

A Machine Learning Framework to Evaluate Transferability of Hospital-Owned Patents: Enhancing Innovation Commercialization in Health Care

Wenjun Gu

Shanghai 6th People's Hospital Affiliated to Shanghai Jiao Tong University

Jinhua Wang

Shanghai Yiji Hui Medical Science and Technology Innovation and Transformation Research Institute

Jiyu Li

lijiyu@fudan.edu.cn

Pudong Hospital Affiliated to Fudan University

Article

Keywords:

Posted Date: May 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9287232/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Machine Learning Framework to Evaluate Transferability of Hospital-Owned Patents: Enhancing Innovation Commercialization in Health Care

Wenjun Gu¹, Jinhua Wang², Jiyu Li^{3*}

1 Shanghai 6th People's Hospital Affiliated to Shanghai Jiao Tong University

2 Shanghai Yiji Hui Medical Science and Technology Innovation and Transformation Research Institute

3 Pudong Hospital Affiliated to Fudan University

* Corresponding authors: Jiyu Li (lijyu@fudan.edu.cn)

Abstract

Hospital-owned patents often encounter obstacles to commercialisation, making reliable assessment tools essential for prioritising high-value innovations. This study presents a machine learning framework that evaluates the transferability of patents using a system of 20 indicators across technical, legal, market and team dimensions. We compare model performance using six algorithms — Support Vector Machine, Decision Tree, Naïve Bayes, K-Nearest Neighbors, Random Forest and AdaBoost — with F1-score and AUC as the key metrics. Our results confirm that the Random Forest model demonstrates superior predictive accuracy in evaluating the transferability of hospital-owned patents. This validated framework can effectively identify high-potential patents held by healthcare institutions, providing data-driven support for optimising innovation investments and commercialisation strategies.

Introduction

Health care delivery is undergoing a transformation driven by technological innovation, yet many medical inventions fail to reach the patients who need them. Hospitals and health systems are increasingly contributing to the innovation pipeline, as evidenced by rising numbers of patent filings. However, the path from discovery to deployment is fraught with barriers—including limited resources, unclear commercial potential, and a lack of systematic methods to prioritize which inventions to pursue [1].

While machine learning (ML) has been applied to evaluate patent transferability in academic and corporate settings (Yan et al,2023;Chen et al,2023;Kim et al,2020), its use in health care—particularly for hospital-owned patents—has been minimal.This gap is significant, as

hospital-generated innovations often address specific clinical needs and hold promise for improving care delivery, yet their unique characteristics are not captured by existing models. In this study, we propose an ML-based framework to assess the transferability of hospital-owned patents—defined as their potential for commercialization through licensing or transaction. Using a global dataset of 57,508 patents and a curated set of 20 indicators across technical, legal, market, and team dimensions, we train and validate multiple ML algorithms. Our goal is to provide a practical, evidence-based tool that helps health care executives, innovation officers, and policymakers identify high-potential patents, allocate resources wisely, and accelerate the translation of inventions into clinical practice.

This work contributes to the growing literature on health care innovation by: (1) Offering a reproducible, data-driven method to evaluate patent transferability;

(2) Identifying key factors that influence commercialization success in the health care context; (3) Supporting strategic decision-making for technology transfer and investment within health systems.

By bridging the gap between invention and implementation, this approach aligns with the mission of health care organizations to deliver meaningful, scalable improvements in patient care.

The rest of the paper is organized as follows. In Section 2, we present the fundamental concepts underlying the terminology employed in this study. In Section 3, a review of existing studies related to patent transfer is conducted. In Section 4, a preliminary index system is developed to assess the commercialization likelihood of medical institutional patents, employing a self-constructed global database of healthcare invention patents to systematically collect indicator data, calculate metrics, and screen indexes. In Section 5 the proposed machine learning approaches are subsequently applied. In Sections 6 and 7, we discuss the results and then present conclusions and suggestions for further research in the final section.

