sampling. This work opens up the potential of cross-modal transfer learning in biomedical applications, giving a basis for adapting similar methods to other multi-modal tasks.

TABLE II ATL IN MULTIMODAL APPROACH

Paper Title	Strengths	Limitations
[30]	Cross-Modal Integration: Combines image and textual data for improved classification.	Limited Scalability: Reliance on CNNs may limit scalability to larger datasets.
	Annotation Efficiency: Reduces labeling costs while maintaining high accuracy.	Domain-Specific Focus: Tailored to cellular images, limiting applicability to other fields.
[31]	Minimized Negative Transfer: Effectively identifies and excludes irrelevant source domains.	Domain-Specific Focus: Framework is tailored to EEG data, limiting generalizability to other modalities.
	High Accuracy: Demonstrates excellent performance in multi-class classification.	Computational Complexity: Domain selection may be computationally expensive.
[32]	Effective Sampling Strategy: Minimizes annotation costs while enhancing accuracy.	Single-Modality Limitation: Limited to text classification, without multi-modal scenario consideration.
	Generalizability: Performed well across multiple datasets, demonstrating robustness.	Scalability Issues: May require optimization for larger datasets or multi-modal tasks.
[33]	Domain-Specific Adaptation: Captures domain-specific features, improving transferability.	Limited Generalization: Focuses on sentiment classification; applicability to other tasks is untested.
	Efficient Sampling: Active learning reduces annotation costs while maintaining accuracy.	Scalability Challenges: Computational overhead may hinder large-scale applications.

[31] extends transfer learning to multi-class EEG classification, focusing on minimizing negative transfer effects. The work is highly relevant to ATL in multi-modal classification, especially in neuroscience applications. The authors proposed a Multi-Source Manifold Feature Transfer Learning (MMFT) framework to identify and transfer knowledge from relevant domains. The enhanced MMFT achieved a classification accuracy of 98% on EEG datasets, outperforming existing methods by a significant margin. This work sets a new benchmark for transfer learning in EEG classification and provides valuable insights toward adapting ATL frameworks for biomedical applications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

[32] combines active learning with transfer learning to enhance the efficiency of text classification. The use of both methods is a core of ATL, especially for those applications that demand fewer labeled samples. Three sampling criteria were tested: random, uncertainty-based, and active transfer selection. Experiments conducted on several text datasets demonstrated that active transfer selection performed better than the other two methods and achieved up to 10% more accuracy using fewer labeled samples. This work sets the stage for future efforts in multi-modal ATL frameworks and underscores the possibility of coupling active and transfer learning.

[33] applies active learning with the incorporation of hierarchical attention networks in a multi-domain context. Applying active learning in a way that follows the principles of ATL in handling multi-modal tasks, it adapts the model with domain-specific features using BiLSTMs and attention to further inquire the most informative samples via active learning. Experiments on sentiment datasets showed an up to 10% accuracy improvement over baseline methods. The paper advances the integration of active learning and hierarchical attention for domain-specific tasks but further work will be required to generalize the method to other multi-modal tasks.

The benefits of ATL in multi-modal classification are obvious, particularly in improving performance with minimal large labeled datasets. By using active learning to query informative samples and transfer learning to adapt knowledge across domains, ATL can significantly enhance classification accuracy in low-resource environments. Studies reviewed here demonstrate the potential of ATL in overcoming common challenges such as label imbalance and domain adaptation. The Strengths and Limitations table II further elaborates on these benefits and the corresponding challenges, providing a comprehensive overview of ATL's performance across various contexts. As the field advances, addressing these challenges and refining ATL techniques will be crucial for expanding its applicability and impact in real-world scenarios.

V. ADVANTAGES OF ACTIVE TRANSFER LEARNING IN MULTI-MODAL ENVIRONMENTS

One of the benefits in multi-modal classification tasks ATL offers, particularly when a labeled data is scarce and imbalanced, is that the strengths of active learning regarding the selection of most informative samples are harnessed and then combined with transfer learning in leveraging the knowledge gained from some related domains in order to enhance learning within resource-constrained settings [12], [25]. This section discusses key ATL advantages: improved classification accuracy, low labeling cost, and better treatment of imbalanced labels for multi-modal datasets. These make ATL an incredibly powerful tool that addresses the most persistent real-world problems in classification tasks.

