From Scalability to Sustainability: A 20-Year Retrospective on Deep Learning and Parameter-Efficient Fine-Tuning for Text Classification

Andry Rachmadany [1,2,*], Ika Safitri Windiarti [3]

[1]*Department of Information Technology, Universiti Muhammadiyah Malaysia (UMAM), Padang Besar 021011, Perlis, Malaysia; ika.windiarti@umam.edu.my or ika.windiarti@umpr.ac.id*

[2] *Department of Digital Business, Universitas Muhammadiyah Sidoarjo, Sidoarjo 61271, Jawa Timur, Indonesia, Faculty of Business and Informatics,*

[3] *Universitas Muhammadiyah Palangkaraya, Palangka Raya 73111, Indonesia*

Email to Correspondence: rachmadany@umsida.ac.id or p5250077@student.umam.edu.my

Abstract: *In the area of natural language processing (NLP), especially regarding text classification, earlier methods that relied on traditional machine learning are being increasingly replaced by neural network designs like convolutional neural networks and recurrent neural networks. Additionally, the rise of transformer-based models has led to considerable improvements in performance, though this comes with higher demands for computing power and energy usage. This paper provides a look back at the development of deep learning and Parameter-Efficient Fine-Tuning (PEFT) methods for text classification from 2005 to 2025. The research explores important technological advancements, evaluates the balance between performance, scalability, and efficient computing, and points out the rising concern for sustainability in the development of artificial intelligence. The findings show a transition from strategies aimed at simply increasing scale to those that focus on more efficiency. In this setting, PEFT has become an important advancement in easing the computing load without greatly impacting performance, although it still faces challenges in flexibility and energy consciousness. These insights are anticipated to lay the groundwork for more research into creating environmentally friendly NLP technologies.*

Keywords: Deep Learning, Natural Language Processing (NLP), Text Classification, Parameter-Efficient Fine-Tuning (PEFT), Technology Evolution, Scalability, Computational Efficiency, Sustainability, Green AI

INTRODUCTION

The advancement of artificial intelligence technology, especially deep learning, has experienced major changes over the last twenty years. Regarding natural language processing, this shift is evident in the enhanced proficiency of systems to comprehend and categorize text with greater precision and relevance. Initially, text classification techniques were primarily influenced by conventional machine learning approaches like Naïve Bayes and Support

Vector Machines, which depend significantly on feature engineering and struggle to grasp deep semantic significanc [6] [11].

With the increase in computing power and the availability of data, techniques relying on neural networks have started to take the place of conventional approaches. Architectures like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), along with their variations like Long Short-Term Memory (LSTM), have shown greater effectiveness in representing textual information [3]. Nonetheless, difficulties in managing prolonged dependencies and obstacles related to scalability continue to hinder its application.

A significant transformation took place with the emergence of the transformer model, enabling the modeling of global context via a self-attention method [15]. Models such as BERT and GPT demonstrate notable enhancements in their performance across numerous NLP activities, including the categorization of text [1] [2]. This improvement in performance, however, comes at the cost of substantially larger model architectures and significantly higher computational demands, which in turn introduces concerns related to efficiency and energy usage [14].

Over the past few years, growing concern over computational efficiency and environmental sustainability has encouraged the development of Parameter-Efficient Fine-Tuning (PEFT) approaches, designed to minimize the number of trainable parameters while maintaining competitive performance [4] [5] [9] [10]. This approach is an important step in addressing the scalability challenges that arise from the use of large-scale models.

Although these advancements have contributed considerable progress, knowledge regarding the long-term evolution of this technology is still relatively limited. Existing studies tend to emphasize recent innovations while giving less attention to the historical progression that has influenced the present state of the field. A retrospective perspective is important for recognizing development trends, evaluating trade-offs, and extracting insights that may guide future technological advancements [12].

The main contribution of this study is the presentation of a structured retrospective literature review on the evolution of deep learning in NLP, accompanied by a critical examination of the transition from scale-oriented approaches toward efficiency and sustainability. The results are anticipated to offer a broader understanding of technological advancements over the last two decades and to provide a basis for future research focused on developing more

efficient and sustainable artificial intelligence systems.

METHODOLOGY

To examine the evolution of deep learning and Parameter-Efficient Fine-Tuning (PEFT) techniques in text classification over the period 2005–2025. This approach was chosen because it allows for chronological tracking of technological developments while providing flexibility in interpreting paradigm shifts in the field of artificial intelligence. Unlike systematic literature reviews, which emphasize rigorous and quantitative study selection, a narrative approach is more appropriate for understanding the dynamics of technological evolution and the relationships between conceptual developments that occur over time [8] [13].