Methods

Patent Transferability Evaluation

The transfer of technology represents a fundamental aspect of technological advancement, encompassing the dissemination of knowledge and expertise from the point of origin to that of utilization [2]. The transfer of technology encourages the accumulation of knowledge within enterprises, thereby assisting small and medium-sized enterprises (SMEs) in reducing the technological disparity that exists between them and leading enterprises [3]. Furthermore, it has been proposed that the mobility of talent represents a significant driver of technology transfer, particularly within knowledge-intensive industries. The movement of individuals with advanced technological skills across organizational boundaries serves to accelerate the transfer of technology among a diverse range of innovation agents [4].

Prior research has examined patent transactions through diverse methodological approaches, primarily focusing on academic and corporate contexts. Key studies include: Coupet and Ba's

(2022) stochastic frontier analysis of university licensing performance [5]; Sapir's (2021) qualitative investigation of Israeli technology transfer processes [6]; Ravi and Janodia's (2022) cross-sectional study of Indian university commercialization [7]; Ma et al.(2021) licensor decision model in China [8]; Cunningham et al. (2020) case study of commercial-experienced PIs in Ireland [9]; and Kim and Geum's (2020) machine learning model predicting patent transactions [10,11]. Collectively, these studies demonstrate methodological diversity in analyzing patent transformation, spanning quantitative, qualitative, and predictive approaches.

Patent indexes

In the extant literature, scholars have employed a variety of metrics to evaluate the transferability of patents, typically considering multiple dimensions such as technology, market, and legal aspects. From a technological perspective, metrics such as standardized patent citation counts [12] the number of priority claims [13], the technology life-cycle [14], the number of cited patents, and non-patent literature citations [15,16], and the number of IPC classifications [17] are widely used to assess technological influence, advancement, maturity, and potential application breadth. In the context of the market dimension, indicators such as the duration of patent maintenance [18], the frequency of patent licensing and transfers [19], and patent generality [20] are frequently utilized to gauge commercial value and market acceptance. From a legal standpoint, metrics such as the number of patent litigation [21,12] and the size of patent families [22] are employed to evaluate legal stability and strength. Furthermore, some studies have introduced interdisciplinary metrics, such as the linkage between patents and scientific literature, to further reveal the foundational research background and potential cross-domain application value [23]. Moreover, certain research has focused on the capabilities of inventor teams [24], suggesting that the more experienced the inventors are in technological development, the higher the value of technology transfer. Despite the heterogeneity of the metrics involved, their common objective is to provide a more comprehensive and objective framework for assessing patent transferability through quantitative analysis.

Methodology and Data Acquisition

This study develops a machine learning-based framework to assess patent transferability using six algorithms (Support Vector Machines, K-Nearest Neighbors, AdaBoost, Random Forest, Naïve Bayes, and Decision Tree) applied to global patent data. The methodology comprises: (1) defining evaluation indicators across technical, legal, economic, and team dimensions, (2) constructing a patent database from healthcare organizations worldwide, (3) selecting optimal indicators through Contribution Rate of Cumulative Information—Grey Relation Analysis (CRCI-GRA) [25], (4) building transferability prediction models using the six machine learning approaches, and (5) validating model performance via F1-score, Accuracy, and AUC-ROC metrics. The procedure is outlined in Figure 1.

Construction and Acquisition of Patent Indicator

According to the contingent effectiveness model of technology transfer constructed by Barry Bozeman [26], this study establishes a comprehensive 20-indicator framework to evaluate healthcare institution patent transferability across four dimensions: (1) technological dimension indicators including standardized citation counts, priority claims, technology life-cycle, the number of non-patent literature references, backward citations, IPC codes, claim quantity and word counts of independent claim, and (2) legal dimension indicators covering patent litigation frequency, publication-to-grant period, patent family size and triadic patent counts, and (3) team dimension factors considering quantity of applicants, applicant category, inventor team size, inventor reputation, hospital bed capacity, and (4) market dimension metric of patent maintenance duration, filing-to-transfer period and patent generality. The detailed definitions and calculations for all indicators are presented in Table 1.

Developing a patent database

This study developed a patent database using incoPat(<https://www.incopat.com>) and the Derwent patent database(<https://derwentinnovation.clarivate.com.cn/>) to assess hospital-owned patent transferability. This study selects the incoPat patent database because it collects more than 190 million pieces of patent information from 157 authorities. All data are collected from official patent offices of different countries and business vendors. Patent data is updated four times every week. Derwent Innovation (Clarivate Analytics) is a premium patent intelligence platform integrating normalized global patent data, DWPI expert annotations, and analytical tools for IP commercialization assessment.