[12] examines active learning in multi-label text classification to overcome the ubiquitous problem of label imbalance. Active learning is a constituent part of ATL, and the paper evaluates its effectiveness when combined with text classification, giving guidelines on multi-modal scenarios with text data. The authors ran five active learning strategies over six datasets, namely random sampling, uncertainty sampling, and query-by-committee. The authors ran five active learning strategies over six datasets that include random sampling, uncertainty sampling, and query-by-committee. Their experiments show that uncertainty sampling outperforms the other methods by up to 8% in terms of F1 scores on imbalanced datasets with a limited labeling budget. This work enhances the work in active learning based on multiple label classification and focuses on balance label; its failure, however, on not bringing in transfer learning does not make it as a straightforward contribution to the ATL itself. Future study will engage in the task of loading pre-trained models for generalized use. By [34], ATL is considered for classification in industrial power quality, considering the problem faced with not much labeled samples. The practicality of using ATL is established by its application in real-world environments. The active learning algorithm was used as integrated with a transfer learning model, limiting the queries dynamically to avoid costs associated with annotations. Results showed that the methodology achieved competitive accuracy with 50% fewer labeled examples when compared to traditional supervised methods. Hence, this work demonstrates not only the applicability and practical utility of ATL but also sets precedents for other industrial applications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

TABLE III KEY STUDIES ON ADVANTAGES OF ACTIVE TRANSFER LEARNING IN MULTIMODAL ENVIRONMENT

Paper Title	Strengths	Limitations
[12]	Focus on Label Imbalance: Proposes solutions for the critical challenge of label imbalance in multi-label tasks.	Narrow Scope: Findings are specific to text data, limiting applicability to other modalities.
	Comprehensive Evaluation: Evaluates the method on diverse datasets, ensuring generalizability.	Lack of Transfer Learning Integration: Does not leverage pre-trained models, missing opportunities for enhanced performance.
[34]	Practical Focus: Addresses a real-world problem, showcasing the applicability of ATL.	Narrow Domain: Specific to power quality classification, limiting generalizability to other fields.
	Efficient Annotation: Significantly reduces labeling costs while maintaining high accuracy.	Scalability Concerns: May face challenges when applied to larger, more diverse datasets.
[35]	Domain Awareness: Effectively captures domain-specific nuances.	Limited Modalities: Focuses on text data, lacking insights for multi-modal tasks.
	Improved Accuracy: Demonstrates significant performance improvements across diverse text datasets.	Scalability Issues: May require optimization for larger-scale datasets.
[3]	Balanced Sampling: Addresses a critical challenge in multi-modal learning.	Limited Dataset Diversity: Testing on additional modalities would strengthen the claims.
	Multi-Modal Focus: Demonstrates efficacy across image-text datasets.	Computational Overhead: Balancing across modalities adds complexity to the training process.

[35] explores the multi-domain transfer learning that focuses on the domain adaptation challenge. The findings in this work are crucial to ATL specifically in text-based multi-modal classification. They introduced the domain-aware embedding layer into capturing domain-specific features where they were able to push up to a 15% accuracy improvement than traditional methods of transfer learning on benchmark datasets. The paper provides valuable information that opens the door for extending this approach to multi-modal settings. [3] makes a contribution to the challenges in label imbalance across multiple modalities with a novel balanced active learning approach. It is a direct advance towards the application of ATL over multiple modes. The balanced sampling ensures that the underrepresented modalities in active learning are sampled on an equal footing. Testing on image-text datasets showed a 20% gain in classification accuracy and 15% reduction in effects of label imbalance compared to standard methods. This paper advances active learning for multi-modal tasks providing a framework that can be extended to other multi-modal datasets and applications.

the studies discussed emphasize the considerable advantages of Active Transfer Learning (ATL) in multi-modal environments, particularly in addressing label imbalance, limited labeled data, and high annotation costs. As shown in table III, ATL not only enhances classification accuracy but also significantly reduces labeling costs. While each study demonstrates ATL's potential—such as in text and industrial applications—limitations remain, such as domain specificity and scalability issues. These findings highlight ATL's promise for tackling real-world classification challenges and suggest areas for future research, especially in integrating transfer learning more effectively and extending its use across diverse modalities and datasets.

