https://doi.org/10.51408/issi2025_030

Balancing Accuracy and Explainability: An Ensemble-KAN Model for Patent Grant Prediction

Jing Shi[1], Xinyi Peng[2], Xizhen Qiao[3], Ye Chen[4], Xiao Liu[5]

[1]*shijing11@smail.nju.edu.cn*
Laboratory of Data Intelligence and Interdisciplinary Innovation, School of Information Management, Nanjing University, No. 163 Xianlin Avenue, Nanjing (China)

[2]*202205571025@smail.xtu.edu.cn*
School of Automation and Electronic Information, Xiangtan University, Yanggulang, western suburb of Yuhu District, Xiangtan (China)

[3]*202205571036@ smail.xtu.edu.cn*
School of Automation and Electronic Information, Xiangtan University, Yanggulang, western suburb of Yuhu District, Xiangtan (China)

[4]*chenye@nju.edu.cn*
School of Information Management, Nanjing University, No. 163 Xianlin Avenue, Nanjing (China)

[5]*liuxiao730@xtu.edu.cn*
School of Automation and Electronic Information, Xiangtan University, Yanggulang, western suburb of Yuhu District, Xiangtan (China)

Abstract

A patent is valuable intellectual property only when granted and held for the long term, and patent grant prediction is a potential strategy for reducing the uncertainty of innovation. Existing machine learning-based prediction models lack interpretability, making it difficult to effectively mitigate innovation risks. This study proposes a novel model for patent prediction that combines high predictive accuracy with strong interpretability. (1) First, we employ the KAN model for prediction, which replaces traditional neural networks with spline functions, endowing the model with interpretability and the ability to generate formula. (2) Additionally, we introduced ensemble learning to enhance the performance of the KAN model, resulting in the development of the EN-KAN model. We tested the model on Electronic Communications datasets and demonstrated strong performance while maintaining high interpretability. EN-KAN directly generates mathematical formulas, providing a more accurate and intuitive representation of the impact of different factors on the prediction results. (3) Moreover, our study reveals that factors at the examiner-level and the patent-level have the greatest impact on patent grants.

Introduction

Patents operate on a fundamental principle of exchanging public disclosure for legal protection, offering innovators a pathway to secure exclusivity, establish technological monopolies, and generate economic returns (Nordhaus, 1969). However, the failure of a patent application to be granted can impose substantial losses on innovators, not only in terms of the time, resources, and financial investment expended but also through the unintended exposure of proprietary technologies, potentially forfeiting competitive advantages (Millar et al., 2018). Early prediction of patent grant outcomes can empower innovators by improving the likelihood of success, informing strategic decision-making in the application process, and guiding investment priorities. Although patent laws mandate that applications meet the criteria of novelty, inventiveness, and utility (Liegsalz & Wagner, 2013), these attributes are often subject to complex and multifaceted influences. The interpretive judgments of patent examiners further complicate the process, as their decisions are neither fully transparent nor easily predictable. Combined with the lengthy application cycles and extensive documentation requirements, these challenges make early prediction of patent grant outcomes a complex and urgent challenge.

To address this challenge, prior research has explored various approaches, including traditional statistical methods and heuristic analyses, to predict patent grant probability (Drivas & Kaplanis, 2020; Gans et al., 2008; D. Yang, 2008; Yao & Ni, 2023). However, these methods often suffer from limitations, such as oversimplification of complex interactions among influencing factors. Machine learning (ML) approaches, which can extract latent patterns from large-scale empirical data, have increasingly been employed to tackle this problem. For instance, ML models have been used to predict the likelihood of innovation failure by identifying significant predictors within voluminous datasets (Yao & Ni, 2023). Despite their promising predictive accuracy, the inherent "black box" nature of most ML algorithms has raised concerns regarding their interpretability, leading to skepticism about their conclusions. This lack of transparency has hindered the dissemination and practical application of ML-based findings. While some researchers have sought to enhance interpretability by appending post hoc explanation models, such methods often yield explanations that are either overly generalized or insufficiently specific to the contexts of patent examinations. Furthermore, prior studies have highlighted the variability in patent grant outcomes across different patent authorities and technological fields (Alcácer et al., 2009), emphasizing that influencing factors are not universally consistent but contingent on the specific jurisdiction and field of innovation. How these contextual factors influence patent grants remains unclear.

This study proposes a novel interpretable machine learning model, Ensemble Kolmogorov-Arnold Network (EN-KAN), to investigate the factors influencing the early prediction of patent grant. This model is designed to achieve two primary research objectives. First, unlike conventional ML models that rely on post hoc interpretability enhancements, KAN incorporates interpretability as a core feature of its design, employing knowledge embeddings and structured influence analysis (Liu et al., 2024). By comparing KAN with several benchmark algorithms, we demonstrate its efficacy and provide visualized explanations of its findings. Our results identify critical predictors of patent grant success elucidating their underlying mechanisms by formula. Second, we examine the differential impacts of patent examination authorities, uncovering jurisdiction-specific patterns and highlighting the role of institutional and procedural variations in shaping grant.

The contributions of this study are twofold. First, we introduce a self-explanatory model that accurately predicts patent grant probabilities while identifying key determinants of patent success. By integrating interpretable methodologies, this research advances the understanding of patent grant processes and provides a robust framework for examining the drivers of patent approval. Second, this study offers comparative insights across diverse technological domains and patent jurisdictions, addressing gaps in the literature regarding the contextual variability of influencing factors. These findings have practical implications for both patent applicants and examiners. For innovators, the results offer actionable guidance for crafting application strategies to maximize the probability of success and minimize uncertainties, ultimately enhancing the commercial value of patents. For patent examiners, the insights enable optimization of examination workflows, improving efficiency by focusing on the most impactful variables. Through these contributions, this research not only advances academic discourse but also supports evidence-based decision-making in the patent ecosystem.

Literature review

The influencing factors of patent grant

The factors influencing patent grant can be categorized into five levels: patent, application, applicant and inventor, examiner, and other factors. Table 1 provides a summary of these levels and their corresponding factors.

Patent Level focuses on the intrinsic characteristics of the innovation, including novelty, innovativeness, and utility. Novelty and innovativeness are fundamental traits of patents and serve as key drivers of technological breakthroughs, playing a decisive role in patent grant. Prior studies have employed various measures to assess novelty, such as the number of International Patent Classification (IPC) categories involved (Harhoff & Wagner, 2009; Liegsalz & Wagner, 2013), the number of references cited (G. Yang et al., 2023), and the Herfindahl index (a measure of concentration) of cited patent classes. Emerging research highlights the role of scientific knowledge in technological innovation, finding that patents utilizing more scientific knowledge exhibit higher innovativeness (C. Lee et al., 2018). Utility reflects the practical applicability or industrial use of an invention. A common metric for utility is the generality index, which measures the breadth of subsequent inventions benefiting from the patent (Niosi, 2006). Public procurement patents tend to have higher generality (Raiteri, 2018) and patents with greater generality demonstrate sustained competitiveness (P.-C. Lee, 2021).