The definition of a medical institution is contingent upon the keywords employed in the patentee search. By the definition of medical institutions, as outlined by the National Health Commission of China, the term encompasses a range of facilities, including hospitals, maternal and child healthcare centers, prevention and treatment centers, community health service centers, clinics, medical centers, rehabilitation institutes, and cancer center. In this study, the aforementioned terms were employed as search patentee in the incoPat and Derwent Innovation patent database to identify granted invention patents in global healthcare organizations.

From 606,180 global patent applications from healthcare institutions (data current through December 2022), we selected 57,508 granted inventions based on rights stability. In this study, we operationalized patent transferability value through the counts of patent transfers and licenses.

Feature selection for indicators

Standardization of indicator data

To address measurement unit heterogeneity in the patent transferability evaluation system, we applied min-max normalization for data standardization. This dimensionless transformation method [37] preserves three key mathematical properties: translation

invariance, ratio consistency, and interval stability. The algorithm linearly re-scales raw values (X) to [0,1] using data-set extrema ($X_{\max}$, $X_{\min}$) by Equation (1)-(2).

$$\text{Positive Indicator: } X' = (X - X_{\min}) / (X_{\max} - X_{\min}) \quad \text{Equation (1)}$$

$$\text{Negative Indicator: } X' = (X_{\max} - X) / (X_{\max} - X_{\min}) \quad \text{Equation (2)}$$

Where $X' \in [0,1]$ represents the normalized value. This preprocessing eliminates scale dependence, enabling cross-indicator integration within our multivariate framework.

Descriptive statistical analysis

The patent data set was categorized into domestic (N=35708) and international (N=18800) groups based on geographic origin. Descriptive statistics (including mean, median, maximum, minimum, standard deviation, and interquartile range) were calculated for all variables. As the data violated the assumptions of normality (Shapiro- Wilk test, $p < 0.05$) and homogeneity of variance (Levene's test, $p=0.003$), the non- parametric Wilcoxon rank-sum test was employed to compare the distribution locations between groups by asterisks ($p < 0.01$).

Feature selection for indicators

Stage 1: The CRCI methodology [25] was systematically applied following Steps 1 through 4 to evaluate feature contributions and enhance model interpretability.

Step 1: Computing relative coefficients of variation (CV) for normalized variables followed by Equation (3).

$$CV_j = \frac{m \sqrt{\frac{1}{m-1} \sum_{i=1}^m \left(x_{ij} - \frac{1}{m} \sum_{i=1}^m x_{ij} \right)^2}}{\sum_{i=1}^m x_{ij}} \quad \text{Equation (3)}$$

Let CV_j be the feature importance vector [38], with each element representing the normalized CV of x_{ij} . We set $n=20$ as the dimensionality of the indicator space.

Step 2: Ranking indicator by information content

The initial set of 20 raw indicators (denoted as $x_1, x_2, \dots, x_{20}$) was systematically ranked in descending order of their information content. The resultant ordered indicators are represented as $x_{n1}, x_{n2}, \dots, x_{nn}$, where x_{ni} denotes the indicator possessing the i-th highest information content among all candidates.

Step 3: Determination of explained variance accumulation using Equation (4).

$$Z_n = \sum_{j=1}^p CV_j^* / \sum_{j=1}^n CV_j \quad \text{Equation (4)}$$

Let CV_j^* ($j=1,2,\dots,p$) represent the top-p indicators $x_{n1}, x_{n2}, \dots, x_{ni}$ with the highest information content (where $p \leq 20$), selected from the original 20 indicators through entropy-based ranking.

Step 4: Screening indicators based on information content

The top-p indicators $\{Z_n\}$ were selected based on their information content, satisfying:

$$Z_n > Z_0 > Z_{(p-1)}$$

Where:

Z_n : Information content of the n-th ranked indicator (measured by variance)

Z_0 : Threshold for cumulative variance contribution ($Z_0=95\%$) [39]

$Z_{(p-1)}$: Information content of the (p-1)-th ranked indicator

Stage 2: Correlation-based centrality assessment coupled with GRA was implemented following Step 5 through 7 to optimize the feature set, effectively minimizing informational overlap and improving metric accuracy.

