Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# TOWARDS EFFICIENT PATENT CLASSIFICATION: KOLMOGOROV ARNOLD NETWORKS AS AN ALTERNATIVE TO MLP

**MINJONG CHEON[1], CHANGBAE MUN[*]**

[1]Center for Sustainable Environment Research, Korea Institute of Science and Technology, 5 Hwarang-ro 14-gil, Wolgok-dong, Seongbuk-gu, Seoul 02792, Republic of Korea

*Department of Electrical, Electronic & Communication Engineering, Hanyang Cyber University, Seoul 04764, Republic of Korea

E-mail: [1]jmj541826@gmail.com, [*]changbae@hycu.ac.kr (Corresponding Author)

**ABSTRACT**

In today's continuously shifting innovation and technological growth environment, effective intellectual property (IP) management and organization have become critical, resulting in more significant patent classification. Moreover, recent advances in natural language processing (NLP) technology have resulted in enhanced patent categorization. However, incorporating multilayer perceptron (MLP) layers in NLP algorithms frequently results in higher memory needs, particularly as network size rises. We suggest using the Kolmogorov Arnold Network (KAN) instead of MLP layers to solve this issue. In this work, we used a dataset from the European Patent Office (EPO) to categorize patents into three groups. We experimented with several KAN setups and discovered that decreasing hidden dimension sizes considerably reduced the number of parameters while keeping good accuracy. The [32, 16, 8] configuration achieved an accuracy of 74.84%, which rose to 75.12% after adjusting crucial hyperparameters such as spline_order and grid_size. Compared to other machine learning models such as MLP (75.83%), Random Forest, and XGBoost, KAN consistently surpassed them in accuracy and efficiency. Our findings broaden the use of KAN to patent classification and offer new avenues for its usage in other text-based classification tasks. KAN's proven efficiency and performance make it a promising alternative to existing machine learning models in this area, emphasizing its potential for further application in patent-related activities.

**Keywords:** *Patent classification, NLP, Kolmogorov Arnold Network (KAN), MLP, Deep Learning*

## 1. INTRODUCTION

Patents play a vital role in preserving ideas by offering legal recognition to inventors and granting them exclusive rights to use, produce, and market their inventions [1][2]. Because their efforts would be protected from illegal use, this exclusivity greatly motivates innovators to invest time, energy, and money in research and development (R&D) [3]. In today's ever-changing innovation and technological growth environment, effective intellectual property (IP) management and organization have become critical [4]. With this trend, patent categorization is an important cornerstone in this context, serving as a systematic framework for classifying discoveries across several industries. Classification methods let inventors, researchers, and examiners navigate a huge collection of technological knowledge more effectively by categorizing patents. To effectively identify originality and innovative steps, new patent applications must be appraised in comparison to earlier art using effective classification. Furthermore, even specialists may struggle to appropriately analyze patent paperwork owing to its length and complex technical and legal vocabulary. Automated solutions are therefore urgently required to assist with the processing and analysis of these articles [5].

The increasing application of natural language processing (NLP) technology may enable us to extract, classify, and analyze complicated information contained in patents more efficiently. Patent categorization, for example, may be easily automated utilizing Transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) and GPT, which help categorize patents by recognizing word context and linkages [6][7]. Because patents usually include a

Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

wide range of technical areas, subject modeling techniques like Latent Dirichlet Allocation (LDA) might identify underlying themes that were not previously discussed [8]. Furthermore, techniques such as Doc2Vec, FastText, and BERT-based sentence transformers offer more complex semantic search by examining the meaning and context of words, resulting in more accurate retrieval than basic keyword matching [9][10]. Inside those algorithms, Multi-Layer Perceptrons (MLPs) remain a foundational building block for many neural network architectures. Despite their simplicity, MLP layers play a crucial role in numerous advanced models. In transformer models, for example, MLP layers are crucial to the feedforward neural networks employed following attention processes, increasing total representational capacity and feature extraction capabilities [11]. Despite their relevance, MLP layers can have a large memory load, particularly as networks grow. In designs like BERT, the vast number of parameters in MLP layers can cause memory consumption and computational efficiency issues, especially when working with large-scale datasets or deploying models in resource-constrained contexts. To address these drawbacks, we present a novel architecture known as the Kolmogorov Arnold Network (KAN), which provides a memory-efficient alternative to typical MLP layers with satisfactory performance [12].

