systems

MDPI

Article

PAI-NET: Retrieval-Augmented Generation Patent Network Using Prior Art Information

Kyung-Yul Lee and Juho Bai *

College of Economics and Business, Hankuk University of Foreign Studies, 81, Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin-si 17035, Gyeonggi-do, Republic of Korea; kyungyul.lee@hufs.ac.kr
* Correspondence: juho@hufs.ac.kr

Abstract: Similar patent document retrieval is an essential task that reduces the scope of patent claimants' searches, and numerous studies have attempted to provide automated patent search services. Recently, Retrieval-Augmented Generation (RAG) based on generative language models has emerged as an excellent method for accessing and utilizing patent knowledge environments. RAG-based patent search services offer enhanced retrieval ranking performance as AI search services by providing document knowledge similar to queries. However, achieving optimal similarity-based document ranking in search services remains a challenging task, as search methods based on document similarity do not adequately address the characteristics of patent documents. Unlike general document retrieval, the similarity of patent documents must take into account prior art relationships. To address this issue, we propose PAI-NET, a deep neural network for computing patent document similarities by incorporating expert knowledge of prior art relationships. We demonstrate that our proposed method outperforms current state-of-the-art models in patent document classification tasks through semantic distance evaluation on the USPD and KPRIS datasets. PAI-NET presents similar document candidates, demonstrating a superior patent search performance improvement of 15% over state-of-the-art methods.

Keywords: generative artificial intelligence; knowledge database retrieval system; document similarity; semantic search

check for updates

Academic Editor: Halil Zaim

Received: 25 January 2025
Revised: 21 March 2025
Accepted: 25 March 2025
Published: 7 April 2025

Citation: Lee, K.-Y.; Bai, J. PAI-NET: Retrieval-Augmented Generation Patent Network Using Prior Art Information. *Systems* **2025**, *13*, 259. https://doi.org/10.3390/systems13040259

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Knowledge retrieval systems in database-driven environments prioritize the delivery of query-relevant information. Within the patent domain, the integration of deep learning methods has accelerated the evolution of similar patent search capabilities. Although GPT-style decoder-only transformers demonstrate robust query responsiveness, they remain susceptible to hallucination. Retrieval-Augmented Generation (RAG) systems have emerged as a preferred solution in knowledge-based expert systems, effectively mitigating hallucination while maintaining high-quality knowledge services. These systems continue to advance, leveraging patent databases to identify query-relevant documents and generate appropriate responses based on validated source material.

For patent document retrieval systems, selecting patent document candidates relevant to queries is crucial for service quality. Traditional patent search systems focus on applying deep learning models to patent classification systems based on registration criteria. Primarily, encoding model-type deep learning methods are being applied to classification tasks for automating similar patent search processes. Patent document classification methods based on Convolutional Neural Networks (CNN), Transformer–Encoder-based

patent document classification methods, and patent classification methods utilizing hierarchical structures for large-scale patent documents have been proposed. Recent deep learning methods demonstrate superior performance in patent classification due to patent documents' typical textual document information characteristics. Despite the performance achieved through deep learning in patent classification tasks, similar patent searches still require expert knowledge for similarity evaluation tasks. RAG systems select similar document candidates through similarity evaluation. While cosine similarity is typically used for general similarity assessment, showing high similarity for content with similar structure and content, patent document similarity evaluation must consider technical relationships with prior art, which fundamentally differs from general document similarity evaluation. For example, an Apple touch interface patent (US 7,469,381) may be technically related to an older Microsoft scrolling patent (US 5,495,566) despite minimal textual overlap due to their functional equivalence, while having high textual similarity with contemporaneous Apple patents in unrelated technical domains. Although image classification methods have been proposed to reduce similarity distances between similar information for robust query handling, similarity distance learning methods are not effectively utilized in document classification due to rare significant differences in document representation similarity. Meanwhile, sematic search tasks are characterized by the difficulty of determining semantic similarity between documents based solely on document class information. Experts traditionally use ontological models like synonym dictionaries for similarity evaluation. However, ontological models present challenges in continuously expanding environments like patent information due to the ongoing expert resource update costs.

This paper proposes a novel model for selecting appropriate candidates for RAG-based patent document knowledge retrieval services. We utilize prior art relationship information, bypassing the need for expert-developed knowledge structures to reflect expert knowledge in similarity evaluations. Similar Patent Search Network using Prior Art Information (PAI-NET) is a contrastive learning-based multi-tasking model that applies similarity to prior arts in classification task models. Figure 1 illustrates the architecture of our Semantic Patent Knowledge RAG (Retrieval-Augmented Generation) system, which emphasizes semantic similarity analysis between priority art documents. The system comprises a retrieval agent, PAI-NET, prompt engineering module, and Large Language Model (LLM), with a particular focus on analyzing deep relationships between prior art documents.

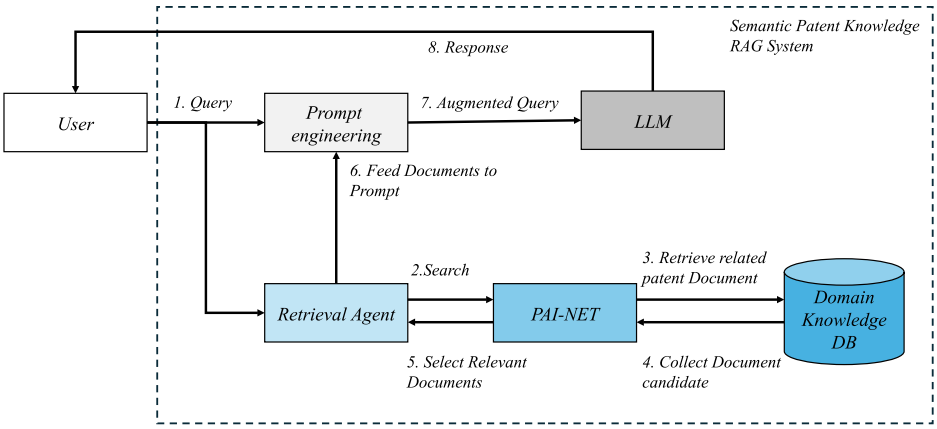


Figure 1. Concept of the RAG (Retrieval-Augmented Generative) system for patent knowledge querying.

The core innovation lies in the interactions between steps 2–5, where the retrieval agent and PAI-NET collaborate to establish semantic relationships between prior art documents. When PAI-NET searches for prior art documents in the Domain Knowledge DB, it

goes beyond simple keyword matching to understand the semantic connections between documents. Specifically, it calculates similarities by considering relationships between patent claims and specifications, citation networks, and technical field hierarchies. This approach ensures that retrieved prior art documents form an interconnected knowledge network rather than a collection of isolated documents.

The core innovation lies in the interactions between steps 2–5, where the retrieval agent and PAI-NET collaborate to establish semantic relationships between prior art documents. In step 2, the retrieval agent transforms the user query into an embedding representation that captures the technical essence of the patent information being sought. During step 3, PAI-NET executes a semantic search in the Domain Knowledge DB, applying specialized embeddings that prioritize technical relationships over superficial textual similarities. In step 4, the Domain Knowledge DB returns candidate patent documents based on these specialized search parameters, providing the raw material for technical analysis. Finally, in step 5, PAI-NET evaluates and ranks the retrieved documents by computing similarity scores that specifically reflect technical relevance rather than textual similarity. This comprehensive approach ensures that retrieved prior art documents form an interconnected knowledge network rather than a collection of isolated documents.

The Domain Knowledge DB functions as a knowledge graph that encodes complex relationships between prior art documents, rather than serving as a simple document repository. PAI-NET leverages this structured information to identify the most relevant set of prior art documents for a given query. This process not only considers technical features of patents but also filing dates, citation relationships, and technological evolution patterns within the field. Before being passed to the LLM (large language model) through the prompt engineering module (steps 6–7), the retrieved documents are structured according to their semantic relevance. This enables the LLM to comprehend the complex relationships between prior art documents and generate more accurate and contextually appropriate responses (step 8).

Figure 2 depicts the core components and process flow of the PAI-NET-based patent RAG system. This system is fundamentally divided into two phases: indexing and retrieval. In the indexing phase, each document from the patent document database is processed through the PAI-NET model. Documents are transformed from textual data into embedded tokens, reflecting the semantic characteristics of the documents. The converted embeddings are indexed in a latent space, where prior art relationship information is incorporated such that technically similar documents are positioned in proximity to one another. In the retrieval phase, user query prompts are converted into embeddings through the identical PAI-NET model. These generated query embeddings are utilized for similarity searching, calculating distances between the query and previously indexed document embeddings in the latent space. Rather than relying on simple document representation matching, the search is based on the distance-based technical similarity inherent in patents, with particular emphasis on similarity measurements that reflect prior art relationships. Similarity measurement in the latent space differs from conventional document similarity approaches by considering the technical relationships between patent documents and prior art information. This overcomes the limitations of ontology-based models and significantly enhances the accuracy and relevance of patent searches.