Step 5: Inter-indicator correlation coefficients were computed by Equation (5).

$$\zeta_{ij}(k) = \frac{\min_j \min_k |x_i(k) - x_j(k)| + \rho \max_j \max_k |x_i(k) - x_j(k)|}{|x_i(k) - x_j(k)| + \rho \max_j \max_k |x_i(k) - x_j(k)|} \quad \text{Equation (5)}$$

Let $\Delta_{\min}$ and $\Delta_{\max}$ represent the minimum and maximum absolute differences of the

remaining two indicators, respectively, where $\Delta_{\min} = \min_j \min_k |x_i(k) - x_j(k)|$ and $\Delta_{\max} =$

$\max_j \max_k |x_i(k) - x_j(k)|$ for $i, j = 1, 2, \dots, n, k = 1, 2, \dots, 20$. The resolution coefficient ρ , which serves

to modulate the output disparity and amplify the discriminative significance between correlation coefficients, is set within the range (0,1). Following common practice in sensitivity analysis, we adopted $\rho = 0.5$ [40] in this study to maintain balanced discrimination.

Step 6: Inter-variable correlation matrix was derived by Equation (6).

Given the observed data $\zeta_{ij}(k)$ dispersion in the correlation coefficients, a temporal averaging approach was adopted to consolidate the information, thus facilitating systematic comparison.

$$\gamma_{ij} = \frac{\sum_{k=1}^m \zeta_{ij}(k)}{m} \quad \text{Equation (6)}$$

In the equation, γ_{ij} represents the correlation degree between X_i and X_j , where m denotes the number of data points.

Calculate the correlation coefficients γ_{ij} to obtain an inter-indicator correlation matrix R_{ij} by Equation (7).

$$R_{ij} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} \\ \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix} \quad \text{Equation (7)}$$

Step 7: Identify indicators with grey relational grades ($\gamma_{ij} > 0.900$) [25], compare their relationships, and subsequently eliminate those with lower information content according to the ranking established in Step 2.

Proposing a Patent Transferability Evaluation Model

To evaluate algorithm applicability, we developed a machine learning framework for predicting healthcare patent conversion using 57,508 samples. Data were categorized as converted/non-converted, with comprehensive descriptive statistics (Mean, Variance,

Extrema, Median, Quartiles) computed for selected features. Group distribution differences were assessed via Wilcoxon rank-sum tests to validate machine learning appropriateness. Then, this study employed the Scikit-learn machine learning library (random_state=1) to conduct a comparative performance analysis of six classification algorithms: AdaBoost, k-nearest neighbors (KNN), Naïve Bayes, random forest, decision tree, and support vector machine (SVM). The optimal algorithm was selected to construct a predictive model for assessing hospital-owned patent transferability potential. Standardized patent data were binarized (0/1 labels) to mitigate algorithm-specific overfitting risks. The dataset (N = 57,508) was randomly split into training and testing subsets (8:2 Ratio), with model performance evaluated using the receiver operator characteristics and area under the curve (ROC and AUC) metrics [41]. Additionally, we estimated *Accuracy* and *F1-score*, using the following Equations 8 and 9, to verify the predictive power and validation of the estimated model.

$$Precision = \frac{TP}{TP+FP}$$

Where *TP* (true positive) denotes the count of accurately categorized transferred healthcare patents with confirmed technology commercialization, while *FP* (false positive) represents misclassified cases within the successfully transferred patent cohort.

$$Recall = \frac{TP}{TP+FN}$$

Here, *FN* denotes false negative instances where hospital-owned patents were both incorrectly categorized and failed to achieve technology transfer.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{Equation (8)}$$

Here, *TN* denotes true negative instances where hospital-owned patents were both correctly categorized and failed to achieve technology transfer.

$$F1\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad \text{Equation (9)}$$

F1-score, defined as the harmonic mean of precision and recall, serves as a comprehensive metric for evaluating classification model performance. This weighted harmonic mean (ranging from 0 to 1, with higher values indicating better predictive accuracy) balances the trade-off between precision and recall in practical applications.