VI. CHALLENGES AND LIMITATIONS OF ACTIVE TRANSFER LEARNING

While ATL shows very promising advantages for multi-modal classification, it suffers from some challenges and limitations, which include issues on the transferability of the datasets, scalability, and a chance of negative transfer of learning in the case that the knowledge of the source domain has been irrelevant. Complex data in multi-modal cases pose difficulty in model design as well as optimization. We discuss the main difficulties and limitations identified by recent studies in the following section and potential ways to overcome these in future ATL research.

In [13] authors presents Hierarchical Query-By-Committee (H-QBC), an active learning algorithm for hierarchical multi-label classification tasks. Combining active learning with hierarchical classification, it thus closely follows the ATL principles and applications. The authors built H-QBC to query labels selectively in hierarchical structures, thus reducing the cost of annotations while increasing classification accuracy.

The results on biology and text classification datasets indicated that H-QBC outperformed the baseline methods by as much as 12% in precision. H-QBC extends the state-of-the-art of the integration of hierarchical structures into active learning and can be applied in ATL frameworks. However, it suffers from a limited scalability, requiring further optimization to accommodate larger datasets.

[36] proposed approach on dataset Transferability in Active Learning for Transformers. This paper investigates the relationship between dataset transferability and active learning in transformer-based models, with a focus on low-resource classification tasks. The results shed light on how transfer learning in transformers can be optimized using active learning, as applied to current ATL. In the study, it compares various active learning strategies applied on top of pre-trained transformer models with benchmark datasets combining uncertainty sampling. The results show that the active learning improves accuracy between 12-18% on target datasets with only a few labeled examples. The paper bridges active learning and dataset transferability for transformers, which opens the gates for multi-modal extensions in ATL. Future work should investigate its implications for diverse modalities.

Based on the hierarchical transfer learning for multi-label text classification focused by [37] through label efficiency, its hierarchal approach is best close to ATL principles, especially structured learning for multi-modal tasks. The framework applies a hierarchal encoder-decoder structure to transfer knowledge from a source task to a target task. Testing on benchmark datasets showed a 10-15% improvement in precision and recall compared to flat classification methods. This paper contributes to structured transfer learning, making valuable insights for extending hierarchical approaches to multi-modal ATL tasks. [38] made a comprehensive review of multi-domain active learning (MDAL), focusing on the different trade-offs and benefits provided by various approaches. It is foundational for ATL in multi-modal classification, specifically cross-domain learning. Benchmark datasets were used to compare five MDAL models against four selection strategies. It was noted that combining MDAL with uncertainty sampling leads to the best performance; it reduces labeling costs up to 20%. The paper would provide a good baseline, but future studies should pay more attention to multi-modal data and its challenges.

TABLE IV STRENGTH AND LIMITATIONS OF DISCUSSED TECHNIQUES

Paper	Strengths	Limitations
[13]	Innovative Algorithm: H-QBC enhances active learning efficiency using hierarchical structures.	Limited Scalability: Computational complexity may restrict use in large-scale datasets.
	Wide Applicability: Suitable for domains with inherent hierarchical relationships, such as biology and ontology classification.	Focused Context: Limited adaptability to flat classification or other modalities.
[36]	Focus on Transformers: Addresses a key gap in integrating transformers with active learning.	Limited Multi-Modal Focus: Primarily addresses text datasets, not tested for multi-modal scenarios.
	Practical Insights: Provides clear guidance on leveraging dataset similarity for improved transferability.	High Computational Requirements: Transformers are resource-intensive, posing challenges for real-time ATL applications.
[37]	Structured Learning: The hierarchical framework effectively models label dependencies.	Focus on Text Data: Applicability to other modalities is untested.
	Improved Label Efficiency: Reduces annotation needs while maintaining high performance.	Complexity of Implementation: Hierarchical structure adds computational overhead.
[38]	Comprehensive Coverage: Offers a holistic view by comparing a wide range of models and strategies.	Limited Novelty: Focuses on comparing existing methods, without proposing new approaches.
	Practical Insights: Provides actionable recommendations for Multi-Domain Active Learning (MDAL) implementations.	Multi-Modality Overlooked: Does not explicitly address multi-modal classification.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