The data sources for this study were obtained from various reputable scientific databases, such as Scopus, Web of Science, IEEE Xplore, ScienceDirect, and ACL Anthology. The search process was conducted using a combination of keywords relevant to the research topic, including text classification, deep learning (NLP), transformer models, BERT, and various parameter-efficient fine-tuning techniques such as LoRA, adapter tuning, and prompt tuning. The selected literature included journal articles and international proceedings published between 2005 and 2025, with a focus on research discussing the development of text classification models, deep learning architectures, and computational efficiency. Studies that were irrelevant, opinionated without empirical basis, or unrelated to the context of NLP and model efficiency were excluded from the analysis.

The analysis process was conducted in stages, combining chronological mapping and thematic synthesis. Initially, the literature was grouped by time period to identify stages of technological development, from classical machine learning approaches to the emergence of transformer models and PEFT techniques. Next, thematic synthesis was conducted to group the research based on its primary focus, such as model architecture, text classification approaches, and aspects of efficiency and sustainability. The analysis then continues by comparing these various approaches to identify trade-offs between performance, scalability, and computational efficiency that have occurred throughout the technology's evolution.

This study uses a retrospective approach to understand patterns of technological development and the factors driving these changes. By reviewing key milestones in the evolution of deep learning, this study seeks to identify the relationship between improvements in model

performance and the consequences for complexity and resource consumption. This approach does not aim to predict the future, but rather to provide a deeper understanding of how historical developments shape the current state of technology.

To ensure the validity and reliability of the study's results, this study applies source triangulation techniques by comparing findings from various relevant literature. Furthermore, only publications from reputable sources are used to ensure the quality of the analysis. The synthesis process is conducted systematically to minimize subjective bias in interpretation, allowing the research results to provide a comprehensive and reliable picture of the evolution of deep learning in text classification.

1. Evolution of Deep Learning in Text Classification

The development of text classification in the last two decades has shown a significant transformation, starting from traditional machine learning-based approaches to large-scale deep learning models. In the early period around 2005–2010, text classification methods were dominated by algorithms such as Naïve Bayes and Support Vector Machines (SVM), which relied on feature representations based on bag-of-words or term frequency-inverse document frequency (TF-IDF). These approaches were quite effective for simple classification tasks, but had limitations in capturing context and semantic relationships between words in depth [6] [14].

Entering the 2010–2015 period, the development of neural networks began to provide a more powerful alternative for modeling text data. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), especially Long Short-Term Memory (LSTM), are able to automatically learn text representations without relying entirely on manual features [3]. These models improve their ability to capture local patterns and sequential dependencies in text, although they still face challenges in terms of computational efficiency and limitations in optimally handling long-term dependencies.

The most significant transformation occurred in the 2017–2022 period with the introduction of the transformer architecture which utilizes self-attention mechanisms to understand context globally [15]. Models like BERT and GPT show very significant performance improvements in a variety of NLP tasks, including text classification [1]. The main advantage of transformers lies in their ability to process information in parallel and capture complex relationships in text. However, this increased performance comes at the cost of a significantly increased number of model parameters, drastically increasing

computational requirements, energy consumption, and training costs [14].

As model complexity increases, the 2022–2025 period is marked by the emergence of Parameter-Efficient Fine-Tuning (PEFT) techniques as a solution to overcome efficiency limitations. Approaches such as adapters [4], Low-Rank Adaptation (LoRA) [5], prefix tuning [10], and prompt tuning [9] allows fine-tuning to be performed by training only a small number of additional parameters. This significantly reduces computational and memory requirements while maintaining competitive model performance.

Overall, the evolution of deep learning in text classification shows a shift from feature-engineering-based approaches to more complex and automated representation learning. However, the performance gains achieved through model scaling also come at the cost of increased resource requirements. In this context, the emergence of PEFT techniques marks the beginning of a shift toward more efficient approaches. However, most existing methods still focus on parameter efficiency and have not fully integrated aspects of adaptivity or energy efficiency [12].