This architecture overcomes the limitations of traditional RAG systems in patent search applications. By performing sophisticated analysis of semantic similarities between prior art documents, it provides more comprehensive and accurate patent search results. Furthermore, this deep document analysis significantly contributes to the thorough identification of relevant prior art documents during patent examination processes. In this process, our model applies a method of generating similar document groups based on

prior art information between patents to extract query-relevant candidates. When these similar document candidates are presented, the service is provided by inputting prompts combining queries and candidate documents into generative language models to compose appropriate responses from candidate documents.

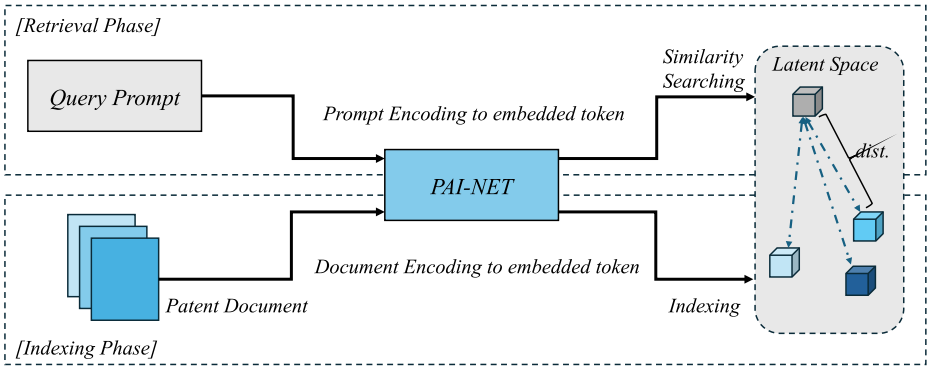


Figure 2. Indexing and retrieval processes using PAI-NET.

Our main contributions are summarized as follows:

- PAI-NET is a novel model for similar patent search that improves document similarity evaluation performance by incorporating expert knowledge into similarity metrics.
- We demonstrate that prior art information can enhance similarity search performance without independent manual ontology construction.
- PAI-NET performs both classification and similarity learning tasks while maintaining computational costs comparable to traditional classification-only models.
- We analyze and evaluate PAI-NET through extensive experiments on real patent datasets, demonstrating significant performance improvements in similar patent search tasks.

2. Related Work

2.1. Applying Retrieval-Augmented Generation to Expert Domains

Domain adaptation of Retrieval-Augmented Generation (RAG) for expert knowledge areas has emerged as a critical research direction, as conventional LLMs often struggle with specialized knowledge. Early approaches focused on joint training of retriever and generator components to adapt to domain-specific knowledge bases, showing notable improvements in specialized areas such as healthcare and news domains [1]. Building on this foundation, researchers have developed more sophisticated frameworks like PEER [2], which systematizes domain-specific tasks through precise question decomposition and advanced information retrieval while maintaining data privacy and cost efficiency. Despite these advances, such prompt engineering-based approaches primarily focus on improving question-answering performance and are insufficient for domains like patent retrieval, where implicit technical relationships with prior art must be considered.

The challenge of professional knowledge services has led to innovative approaches such as Knowledge-Augmented Generation (KAG), which addresses the limitations of traditional RAG systems by incorporating knowledge graphs and enhancing logical reasoning capabilities. This approach has demonstrated significant improvements in professional domains across multiple benchmark datasets; however, it has limitations in that it can only be used in expert domains where ontologies are clearly structured, as it is designed to construct systems based on existing knowledge graphs [3]. This constraint fails to address the existing issues in patent systems, which are large-scale document management environments with continuously registered new documents, when considering the costs

of constructing and maintaining ontologies. For specialized domains with unique characteristics, such as Korean Medicine, researchers have developed Prompt-RAG, a vector embedding-free approach that has shown superior performance in terms of relevance and informativeness compared to conventional RAG models [4]. While Prompt-RAG can provide information without constructing ontologies by utilizing document table of contents information, it can only be used in document environments configured as table of contents and content sets, making it difficult to apply directly to patent documents structured in the abstract-claims-description format.

Recent research has also focused on specific challenges in handling tabular data within expert domains. The Tabular Embedding Model (TEM) has been developed to address the limitations of standard embedding models when dealing with complex numerical and tabular data, providing a more efficient solution for domain-specific applications [5]. Additionally, studies have shown that combining domain-specific fine-tuning with iterative reasoning mechanisms can significantly enhance question-answering accuracy in professional contexts [6]. While there have been studies [7,8] showing that RAG systems outperform general-purpose AI systems like GPT-4 in patent search applications, there remains a persistent demand for systems that better reflect the unique similarity characteristics of patent knowledge information.

2.2. Automated Patent Classification Methods

Patent documents explicitly include classification information for relevant fields during both the examination process and registration. Due to this characteristic, classification categories have traditionally been used to select candidate lists when searching for related patents for new patent registrations. Since accurately classifying the category of a query makes it easy to identify appropriate candidates, deep learning research has also primarily focused on automating patent document retrieval systems by utilizing text classification techniques. Among the various deep learning studies, models that show superior performance in text classification tasks are based on the Transformer method [9–14]. Deep learning approaches are also making valuable progress in classifying patent documents. DeepPatent [15] uses word2vector embedding methods [16] and CNN methods at the sentence level [17]. This method has the advantage of applying visual representation methods to the features of words. The RNN-based method [18] utilizes a bidirectional recurrent model [19], which performs document classification tasks considering the flow of words. Meanwhile, PatentBERT [20] is designed to take into account essential parts of the input document using the self-attention technique [9,10], which can be weighted to important keywords to reflect the interests of the document. Hierarchical Feature Extraction Model (HFEM) [21] and a Multi-Stage Extraction Network for Patent Document Classification (MEXN) [22] adopts a method of dealing with the entire document as a way of extracting and summarizing the features of the document portion hierarchically. Using hierarchical models is a helpful method for considering whole data. However, it is unlikely that the evaluation will be reversed in the next step for the undervalued part in the previous step. The value of the input part may not be adequately evaluated depending on the size of the input range considered in the first step. Meanwhile, there are studies that classify various forms of drawing images [23,24] and studies that find drawing images in patent documents [25] at the individual classification level. Although deep learning research on patent documents has primarily focused on classification tasks, a challenge persists: expert intervention is still required to evaluate similarities among classified patent documents in a multi-label classification environment. Our study introduces a RAG-based patent similarity search that incorporates prior art considerations. This approach aims to reduce the need

for expert involvement by integrating prior art relationships into similarity metrics while maintaining effective classification performance.

2.3. Document Similarity

Latent semantic indexing (LSI) is one of the classic methods for document retrieval [26]. Search using LSI is a method of mapping the contents of a document to a latent space and searching for similar documents using euclidean operations between mapped features. Unlike methods with keywords as queries, these methods return results with close euclidean distances even if there are no accurately matched values. In contrast, features with close distances in potential space have the advantage of returning documents with similar meanings even if there are no identical keywords in real documents. Using these points, we can also perform nearest neighbor queries on similar documents using similarity metrics, such as cosine similarity, to provide results for nearby candidates, even in latent spaces, where we project the features of documents using deep learning models. In this case, similar distance measurements between similar documents can be used as a measure of service quality [27], as high-similarity documents must be kept at a close distance. This approach can be particularly effective in measuring patent document similarity, where technical concepts and innovations may be described using different terminology while maintaining similar underlying meanings. The ability to capture semantic relationships in the latent space makes it especially valuable for patent analysis and retrieval tasks.

2.4. Contrastive Learning Approaches

Studies of contrastive learning methods on image information [28] have been very meaningful in creating robust predictive models in the field until recently. Contrastive learning methods are metric learning methods that learn by comparing similar data, which can be mainly utilized in self-supervised learning. They have been applied in a variety of approaches.