Our study intends to evaluate if KAN can successfully replace MLP layers for patent categorization by adding it into neural network designs, including those that generally rely on them, such as Transformer. The fundamental goal of this study is to do a series of tests to compare KAN's performance against MLPs, particularly in jobs involving huge patent datasets and extensive feature extraction. We want to see if KAN can save considerable amounts of RAM while maintaining or improving classification accuracy and overall model performance. Furthermore, KAN's modular structure will be evaluated for its potential to integrate smoothly with current systems, as well as its applicability to a wide range of applications. Our study aims to determine whether the Kolmogorov Arnold Network (KAN) can effectively replace MLP layers for patent classification. Specifically, we aim to:

1. Evaluate the performance and efficiency of KAN compared to conventional models like MLP, XGBoost, and Logistic Regression.
2. Investigate how KAN's architecture impacts memory consumption and accuracy.
3. Explore hyperparameter tuning (e.g., spline_order, grid_size) and its impact on the performance of KAN.

We end this section by detailing the paper's structure. In Section 2: Related Works, we review existing research on patent classification and neural network architectures. Section 3: Methodology details the KAN and our experimental setup. Section 4: Results presents the outcomes of our experiments, while Section 5: Discussion provides insights into the findings. Finally, Section 6: Conclusion summarizes the contributions of this work and outlines future research directions.

## 2. RELATED WORKS

Lee and Hslang utilized a fine-tuned BERT model, focusing specifically on patent claims. The model, referred to as PatentBERT, was trained on the USPTO-3M dataset and compared against DeepPatent, a CNN-based model. PatentBERT achieved superior performance, with a top-1 precision of 81.75% and an F1 score of 65.87%, significantly outperforming DeepPatent. The study also found that using the Cooperative Patent Classification (CPC) system resulted in better performance than the International Patent Classification (IPC) system. Overall, the study demonstrates that fine-tuning BERT on patent claims alone can achieve state-of-the-art results, making it an efficient method for patent classification [13].

Roudsari et al. focused on patent categorization at the difficult subgroup level when they introduced the AttentionXML deep learning model with the DistilBERT pre-trained language model. The challenge of categorizing patents into lower-level CPC and IPC hierarchies—which have many categories, significant overlap, and unbalanced data distribution—is the subject of this study. The authors compare DistilBERT's performance to earlier deep learning techniques like DeepPatent and optimize it to increase classification accuracy. Findings indicate that combining AttentionXML with DistilBERT for patent subgroup classification can lead to higher accuracy, suggesting that this approach is a viable solution for extremely complex multi-label classification issues in the patent field. They also emphasize the importance of recent advancements in NLP, such as pre-trained language models, to improve patent classification performance at lower hierarchical levels [14].

Otiefy and Endres concentrated on IPC and CPC subclass patent classification. Using patent claims, they produced a new USPTO 2.8M dataset, and they contrasted the outcomes of BERT-for-Patents with those of earlier models such as PatentBERT and DeepPatent. By obtaining a micro-F1 score of 69.39% with a threshold of 0.4, they showed that fine-tuning BERT-for-Patents on this dataset improves classification performance, surpassing state-of-the-art models. The major conclusions indicate that domain-specific language models, like BERT-for-Patents, perform very well in tasks involving the classification of patents, especially when adjusted for sub-class categorization. The study also emphasizes how resilient traditional models, such as NBSVM, are under certain conditions. The newly created USPTO 2.8M dataset, models, and code are released to the public to further enhance research in patent classification [15].