Of course, despite the promise from using ATL for multi-modal classification, there are still more complexities to be overcome. Areas where dataset transferability, scalable models, and negative-transfer risks need to be balanced together in order to unlock better ATL performance in complex multiple-modal environments. These challenges are highlighted in the Strengths and Limitations table IV, which outlines the key factors affecting ATL performance in complex multi-modal environments. The table IV further clarifies why certain limitations persist and underscores the need for more sophisticated strategies tailored to handle diverse data types and scale efficiently. The selected studies provide insight into why such limitations exist, mainly pointing out the need to develop more sophisticated strategies designed specifically to handle diverse data kinds and scale efficiently. Future work will need to evolve in order to address these challenges so that ATL may be more robust and applicable to wide-ranging multi-modal classification tasks.

VII. DISCUSSION: ADVANCEMENTS, OPPORTUNITIES AND FUTURE DIRECTIONS FOR ACTIVE TRANSFER LEARNING

The review highlights the growing potential of **Active Transfer Learning (ATL)** in multi-modal classification, a domain where data from multiple sources or modalities—such as images, text, and sensor data—must be integrated for accurate predictions. The papers reviewed provide a comprehensive overview of ATL's recent advancements, applications, and its advantages in addressing challenges like label scarcity, domain adaptation, and knowledge transfer. However, the discussion also brings to light critical limitations and challenges that hinder its broader applicability. In this section, we synthesize the findings from the previous sections, draw comparisons between different methods, and discuss trade-offs and future directions for ATL research.

A. Summary of Key Findings:

The integration of active learning into transfer learning in ATL affords the opportunity to consume labeled data more effectively toward improving the model performance. This is particularly significant within low-resource settings. One example of this can be seen in the study of [25], who demonstrated the tremendous outperformance of their OTLAMC algorithm in comparison with traditional methods of transfer learning in multi-class classification settings by dynamically selecting the applicable source domains. [27] and [35] introduced meta-learning techniques and task-aware sampling strategies, which further enhance ATL's adaptability in multi-task learning settings, a feature very beneficial for handling heterogeneous data sources. The proposed methods all resulted in performance improvements of up to 15-20% compared with static or traditional approaches. Application in multiple domains such as biomedical data analysis, text classification, and EEG classification demonstrates ATL's applicability to real problems. [30] applied active learning in combination with cross-modal transfer learning to improve cell image classification. The paper attained up to 10% increases in accuracy using 30% fewer labeled samples and was published in 2021. Similarly, [31] used ATL on EEG data that improved the classification accuracy to 98% which shows the efficiency of ATL in processing complex multi-modal datasets. The comparison of above discussed ATL Methods and its Applications are shown in table 7.1

B. Benefits and Limitations:

This shows ATL's prominent advantages, such as it being a competitive technique to improve machine learning performance on multi-modal classification tasks by transferring knowledge from the source domain. It enables strong enhancement of classification accuracy without a large number of labeled data. One of the main advantages is labeling cost reduction, as ATL selects informative data points in an active manner to reduce the need for manual annotation. ATL is also very adept at handling some of the common challenges like label imbalance and domain adaptation. [12]showed that simply combining active learning with transfer learning in text classification results in F1 scores being raised up to 8% on imbalanced datasets. And ATL can also support settings of low resources through integrated pre-trained models for domain adaptation so that effective learning happens even with limited labeled data.