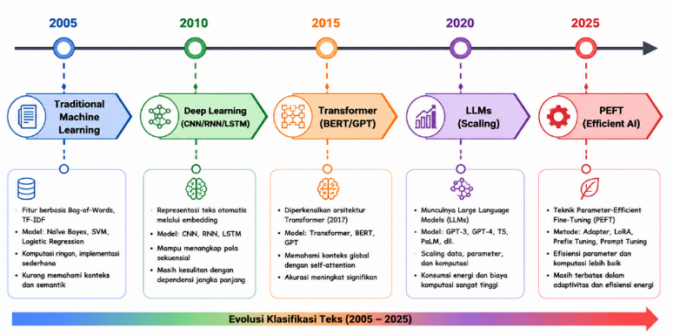

Figure 1. Timeline Deep Learning

Thus, the evolutionary journey over the past two decades not only demonstrates technological progress but also reveals trade-offs between performance, scalability, and efficiency. Understanding these evolutionary patterns is crucial as a basis for analyzing comparisons between approaches and drawing lessons, which will be discussed in the following sections.

RESULT AND DISCUSSION

Comparative Analysis and Trade-offs

Developments in text classification over the past two decades have demonstrated a paradigm shift that involves not only improving model performance but also a trade-off between accuracy, scalability, and computational efficiency. Each generation of technology brings specific advantages, but also limitations that encourage new approaches.

Traditional machine learning approaches, such as Naive Bayes and Support Vector Machines, have a primary advantage in computational efficiency and ease of

implementation. These models are relatively lightweight and do not require extensive resources, making them suitable for a variety of computationally limited applications [6]. However, the model performance is highly dependent on the quality of manually designed features, making it less capable of capturing complex semantic contexts.

The shift to CNN- and RNN-based deep learning has brought significant improvements in data representation capabilities. These models are able to automatically learn features and more effectively capture patterns in text [7]. However, this performance increase is accompanied by increased computational requirements and training time, and there are still limitations in optimally handling long-term dependencies (Hochreiter & Schmidhuber, 1997).

Table 1. Comparison of Text Classification Evolution

Period	**Main Approach**	**Model**	**Advantages**	**Disadvantages**
2005–2010	Machine Learning	Naïve Bayes, SVM	Light computing, simple implementation	Depends on feature engineering, lacks understanding of context
2010–2015	Deep Learning	CNN, RNN, LSTM	Automatic text representation, capable of capturing sequential patterns	More expensive training, long dependency difficulties
2017–2022	Transformer & Large Language Models	Transformer, BERT, GPT	High accuracy, understanding global context	Very large parameters, high energy consumption
2022–2025	Parameter-Efficient Fine-Tuning (PEFT)	Adapter, LoRA, Prefix Tuning, Prompt Tuning	Better parameter and computational efficiency	Still limited in adaptability and energy efficiency

The emergence of the transformer architecture marked a major leap in NLP model performance. With self-attention mechanisms, models like BERT and GPT are able to understand context more comprehensively and achieve significantly higher performance than previous approaches [1] [2]. However, these advantages come at a very high cost in terms of the number of parameters, energy consumption, and computing infrastructure requirements [14]. This raises new issues regarding scalability and sustainability, especially in the context of widespread model use.

In response to these challenges, the Parameter-Efficient Fine-Tuning (PEFT) technique emerged as a compromise between performance and efficiency. By training only a small number of additional parameters, this approach can maintain the performance of large models while significantly reducing computational requirements [4] [5]. However, this approach still has limitations, especially in terms of adaptability to various task conditions and the lack of explicit integration of energy efficiency.

Overall, there is a consistent pattern of trade-offs in the evolution of text classification technology. Early approaches offered high efficiency but limited performance, while newer approaches offer high performance at a significant computational cost. PEFT emerged as an attempt to balance these two aspects, but it has not yet fully addressed sustainability concerns.

This analysis shows that improving model performance cannot be separated from the consequences of increased complexity and resource consumption. Therefore, future technological developments need to consider the balance between accuracy, efficiency, and environmental impact. Understanding these trade-offs provides an important foundation for drawing lessons from technological developments over the past two decades, which will be discussed in the next section.

Lessons Learned from Deep Learning Evolution

An analysis of the development of text classification over the past two decades provides several important lessons that not only reflect technological progress but also reveal the patterns and consequences of each stage of evolution. One key lesson is that improved model performance is often achieved through increasing complexity and scale. The transformation from traditional machine learning to deep learning, to transformer-based models, shows that higher accuracy generally comes at a higher computational cost [1] [15].