Roudsari et al. aimed to improve multi-label patent classification by using pre-trained language models such as BERT, XLNet, RoBERTa, and ELECTRA. These models were tuned for patent classification, and their performance was compared to other deep learning models such as CNNs, LSTMs, and the DeepPatent-based word embedding model. The studies were conducted using two datasets: USPTO-2M and M-Patent. XLNet beat the other models and achieved cutting-edge performance, with a micro-F1 score of 0.572 on the USPTO-2M dataset and 0.736 on the M-Patent dataset. The main discovery is that fine-tuning XLNet greatly enhances multi-label patent classification, owing to its bidirectional context learning and permutation language modeling capabilities, which make it more suited to the complex language and structure of patent documents [16].

Yoo et al. aimed to enhance the automatic multi-label categorization of patents linked to artificial intelligence using a Sentence Attention mechanism in conjunction with a modified D2SBERT model. The study addressed the difficulties associated with manually categorizing complicated AI-related patents, which included both legal and technological terminology and were thus labor-intensive and prone to mistakes. Yoo et al. sought to address the drawbacks of earlier models, notably BERT's sequence length restriction, by modifying the D2SBERT model to support longer sequences and applying sentence attention to extract significant information from patent documents. Results from experiments on AI-related patents from the USPTO showed that the proposed model outperformed traditional deep learning methods, such as CNN, LSTM, and previous versions of BERT, achieving higher macro and micro F1 scores. The principal finding was that the combination of D2SBERT and Sentence Attention significantly enhanced the accuracy of multi-label classification in AI-related patents, demonstrating its potential for automating patent classification in technologically complex fields [17].

We reviewed recent papers from peer-reviewed journals, focusing on Transformer models, and MLP-based classifiers. Patent classification is essential for managing intellectual property but requires extensive computational resources. Traditional MLP-based methods, although effective, are resource-intensive. This motivated us to explore KAN as a memory-efficient alternative. The adaptability of the Kolmogorov Arnold Network (KAN), which not only offers a singular solution to memory efficiency but also demonstrates exceptional universality, sets our method apart even further. The modified D2SBERT with Sentence Attention by Yoo et al. is one of the numerous earlier models that is tailored for certain tasks or datasets; in contrast, the KAN-based approach may be easily integrated into these and other frameworks. Our solution can enhance performance and memory economy in a variety of applications by substituting KAN for conventional MLPs. Because of its versatility, KAN is a desirable option for research on patent categorization in the future and beyond. It offers a scalable and adaptable method that can be used with little modification to various contemporary, cutting-edge models.

## 3. Materials and Methods

### 3.1 Sections and Subsections

Simple Contrastive Sentence Embedding (SIMCSE), which was developed by Gao et al., utilizes contrastive learning to align semantically similar sentences close together in the embedding space while pushing apart semantically dissimilar sentences. It consists of two different methods: unsupervised and supervised. Supervised SimCSE is a contrastive learning-based approach for generating high-quality sentence embeddings by leveraging labeled sentence pairs, where semantically similar sentences (e.g., premise-hypothesis pairs from natural language inference datasets) are treated as positive pairs and dissimilar sentences are used as

Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

negatives. While the supervised version relies on explicit sentence pairs to guide the learning process, the unsupervised version generates positive pairs by applying different dropout masks to the same sentence and treats other sentences in the batch as negatives, relying on random augmentations instead of explicit labels for training [18].

### 3.2 Kolmogorov–Arnold Network

Kolmogorov–Arnold Network (KAN), inspired by the Kolmogorov–Arnold representation theorem, is a novel type of neural network with learnable activation functions on edges instead of fixed activations on nodes, as seen in traditional MLPs. By leveraging the theorem, which states that any multivariate function can be broken into univariate functions and sums, KANs can effectively handle high-dimensional data and overcome the curse of dimensionality. The activation functions are parameterized using B-splines, which are controlled by adjustable points and knots, and the network's architecture aggregates results at nodes through these transformations. Unlike MLPs, KANs do not require traditional linear transformations with weights and biases, improving both accuracy and interpretability while reducing the number of parameters. KANs excel in tasks like data fitting and solving partial differential equations (PDEs), making them suitable for scientific applications, and they can prevent catastrophic forgetting. However, KANs have slower training speeds while it could offer superior performance and flexibility compared to MLPs [12].