Results and Discussion

Descriptive analysis

The descriptive analysis results of the initial patent transferability value assessment for healthcare institution-owned patents are systematically compared between geographical regions in Table 2. The Wilcoxon rank-sum test results, evidenced by consistently negative Z-statistics, reveal that internationally granted invention patents generally exhibit higher indicator values in the overall distribution compared to domestic ones.

Regarding the negative-polarity indicator of the technology life cycle, domestic patents demonstrated significantly higher mean ranks ($Z_{Domestic} = 32,613.765 > Z_{International} = 22,433.094$). Conversely, for three positive indicators - number of priority claims ($Z_{Domestic} =$

40281.314 < $Z_{International} = 21717.294$), patent claims ($Z_{Domestic} = 21323.312 < Z_{International} = 40926.650$), and patent maintenance years ($Z_{Domestic} = 21426.630 < Z_{International} = 40757.417$)- internationally granted patents showed substantially greater mean ranks, indicating their superior performance in these dimensions. Notably, domestic patents outperformed in inventor reputation ($Z_{Domestic} = 32729.940 > Z_{International} = 22242.803$) and healthcare institution bed capacity ($Z_{Domestic} = 32722.167 > Z_{International} = 22255.533$). These findings align with empirical observations that: (1) domestic institutions maintain higher patent grant volumes, and (2) possess significantly greater bed capacities than international counterparts.

The Wilcoxon rank-sum test results suggest that stratified analysis between domestic and internationally granted invention patents should be implemented when applying CRCI-GRA for indicator screening.

Indicator significance evaluation

The analysis employed a 95% cumulative contribution rate threshold for indicator retention. As evidenced in Table 3, triadic patent count (L_{22}) was retained as the terminal indicator (cumulative rate = 95.427%) in domestic samples. Similarly, Table 4 demonstrates that patent family size (L_{21}) represented the cutoff point for indicators (cumulative rate = 95.404%) in international samples.

The indicator selection results from Table 3 and 4 reveal that the evaluation system for internationally granted invention patents retained 12 significant indicators, encompassing all 10 indicators selected for domestic ones. Notably, the international system incorporated two additional indicators: claim quantity and patent maintenance duration.

As evidenced by both tables, three indicators demonstrated particularly high importance across both domestic and international evaluation systems for hospital- owned patent transferability value: (1) patent litigation frequency ($CV_{f-Domestic} = 63.441$ and $CV_{f-International} = 53.874$), (2) standardized citation counts ($CV_{j-Domestic} = 75.551$ and $CV_{j-International} = 60.845$), and (3) non-patent literature references ($CV_{l-Domestic} = 82.731$ and $CV_{l-International} = 67.502$).

To reconcile indicator selection between domestic and international granted invention patents transferability value evaluation systems, this study retained 12 key indicators under a 95% cumulative contribution rate threshold: patent litigation frequency (L_{11}), standardized citations (T_{11}), non-patent literature references (T_{32}), inventor reputation (H_{21}), triadic patents (L_{22}), technology life-cycle (T_{22}), patent family size (L_{21}), IPC codes (T_{41}), claim quantity (T_{42}), hospital bed capacity (H_{22}), priority claims (T_{21}), and patent maintenance duration (E_{11}).

Indicator correlation analysis

The 12 patent transferability value indicators pre-selected through CRCI methodology were subsequently analyzed using GRA to evaluate pairwise correlations for precision screening. The resultant inter-indicator relationships are visualized in the grey relational heat-map (Figure 2).

The heat-map analysis indicates that patent family size (L_{21}) and triadic patent count (L_{22}) show a correlation coefficient of $\gamma = 0.619$, which, despite being statistically significant, did not

exceed our conservative retention threshold ($\gamma < 0.900$). Thus, both variables were included in the final model.

Application of the proposed approach

Descriptive analysis of the 12 CRCI-GRA-filtered indicators from a translational perspective revealed significantly different distributions between converted and non-converted groups (Table S1). The Wilcoxon rank-sum tests consistently showed negative Z-values (all $Z < 0$), indicating systematically higher value distributions in the converted group. This distributional divergence confirms the appropriateness of applying machine learning classification algorithms to develop the hospital-owned patent transferability valuation model.