In this regard, studies applying these methods in NLP have also been tried [29–32], which is difficult to distinguish compared to image information because of its different meaning. Relevant studies use fine-tuning methods using pre-trained models to compensate for these problems also seek to improve performance by following this trend. In particular, contrast learning methods utilizing similar pairs of documents are an important topic in this work. They can reflect users' perspectives on similarity in screening similar documents in more detail within the prediction category of classification tasks. In this regard, recent work [31] proposes a method to reduce similarity distances based on classification predictions. However, in our work, we propose a model more suitable for real-world similar search environments by leveraging information from the prior art that reflects direct patent expert knowledge.

While there are patent document classification studies using various deep learning methods, methods to yield candidates for similar documents in patent domains are still challenging. The problem is that patent documents have an imbalanced data distribution between labels in a multi-label environment. In addition, the meaning of patent similarity can be applied differently from the similarity of general-purpose documents. We consider the characteristics for similarity evaluation. Our work aims to design a robust deep learning model that can improve classification task performance and recommend highly similar document candidates even in the imbalanced dataset domain [33].

3. PAI-NET: RAG Patent Network Using Prior Art Information

RAG Patent network using Prior Art Information (PAI-NET) is designed for similar patent search services. The PAI-NET framework has a multi-tasking structure that learns the

similarity distance between documents and classification tasks together. To this end, PAI-NET has a conjured triple-encoder structure for multi-tasking tasks and utilizes objective functions for similarity distance learning tasks and classification tasks.

Figure 3 shows how the proposed method utilizes prior art information. Figure 3a shows the relationship between the prior art document and the claimed patent (target document) as domain knowledge, the most technically similar relationship with 1 to 1 matching. Meanwhile, Figure 3b shows how to generate traditional ontological structures. In Figure 3b, in addition to prior art relationships, patent experts leverage domain knowledge to further apply the linkages between each document to generate a searchable graph structure. On the other hand, in Figure 3c, the proposed method combines non-similar patent document information with the prior art pair. It uses it to consider the similarity distance between the prior art and the target patent document relatively in the learning task.

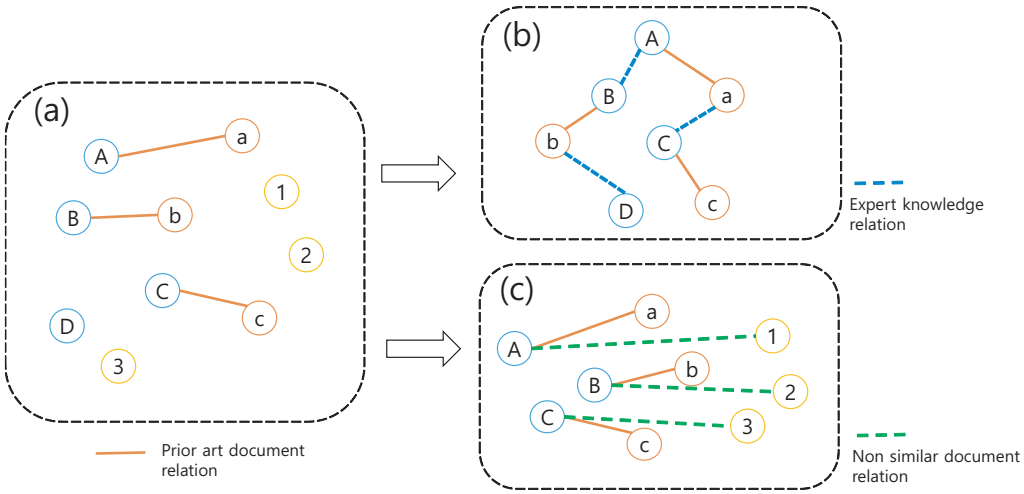


Figure 3. Concept of patent document connections using domain knowledge information relationships. Capital letters (A, B, C, D) represent target patent documents, lowercase letters (a, b, c) indicate prior art documents, and numbers (1, 2, 3) denote non-similar patent documents: (**a**) Prior art and target document relationships. (**b**) Traditional ontology structure with expert knowledge. (**c**) Contrastive triple pair with prior art and non-similar document.

This approach can be illustrated through a real-world example involving Apple's "bounce-back" effect patent (US 7,469,381), its family patent US 7,864,163 (covering double-tap zoom functionality), and its prior art, Microsoft's scrolling patent (US 5,495,566). As shown in Figure 3c, PAI-NET does not simply learn the relationship between a target patent (A) and its prior art patent (P); it also takes non-prior art patent documents (N) into consideration. This represents a significant innovation in learning patent document similarities. Family patents US 7,469,381 and US 7,864,163 share the same priority application (January 2007) and use similar terminology and expressions, thus showing high similarity even in text-based similarity measurements. In contrast, Apple's patent and Microsoft's prior art patent both concern scrolling functionality, but due to their more than 10-year difference in filing dates and different writing styles, they would be evaluated as distant based on text-based similarity metrics. PAI-NET accurately reflects the technical relevance of prior art in the embedding space despite these differences in document representation. PAI-NET's contrastive learning approach simultaneously considers three document relationships, as depicted in Figure 3c. It learns prior art relationships (A-P) as positive pairs while ensuring distance from non-prior art patent documents (N). This approach relatively evaluates the similarity distance between patents A and P, enabling effective patent similarity learning without the complex expert knowledge graphs required

by traditional ontological structures (Figure 3b). By leveraging prior art relationships directly, PAI-NET offers a solution to the cost challenges associated with constructing and maintaining comprehensive ontological structures, which typically require significant expert resources and continuous updates to remain relevant in the evolving patent landscape.

Our framework uses the set S of three different documents in parallel in one input step. The three document inputs are used as anchors S_a, positives S_p, and negatives S_n, respectively. Each document is transformed into a set of embedding tokens through an embedding process [10]. The encoder $\mathcal{F}$ uses three conjured Transformer encoders that share weights, and the embedded document set S_i entered is converted into document feature set h_i of h_a, h_p, h_n through the encoder corresponding to each location. Each transformed document feature h_a, h_p, h_n is used for classification tasks and similarity distance learning tasks. Document features that pass through the anchor encoder are used as inputs to the classifier $\mathcal{G}$ for classification objective function computing. The anchor document feature h_a is also used for similarity distance learning, where it is used for objective function computation with the positive document feature h_p and the negative document feature h_n. We design our framework without increasing computational time cost compared to existing single classification tasks using a parallel encoder batch process and share the weights of each encoder so that the encoders can focus on achieving the goal of the objective function.

Figure 4 and the following equations represent the overall process:

$$h_i = \mathcal{F}(S_i), \quad S_i \in \{S_{ai}, S_{pi}, S_{ni}\}, \tag{1}$$

$$v = \mathcal{G}(h_{ai}), \quad h_i \in \{h_{ai}, h_{pi}, h_{ni}\}, \tag{2}$$

$$p = Softmax(W_v v + b_v). \tag{3}$$

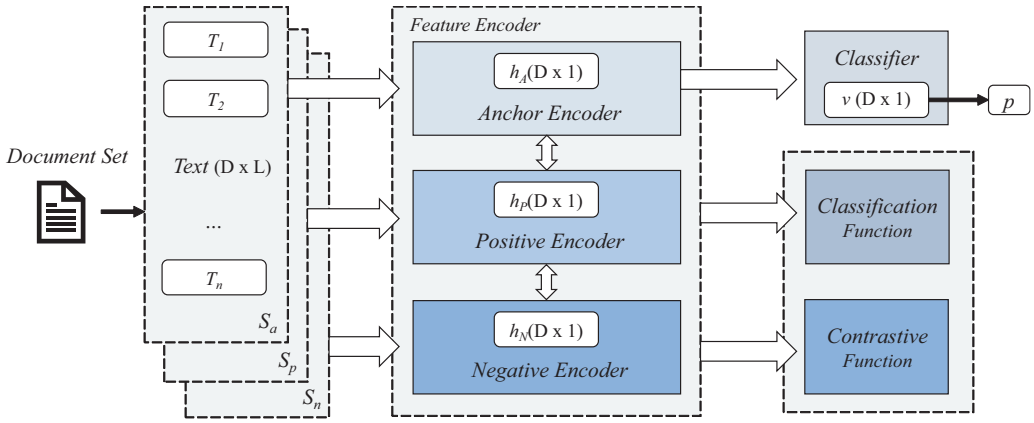


Figure 4. Overall concept of PAI-NET.

We describe how to combine pairs of documents in the preprocessing step and the embedding process in Section 3.1. The document embeddings after the preprocessing process are used as input to the encoder. Equation (1) shows the feature encoding process for the multi-tasking process described in Section 3.2. Equation (2) presents the classification task in Section 3.3, and the description of the objective function for multi-tasking is covered in Section 3.2. Lastly, the aggregated features pass through a final fully connected layer and softmax for label prediction learning in Equation (3) while being utilized in parallel as features for document similarity evaluation, as shown in Algorithm 1.