Application Level emphasizes the quality of the application documents, including indicators such as the number of pages, titles, abstracts, claims, and the length of claims. Claims delineate the scope of the patent. While a higher number or broader scope of claims increases examination complexity and may prolong the review process (Liegsalz & Wagner, 2013), research also suggests a positive relationship between the number of claims and patent grant. A patent with numerous independent claims is perceived as robust in legal terms (Harhoff & Wagner, 2009; Y.-G. Lee & Lee, 2010). The word count of the first claim is another commonly used indicator, reflecting the patent's protection scope (Sampat & Williams, 2019). Moreover, particular attention is given to Patent Cooperation Treaty (PCT) applications. PCT filings, which enable the extension of patent protection to multiple countries while minimizing costs and complexities, positively impact patent grant rates (Harhoff & Wagner, 2009)

Applicant and inventor level explores the influence of applicant and inventor characteristics, such as quantity, nationality, and historical experience. Analysis of USPTO data reveals that U.S. nationality increases the likelihood of patent approval, whether as applicants or inventors (Drivas & Kaplanis, 2020). Some patent office's exhibit preferential treatment toward domestic applicants (D. Yang, 2008), leading to higher granting probabilities for local inventors. Additionally, in areas of technological specialization, domestic inventors show stronger positive effects (Webster et al., 2014). However, excessive domestic collaboration may reduce the probability of patent grants. In contrast, international collaborations tend to confer advantages (Guellec & de la Potterie, 2000). Applicants with prior success in securing patents are more likely to achieve subsequent grants (Liegsalz & Wagner, 2013). Persistent efforts in filing patents also significantly enhance granting probabilities (Drivas & Kaplanis, 2020).

Examiner Level addresses the role of patent offices and examiners. Decisions on patent grant are heavily influenced by individual examiners (Lemley & Sampat, 2012), and examiner biases can distort patent allocation. For instance, examiners may be less likely to grant patents to inventors outside their social group (Desai, 2019). They also demonstrate a tendency to approve patents for applicants of the same gender (Shen & Zingg, n.d.). Examiners' behaviors are influenced by their peers, particularly when in close physical proximity (Frakes & Wasserman, 2021). These dynamics underscore the subjective aspects of the patent examination process. *Other Factors*. Additional factors include the technological field, patent application strategies, and the number of related patent filings. Comparative analyses of 30 technological fields reveal significant differences in patent review durations across domains (Liegsalz & Wagner, 2013). A Difference-in-Differences (DID) analysis by Bekkers demonstrated that increased awareness of earlier related technologies among examiners reduces patent grant probabilities (Bekkers et al., 2020).

Table 1. The relevant influencing factors of patent grant.

Dimension	*Factors*	*Sources*
Patent level	Novelty	Harhoff & Wagner, 2009; Liegsalz & Wagner, 2013; C. Lee et al., 2018; G. Yang et al., 2023
	Utility	Niosi, 2006; Raiteri, 2018; P.-C. Lee, 2021
Application level	the number of pages of application file	Yao & Ni, 2023
	the number of claims	Harhoff & Wagner, 2009; Y.-G. Lee & Lee, 2010; Liegsalz & Wagner, 2013; Marco et al., 2019
	the word count of title	Yao & Ni, 2023
	the word count of abstract	Yao & Ni, 2023
	the word count of claims	Marco et al., 2019; Sampat & Williams, 2019
	whether submit PCT application or not	Harhoff & Wagner, 2009
Applicant & inventor level	whether local applicant/inventor or not	D. Yang, 2008; Guellec & de la Potterie, 2000; Drivas & Kaplanis, 2020

	the number of applicants/inventors	C. Lee et al., 2018; Yao & Ni, 2023
	applicant's experience	Harhoff & Wagner, 2009; Liegsalz & Wagner, 2013
	the nationality of applicant	D. Yang, 2008; Webster et al., 2014; Drivas & Kaplanis, 2020
Examiner level	Examiner	Lemley & Sampat, 2012; Desai, 2019; Shen & Zingg, n.d.; Frakes & Wasserman, 2021
	the country of prior right	Guellec & de la Potterie, 2000; Yao & Ni, 2023
	The duration of examine	Harhoff & Wagner, 2009
Others	technological field	Guellec & de la Potterie, 2000; Liegsalz & Wagner, 2013
	the strategy of application	Guellec & de la Potterie, 2000
	the number of relevant applications	Bekkers et al., 2020

Interpretable Machine Learning Research

Interpretable Machine Learning (IML) seeks to provide insights into machine learning models that are understandable to humans. IML encompasses understanding data, the internal structures of models, and interpreting the results produced by these models (Allen et al., 2024; Lipton, 2018). The applications of IML span various stages of the machine learning pipeline, including the explanation of input data, the elucidation of model mechanisms, and the interpretation of output outcomes. Explanation techniques in IML can be categorized along three dimensions: intrinsic interpretability versus post-hoc interpretability, model-specific explanations versus model-agnostic explanations, and global explanations versus local explanations.

Intrinsic Interpretability vs. Post-hoc Interpretability. Intrinsic interpretability refers to the inherent transparency of a model, allowing users to understand its behavior directly through the training process. Examples of intrinsically interpretable models include decision trees (Costa & Pedreira, 2023), additive models (Agarwal et al., 2021), and models enhanced with regularization techniques such as sparsity (Hoefler et al., 2021) or smoothness (Crawshaw et al., 2022), which naturally provide high levels of interpretability (Rudin, 2019). Recent advancements have further improved the intrinsic interpretability of deep neural networks by integrating prototypes or specific interpretability constraints into their final layers (Dong et al., 2017). In contrast, post-hoc interpretability involves applying additional methods to interpret the model or its outputs after the training phase. These methods include feature importance scoring based on backpropagation and Local Interpretable Model-agnostic Explanations (LIME) (Molnar, 2020). LIME, for example, constructs simplified surrogate models around specific input points to approximate the behavior of complex models, making it applicable to various pre-trained models and providing additional insights into their decision-making processes (Molnar, 2020).