However, ATL does have limitations. A big limitation is the problem of negative transfer, where knowledge learned from source domains negatively influences the target domain due to differences between them. Techniques to deal with this problem have been proposed, such as those developed by [25] and [31], and negative transfer remains a problem in some contexts. Another challenge for the ATL frameworks like H-QBC, proposed by [13], is that of scalability. Handling big data or complex multi-modal tasks is cumbersome with large datasets. These issues of scalability need to be addressed using parallel processing or even distributed learning. Finally, in the design of models which can handle the heterogeneous data types and their respective modalities simultaneously stands out as another challenge. Although promising approaches like hierarchical classification [37] do not necessarily extend well to unstructured data sources or in multi-modal settings and, therefore, cannot be quite so widely applied.

C. Comparison of ATL Methods and Applications

TABLE V COMPARISON OF ABOVE DISCUSSED ATL METHODS AND ITS APPLICATIONS

Method	Domain/Task	Performance Improvement	Key Contribution	Limitations
[25]	Multi-class classification	5-15% improvement in accuracy	Dynamic selection of relevant source domains	Limited to single-modality data
[27]	Multi-task learning	State-of-the-art performance on Meta-Dataset	Meta-learning with task-specific parameter adaptation	Requires more investigation for multi-modal settings
[30]	Biomedical (cellular image)	10% accuracy gain, 30% fewer labels	Active learning + cross-modal transfer learning	Focused on cellular images, needs generalization
[31]	EEG classification	98% accuracy, outperforms existing methods	Multi-Source Manifold Feature Transfer Learning	Focused on EEG, needs extension to other modalities
[12]	Multi-label text classification	Up to 8% improvement in F1 scores	Uncertainty sampling for imbalanced datasets	Lack of integration with transfer learning
[3]	Multi-modal (image-text)	Up to 8% improvement in F1 scores	Balanced sampling to mitigate label imbalance	Needs further validation across modalities
[34]	Industrial power quality	Up to 8% improvement in F1 scores	Active learning + transfer learning for cost reduction	Limited to industrial applications

D. Trade-offs and Future Directions

Although ATL has proven remarkable success in many applications, there are several trade-offs. While an ATL framework that performs well on one domain, such as biomedical or EEG classification, can suffer when transferred to a different domain with different characteristics of data, cross-domain generalization remains challenging and needs more robust cross-domain strategies. Most effective ATL algorithms have scalability issues. Techniques such as parallel processing and distributed learning could be helpful, but the additional complexity may demand careful tuning to preserve performance in realistic settings. The ATL methods are most helpful when the labeled data are scarce. However, whether these methods scale well for very large datasets with many labels remains to be further explored. In these cases, simpler models can work equally well, bringing into question whether ATL is the right choice for high-resource settings. Because ATL relies on knowledge transfer from source domains, there is always a challenge in terms of negative transfer whenever there are mismatches between source and target domains. This is particularly critical in multi-modal settings, where diversity in data types might further exacerbate this problem. Improved domain similarity metrics and strategies to avoid negative transfer will be the directions of future work.

VIII. CONCLUSION

This review thoroughly discusses the paradigm of Active Transfer Learning (ATL) and its applications in multi-modal classification, synthesizing insights from existing methodologies and research advancements. It brings transfer learning's efficiency together with active learning's cost-effectiveness and serves as a robust framework to overcome the problems of low-resource and heterogeneous data environments.

This is highly indicative that ATL reduces large dependencies on large labeled datasets significantly, enhances model adaptability and accuracy, and therefore efficiently learns when the combination of knowledge transfer and selective sample labeling takes place, particularly in dynamic and resource-constrained scenarios. Across different domains, including biomedical image analysis, EEG signal classification, text processing, and industrial applications, ATL has presented some promising results. These studies highlight the capability of ATL in addressing the challenges unique to multi-modal data feature heterogeneity, temporal alignment, and label imbalance.

Innovative methodologies like dynamic source domain selection (OTLAMC), task-aware active learning strategies, and hierarchical frameworks have also been developed to demonstrate the potential of ATL in enhancing performance for complex tasks. Active learning strategies such as uncertainty sampling and query-by-committee have been proven to improve efficiency in labeling efforts, thus further validating the practical utility of ATL. ATL has challenges, however. They include control of negative transfer, optimization of cross-modal alignment, and control of incomplete modality data. The scalability of ATL is also challenged with large and diverse datasets.