Algorithm 1 Pseudo-code of PAI-NET

Input: Document Set $S = [S_a, S_p, S_n]$
Output: Document feature v_a
Initialization: initialize shared weight w of document encoders $\mathcal{F}_a$, $\mathcal{F}_p$, $\mathcal{F}_n$ and set α value
○ ***Training phase***
S_a = target Document set (**a**nchor)
S_p = the prior art document pair set of S_a (**p**ositive)
S_n = non relevant document set of S_a (**n**egative)
{▷process mini-batch size learning}
for each batch size s of S **do**
  $v_a = \mathcal{F}_a(S_a)$
  $v_p = \mathcal{F}_p(S_p)$
  $v_n = \mathcal{F}_n(S_n)$ {▷each v encoded by parallel process of $\mathcal{F}$}
  $\nabla w_{Classification} = L_{Classification}\ (v_a)$ by (8)
  $\nabla w_{Margin} = L_{Margin}\ (v_a, v_p, v_n)$ by (9)
  $\nabla w_{Total} = \nabla w_{Classification} + \alpha \cdot \nabla w_{Margin}$
  $w = w + \nabla\ w_{Total}$
  {▷update shared w of document encoders}
end for
○ ***Similar search phase***
{ ▷find similar candidate documents S_c with target S_a}
$v_a = \mathcal{F}_a(S_a)$
for all Candidate Set S_c **do**
  $v_c = \mathcal{F}_a(S_c)$
  get Top-K v_k of MAX($v_a \cdot v_c$)
end for
return Candidate S_c by using index of v_k

3.1. Preprocessing

Patent documents have hierarchical classification labels. Patent documents are organized using the Cooperative Patent Classification (CPC) system, which has a five-level hierarchical structure: section (e.g., G—Physics), class (e.g., G06—Computing), subclass (G06F—Electric Digital Data Processing), group (G06F3—Input/Output arrangements), and subgroup (G06F3/0488—Interaction with touch-sensitive screens). As classification specificity increases from section to subgroup, patents sharing deeper-level classifications typically exhibit greater technical similarity. However, classification alone is insufficient for determining technical relevance, as patents with identical classifications may implement fundamentally different technical approaches. In general, patent documents with the same classification label have higher similarity as the classification level becomes deeper. However, having the same label does not mean they are similar, because each patent document has different technical claims, even if it has the same classification category. We take this into account and utilize the records of prior art information that experts consider most similar to the target patent document in the patent registration or prior art investigation process. We design a method to compute the relative similarity distance of document features closely by using document information and classification information from each patent document to learn the model, and we use leading literature similar to the target document as input. Also, while patent document collections can be classified into multiple overlapping classes in a multi-class format, as shown in Figure 5, they have a highly skewed distribution, making it difficult to select similar documents based solely on similar classification attributes.

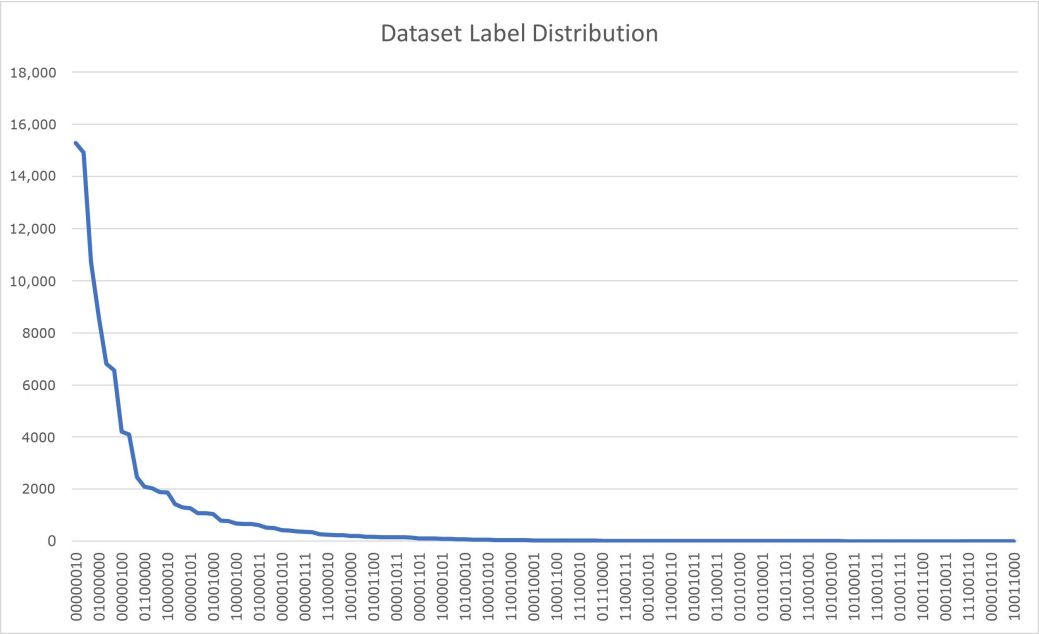


Figure 5. Dataset label distribution: the x-axis is the label name(one-hot encoded) and the y-axis is the number of labeled documents.

In our setting, we construct a triple-input document dataset by adding non-similar documents to consider the similarity between the prior art document and the target document in classification task learning. Given the challenges of quantitatively evaluating the technical similarity between patent documents, a method that considers relative similarity distances in learning is a detour, yet it is reasonable method for searching similar documents.

Let $\mathcal{T} = [T_1, T_2, ..., T_N]$ and $\mathcal{L} = [L_1, L_2, ..., L_N]$ be the set of documents S and their labels. We use a triple-pair dataset of inputs:

$$S = [S_a, S_p, S_n], \tag{4}$$

In Equation (4), we group the target document as S_a (anchor), the prior art document as S_p (positive), and documents that are not similar to the target document as S_n (negative) into one input pair.

In our contrastive learning framework, we structure each training instance as a triplet of patent documents. The target document S_a serves as the reference patent. The prior art document S_p is selected from specific prior art patents cited within the target document itself, representing expert-validated technical relationships. These citations, formally documented during patent registration, provide reliable technical similarity information. The non-relevant document S_n is chosen from patents that may share the same classification category but have no direct technical relevance to the target. This approach leverages existing expert knowledge embedded in citation relationships without requiring additional human annotation, enabling our model to distinguish between superficial category-based similarities and meaningful technical relationships.

Before feeding the data, we split each document into word tokenizing with a fixed-size word length l, and all words are embedded to D dimensional features as $S_{a,p,n} \in \mathbb{R}^{D \times l}$. We add a classification token $[CLS]$ to the front of the token array of each embedded document. Using a classification token $[CLS]$ for classification tasks is an easy way to summarize the features of a document as a single length dimension.

We design a triplet-conjured encoder as an encoding task method that extracts document features for document classification and similarity learning tasks. Each encoder

uses a stacked encoder such as a Transformer mechanism [9], which extracts the document feature as follows:

$$\mathcal{F}(S_i) = \sum_{l=L-s}^{L} \mathcal{A}^l{}_S, \tag{5}$$

$$[CLS]_i^l = \mathcal{A}_S^l(Q_i^{l-1}, K_i^{l-1}, V_i^{l-1}), \tag{6}$$

where Q, K, and V indicate the query, key, and value for the Transformer encoder, respectively. These feature vectors are also embedded by concatenating the $[CLS]$ token and S_i as $Q_i = K_i = V_i = \{[CLS], S_i\}$. Here, $\{\cdot, \cdot\}$ denotes concatenation. L is the total number of stacked encoders, and l is the index of the stacked encoder. L and t determine how many encoder layers are used to summarize paragraph S_i. Considering the efficiency of learning, we use the contextualized word embedding method [34] and a pre-trained language model, as used by the encoding Transformer [10].

Each encoder receives input from the anchor S_a, positive S_p, and negative S_n embedded document and outputs summarized document features as h_a, h_p, and h_n. The summarized document features are computed from the objective function and used for learning similarity distancing. Meanwhile, the anchor document feature h_a is used for classification tasks. Each encoder performs operations independently and in parallel. However, the three encoders share the same weight, which is inspired by a relevant study [35]. In the encoding process of PAI-NET, the anchor encoder can consider the learning experience of each encoder in regulating the similarity distance between documents.