Table 5 compares the six classifiers' performance in terms of *AUC-ROC*, *F1-score*, and *Accuracy*. The results demonstrate that Support Vector Machine (SVM) and Naïve Bayes classifiers achieved *AUC* values of 0.600-0.650, outperforming random models with moderate accuracy. In contrast, Decision Tree, K-Nearest Neighbors (KNN), AdaBoost, and Random Forest algorithms exhibited superior discriminative ability with *AUCs* ranging from 0.700 to 0.900. Notably, Random Forest emerged as the optimal model among the six approaches, highlighting its distinctive value for patent transferability valuation with an *AUC* of 0.900, *F1-score* = 0.869, and *Accuracy* = 0.871. Feature importance analysis revealed claim quantity, patent maintenance duration, and inventor reputation as the strongest predictors according to Figure 3.

Consequently, we selected random forest to construct the hospital-owned patent transferability valuation model.

The parameter configuration of the Random Forest algorithm critically determines both predictive accuracy and training efficacy. Parameter optimization involves two hierarchical levels: (1) framework-level hyperparameters and, (2) decision tree-specific parameters. To achieve optimal performance, we employed manual tuning through 10-fold cross-validation with *AUC* as the evaluation metric, ultimately deriving the ideal parameter combination.

Figure 4 indicates that the choice of *mtry* and the number of trees significantly impacts the performance of the RF model. Specifically, a moderate *mtry* value (around 7) combined with 130 trees optimizes the model's predictive ability, as evidenced by the highest *AUC*.

Using the established random forest-based evaluation model for assessing the transferability of hospital-owned patents, we predicted the commercialization potential of non-transformed granted invention patents in both domestic and international samples. The model generated individual predictions for each patent, visualized through a transferability histogram with logarithmic curve fitting, illustrated in Figures 5 and 6.

Figures 5 and 6 show the transfer probability distributions of non-transferred patents from both domestic and international sources followed by logarithmic curves. Both R^2 values exceeding 0.900 indicate excellent curve fitting, suggesting that healthcare institution-owned patents face similar transfer challenges globally, with the majority demonstrating low commercialization potential. These findings align with Nair S.S.'s research demonstrating that patent value typically follows a log-normal distribution pattern [42].

The findings highlight several key indicators—claim quantities, patent maintenance duration, and inventor reputation—as pivotal determinants of transferability, aligning with existing theories on patent value while introducing healthcare-specific metrics. This not only validates the model's theoretical grounding but also provides practical insights for hospital administrators to prioritize high-potential patents for commercialization, thereby optimizing resource allocation. The study's reliance on publicly accessible data further enhances its reproducibility and applicability across global healthcare systems.

This work significantly advances the field by: (1) establishing the first evidence-based patent transferability assessment tool tailored for healthcare institutions; and (2) providing a reproducible methodology that bridges the gap between innovation and commercialization in medical research.

Data availability

Upon formal acceptance of the Stage 2 manuscript for publication, we commit to promptly and permanently making the underlying raw data and materials publicly available.

Code availability

Upon formal acceptance of the Stage 2 manuscript, we will publicly share all the code necessary to reproduce the findings of this study. The machine learning model was developed and executed within the ENVI software environment. The key analytical scripts and process files (e.g. ENVI batch scripts, .sav state files and detailed procedural documentation) will be deposited in a public repository .

References

1. Vega RJ, Kizer KW. VHA' s Innovation Ecosystem: Operationalizing Innovation in Health Care. *NEJM Catal Innov Care Deliv* 2020;1(6)
<https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0263>.
2. Glass AJ, Saggi K. International technology transfer and the technology gap. *Journal of Development Economics* 1998;55(2):369-398
<https://www.sciencedirect.com/science/article/abs/pii/S0304387898000418>.
3. Teece DJ. Technology transfer by multinational firms: the resource cost of transferring technological know-how. *The Economic Journal* 1997;87(346):242 – 261
<https://www.jstor.org/stable/2232084> <https://doi.org/10.2307/2232084> .
4. Yasar M, Rejesus RM. International linkages technology transfer and the skilled labor wage share: evidence from plant-level data in indonesia. *World Development* 2020;128
<https://www.sciencedirect.com/science/article/abs/pii/S0305750X19304966>
<https://doi.org/10.1016/j.worlddev.2019.104847>.
5. Coupet J, Ba Y, Link AN, Siegel DS, Bozeman B, Mosey S. Benchmarking university technology transfer performance with external research funding: a stochastic

frontier analysis. *J Technol Transf* 2020;47:605 – 620
<https://link.springer.com/article/10.1007/s10961-021-09856-3>
<https://doi.org/10.1007/s10961-021-09856-3>.