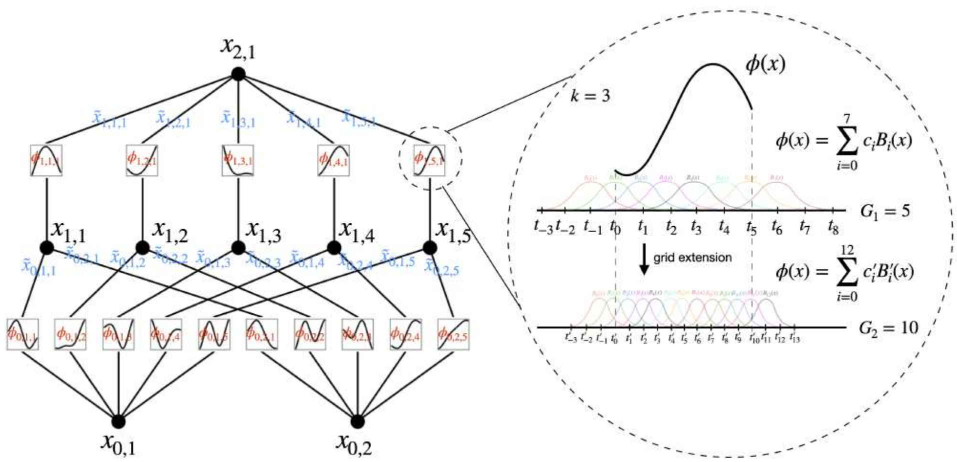


*Figure 1: The overall architecture of Kolmogorov-Arnold Networks (KANs)*

### 3.3 Fast KAN

FastKAN, developed by Ziyao Li, is an accelerated version of Kolmogorov-Arnold Networks (KANs), which traditionally use B-splines to decompose complex functions into simpler components for more efficient learning. However, the computational inefficiency of B-spline calculations led to the development of FastKAN, which replaces B-splines with Gaussian radial basis functions (RBFs), significantly simplifying the model without losing accuracy. By incorporating layer normalization to keep inputs within the valid range, FastKAN improves computational speed by 3.33x over an efficient KAN implementation, while maintaining or improving accuracy on tasks like MNIST classification. This makes FastKAN a faster and simpler alternative to KANs, using Gaussian RBFs for function approximation [19].

### 3.3 FasterKAN

In the FasterKAN model, the goal is to experiment with different bases, exponent values, h values, and the removal of the SiLU activation function. The Reflectional Switch Activation Function (RSWAF) has shown the most promise, as it can approximate the behavior of B-spline basis functions and is computationally efficient with uniform grids. FasterKAN also adopted Layer Normalization which could reduce the need to adjust the grids. Additionally, the latest version includes an experimental feature that allows the inverse of the denominator (1/h) to be a learnable parameter, providing more flexibility in the model [20].

Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

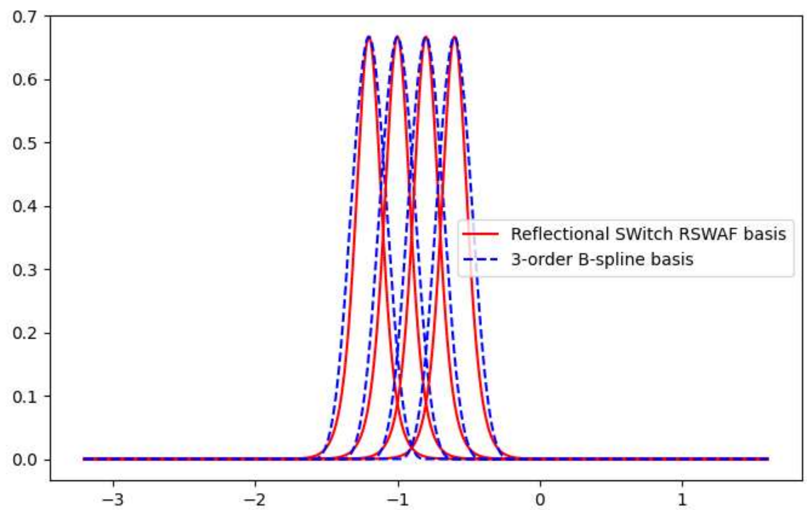


*Figure 2: Reflectional Switch RSWAF basis function from FasterKAN [20]*

### 3.5 Data Description

We obtained a dataset from the European Patent Office (EPO) for the experiment, classified according to the Cooperative Patent Classification (CPC) system. The CPC system categorizes patents into nine broad sections: A (Human Necessities), B (Performing Operations; Transporting), C (Chemistry; Metallurgy), D (Textiles; Paper), E (Fixed Constructions), F (Mechanical Engineering; Lighting; Heating; Weapons; Blasting), G (Physics), H (Electricity), and Y (General Tagging of New or Cross-Sectional Technologies like climate change mitigation). However, we focused specifically on patents classified under sections A (Human Necessities), B (Performing Operations; Transporting), and C (Chemistry; Metallurgy) of the CPC system since we aim to compare the KAN-based classifier to the MLP-based one [21]. The word cloud visualizations highlight the most common terms in different patent categories, each shown with a distinct color. The A cloud (Figure 3) in blue emphasizes terms related to human needs. The B cloud (Figure 4) in green showcases terms about operations and transport. The C cloud (Figure 5) in red reveals chemical and metallurgical terms.

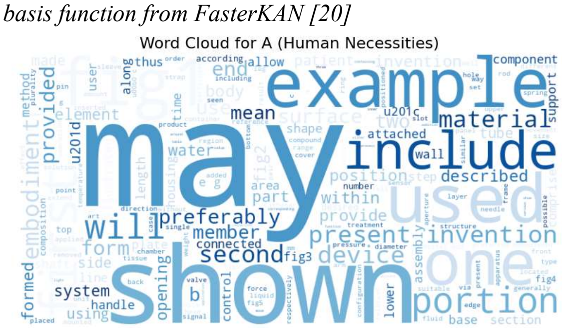


*Figure 3: Blue-themed word cloud for patents addressing human necessities*


*Figure 4: Green-themed word cloud for patents on operations and transport*

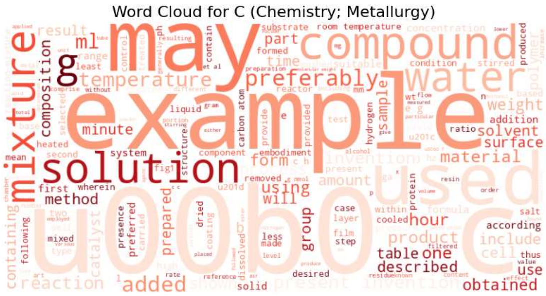


*Figure 5: Red-themed word cloud for patents related to chemistry and metallurgy*

## 4. RESULTS

We used an L4 GPU from Google's Colab environment to speed up the computations. First, we configure the device to use the GPU when available. We then loaded the tokenizer and SimCSE's pre-trained model. Next, we divided the dataset into three parts: training, validation, and testing. The data was divided into a stratified 6:2:2 ratio to ensure a class balance between splits. Initially, the dataset was split into 60% training and 40% temporary data. The temporary set was divided equally to form the validation and test sets. This enabled balanced and representative data splits during the training, validation, and testing stages.

First, MLP and KAN-based classifiers were applied to the provided datasets. For both models, the input dimension was set to 768, which corresponds to the size of the embedding. The model comprises three hidden layers with sizes of 512, 256, and 128, and it is intended to categorize three classes. The loss function was set to CrossEntropyLoss, and the optimizer was Adam, with a learning rate of 1e-4. The MLPClassifier obtained an accuracy of 75.83%, whereas KAN achieved 75.91%. Furthermore, the FastKAn and FasterKAN yielded 72.37% and 73.75%, respectively, demonstrating that the KAN's performance is quite like the MLP.