3.2. Objective Function

3.2.1. Total Loss

We use a sum of functions for classification tasks and functions for similarity distance computing to perform multi-tasking tasks. Recent studies [36,37] have used the sum of classification and margin losses for the robustness of classification queries. Inspired by these parts, we use them for the robustness of similar document search queries. Total loss is expressed in a way that adds a percentage of margin losses to classification losses as follows:

$$L_{Total} = L_{Classification} + \alpha \cdot L_{Margin}, 0 \leq \alpha \leq 1, \tag{7}$$

This multi-task learning approach can be demonstrated using our patent document example. Consider the embedding space shown in Figure 4, where we have Apple's "bounce-back" patent (US 7,469,381) as our anchor document (A), Microsoft's scrolling patent (US 5,495,566) as the prior art positive example (P), and Apple's zoom functionality patent (US 7,864,163) as a patent within the same classification but not technically related as the negative example (N). Initially, the embedding distances between these documents have the following relationships:

- Prior art distance: $D_P = d(A, P)$
  (technically related but textually different)
- Similar classification patent distance: $D_N = d(A, N)$
  (close due to similar representation and classification)

Here, the initial distance relationship is often $D_N < D_P$ because patents in the same classification typically share vocabulary and structure. During training, PAI-NET applies the total loss function to adjust these distances. The classification loss ($L_{Classification}$) works to ensure documents are correctly classified into their respective patent categories. Meanwhile, the margin loss (L_{Margin}) specifically focuses on reducing the distance between the anchor and positive prior art document while increasing the distance to negative examples according to Equation (9). Using the optimal α value, the model balances these objectives.

After training, remarkably, the distance between Apple's patent and its prior art decreases substantially despite their textual differences, resulting in the final relationship $D_P \ll D_N$. This demonstrates how PAI-NET can learn similarity based on actual technical relevance beyond simple textual similarity or classification-based similarity.

3.2.2. Classification Loss

The classification task of patent documents is a multi-class multi-label problem, so we use the BCEWithLogitsLoss function for this as follows:

$$\begin{aligned} l(x, y) = L_{Classification} = \{l_1, ..., l_N\}, \\ l_n = -[y_n \cdot \log \sigma(x_n) + (1 - y_n) \cdot \log(1 - \sigma(x_n))], \end{aligned} \tag{8}$$

where target x and predicted label y are for coordinate training loss.

3.2.3. Margin Loss

We use triplet loss to bring the similarity distance between the target document and the prior art document close. Triple loss controls the distance between the prior art and non-similar documents as follows:

$$L_{Margin}(a, p, n) = max\{d(a_i, p_i) - d(a_i, n_i) + margin, 0\}, \tag{9}$$

$$d(x_i, y_i) = \|x_i - y_i\|_p \tag{10}$$

where a_i, p_i, and n_i indicate the target document (anchor), prior art document (positive), and non-similar document (negative), respectively. The *margin* is a constant factor to control the distance between positive and negative features. If the margin value increases, the robustness of the similar search query is enhanced. However, the size of the latent feature space is limited, and we need to determine the range of a margin factor to determine the query performance. In (10), the distance function uses the euclidean distance.

3.3. Document Classification

The end of the network is the single fully connected (FC) network for document classification. The output of the anchor encoder is fed into an 8-way $softmax$ function, which produces the distribution over the size of the section category (CPC).

3.4. Evaluation Metrics

To quantitatively assess the model's ability to rank similar documents highly, we employ the Mean Reciprocal Rank (MRR) metric. MRR effectively measures how well the model positions pre-defined relevant documents at higher ranks in a retrieval task. For a set of queries Q, MRR is formally defined as follows:

$$\textbf{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i} \tag{11}$$

where $|Q|$ is the number of queries, and $rank_i$ is the position of the first relevant document for the i-th query. The reciprocal rank $\frac{1}{rank_i}$ assigns higher scores when relevant documents appear at higher positions (e.g., 1.0 for the first position, 0.5 for the second position, etc.).

Additionally, we employ the Jaccard Index as a supplementary metric to maintain classification performance. The Jaccard Index measures the overlap between two document multi-label classification term sets and is formulated as follows:

$$\textbf{Jacc} = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{12}$$

where A and B represent the term sets of two documents, $|A \cap B|$ denotes the number of common terms, and $|A \cup B|$ represents the total number of unique terms. This metric produces values between 0 and 1 and is used to optimize document similarity without compromising the primary classification objective.

In our evaluation framework, each query document is paired with one known similar document (positive) and multiple dissimilar documents (negatives). The MRR score reflects the model's ability to consistently rank the positive document above the negative documents. A higher MRR value indicates better performance, with a perfect score of 1.0 indicating that the model always ranks the similar document in the first position. This metric is particularly suitable for our task, as it directly quantifies the model's effectiveness in identifying and prioritizing semantically similar patent documents in a retrieval context.

3.5. Implementation Details

To establish our experimental environment, we prepared two datasets. First, we randomly selected 100,000 domestic and international patent documents (patent registrations only, excluding utility models) registered in the Republic of Korea (KPRIS, https://plus.kipris.or.kr, accessed on 2 April 2025) from 2020 to 2023 for training, and 20,000 for testing. This dataset was used for performance comparison with other models. Figure 5 shows the distribution of multi-labels across eight sections based on the CPC classification system for this dataset. To validate performance in a specific domain environment, we also utilized the Artificial Intelligence Patent Dataset (AIPD, https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset, accessed on 2 April 2025) provided by the USPTO. For the AIPD, we randomly selected 10,000 documents from approximately 1 million documents in the seed group and split them in an 8:2 ratio for training and testing as anchor document data. This dataset was employed in Section 3.9 to evaluate how the proposed technique's training affected the similarity distance adjustment compared to the pre-trained model environment. Table 1 shows the AI classification descriptions and their proportions in this dataset, where multiple classifications can overlap for a single patent due to the multi-label attribute of the data. For contrastive learning, each dataset was structured as triplets consisting of a reference patent, a registered patent explicitly cited as prior art for that patent, and a non-prior art patent document.

We trained the networks for 10 epochs with an initial learning rate set to 0.001 and divided by 0.1 every 10 epoch. The datatset consisted of 100,000 training and 20,000 validation documents. In our experiments, we used 768 dimensions of embedding vector size D for each word, and we set the input size to l to 100 words. We used a fine-tuning method with a pre-trained Bert model [10] for efficient learning.

We evaluated PAI-NET 10 times with difference random seeds, then we reported the average performances. We used the Adam optimizer [38] and applied the early stopping method based on the EM score. We used a 7800X3D AMD CPU, 192 GB of RAM, and RTX-4090 24GB GPU RAM and A100 cloud system for the implementation.

Table 1. Artificial Intelligence Patent Dataset (AIPD2023, USPTO) sampling data distribution.

Label	Distribution (%)	Description
hardware	44.57	AI hardware implementations, accelerators, specialized processors, and chips
planning	38.43	Decision-making systems, automated planning, scheduling, and optimization
vision	31.69	Computer vision, image recognition, object detection, and video analysis
ml	22.06	General machine learning algorithms, techniques, and methodologies
kr	20.74	Knowledge representation, ontologies, knowledge graphs, and reasoning systems
speech	17.97	Speech recognition, processing, synthesis, and speaker identification
nlp	14.22	Natural language processing, text analysis, and language understanding
evo	12.32	Evolutionary and genetic algorithms for optimization and learning

3.6. Ablation Studies

In this work, we have taken steps to determine the degree of margin distance and the ratio of margin loss values to combine classification and margin functions. To this end, we first use a method to determine the margin distance with low losses for classification tasks and then determine the ratio sequentially.