Model-specific Explanations vs. Model-agnostic Explanations. Model-specific explanation methods are designed for types of models and do not generalize well across different model architectures. Examples include regression coefficients in generalized linear models (Rong & Bao-Wen, 2018), feature importance scores in tree-based models (Zhou & Liu, 2021), and techniques such as backpropagation or layer-wise relevance propagation in deep learning (Zhou & Liu, 2021). Conversely, model-agnostic explanation methods are applicable to a wide range of model types, offering a unified framework for interpretation. Common model-agnostic methods include Shapley values (Fryer et al., 2021),feature permutation (Covert et al., 2021), feature masking (J. Dai et al., 2015), and LIME (Molnar, 2020), which provide consistent explanatory effects across different models. It is important to note that model-specific explanation methods do not necessarily provide intrinsic interpretability. For instance, feature importance scores in decision trees and feature attribution via backpropagation are model-specific yet fall under post-hoc interpretability. Most model-agnostic explanation methods are inherently post-hoc in nature.

Global Explanations vs. Local Explanations. Global explanations aim to reveal the overall structure of the model and the general importance of all features. Examples include coefficients in linear or additive models, feature importance scores in tree-based models, and global feature attribution methods, which reflect each feature's role in the model's overall predictions. On the other hand, local explanations focus on specific inputs or subsets of inputs, providing targeted interpretations. For example, LIME and saliency map methods concentrate on individual test instances or the significant features of specific observations (Ribeiro et al., 2016). In unsupervised learning, local embedding methods such as t-SNE (t-distributed Stochastic Neighbor Embedding) (Van der Maaten & Hinton, 2008) and UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2018) analyze data patterns and relationships within specific neighborhoods to explain local data.

Despite significant advancements in enhancing model transparency, current IML approaches exhibit several limitations. Firstly, there is considerable technical heterogeneity among existing methods, with each approach typically catering to specific interpretative needs and lacking generalizability. This fragmentation leads to inconsistent explanatory outcomes across different methods, thereby complicating users' understanding of model behavior. For instance, some methods emphasize global feature importance while others focus on local instance explanations; employing multiple methods simultaneously may yield conflicting conclusions. Additionally, varying assumptions and focal points among different methods result in a lack of unified evaluation standards, undermining the reliability and consistency of explanations. Such inconsistencies not only increase the difficulty for users to comprehend and trust the models but also risk misleading decision-making processes, thereby reducing the practical effectiveness of interpretability techniques. Consequently, there is an urgent need to develop more unified and coordinated interpretability frameworks to mitigate methodological discrepancies, enhance the consistency of explanatory outcomes, and bolster user trust.

Methodology

Data collection

We select patents in the fields of Electronic Communications (EC) for empirical analysis and comparison due to their pivotal roles in driving technological progress and economic growth. EC, as a mature and highly competitive sector, presents unique challenges in balancing innovation with the standardization of technologies. Invention patents are selected for analysis due to their emphasis on groundbreaking innovations and their rigorous examination standards. Invention patents are emphasized because they represent substantive technological innovations and generally possess higher overall market value. Moreover, the examination process for invention patents is more rigorous, with clearer and more consistent decision-making criteria, making them more predictable. Finally, invention patents offer higher data quality and richer textual information, making them particularly well-suited for training patent grant prediction models. The process of obtaining an invention patent typically involves several key stages, beginning with the filing of a patent application. After filing, the application undergoes a formal examination and the substantive examination phases. If approved, the patent is granted and published, providing the inventor with exclusive rights to the invention, typically having a protection period of up to 20 years.

The patent examination process generally spans 2 to 5 years, with an average duration of approximately 4 years, supporting the selection of a five-year observation window. Thus, invention patent applications in 2017 of the EC fields are chosen, enabling an evaluation of whether these patents were successfully granted within 5 years. Our patent data are collected from PATSTAT (Worldwide Patent Statistical Database) and the final dataset contains 299,912 patent applications (137,257 patents are granted).

Influencing factors extraction and description

The grant status of a patent is operationalized as a binary variable, where granted patents are assigned a value of 1, and non-granted patents are assigned a value of 0. This study selects patent features as influencing factors at five levels, and the final factors and measurement methods are detailed in Table 2.

Table 2. Influencing factors selected.

Dimension	*Factors*	*Measurement*
Patent level	backward_citation	The number of backward citations.
	family_size	The family size of focal patent.
	nb_claims	The number of claims.
	nb_title_char	The word count of patent applications' title.
	nb_abstr_char	The word count of patent applications' abstract.
	is_PCT	Whether the patent is filed as a PCT application: 1 for Yes, 0 for No.
Applicant & inventor level	nb_inventors	The number of inventors.
	nb_applicants	The number of applicants.
	nb_applications	The total number of patent applications of all applicant and inventors of focal patent in 2017.
	ratio_granted	The granting rate of the applicant's patent applications in 2016.
	ctry_first_applicant	The nationality of the first applicant.
	nb_local_applicant	The number of local applicants.
	nb_foreign_applicant	The number of foreign applicants.
	nb_local_inventor	The number of local inventors.
	nb_foreign_inventor	The number of foreign inventors.
	appln_auth	The examination authority of the focal patent.
Examiner level	int_phase	Whether the patent entered the international phase: Y = 1; N = 0.
	reg_phase	Whether the patent entered the regional

		phase: Y = 1; N = 0.
	nat_phase	Whether the patent entered the national phase: Y = 1; N = 0.
	duration	The number of years from the initial patent application filing to the final decision.
Others	tech_field	The 3_digit IPC code which focal patent belongs to.
	nace_code	The NACE[1] code of focal patent.
	nb_relevant_patent	The number of relevant applications[2].

Table 3. The patent features' description.

Factors	*Mean*	*SD*	*Factors*	*Mean*	*SD*
backward_citation	7.68	38.39	nb_foreign_applicant	0.90	0.58
family_size	3.84	4.79	nb_local_inventor	0.45	1.29
nb_claims	13.26	37.4	nb_foreign_inventor	2.39	2.17
nb_title_char	8.5	4.24	appln_auth	NA	NA
nb_abstr_char	134.64	52.41	int_phase	0.37	0.48
is_PCT	0.26	0.44	reg_phase	0.08	0.27
nb_inventors	2.78	2.12	nat_phase	0.81	0.39
nb_applicants	1.08	0.48	duration	1.73	1.11
nb_applications	1095.9	1956.03	tech_field	NA	NA
ratio_granted	0.47	0.35	nace_code	NA	NA
ctry_first_applicant	NA	NA	nb_relevant_patent	0	0.03
nb_local_applicant	0.20	0.47			

Model construction

This paper proposes an ensemble learning approach based on the ENsemble Kolmogorov-Arnold Network (EN-KAN) for predicting patent grant outcomes. The proposed method enhances prediction accuracy and model generalization through systematic data preprocessing, the design and training of the KAN model, and the implementation of an ensemble learning strategy.