6. Sapir A. Brokering knowledge, monitoring compliance: technology transfer professionals on the boundary between academy and industry. *Journal of Higher Education Policy and Management* 2020;1-16
<https://doi.org/10.1080/1360080X.2020.1804657>.
7. Ravi R, Janodia MD, Carayannis EG. Factors affecting technology transfer and commercialization of university research in india: a cross-sectional study. *Journal of the Knowledge Economy* 2022;13:787-803
<https://link.springer.com/article/10.1007/s13132-021-00747-4>
<https://doi.org/10.1007/s13132-021-00747-4>.
8. Rongkang M, Shanshan Z, Fengchao L. How does firm innovation size affect the timing of technology licensing? theory and evidence from china. *Industrial and Corporate Change* 2021;(3)3 <https://doi.org/10.1093/icc/dtaa037>.
9. Cunningham JA, Dolan B, Menter M, O'Kane C, O'Reilly P. How principal investigators' commercial experience influences technology transfer and market impacts. *Research-Technology Management* 2020;63(5):49 – 58
<https://doi.org/10.1080/08956308.2020.1790244>.
10. .Chen H, Deng, WW. Interpretable patent recommendation with knowledge graph and deep learning. *Scientific Reports* 2023;13:2586 <https://doi.org/10.1038/s41598-023-28766-y>.
11. Kim S. Innovating knowledge and information for a firm-level automobile demand forecast system: a machine learning perspective. *Journal of Innovation & Knowledge* 2023 <https://doi.org/10.1016/j.jik.2023.100355>.
12. Méndez-Morales A, Ochoa-Urrego R, Randhir TO. Measuring the quality of patents among Latin-American universities. *Studies in Higher Education* 2021;47(11):2174-2189 <https://doi.org/10.1080/03075079.2021.2020749>.
13. Grimaldi M, Cricelli L. Indexes of patent value: a systematic literature review and classification. *Knowledge Management Research & Practice* 2019;(1):1-20
<https://doi.org/10.1080/14778238.2019.1638737>.
14. Peng L. Research on the Investment Value Evaluation of Technology-based Small and Medium Enterprises Based on the Technological Life Cycle Shanghai: Shanghai University of Engineering Science 2019;(6)
<https://doi.org/10.27715/d.cnki.gshgj.2019.000282>.
15. Federico C, Giuseppe S, Elisa U. Lenders' selection capabilities patent quality and the outcome of patent-backed loans. *Industrial and Corporate Change* 2019;(1)1
<https://doi.org/10.1093/icc/dtz029>.
16. Shu-Hao C, Hsin-Yuan C, Chin-Yuan F. Structural model of patent quality applied to various countries. *International Journal of Innovation Science* 2018;10(3):371-384
<https://doi.org/10.1108/IJIS-05-2017-0036>.
17. Grimaldi M, Cricelli L, Rogo F. Valuating and analyzing the patent portfolio: the patent portfolio value index. *European Journal of Innovation Management*, European

Journal of Innovation Management 2018;21(2):174 – 205
<https://doi.org/10.1108/EJIM-02-2017-0009>.