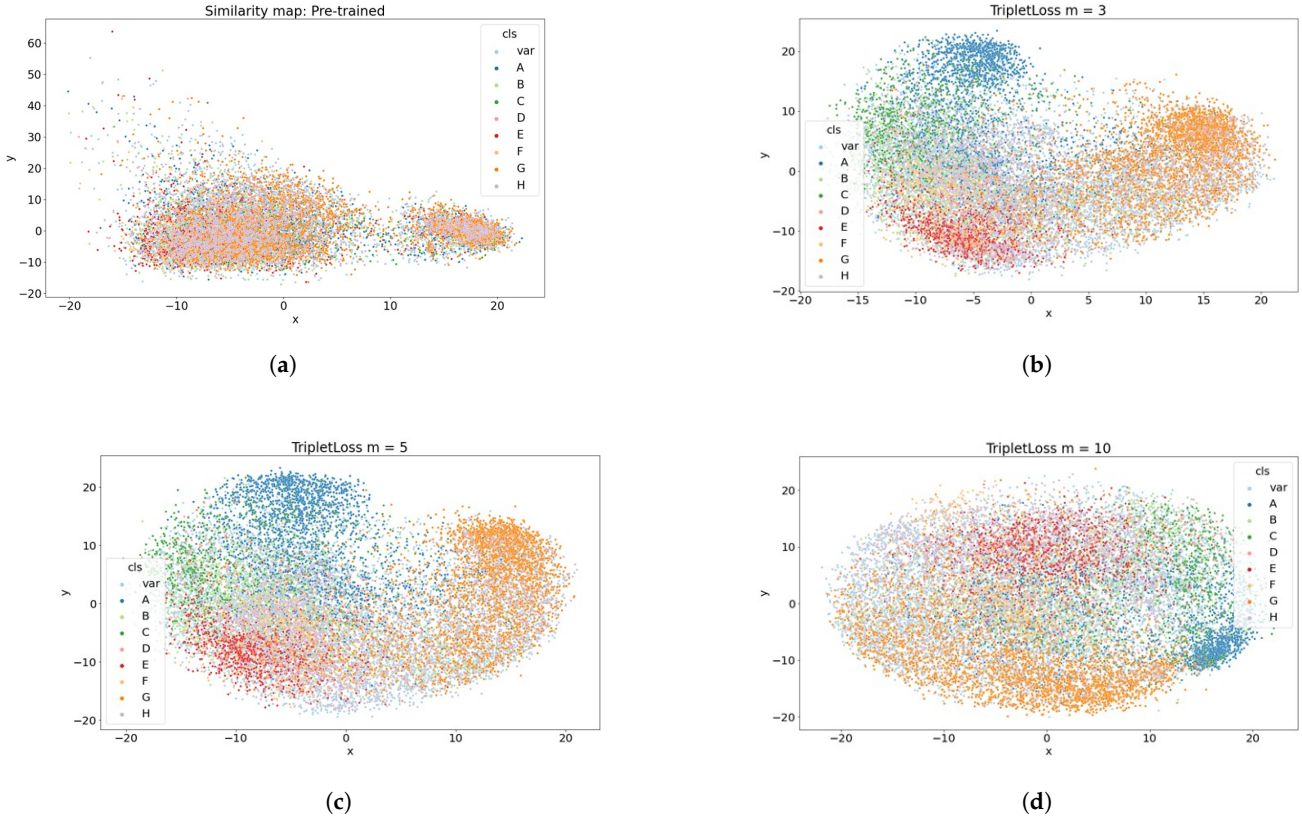
3.6.1. Margin Distance

When a document feature is projected into a latent region, the larger the distance between the classification sets of each document feature, the more robust the model can be seen. Thus, theoretically, the greater the margin distance, the greater the degree of robustness. However, since the margin distance cannot be increased indefinitely within a limited area, the experimental margin distance between the classification sets must be maintained appropriately, and the closer the margin distance between similar documents. In this regard, Table 2 demonstrates the rationale for selecting appropriate coefficients for the margin distance based on the classification performance of the document. Evaluation of the classification performance of documents is not used when training margin distances. The negative values for Dn (e.g., -0.009 at margin = 5) represent positioning in the embedding space where negative examples are placed in the opposite direction from anchor documents relative to the origin. This arrangement creates greater distances between similar and dissimilar patents, with larger absolute values of $|Dp - Dn|$ reflecting increased separation in the vector space. However, it is reasonable to select margin distances that have the most negligible impact on the classification results performance and combine them with classification tasks in a multi-task environment. Therefore, in this work, we compute the margin distance based on the label classification performance and choose '3' as the coefficient.

Table 2. Margin constant evaluation.

Margin Distance	MRR	D_p	D_n	$\lvert D_p - D_n \rvert$
2	0.8914	0.881	0.629	0.252
3	0.8913	0.713	0.022	0.691
3.5	0.8939	0.725	0.019	0.706
4	0.8889	0.756	0.031	0.725
5	0.8847	0.775	−0.009	0.784

Meanwhile, visualizing the dispersion of document features over margin distances in Figure 6, as previously described, the degree of overlap of the features is high, making it challenging to evaluate the similarity or classification criteria between documents. On the other hand, if the margin distance is 10, it can be seen that document features are spread evenly throughout the area. However, there is a problem of poor cohesion according to the classification boundary.



(**a**) (**b**)

(**c**) (**d**)

Figure 6. t-SNE visualization of document embeddings with different margin values: (**a**) m = 0, (**b**) m = 3, (**c**) m = 5, (**d**) m = 10. The x and y axes represent dimensionality-reduced and normalized embedding values from the high-dimensional vector space, projected in 2D for visualization. Optimal margin values balance separation between document types while preserving class structure.

3.6.2. Margin Loss Ratio

Margin loss functions require expert similarity assessments within the training data to effectively learn similarity relationships. In patent datasets with imbalanced label distributions, relying solely on similarity learning would necessitate extensive training data for each classification label. To address this challenge, we combine margin loss with classification functions to achieve learning efficiency. We conducted experiments varying the margin loss ratio α from 0 to 0.5, with the results presented in Table 3. In Table 3, the EM score (Exact Match score) represents the percentage of test cases where all predicted

classification labels exactly match the ground truth labels. As a strict evaluation metric for multi-label classification, it only considers a prediction correct when there is 100% agreement across all possible labels, with no partial credit for partially correct predictions.

Table 3. Margin loss ratio.

α	EM Score	Jacc	MRR	D_p	D_n	$\lvert D_p - D_n \rvert$
0.0	60.935	75.388	0.5856	0.653	0.006	0.647
0.2	60.545	75.389	0.7417	0.733	−0.044	0.777
0.4	57.231	72.654	0.7965	0.754	−0.052	0.806
0.5	56.735	72.106	0.8124	0.768	−0.292	1.058

The selection of an appropriate α value requires balancing multiple performance objectives. At α = 0.0 (baseline classification-only model), we observe the highest Exact Match (EM) score of 60.935% but the poorest retrieval performance (MRR = 0.5856), with minimal differentiation between similar and dissimilar documents ($|D_p - D_n| = 0.647$). Conversely, at α = 0.5, the model achieves the highest MRR (0.8124) and maximum separation between positive and negative examples ($|D_p - D_n| = 1.058$), but at the cost of substantially reduced classification accuracy (EM = 56.735%).

Our analysis identified α = 0.2 as the optimal value, representing a strategic balance point where classification performance remains robust (EM = 60.545%, just 0.39% below the baseline), the Jaccard index reaches its maximum value (75.389%), retrieval performance shows significant improvement (MRR = 0.7417, a 26.7% increase over baseline), and a sufficient margin between similar and dissimilar documents is maintained. When α exceeds 0.2, both the Jaccard index and EM scores decline more rapidly, indicating that higher weights on the margin loss begin to introduce noise into the classification task as, similarity pairs are constructed independently of classification criteria. The empirical evidence demonstrates that α = 0.2 provides the optimal trade-off: maintaining high-quality classification while significantly enhancing the model's ability to recognize technical relationships between patent documents.

3.7. Using Cosine Distance for Loss Function

The choice of distance metric in margin loss calculations significantly impacts model performance. While document similarity is traditionally measured using cosine distance in information retrieval, we investigated whether euclidean distance might be more effective for patent document relationships in our contrastive learning framework. In Table 4, we compare four different distance metric configurations: pure euclidean distance, pure cosine distance, a combination of both (Euc + Cos), and a hybrid approach that uses euclidean distance globally while applying cosine distance specifically to anchor-positive pairs $(Euc + a - pCos)$. The results reveal distinct performance patterns across these configurations. The highest Mean Reciprocal Rank (MRR) of 0.7627 was achieved using the hybrid approach $(Euc + a - pCos)$, which applies cosine distance specifically to anchor-positive document pairs while using euclidean distance elsewhere. Pure cosine distance achieves the largest separation between positive and negative document pairs ($|D_p - D_n| = 0.89$), indicating its superior discriminative power in embedding space. However, this comes at the cost of reduced MRR (0.6865), suggesting that while it creates clearer boundaries between similar and dissimilar documents, it may not align optimally with the ranked retrieval task. These findings demonstrate that euclidean distance is more effective for maintaining classification performance in our multi-task learning framework, likely because it better preserves the geometric relationships needed for classification boundaries. Meanwhile, cosine distance excels at capturing semantic similarity between technically related documents, particularly between anchor patents and their prior art. The hybrid

approach $(Euc + a - pCos)$ successfully balances these complementary strengths, resulting in the best overall retrieval performance.