[1] NACE: Statistical Classification of Economic Activities in the European Community is the statistical classification system of economic activities in the European Union (EU).

[2] Technical relations are "priority-like" relations between applications which have been detected by EPO examiners, but which have not been published by a patent office.

(a) Base model

The Ensemble-KAN utilizes the Kolmogorov-Arnold Network (KAN) as the foundational model for patent grant prediction. KANs, based on the Kolmogorov-Arnold theorem, are emerging machine learning architectures recognized as powerful alternatives to multilayer perceptrons (MLPs). The KAN network exhibits significant advantages over traditional MLPs in several key aspects, particularly in weight parameter representation and function approximation methods.

According to the Kolmogorov-Arnold theorem, for any continuous multivariate real function $f:[0,1]^n \to R$, there exists a set of univariate continuous functions $\{\phi_k\}$ and $\{\psi_{k,i}\}$ such that f can be expressed as a finite nested and summative form:

$$f(x_1,x_2,\dots,x_n) = \sum_{k=1}^{2n+1} \phi_k\left(\sum_{i=1}^{n} \psi_{k,i}(x_i)\right).$$

This theorem theoretically demonstrates that multivariate continuous functions can be decomposed into a weighted sum of univariate nonlinear functions. Unlike traditional MLPs, which employ fully connected linear transformations combined with fixed activation functions, KAN networks represent each channel with learnable univariate nonlinear functions. This alignment with the Kolmogorov-Arnold decomposition enhances the function representation's conformity to the theorem's decomposition principle.

Specifically, the KAN aims to approximate a target function $f(\mathbf{x}) = f(x_1,\dots,x_n)$ as:

$$\hat{f}(x) = \sum_{k=1}^{K} g_k\left(\sum_{i=1}^{n} h_{k,i}(x_i)\right),$$

where $g_k(\cdot)$ and $h_{k,i}(\cdot)$ are learnable univariate nonlinear functions. To enhance the function space's representation capability, KAN networks incorporate learnable B-splines as the base functions, parameterizing both $h_{k,i}$ and g_k. For example, the B-spline basis functions for $h_{k,i}$ are expressed as:

$$h_{k,i}(x_i) = \sum_{j=1}^{J} \alpha_{k,i,j} B_j(x_i).$$

Similarly, for $g_k(u)$:

$$g_k(u) = \sum_{j=1}^{J'} \beta_{k,j} B_j(u),$$

where $\{\alpha_{k,i,j}\}$ and $\{\beta_{k,j}\}$ are trainable parameters. The incorporation of learnable

B-spline activation functions allows the model to adaptively adjust the univariate nonlinear mappings during training, thereby shaping the function forms according to the data distribution characteristics and enhancing the model's ability to capture complex data patterns.

Furthermore, the univariate learnable nonlinear function structure of the KAN network improves model interpretability. Since the function is explicitly decomposed into a finite sum of univariate nonlinear functions, it facilitates the analysis of input variables' individual contributions to the output, providing more intuitive explanations for the decision-making process in the task.

(b) ENsemble Kolmogorov-Arnold Network

In this paper, we propose an *ENsemble Kolmogorov-Arnold Network (EN-KAN)*, by centrally training multiple Kolmogorov-Arnold Network (KAN) models and generating combined prediction results. EN-KAN mitigates individual model biases, significantly enhancing the overall model's generalization capability. The core idea is to leverage the diversity of multiple independently trained KAN models and integrate their predictions through an ensemble decision mechanism to achieve more robust and accurate classification performance.

Figure 1 illustrates the structure of the proposed EN-KAN. The process begins with the data preprocessing stage, which includes three main steps: Data Cleaning, Normalization, and Feature Selection. These steps work together to produce a high-quality training dataset. Once the data is preprocessed, it is fed into the EN-KAN module. This module is made up of several KAN. Each KAN network starts by fitting an explainable spline function to capture the nonlinear patterns in the data. After fitting the spline functions, they are combined to form a complete KAN network. During the prediction phase, each individual KAN network makes its own prediction based on the input data. These predictions are then collected through a voting mechanism, where each KAN network casts a vote for its predicted outcome. Finally, the EN-KAN algorithm uses a Model Ensemble process to merge all the votes from the KAN networks, resulting in the final output. This structure not only enhances the prediction accuracy of the model but also maintains the interpretability of the results.

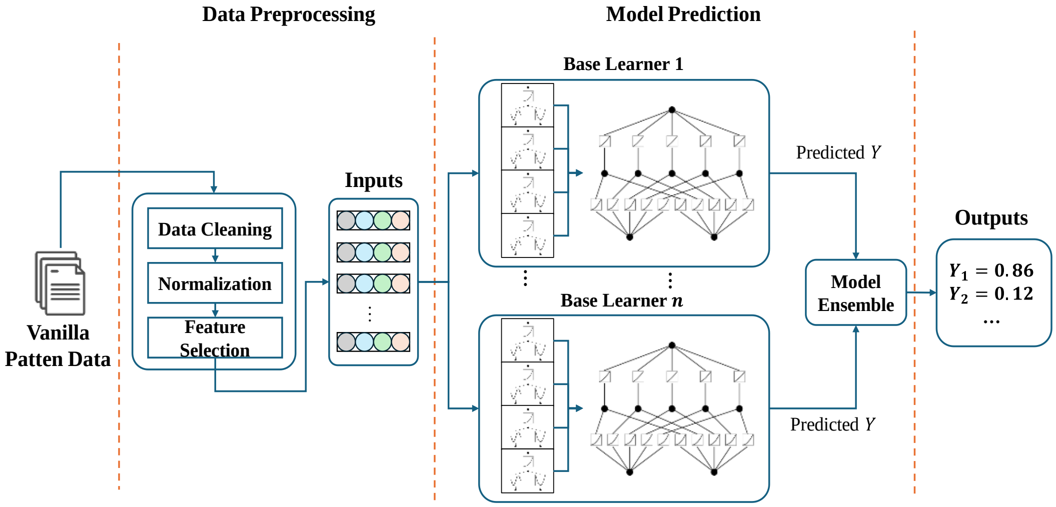

Figure 1. A high-level structure of the proposed EN-KAN.