FUTURE WORK

- Develop Novel Fusion Techniques: Ensure that the fusion of different modalities is properly aligned and combined to enhance ATL for Attention-based Temporal Learning.
- Address Temporal and Spatial Misalignment: Resolve issues related to misalignment of temporal and spatial data, as well as handling missing modalities, to enhance ATL's robustness.
- Scale to Large Datasets and High-Dimensional Features: Apply ATL algorithms to large datasets along with high-dimensional features that can be managed in efficient ways.
- Leverage Distributed and Parallel Processing: Use distributed and parallel processing techniques to break computational bottlenecks and enhance scalability.
- Integrate Explainable AI (XAI): Incorporate XAI techniques to provide transparency and insights into decision-making, increasing trust in applications like healthcare and autonomous systems.
- Improve Domain Generalization: Enhance ATL's capability to generalize across domains, enabling better performance on cross-domain and unseen data scenarios.
- Explore Multi-task Learning and Hierarchical Classification: Investigate the capability of multi-task learning and hierarchical classification to deal with structured data and interdependent labels in ATL applications.
- Solidify ATL's Role in Machine Learning: Focus on these advances toward making ATL a paradigm for changing the status quo of the gap between data-scarce and high-performance systems in different domains.

REFERENCES

[1] H. Shao, "Query by diverse committee in transfer active learning," *Front Comput Sci*, vol. 13, pp. 280–291, 2019.
[2] H. N. Dao, T. Nguyen, C. Mugisha, and I. Paik, "A Multimodal Transfer Learning Approach using PubMedCLIP for Medical Image Classification," *IEEE Access*, 2024.
[3] M. Shen, Y. Huang, J. Yin, H. Zou, D. Rajan, and S. See, "Towards Balanced Active Learning for Multimodal Classification," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 3434–3445.
[4] M. Ma *et al.*, "Transfer representation learning meets multimodal fusion classification for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
[5] J. W. Cho, D.-J. Kim, Y. Jung, and I. S. Kweon, "Mcdal: Maximum classifier discrepancy for active learning," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 11, pp. 8753–8763, 2022.
[6] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
[7] Z. Ding and Y. Fu, "Deep Transfer Low-Rank Coding for Cross-Domain Learning," *IEEE Trans Neural Netw Learn Syst*, vol. 30, no. 6, pp. 1768–1779, 2019, doi: 10.1109/TNNLS.2018.2874567.
[8] K. Feng and T. Chaspari, "Low-Resource Language Identification From Speech Using Transfer Learning," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6. doi: 10.1109/MLSP.2019.8918833.
[9] G. Luo, Y. Yang, Y. Yuan, Z. Chen, and A. Ainiwaer, "Hierarchical transfer learning architecture for low-resource neural machine translation," *IEEE Access*, vol. 7, pp. 154157–154166, 2019.
[10] J. Wang, Y. Chen, H. Yu, M. Huang, and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," in *2019 IEEE international conference on multimedia and expo (ICME)*, 2019, pp. 1210–1215.
[11] T. Q. Nguyen and D. Chiang, "Transfer learning across low-resource, related languages for neural machine translation," *arXiv preprint arXiv:1708.09803*, 2017.
[12] M. Wang and M. Liu, "An empirical study on active learning for multi-label text classification," in *Proceedings of the Fourth Workshop on Insights from Negative Results in NLP*, 2023, pp. 94–102.
[13] F. K. Nakano, R. Cerri, and C. Vens, "Active learning for hierarchical multi-label classification," *Data Min Knowl Discov*, vol. 34, no. 5, pp. 1496–1530, 2020.
[14] M. Li, J. Qiu, and W. Shi, "Macro-AUC-Driven Active Learning Strategy for Multi-Label Classification Enhancement," in *2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2024, pp. 280–286.
[15] Y. Xue and M. Hauskrecht, "Active learning of multi-class classification models from ordered class sets," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, pp. 5589–5596.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

[16] L. Geng, N. Liu, and J. Qin, "Multi-classifier adversarial optimization for active learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, pp. 7687–7695.