*Table 1. Accuracy scores of different models: MLP, KAN, FastKAN, and FasterKAN*

| Models | Accuracy Score (%) |
|---|---|
| MLP | 75.83 |
| KAN | 75.91 |
| FastKAN | 72.37 |
| FasterKAN | 73.75 |

In the experiments we conducted with the KAN algorithm, we investigated the effect of lowering the hidden dimension sizes over three layers, utilizing combinations like [128, 64, 32], [64, 32, 16], and [32, 16, 8]. By reducing these sizes, we considerably decreased the number of parameters in the KAN model. For example, the [32, 16, 8] setup generated just 25,411 total parameters, proving the model's efficiency. The equivalent accuracy ratings were 74.89%, 75.73%, and 74.84%, demonstrating that even with fewer dimensions, the model remained competitive.

Additionally, we investigated the role of key hyperparameters like grid_size and spline_order, which influence the number of grids representing the spline and the order of the spline, respectively. By setting the spline_order to 7 and grid_size to 32, the [32, 16, 8] configuration achieved a slightly improved accuracy of 75.12% without increasing the number of parameters. This indicates that KAN can perform well without extensive hyperparameter tuning, offering a more cost-efficient alternative to traditional MLP classifiers, both in terms of computational resources and tuning efforts.

Overall, our findings illustrate KAN's ability to lower computational costs while retaining great performance, making it a more efficient alternative than MLP classifiers, especially where parameter optimization is a priority.

*Table 2. Comparison of model accuracy, hidden dimensions, and total parameters across different layer configurations*

| Hidden Dimensions (Layers) | Total Parameters | Accuracy |
|---|---|---|
| [128, 64, 32] | 1,010,048 | 74.89% |
| [64, 32, 16] | 505,024 | 75.73% |
| [32, 16, 8] | 252,512 | 74.84% |
| [32, 16, 8](tuned) | 252,512 | 75.12% |

In addition to experimenting with the KAN method, we used numerous classic machine learning models to compare. The findings provide a benchmark for comparing the KAN model's efficacy to other frequently used classifiers. These findings reveal that the KAN model, with a tuned accuracy of 75.12%, exceeds all the machine learning models tested, including XGBoost and Logistic Regression, both of which are benchmarks. This emphasizes KAN's promise as an efficient and accurate alternative to existing approaches, particularly given its lower computing cost and fewer parameters. Table 3 lists detailed accuracies. The studies reveal that the KAN model, with just 25,411 parameters, achieved an accuracy of 74.84%, which increased to 75.12% after modifying hyperparameters such as spline_order and grid_size. When compared to standard machine learning models, KAN beat a variety of machine learning models, highlighting KAN's strength not just in terms of accuracy but also in lower computing needs. These findings demonstrate that KAN is an efficient and effective alternative to conventional machine learning models, making it an appropriate solution for applications that need both performance and resource efficiency.

*Table 3. Accuracy scores of different models: Logistic Regression, Random Forest, XGBoost, and Extra Trees*

| Models | Accuracy Score (%) |
|---|---|
| Logistic Regression | 74.89 |
| Random Forest | 72.42 |
| XGBoost | 74.33 |
| Extra Trees | 72.24 |