Specifically, let there be M independent KAN models, each model m characterized by a unique parameter set θ_m. Due to different initializations and the stochastic nature of the training process, the parameter sets θ_m exhibit diversity, which is crucial for the ensemble method to improve generalization.

Formally, for each sample $\mathbf{x_i} \in \mathbf{X}_{\text{ts}}$ in the test dataset, each KAN model m generates a prediction probability vector $\hat{\mathbf{y}}_i^{(m)}$ as follows:

$$\hat{\mathbf{y}}_i^{(m)} = f_m(\mathbf{x_i}; \theta_m),$$

were, $\hat{\mathbf{y}}_i^{(m)}$ represents the predicted probabilities of sample $\mathbf{x_i}$ belonging to each class by model m. For each model m, the predicted class label $\hat{y}_i^{(m)}$ is determined by selecting the class with the highest probability:

$$\hat{y}_i^{(m)} = \arg\max_c \left(\hat{\mathbf{y}}_i^{(m)}\right)_c,$$

where c denotes the class index. The final ensemble prediction label $\hat{y}_i^{(\text{ensemble})}$ is obtained by majority voting among all M models:

$$\hat{\mathbf{y}}_i^{(ensemble)} = \text{mode}\left(\{\hat{\mathbf{y}}_i^{(m)}\}_{m=1}^{M}\right).$$

The mode function returns the class that appears most frequently among the predictions of the individual models. By integrating multiple diverse KAN models, Ensemble-KAN (E-KAN) effectively reduces the risk of overfitting inherent in single models, thereby enhancing the system's overall generalization capability.

(c) Model Training

Ensemble-KAN (E-KAN) optimizes multiple KAN networks collectively. The overall training loss $L_{\text{E-KAN}}$ is defined as the sum of the loss functions of all M models:

$$L_{\text{E-KAN}} = \sum_{m=1}^{M} \mathcal{L}_m ,$$

where $\mathcal{L}_m$ represents the loss function of the m-th KAN model, defined as:

$$\mathcal{L}_m = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{\mathbf{y}}_i^{(m)}\right) + (1 - y_i)\log\left(1 - \hat{\mathbf{y}}_i^{(m)}\right)\right],$$

were, N is the number of samples in the training set, y_i is the true label of the i-th sample, and $\hat{\mathbf{y}}_i^{(m)}$ is the predicted probability by the m-th KAN model for the i-th sample. Thus, the overall training loss can be expressed as:

$$L_{\text{E-KAN}} = -\sum_{m=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{\mathbf{y}}_i^{(m)}\right) + (1 - y_i)\log\left(1 - \hat{\mathbf{y}}_i^{(m)}\right)\right]\right).$$

The objective is to minimize the overall training loss $L_{\text{E-KAN}}$. By optimizing multiple Kolmogorov-Arnold networks simultaneously and employing an ensemble decision mechanism, Ensemble-KAN (E-KAN) effectively enhances model performance and generalization in patent grant prediction tasks, offering a robust and efficient solution.

Result

Prediction results

Table 5 presents a comparison of the performance of our model with other models. The primary evaluation metrics include precision (P), recall (R), and F1-score. Overall, the EN-KAN, Random Forest, and KNN models demonstrated better performance compared to traditional models. EN-KAN model showed best performance, with F1-score 0.89.

Table 5. Results of different models.

Model	**P (%)**	**R (%)**	**F1 (%)**
EN-KAN	0.8946	0.8975	**0.8949**
RandForest	0.8643	0.8640	**0.8638**
KNN	0.8547	0.8546	0.8544
LASSO	0.7125	0.7125	0.7117
Logistics	0.7921	0.7904	0.7894

In the Figure 2, the orange curve represents the ROC curve of the EN-KAN model. The EN-KAN, Random Forest, and KNN models showed strong performance, while the LASSO model performed the worst. The Random Forest model, through the integration of multiple decision trees, effectively handles noise and feature correlations within the data, achieving performance comparable to the EN-KAN model on the dataset. However, the high performance of Random Forest comes at the cost of interpretability, as its results are often considered a "black box." In contrast, EN-KAN strikes an optimal balance between predictive performance and interpretability, making it a more suitable choice for applications requiring both robust predictions and explainable outcomes.

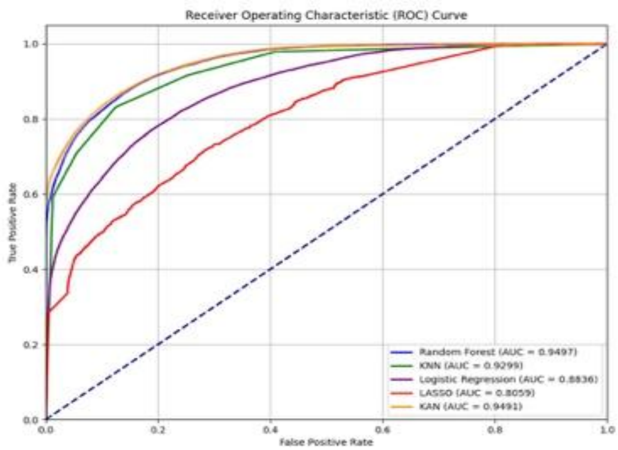

Figure 2. The ROC curve.

Figure 3 illustrates the trade-off between model interpretability and predictive accuracy, helping us understand the relative positions of different machine learning models along these two dimensions. The x-axis represents model interpretability, with models positioned further to the right being more understandable to humans. The y-axis indicates predictive accuracy, with higher positions corresponding to better performance on the patent grant prediction task. Models in the lower right red circle are intrinsically interpretable but demonstrate lower predictive accuracy. Models in the upper left blue region achieve higher accuracy but require post-hoc interpretation methods such as SHAP and LIME to explain their predictions (Lundberg & Lee, 2017; Ribeiro et al., 2016). In contrast, models in the upper right black region—including the EN-KAN proposed in this study and its base model KAN—represent a class of neural network architectures that combine high interpretability with strong performance. These models are inherently interpretable and do not rely on external tools for post-hoc explanations. Among them, KAN provides the most transparent model structure, although its predictive performance is slightly lower than that of Random Forest. After incorporating ensemble learning, EN-KAN not only surpasses RF in accuracy but also offers superior interpretability compared to other models. The green dashed line in the figure denotes the signal-to-noise ratio (SNR), with higher values indicating that the model can more effectively capture underlying patterns, leading to improved accuracy. The transition from models in the red region to those in the blue region reflects the evolution from traditional statistical models to high-performance nonlinear models. While increased SNR supports the performance of such complex models, it often comes at the cost of reduced interpretability. The EN-KAN model introduced in this study seeks to break this trade-off by achieving an optimal balance between interpretability and predictive power.