Table 4. Comparison of distance metrics for loss function. 'Euclidean' represents standard euclidean distance; 'Cosine' indicates cosine distance used for all similarity calculations; 'Euc + Cos' denotes a combined approach using both metrics equally weighted; 'Euc + a-p Cos' represents our hybrid approach that uses euclidean distance for the general loss calculation while specifically applying cosine distance for anchor-positive pair similarity measurement.

	MRR	D_p	D_n	$\lvert D_p - D_n \rvert$
Euclidean	0.7417	0.713	−0.040	0.753
Cosine	0.6865	0.820	−0.071	0.89
$Euc + Cos$	0.7618	0.553	0.005	0.548
$Euc + a - pCos$	0.7627	0.708	−0.037	0.745

3.8. Episodic Training

We evaluated performance as a way to regulate the learning dataset. In this regard, Table 5 compares PAI-NET with PAI-NET using the underlying dataset, a set of documents with only the same label compared to the baseline document, and PAI-NET$_n$ with the same label without considering similarity. We reveal that pairs of similar document sets with only the same label have lower classification performance or similarity than baseline datasets. They were insufficient to be written as appropriate learning sets for pairs of similar document sets with different labels.

Table 5. Comparisons with episodic training.

	MRR	D_p	D_n	$\lvert D_p - D_n \rvert$
PAI-NET	0.7417	0.713	−0.040	0.753
PAI-NET$_p$	0.7499	0.744	0.016	0.728
PAI-NET$_n$	0.5625	0.619	0.003	0.616

On the other hand, the classification performance was best when using different datasets with the same label without considering similarity, which can be seen as reflected in the classification feature even if it is not classified as data written in the existing self-supervision method.

3.9. Comparisons with Pretrained Model

To evaluate the performance enhancement of our proposed model, we conducted comprehensive experiments comparing PAI-NET with its pretrained baseline model through cosine similarity distances and Mean Reciprocal Rank (MRR) analysis. Our experimental findings demonstrate substantial improvements achieved by PAI-NET.

Specifically examining the cosine similarity distributions shown in Figure 7, the pretrained model exhibits minimal discriminative capability, with nearly indistinguishable similarity scores between positive (0.998 ± 0.001) and negative document pairs (0.997 ± 0.001). In contrast, PAI-NET demonstrates significantly enhanced discriminative power by establishing a clear demarcation between positive pairs (0.708 ± 0.199) and negative pairs (0.095 ± 0.216). This pronounced separation in similarity metrics indicates PAI-NET's superior ability to effectively differentiate between semantically similar and dissimilar patent documents.

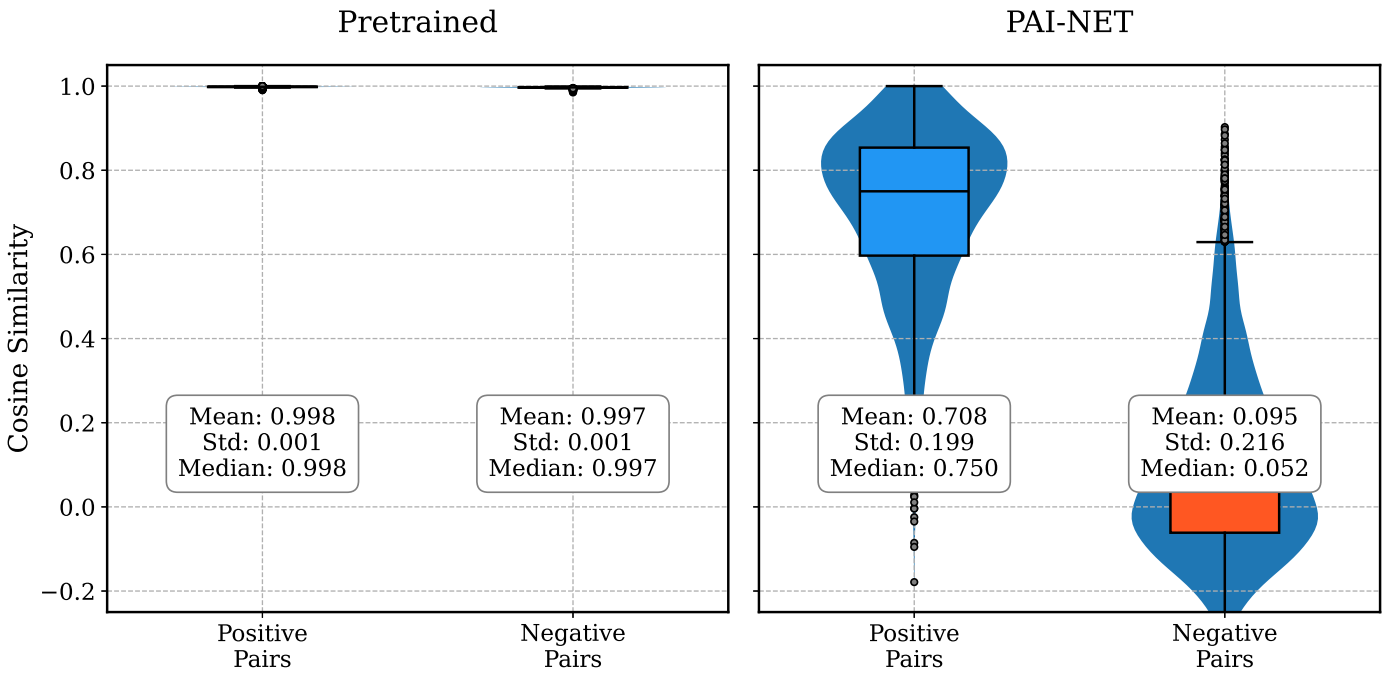


Figure 7. Cosine similarity distribution comparison with pretrained model.

The quantitative assessment through MRR metrics, as presented in Table 6, further validates PAI-NET's enhanced performance. Our model achieves an MRR score of 0.954, representing a 9.8% improvement over the pretrained model's score of 0.869. This improvement in MRR metrics substantiates PAI-NET's enhanced capability in accurately positioning relevant documents at higher ranks during retrieval operations, thereby delivering more precise and reliable search results.

Table 6. Comparison of pretrained and fine-tuned models.

Model	Positive Pairs	Negative Pairs	MRR
Pretrained	0.998 ± 0.001	0.997 ± 0.001	0.869
Finetuned	0.708 ± 0.199	0.095 ± 0.216	0.954

This analysis is particularly relevant to domains with limited prior art information, as it demonstrates a key advantage of our approach. Unlike legacy ontology-based systems that require comprehensive knowledge structures covering all possible technical relationships, PAI-NET's contrastive learning method enables the model to learn generalizable patterns of technical similarity from a limited set of examples. Our deep learning encoder, trained on existing prior art relationships, develops the ability to recognize technical relevance patterns that extend beyond the specific documents used in training. This allows PAI-NET to effectively handle novel documents or queries without established prior art connections. The model can flexibly apply learned similarity principles to emerging technical domains where explicit citations are sparse and traditional ontological frameworks would be impractical to maintain. This adaptability represents a significant advancement in making patent retrieval systems more resilient and scalable in real-world applications where complete knowledge structures are often unavailable.

3.10. Comparisons with State-of-the-Art

We measure classification performance and cosine similarity distances together with recent studies showing good performances [20,30,31] (SOTA, state-of-the-art). Evaluating the degree to which the patent document classification and similarity distance performance

are satisfied simultaneously, the performance of this study, as observed in Table 7, shows superior results compared to existing studies. Among the relevant studies, CL+SCL [31] adds sampling considering distribution at the dataset stage because the model is constructed on the premise of being self-supervised, which makes it challenging to create a suitable candidate group for classification labels within batch sizes. In PAI-NET, using triple-pair document datasets that did not consider label values, the similarity distance between similar and non-similar documents was the highest, indicating that the distance from similar documents became relatively close. We observed that our proposed method showed more than 15% improvement in MRR performance over the related SOTA methods.

Table 7. Comparisons with state-of-the-art.

	Baseline			**PAI-NET**					
	patentBert [20]	**SentenceBert (cls) [30]**	**CL + SCL [31]**	***Euc***	***Cos***	***Euc* + *Cos***	***Euc* + *apCos***	***Positive***	***Negative***
MRR	0.5856	0.5617	0.5763	0.7417	0.6865	0.7618	0.7627	0.7499	0.5810

4. Conclusions

We have presented PAI-NET, a novel framework for improving similarity search performance in Retrieval-Augmented Generation (RAG) systems for patent knowledge management. Our approach effectively addresses the unique challenges of patent document similarity search and makes several significant contributions to patent information systems and knowledge management.