## 5. DISCUSSION

Our investigation provides a major contribution because it is the first to use the KAN algorithm for patent categorization. While KAN is a relatively new product, it has already been used to test its performance on a variety of tasks. Previous applications include vision datasets like MNIST and ImageNet for classification, remote sensing, time-series forecasting, segmentation, and medical datasets [22][23][24][25][26][27][28]. However, to our knowledge, this is the first time KAN has been explored within the context of patent classification. By extending KAN to this new area, we not only broaden the algorithm's utility but also propose the direction for future study into its possibilities for additional text-based categorization problems. Furthermore, while existing studies demonstrated improvements in patent classification, they relied on large Transformer models, which may overburden memory. Our work differs by introducing KAN, which reduces the parameter count while maintaining competitive accuracy however, there are still limits to our existing technique. However, there still exists a limitation on our study; we were unable to fully capitalize on two of KAN's primary strengths: its ability to prevent catastrophic forgetting during fine-tuning, and its interpretability, which can assist address the black box problem typically associated with deep learning models. For example, KAN's interpretability might lead to more transparent decision-making processes, providing insights into why certain classifications were selected, which is extremely advantageous in areas where comprehending model conclusions is critical.

## 6. CONCLUSION

The present study exhibited the initial implementation of the KAN algorithm for patent classification. The study concentrated on patents classified into three distinct parts of the Cooperative Patent Classification (CPC) system. Our tests showed that KAN outperformed conventional machine learning models like MLP, XGBoost, and Logistic Regression, albeit with fewer parameters and less computing complexity. Three distinct KAN types—KAN, FastKAN, and FasterKAN—were used. Additionally, we investigated several KAN configurations and discovered that, while keeping good accuracy, lowering hidden dimension sizes greatly reduced the number of parameters. In particular, the accuracy of the [32, 16, 8] configuration was 74.84%; however, after adjusting important hyperparameters such as spline_order and grid_size, the accuracy increased to 75.12%. This demonstrates KAN's ability to deliver strong performance without extensive hyperparameter tuning, making it a cost-efficient alternative to other models. Compared to machine learning models like MLP (75.83%), Random Forest, and XGBoost, the KAN model consistently outperformed them in accuracy and efficiency. These results highlight KAN's potential in patent classification tasks, where both model performance and resource efficiency are critical. Our approach could not fully make use of two of KAN's distinctive qualities, despite its promising performance: its interpretability, which might help minimize the black box problem frequently associated with deep learning models, and its ability to prevent catastrophic forgetting during fine-tuning. To increase KAN's usefulness in text-based tasks like patent categorization and others where interpretability and flexibility are critical, future studies may concentrate on examining these facets of the technology. Finally, our approach creates new opportunities for investigating KAN's wider application in text-based classification tasks and expands its applicability into the patent classification arena. In this arena, KAN presents a possible alternative to conventional machine learning methods because of its proven efficiency and performance.

In conclusion, this study indicates KAN's efficiency and potential as a reliable and efficient substitute for conventional machine learning methods for patent categorization. With future improvement and broader applications, KAN might play a critical role in creating automated solutions for managing intellectual property and other text-based categorization activities. The findings of this study lay the path for future patent categorization advancements, making it an important step toward developing smarter, more efficient NLP models for legal and technical areas.

## REFERENCES:

[1] Mayerhoff ZDVL, Von Der Weid I, Valladão ABG. Repositories for patented and safeguarded material. In: Springer eBooks [Internet]. 2013. p. 305–14. Available from: https://doi.org/10.1007/978-3-642-30194-0_116.

[2] Eisenberg RS. Patents and the Progress of Science: Exclusive rights and experimental use. The University of Chicago Law Review [Internet]. 1989 Jan 1;56(3):1017. Available from: https://doi.org/10.2307/1599761.

[3] Mireles MS. An examination of patents, licensing, research tools, and the tragedy of the anticommons in biotechnology innovation. U. Mich. JL Reform. 2004;38:141.

[4] Hanel P. Intellectual property rights business management practices: A survey of the literature. Technovation [Internet]. 2006 Aug 1;26(8):895–931. Available from: https://doi.org/10.1016/j.technovation.2005.12.001

[5] Li S, Hu J, Cui Y, Hu J. DeepPatent: patent classification with convolutional neural networks and word embedding. Scientometrics [Internet]. 2018 Sep 6;117(2):721–44. Available from: https://doi.org/10.1007/s11192-018-2905-5.