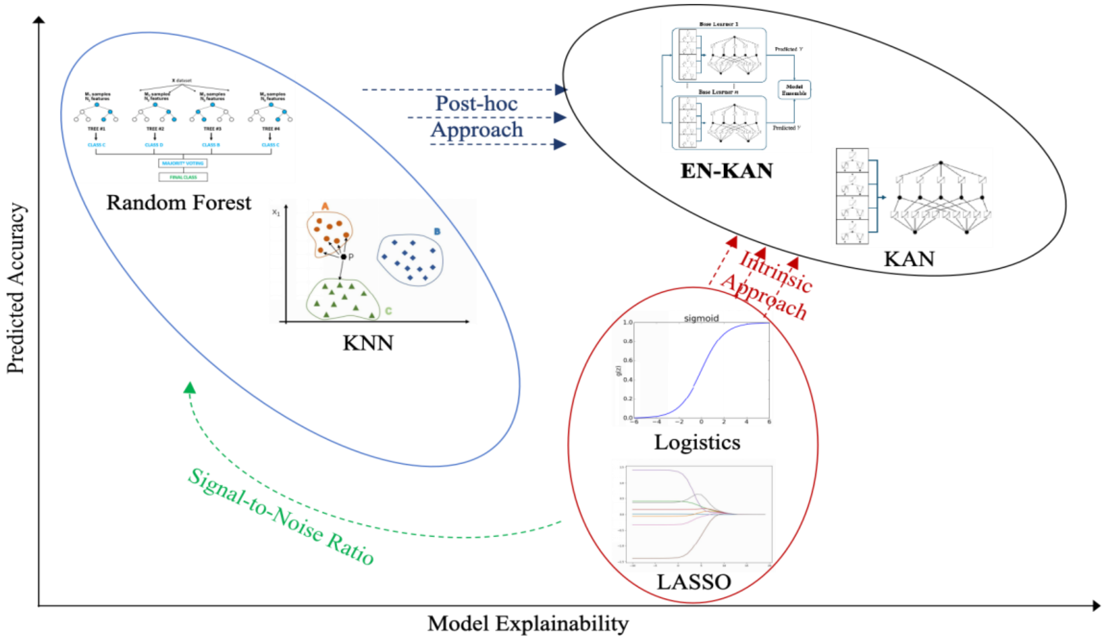

Figure 3. Explainability and predicted accuracy of different models.

Which patents are granted?

For patents in the EC field, the most significant factors are *nb_claims*, *nat_phase*, and *int_phase*. Similar to the AI field, the number of claims is the most impactful factor among all, far surpassing others. However, a key difference lies in the substantial influence of different examination phases on EC patent approvals. This may be related to the stronger global nature of EC technologies. For innovators in the EC field, participating in international patent examination procedures not only enhances the global competitiveness of their technologies but also reduces the risk of infringement by meeting international examination standards. Moreover, the examination processes at various stages are more standardized and systematic, making them critical determinants of patent approval.

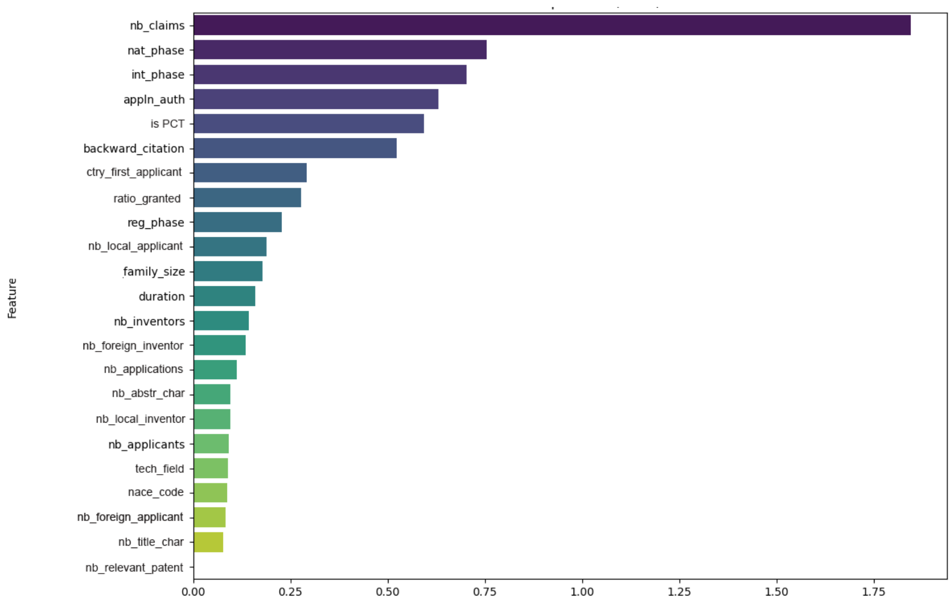

Figure 3. Feature Importance Analysis for EC Patent Grants.

Specifically, the influence of different factors on patent grants varies. First, filing a PCT application, entering the international phase, and having a higher number of claims are positive indicators of patent grants. Second, it is observed that for EC patents, a larger number of local applicants and inventors is more favorable for patent grant. Local innovators are likely to have a better understanding of the local market and regulatory environment, enabling them to submit patent applications that align more closely with examination requirements. Moreover, the involvement of local inventors may signify the practical feasibility and localized value of the technological innovation, thereby garnering greater recognition. Interestingly, unlike the other two fields, EC patent grants appear to be unrelated to backward citation. A possible explanation is that the EC field is characterized by mature technologies with rapid innovation cycles. Innovations in this domain are often driven by new application scenarios or cross-disciplinary integration, rather than heavy reliance on existing technological foundations. Consequently, examination authorities may focus more on the practical utility of the patent rather than its connections to prior technologies.