However, this pattern also reveals the limitations of scaling-based approaches. Reliance on models with a very large number

of parameters leads to increased energy consumption and carbon emissions, ultimately posing sustainability challenges in artificial intelligence development [12] [14]. This shows that performance improvements cannot continue to rely on increasing scale without considering efficiency and environmental impact.

Another important lesson is the shift from feature-engineering-based approaches to automated representation learning. Deep learning allows models to directly learn features from data, reducing reliance on manual processes [7]. However, this convenience also comes at the cost of increased model complexity and resource requirements, creating a trade-off between ease of use and computational efficiency.

The emergence of Parameter-Efficient Fine-Tuning (PEFT) techniques provides insight into how efficiency can be achieved without significantly sacrificing performance. This approach demonstrates that not all parameters in a model need to be adjusted to achieve optimal results [4] [5]. However, the lesson to be learned is that parameter efficiency alone is not enough, as most approaches still do not explicitly consider adaptivity and energy efficiency.

Furthermore, technological evolution also demonstrates the importance of balancing innovation and sustainability. Developments that focus too much on performance without considering long-term impacts can create new problems, such as high operational costs and limited access to technology. Therefore, future AI development needs to consider a more holistic approach, emphasizing not only accuracy but also efficiency and environmental responsibility.

Overall, the lessons learned from the evolution of deep learning over the past 20 years demonstrate that technological progress is determined not only by performance improvements, but also by the ability to effectively manage complexity and resources. Understanding these lessons is crucial as a foundation for formulating a more efficient, adaptive, and sustainable direction for future technology development.

CONCLUSION

This paper presents a retrospective literature review of the evolution of deep learning in text classification over the period 2005–2025, focusing on the development of model architectures and Parameter-Efficient Fine-Tuning (PEFT) techniques. The study shows that technological development is moving from traditional machine learning-based approaches to increasingly complex and large-scale deep learning models, which significantly improve performance in various NLP tasks.

However, these performance improvements come at the cost of increased computational requirements, energy consumption, and model complexity. A comparative analysis reveals a consistent trade-off between performance, scalability, and computational efficiency. In this context, the emergence of the PEFT technique is a crucial step in reducing the computational burden without significantly sacrificing performance, although it still has limitations in terms of adaptability and energy efficiency.

Furthermore, a retrospective analysis reveals that the evolution of deep learning is driven not only by performance improvements but also by the need to manage resources more efficiently. The shift from a scale-up-based approach to efficiency demonstrates that sustainability is becoming a critical factor in the development of artificial intelligence technology.

Overall, this study provides a comprehensive understanding of the development of text classification technology over the past two decades and identifies key lessons that can form the basis for the development of more efficient and sustainable AI. These findings are expected to support further research in designing NLP systems that not only excel in performance but also are responsible in terms of resource use and environmental impact.

REFERENCES

[1]. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

[2]. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

[3]. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[4]. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., Laroussilhe, Q. de, Gesmundo, A., Attariyan, M., & Gelly, S. (2019). *Parameter-Efficient Transfer Learning for NLP* (arXiv:1902.00751). arXiv. https://doi.org/10.48550/arXiv.1902.00751

[5]. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *LoRA: Low-Rank Adaptation of Large Language Models* (arXiv:2106.09685). arXiv. https://doi.org/10.48550/arXiv.2106.09685

[6]. Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. In C. Nédellec & C. Rouveirol (Eds.), *Machine Learning: ECML-98* (Vol. 1398, pp. 137–142). Springer

Berlin Heidelberg. https://doi.org/10.1007/BFb0026683

[7]. Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. https://doi.org/10.3115/v1/D14-1181

[8]. Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, *51*(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

[9]. Lester, B., Al-Rfou, R., & Constant, N. (2021). *The Power of Scale for Parameter-Efficient Prompt Tuning* (arXiv:2104.08691). arXiv. https://doi.org/10.48550/arXiv.2104.08691

[10]. Li, X. L., & Liang, P. (2021). *Prefix-Tuning: Optimizing Continuous Prompts for Generation* (arXiv:2101.00190). arXiv. https://doi.org/10.48550/arXiv.2101.00190

[11]. McCallum, A., & Nigam, K. (2019). *A Comparison of Event Models for Naive Bayes Text Classification*.

[12]. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54–63. https://doi.org/10.1145/3381831

[13]. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

[14]. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for*

Computational Linguistics, 3645–3650. https://doi.org/10.18653/v1/P19-1355

[15]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relation- ships that could be construed as a potential conflict of interest.

Article History:

Received: 06 March 2026 | Accepted: 27 April 2026 | Published: 30 April 2026