[6] Kenton JD, Toutanova LK. Bert: Pre-training of deep bidirectional transformers for language understanding. InProceedings of naacL-HLT 2019 Jun 2 (Vol. 1, p. 2).

[7] Radford A. Improving language understanding by generative pre-training.

[8] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research. 2003;3(Jan):993-1022.

[9] Le Q, Mikolov T. Distributed representations of sentences and documents. InInternational conference on machine learning 2014 Jun 18 (pp. 1188-1196). PMLR.

[10] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics [Internet]. 2017 Dec 1;5:135–46. Available from: https://doi.org/10.1162/tacl_a_00051.

[11] Vaswani A. Attention is all you need. Advances in Neural Information Processing Systems. 2017.

[12] Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, Hou TY, Tegmark M. Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756. 2024 Apr 30.

[13] Lee JS, Hsiang J. Patent classification by fine-tuning BERT language model. World Patent Information [Internet]. 2020 Jun 1;61:101965. Available from: https://doi.org/10.1016/j.wpi.2020.101965

[14] Roudsari AH, Afshar J, Lee CC, Lee W. Multi-label patent classification using attention-aware deep learning model. In2020 IEEE international conference on big data and smart computing (BigComp) 2020 Feb 19 (pp. 558-559). IEEE.

[15] Chikkamath R, Parmar VR, Otiefy Y, Endres M. Patent Classification Using BERT-for-Patents on USPTO. InProceedings of the 2022 5th International Conference on Machine Learning and Natural Language Processing 2022 Dec 23 (pp. 20-28).

[16] Roudsari AH, Afshar J, Lee W, Lee S. PatentNet: multi-label classification of patent documents using deep learning based language understanding. Scientometrics [Internet]. 2021 Dec 18;127(1):207–31. Available from: https://doi.org/10.1007/s11192-021-04179-4.

[17] Yoo Y, Heo TS, Lim D, Seo D. Multi label classification of Artificial Intelligence related patents using Modified D2SBERT and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Sentence Attention mechanism. arXiv preprint arXiv:2303.03165. 2023 Mar 3.

[18] Gao T, Yao X, Chen D. Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821. 2021 Apr 18.

[19] Li Z. Kolmogorov-arnold networks are radial basis function networks. arXiv preprint arXiv:2405.06721. 2024 May 10.

[20] AthanasiosDelis. GitHub - AthanasiosDelis/faster-kan: Benchmarking and Testing FastKAN [Internet]. GitHub. Available from: https://github.com/AthanasiosDelis/faster-kan

[21] European Patent Classification [Internet]. Kaggle. 2024. Available from: https://www.kaggle.com/datasets/gauravduttakiit/european-patent-classification

[22] Wang J, Cai P, Wang Z, Zhang H, Huang J. CEST-KAN: Kolmogorov-Arnold Networks for CEST MRI Data Analysis. arXiv preprint arXiv:2406.16026. 2024 Jun 23.

[23] Cheon M. Demonstrating the efficacy of kolmogorov-arnold networks in vision tasks. arXiv preprint arXiv:2406.14916. 2024 Jun 21.

[24] Cheon M, Mun C. Combining KAN with CNN: KonvNeXt's Performance in Remote Sensing and Patent Insights. Remote Sensing. 2024 Sep 14;16(18):3417.

[25] Vaca-Rubio CJ, Blanco L, Pereira R, Caus M. Kolmogorov-arnold networks (kans) for time series analysis. arXiv preprint arXiv:2405.08790. 2024 May 14.

[26] Drokin I. Kolmogorov-Arnold Convolutions: Design Principles and Empirical Studies. arXiv preprint arXiv:2407.01092. 2024 Jul 1.

[27] Kashefi A. Kolmogorov-Arnold PointNet: Deep learning for prediction of fluid fields on irregular geometries. arXiv preprint arXiv:2408.02950. 2024 Aug 6.

[28] Cheon M, Mun C. Combining KAN with CNN: KonvNeXt's Performance in Remote Sensing and Patent Insights. Remote Sensing. 2024 Sep 14;16(18):3417.