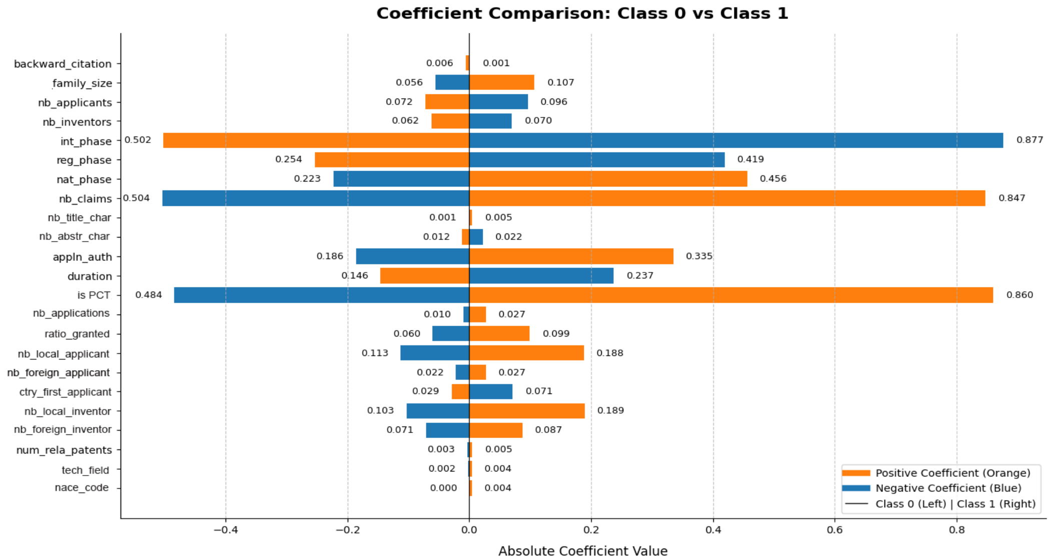

Figure 4. The coefficient comparison of influencing factors.

$$f(not\ \ granted) = -0.504 * nb_{claims} + 0.502 * int_{phase} - 0.484 * is_{PCT} + 0.254 * reg_{phase} - 0.223 * nat_{phase} - 0.186 * appln_{auth} + 0.146 * duration - 0.113 * nb_{local_{applicant}} - 0.103 * nb_{local_{inventor}} + 0.072 * nb_{applicants} - 0.071 * nb_{foreign_{inventor}} + 0.062 * nb_{inventors} - 0.060 * ratio_{granted} - 0.056 * family_{size} + 0.029 * ctry_{first_{applicant}} - 0.022 * nb_{foreign_{applicant}} + 0.012 * nb_{abstr_{char}} - 0.010 * nb_{applications} + 0.006 * backward_{citation} - 0.003 * nb_{relevant_{patent}} - 0.002 * tech_{field} - 0.001 * nb_{title_{char}} + 0.000 * nace_{code} + 0.776$$

Formula 1

$$f(granted) = -0.877 * \mathrm{int_{phase}} + 0.860 * \mathrm{is_{PCT}} + 0.847 * \mathrm{nb_{claims}} + 0.456 * \mathrm{nat_{phase}} - 0.419 * \mathrm{reg_{phase}} + 0.335 * \mathrm{appln_{auth}} - 0.237 * \mathrm{duration} + 0.189 * \mathrm{nb_{local_{inventor}}} + 0.188 * \mathrm{nb_{local_{applicant}}} + 0.107 * \mathrm{family_{size}} + 0.099 * \mathrm{ratio_{granted}} - 0.096 * \mathrm{nb_{applicants}} + 0.087 * \mathrm{nb_{foreign_{inventor}}} - 0.071 * \mathrm{ctry_{first_{applicant}}} - 0.070 * \mathrm{nb_{inventors}} + 0.027 * \mathrm{nb_{applications}} + 0.027 * \mathrm{nb_{foreign_{applicant}}} - 0.022 * \mathrm{nb_{abstr_{char}}} + 0.005 * \mathrm{nb_{relevant_{patent}}} + 0.005 * \mathrm{nb_{title_{char}}} + 0.004 * \mathrm{tech_{field}} + 0.004 * \mathrm{nace_{code}} + 0.001 * \mathrm{backward_{citation}} - 1.270$$

Formula 2

Discussion and conclusion

This study introduces a novel algorithm for patent grant prediction based on the Kolmogorov-Arnold Network (EN-KAN), which enhances interpretability while maintaining superior performance. Unlike traditional multilayer perceptions, the proposed model leverages the Kolmogorov-Arnold theorem to overcome the limitations of conventional methods that rely on linear transformations combined with activation functions. By allowing the use of nonlinear functions, this approach provides a more detailed analysis of the nonlinear impacts of input variables on outputs, offering intuitive insights into decision-making processes. To validate the proposed model, we collected patent datasets from Electronic Communication fields and extracted potential influencing factors at different levels. To further improve the predictive performance, ensemble learning strategies were employed to enhance the model's generalization ability. The final trained model consistently outperformed traditional machine learning algorithms across multiple datasets, achieving performance levels comparable to neural networks. More importantly, the model provides feature importance rankings and directly generates equations, offering precise explanations for influential factors.

The findings reveal that the factors influencing patent grant exhibit significant consistency across fields, with examination-level and patent-level factors playing pivotal roles. Among examination- level factors, the submission of a PCT application shows a strong positive correlation with patent grants. This relationship is closely tied to the international, national, and regional phases, each of which serves distinct purposes in the patenting process. The international phase primarily focuses on patentability searches, providing applicants with more time to determine target markets. In contrast, the national and regional phases involve substantive reviews to secure patent protection in individual jurisdictions or regional organizations. Patent-level factors also significantly influence granting outcomes, with backward citation and the number of claims standing out as critical variables. Backward citation, which reflects the foundational knowledge underlying the innovation, is positively associated with patent grants, corroborating prior studies that link it to patent value (Junbyoung Oh & Wonchang Hur, 2018). The number of claims, often considered an indicator of patent scope (Novelli, 2015), displays an unexpected positive correlation with patent granting probabilities. This finding challenges the conventional view that more claims result in stricter examination processes and lower grant rates (Marco et al., 2019). Instead, the study aligns with recent research suggesting that the number of claims represents not only the scope but also the comprehensiveness and innovativeness of a patent, thereby highlighting its potential value (Kuhn & Thompson, 2019; Yao & Ni, 2023).

This study introduces the EN-KAN model, which combines interpretability with high predictive performance. By leveraging the Kolmogorov-Arnold theorem instead of traditional multilayer neural network methods, the model not only identifies the key factors influencing patent granting but also provides mathematical formulas with coefficients. This approach addresses the “black box” problem inherent in neural network algorithms, further enhancing the interpretability of the predictive model. From a practical application perspective, these findings can assist innovative entities in optimizing their patent application strategies. Innovators in different fields can tailor their patent documentation based on their specific key factors, refine their patent portfolios, and significantly improve the likelihood of granting. For examination authorities, understanding the critical factors influencing patent granting enables a more focused review process, enhancing examination efficiency and refining patent review rules. Lastly, these conclusions can also guide research and market strategies. Considering patent grant factors during the research and development phase can facilitate the creation of technologies that are not only more patentable but also have higher market potential.