The core strength of PAI-NET lies in its ability to incorporate expert domain knowledge through prior art relationships, resulting in superior document recommendation performance. Our experimental results demonstrate a 15% improvement in similarity-based retrieval performance compared to state-of-the-art methods in the patent domain. This substantial enhancement in retrieval accuracy provides particular value for prior art search processes where identifying relevant patent information is crucial.

A key innovation of our approach is the novel solution to knowledge representation in expert systems through the leveraging of prior art information rather than traditional ontological structures. This approach not only provides superior performance but also significantly reduces the costs associated with maintaining expert knowledge bases. This framework demonstrates sophisticated capabilities in understanding and utilizing complex relationships between patent documents, enabling more nuanced and accurate knowledge retrieval. The computational efficiency of PAI-NET's architecture is noteworthy, as it handles similarity learning tasks through fine-tuning without significant overhead. This efficiency is particularly valuable in practical applications, enabling robust service deployment even in environments with continuously accumulating patent document information. The implications of this work extend beyond patent systems to the broader field of expert knowledge systems. The methodology we have developed for incorporating domain expertise through document relationships shows promise for adaptation to other specialized fields where similar expert-verified relationships exist. This approach represents a significant step forward in reducing the human effort required for constructing and maintaining domain-specific knowledge bases.

Looking ahead, our research opens several promising avenues for future investigation. These include the potential adaptation of our framework to general knowledge management systems and the development of more flexible neural network architectures capable of accommodating evolving domain knowledge. Furthermore, we envision extending our approach to other types of expert systems where document relationships play a crucial role in knowledge organization and retrieval.

Author Contributions: Conceptualization, K.-Y.L. and J.B.; methodology, K.-Y.L. and J.B.; software, J.B.; validation, K.-Y.L.; formal analysis, K.-Y.L.; investigation, K.-Y.L. and J.B.; resources, K.-Y.L. and J.B.; data curation, K.-Y.L. and J.B.; writing—original draft preparation, K.-Y.L. and J.B.; writing—review and editing, K.-Y.L. and J.B.; visualization, J.B.; supervision, J.B.; project administration, J.B.; funding acquisition, J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Hankuk University of Foreign Studies Research Fund Of 2025.

Data Availability Statement: The data supporting the reported results can be accessed through the Artificial Intelligence Patent Dataset (AIPD) provided by the United States Patent and Trademark Office (USPTO). This dataset, which was used for our experiments, is publicly available and can be downloaded from https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset (accessed on 24 January 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Siriwardhana, S.; Weerasekera, R.; Wen, E.; Kaluarachchi, T.; Rana, R.; Nanayakkara, S. Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering. *Trans. Assoc. Comput. Linguist.* **2022**, *11*, 1–17. [CrossRef]
2. Wang, Y.; Li, X.; Wang, B.; Zhou, Y.; Ji, H.; Chen, H.; Zhang, J.; Yu, F.; Zhao, Z.; Jin, S.; et al. PEER: Expertizing Domain-Specific Tasks with a Multi-Agent Framework and Tuning Methods. *arXiv* **2024**. [CrossRef]
3. Liang, L.; Sun, M.; Gui, Z.; Zhu, Z.; Jiang, Z.; Zhong, L.; Qu, Y.; Zhao, P.; Bo, Z.; Yang, J.; et al. KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation. *arXiv* **2024**. [CrossRef]
4. Kang, B.; Kim, J.; Yun, T.R.; Kim, C.E. Prompt-RAG: Pioneering Vector Embedding-Free Retrieval-Augmented Generation in Niche Domains, Exemplified by Korean Medicine. *arXiv* **2024**. [CrossRef]
5. Khanna, S.; Subedi, S. Tabular Embedding Model (TEM): Finetuning Embedding Models For Tabular RAG Applications. *arXiv* **2024**. [CrossRef]
6. Nguyen, Z.; Annunziata, A.; Luong, V.; Dinh, S.; Le, Q.; Ha, A.H.; Le, C.; Phan, H.A.; Raghavan, S.; Nguyen, C. Enhancing Q&A with Domain-Specific Fine-Tuning and Iterative Reasoning: A Comparative Study. *arXiv* **2024**. [CrossRef]
7. Chu, J.M.; Lo, H.C.; Hsiang, J.; Cho, C.C. Patent Response System Optimised for Faithfulness: Procedural Knowledge Embodiment with Knowledge Graph and Retrieval Augmented Generation. In Proceedings of the 1st Workshop on Towards Knowledgeable Language Models (KnowLLM 2024), Bangkok, Thailand, 16 August 2024; pp. 146–155.
8. Wang, S.; Yin, X.; Wang, M.; Guo, R.; Nan, K. EvoPat: A Multi-LLM-based Patents Summarization and Analysis Agent. *arXiv* **2024**, arXiv:2412.18100.
9. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
10. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1 (Long and Short Papers), pp. 4171–4186.
11. Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.G.; Le, Q.; Salakhutdinov, R. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 2978–2988.
12. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. Xlnet: Generalized autoregressive pretraining for language understanding. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 5753–5763.
13. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The long-document transformer. *arXiv* **2020**, arXiv:2004.05150.
14. Roudsari, A.H.; Afshar, J.; Lee, C.C.; Lee, W. Multi-label Patent Classification using Attention-Aware Deep Learning Model. In Proceedings of the IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 558–559.
15. Li, S.; Hu, J.; Cui, Y.; Hu, J. DeepPatent: patent classification with convolutional neural networks and word embedding. *Scientometrics* **2018**, *117*, 721–744.
16. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.

17. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:1408.5882.
18. Risch, J.; Krestel, R. Domain-specific word embeddings for patent classification. *Data Technol. Appl.* **2019**, *53*, 108–122.
19. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681.
20. Lee, J.S.; Hsiang, J. PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model. *arXiv* **2019**, arXiv:1906.02124.
21. Hu, J.; Li, S.; Hu, J.; Yang, G. A Hierarchical Feature Extraction Model for Multi-Label Mechanical Patent Classification. *Sustainability* **2018**, *10*, 219. [CrossRef]
22. Bai, J.; Shim, I.; Park, S. MEXN: Multi-Stage Extraction Network for Patent Document Classification. *Appl. Sci.* **2020**, *10*, 6229. [CrossRef]
23. Song, G.; Huang, X.; Cao, G.; Liu, W.; Zhang, J.; Yang, L. Enhanced deep feature representation for patent image classification. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP 2018). International Society for Optics and Photonics, Chengdu, China, 12–14 December 2018; Volume 11069, p. 110690P.
24. Jiang, S.; Luo, J.; Pava, G.R.; Hu, J.; Magee, C.L. A CNN-based Patent Image Retrieval Method for Design Ideation. *arXiv* **2020**, arXiv:2003.08741.
25. Csurka, G. Document image classification, with a specific view on applications of patent images. In *Current Challenges in Patent Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 325–350.
26. Landauer, T.K.; Foltz, P.W.; Laham, D. An introduction to latent semantic analysis. *Discourse Processes* **1998**, *25*, 259–284.
27. Kim, B.T.S.; Hyun, E.J. Mapping the Landscape of Blockchain Technology Knowledge: A Patent Co-Citation and Semantic Similarity Approach. *Systems* **2023**, *11*, 111. [CrossRef]
28. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International conference on machine learning. PMLR, Virtual, 13–18 July 2020; pp. 1597–1607.
29. Fang, H.; Xie, P. Cert: Contrastive self-supervised learning for language understanding. *arXiv* **2020**, arXiv:2005.12766.
30. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3973–3983.
31. Gunel, B.; Du, J.; Conneau, A.; Stoyanov, V. Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 4 May 2021.
32. Li, Z.; Zhou, L.; Yang, X.; Jia, H.; Li, W.; Zhang, J. User Sentiment Analysis of COVID-19 via Adversarial Training Based on the BERT-FGM-BiGRU Model. *Systems* **2023**, *11*, 129. [CrossRef]
33. Lin, W.; Yu, W.; Xiao, R. Measuring Patent Similarity Based on Text Mining and Image Recognition. *Systems* **2023**, *11*, 294. [CrossRef]
34. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75.
35. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
36. Caruana, R. Multitask learning. *Mach. Learn.* **1997**, *28*, 41–75.
37. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv* **2017**, arXiv:1706.05098.
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.