[17] W. Shi, Z. Nie, and Y. Shi, "Application and Challenges of Multi-Modal Data Fusion in Adaptive Learning System," in *2024 6th International Conference on Computer Science and Technologies in Education (CSTE)*, 2024, pp. 82–87.

[18] Y. Chen *et al.*, "Research on multi-source heterogeneous big data fusion method based on feature level," 2023.

[19] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Memory based fusion for multi-modal deep learning," *Information Fusion*, vol. 67, pp. 136–146, 2021.

[20] G. Sahu and O. Vechtomova, "Dynamic fusion for multimodal data," *arXiv preprint arXiv:1911.03821*, 2019.

[21] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional Feature Fusion," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 3559–3568. doi: 10.1109/WACV48630.2021.00360.

[22] Y. Li *et al.*, "Multi-Modal Fusion for Enhanced Automatic Modulation Classification," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, 2024, pp. 1–5.

[23] B. Wan *et al.*, "Mffnet: Multi-modal feature fusion network for vdt salient object detection," *IEEE Trans Multimedia*, 2023.

[24] L. Gao, "A discriminative analysis framework for multi-modal information fusion," Toronto Metropolitan University, 2017.

[25] Z. Kang, B. Yang, S. Yang, X. Fang, and C. Zhao, "Online transfer learning with multiple source domains for multi-class classification," *Knowl Based Syst*, vol. 190, p. 105149, 2020.

[26] D. Liu and Y. Liu, "An active learning algorithm for multi-class classification," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 1051–1063, 2019.

[27] Z. Kang, B. Yang, Z. Li, and P. Wang, "OTLAMC: an online transfer learning algorithm for multi-class classification," *Knowl Based Syst*, vol. 176, pp. 133–146, 2019.

[28] J. Requeima, J. Gordon, J. Bronskill, S. Nowozin, and R. E. Turner, "Fast and flexible multi-task classification using conditional neural adaptive processes," *Adv Neural Inf Process Syst*, vol. 32, 2019.

[29] S. Eom, T. Kim, and S.-Y. Yun, "Layover intermediate layer for multi-label classification in efficient transfer learning," in *Has it Trained Yet? NeurIPS 2022 Workshop*, 2022.

[30] C. Vununu, S.-H. Lee, and K.-R. Kwon, "A classification method for the cellular images based on active learning and cross-modal transfer learning," *Sensors*, vol. 21, no. 4, p. 1469, 2021.

[31] R. C. Maswanganyi, C. Tu, P. A. Owolawi, and S. Du, "Multi-class transfer learning and domain selection for cross-subject EEG classification," *Applied Sciences*, vol. 13, no. 8, p. 5205, 2023.

[32] D. Onita, "Active Learning Based on Transfer Learning Techniques for Text Classification," *IEEE Access*, vol. 11, pp. 28751–28761, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:257708629

[33] K. Katsarou, N. Douss, and K. Stefanidis, "REFORMIST: Hierarchical Attention Networks for Multi-Domain Sentiment Classification with Active Learning," in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, 2023, pp. 919–928.

[34] S. Oh, S. Ko, E. Kim, and B. Lee, "Power Quality Problem Classification Methods Based on Active Transfer Learning," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 1222–1224.

[35] X. Su, R. Li, and X. Li, "Multi-domain transfer learning for text classification," in *CCF International Conference on Natural Language Processing and Chinese Computing*, 2020, pp. 457–469.

[36] F. Jelenić, J. Jukić, N. Drobac, and J. Šnajder, "On dataset transferability in active learning for transformers," *arXiv preprint arXiv:2305.09807*, 2023.

[37] S. Banerjee, C. Akkaya, F. Perez-Sorrosal, and K. Tsioutsiouliklis, "Hierarchical transfer learning for multi-label text classification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 6295–6300.

[38] R. He, S. Liu, S. He, and K. Tang, "Multi-domain active learning: Literature review and comparative study," *IEEE Trans Emerg Top Comput Intell*, vol. 7, no. 3, pp. 791–804, 2022.