In summary, this study proposes EN-KAN as a robust tool for patent grant prediction, yet two limitations should be noted. First, the dataset used in this study is limited to a single technological domain and includes only patents filed in 2017, which may raise concerns regarding the generalizability of the findings. Future research could expand the scope to include multiple domains and application years to enable comparative analysis and enhance the robustness of the results. Additionally, despite efforts to include all relevant influencing factors, certain features, such as patent filing strategies, could not be incorporated due to data limitations. Future research could address this by exploring additional data sources to include a broader range of influencing factors.

Reference

Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., & Hinton, G. E. (2021). Neural additive models: Interpretable machine learning with neural nets. *Advances in Neural Information Processing Systems*, *34*, 4699–4711.

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, *38*(2), 415–427.

Allen, G. I., Gan, L., & Zheng, L. (2024). Interpretable Machine Learning for Discovery: Statistical Challenges and Opportunities. In *Annual Review of Statistics and Its Application*, 11, 97–12.

Bekkers, R., Martinelli, A., & Tamagni, F. (2020). The impact of including standards-related documentation in patent prior art: Evidence from an EPO policy change. *Research Policy*, *49*(7), 104007.

Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review, 56*(5), 4765–4800.

Covert, I. C., Lundberg, S., & Lee, S.-I. (2021). Explaining by removing: A unified framework for model explanation. *J. Mach. Learn. Res., 22*(1), Article 209.

Crawshaw, M., Liu, M., Orabona, F., Zhang, W., & Zhuang, Z. (2022). Robustness to unbounded smoothness of generalized signsgd. *Advances in Neural Information Processing Systems, 35*, 9955–9968.

Desai, P. (2019). Biased Regulators: Evidence from Patent Examiners. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3485965

Dong, Y., Su, H., Zhu, J., & Bao, F. (2017). Towards interpretable deep neural networks by leveraging adversarial examples. *arXiv Preprint arXiv:1708.05493*.

Drivas, K., & Kaplanis, I. (2020). The role of international collaborations in securing the patent grant. *Journal of Informetrics, 14*(4), 101093.

Frakes, M. D., & Wasserman, M. F. (2021). Knowledge spillovers, peer effects, and telecommuting: Evidence from the US Patent Office. *Journal of Public Economics, 198*, 104425.

Fryer, D., Strümke, I., & Nguyen, H. (2021). Shapley values for feature selection: The good, the bad, and the axioms. *Ieee Access, 9*, 144352–144360.

Gans, J. S., Hsu, D. H., & Stern, S. (2008). The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays. *Management Science, 54*(5), 982–997.

Guellec, D., & de la Potterie, B. van P. (2000). Applications, grants and the value of patent. *Economics Letters, 69*(1), 109–114.

Harhoff, D., & Wagner, S. (2009). The duration of patent examination at the European Patent Office. *Management Science, 55*(12), 1969–1984.

Hoefler, T., Alistarh, D., Ben-Nun, T., Dryden, N., & Peste, A. (2021). Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. *Journal of Machine Learning Research, 22*(241), 1–124.

J. Dai, K. He, & J. Sun. (2015). Convolutional feature masking for joint object and stuff segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3992–4000.

Junbyoung Oh & Wonchang Hur. (2018). A Man is Known by the Company He Keeps? : A Structural Relationship Between Backward Citation and Forward Citation of Patents. *Research Policy, 50*(1), 104117.

Kuhn, J., & Thompson, N. (2019). How to Measure and Draw Causal Inferences with Patent Scope. *International Journal of the Economics of Business, 26*, 5–38.

Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change, 127*, 291–303.

Lee, P.-C. (2021). Investigating Long-Term Technological Competitiveness: Originality, Generality, and Longevity. *IEEE Transactions on Engineering Management*, *71*, 20–42.

Lee, Y.-G., & Lee, J.-H. (2010). Different characteristics between auctioned and non-auctioned patents. *Scientometrics*, *82*(1), 135–148.

Lemley, M. A., & Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of Economics and Statistics*, *94*(3), 817–827.

Liegsalz, J., & Wagner, S. (2013). Patent examination at the state intellectual property office in China. *Research Policy*, *42*(2), 552–563.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31–57.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., ... & Tegmark, M. (2024). Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Marco, A. C., Sarnoff, J. D., & Charles, A. W. (2019). Patent claims and patent scope. *Research Policy*, *48*(9), 103790.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv Preprint arXiv:1802.03426*.

Millar, C. C., Groth, O., & Mahon, J. F. (2018). Management innovation in a VUCA world: Challenges and recommendations. *California Management Review*, *61*(1), 5–14.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Niosi, J. (2006). Introduction to the Symposium: Universities as a Source of Commercial Technology. *The Journal of Technology Transfer*, *31*(4), 399–402. https://doi.org/10.1007/s10961-006-0001-0

Nordhaus, W. D. (1969). *Invention growth, and welfare; a theoretical treatment of technological change*. M.I.T. Press Cambridge, Mass.

Novelli, E. (2015). An examination of the antecedents and implications of patent scope. *Research Policy*, *44*(2), 493–507.

Raiteri, E. (2018). A time to nourish? Evaluating the impact of public procurement on technological generality through patent data. *Research Policy*, *47*(5), 936–952.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). " Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Rong, S., & Bao-Wen, Z. (2018). The research of regression model in machine learning field. *MATEC Web of Conferences*, *176*, 01033.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sampat, B., & Williams, H. L. (2019). How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome. *American Economic Review*, *109*(1), 203–236. https://doi.org/10.1257/aer.20151398

Shen, M., & Zingg, R. (n.d.). *Patent Examiners and the Citation Bias in Innovation*.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Webster, E., Jensen, P. H., & Palangkaraya, A. (2014). Patent examination outcomes and the national treatment principle. *The RAND Journal of Economics*, *45*(2), 449–469.

Yang, D. (2008). Pendency and grant ratios of invention patents: A comparative study of the US and China. *Research Policy*, *37*(6–7), 1035–1046.

Yang, G., Lu, G., Xu, S., Chen, L., & Wen, Y. (2023). Which type of dynamic indicators should be preferred to predict patent commercial potential? *Technological Forecasting and Social Change*, *193*, 122637.

Yao, L., & Ni, H. (2023). Prediction of patent grant and interpreting the key determinants: An application of interpretable machine learning approach. *Scientometrics*, *128*(9), 4933–4969.

Zhou, Z. H., & Liu, S. (2021). *Machine Learning*. Springer Nature Singapore. https://books.google.ru/books?id=ctM-EAAAQBAJ