18. Bakker J. The log-linear relation between patent citations and patent value. *Scientometrics* 2017;110(2):1-14 <https://doi.org/10.1007/s11192-016-2208-7>.
19. Li M, Wu Z, Wang L, Zhou K. Does firm's value matter with firm's patent quality in technology-intensive industries?. *IEEE Transactions on Engineering Management* 2020;PP(99):1-18 <https://doi.org/10.1109/TEM.2020.3014228>.
20. Sonmez Z. Interregional inventor collaboration and the commercial value of patented inventions: evidence from the us biotechnology industry. *The Annals of Regional Science* 2018;61(2):1-40 <https://doi.org/10.1007/s00168-018-0874-5>.
21. Tsao CC, Fan CY, Chang PC. A patent quality classification model based on artificial immune system. *Soft Comput* 2015;21:2847 – 2856
<https://link.springer.com/article/10.1007/s00500-016-2212-0>
<https://doi.org/10.1007/s00500-016-2212-0>.
22. Huang, G. L, Huang, C, Shen, H, Mao, H. Assessing the value of china's patented inventions. *Technological Forecasting and Social Change* 2021;170
<https://doi.org/10.1016/j.techfore.2021.120868>.
23. Gu W, Shu L, Chen W, Wang J, Wu D, Ai Z, Li J. Evaluation of Chinese healthcare organizations' innovative performance in the digital health era. *Front. Public Health* 2023;11:1141757 <https://doi.org/10.3389/fpubh.2023.1141757>.
24. Du W, Wang Y, Xu W, Ma J. A personalized recommendation system for high-quality patent trading by leveraging hybrid patent analysis. *Scientometrics* 2021;126(12):9369-9391 <https://doi.org/10.1007/s11192-021-04180-x>.
25. Dong L, Xiaochen QI, Weifeng Z. Resilience evaluation index optimization for combined regional agricultural water and soil resources system based on crci-gra model. *Journal of Northeast Agricultural University* 2019;50(01):59-67
<https://doi.org/10.19720/j.cnki.issn.1005-9369.2019.01.008>.
26. Bozeman B. Technology transfer and public policy: a review of research and theory. *Research Policy* 2000;29(4-5):627-655 [https://doi.org/10.1016/S0048-7333\(99\)00093-1](https://doi.org/10.1016/S0048-7333(99)00093-1).
27. Svensson R. Patent value indicators and technological innovation. *Empirical Economics* 2021;(2) <https://doi.org/10.1007/s00181-021-02082-8>.
28. Hsu D. H Hsu PH, Zhou T, Ziedonis AA. Benchmarking u.s. university patent value and commercialization efforts: a new approach. *IRPN: Innovation Management(Topic)* 2019;50(1):104076 <https://doi.org/10.2139/ssrn.3264090>.
29. Lv X, Luo Y, Li J, You X. Review of studies on detecting chinese patent infringements. *Data Analysis and Knowledge Discovery* 2020;5:60-68
<https://doi.org/10.11925/INFOTECH.2096-3467.2020.1028>.
30. Yang S, Hao Y, Dai L. Evaluation and Empirical Research on the Value of Patent Technology. *Forum on Science and Technology in China* 2017;(09):146-152
<https://doi.org/10.13580/j.cnki.fstc.2017.09.018>.
31. Sergio Gómez Arroyave, Eliana María Villa Enciso, Claudia Nelcy Jiménez Hernández. Minimum qualitative variables to value patents by technology-based firms. *Journal*

- of Technology Management and Innovation 2017;14(2):21-32
<https://doi.org/10.4067/S0718-27242019000200021>.
32. .Liu W, Liu X, Qiao W. Probabilistic graph-based valuation model for measuring the relative patent value in a valuation scenario. Pattern Recognition Letters 2020;138(3) <https://doi.org/10.1016/j.patrec.2020.07.023>.
 33. Raan V, Anthony FJ. Sleeping beauties cited in patents: is there also a dormitory of inventions?. Scientometrics 2017;110(3):1123-1156
<https://doi.org/10.1007/s11192-016-2215-8>.
 34. Tahmooresnejad, L, Beaudry, C. Capturing the economic value of triadic patents. Scientometrics 2018;118 <https://doi.org/10.1007/s11192-018-2959-4>.
 35. Xu X, Zhao Wh, Yao Z. Identification and Application Research of High-Quality Patents from a Multidimensional Perspective. Journal of Modern Information 2019;39(11):13-22,45 <https://doi.org/10.3969/j.issn.1008-0821.2019.11.002>.
 36. Yan Y. Research on the Current Situation and Countermeasures of Traditional Chinese Medicine Patent Quality in Guangzhou. Guangzhou: Guangzhou University of Chinese Medicine 2014.
 37. .Liu D, Gong FH, Fu Q, FAIZ MA, Li XT, Cei S. Game theory-weighted GRA-TOPSIS evaluation model for irrigation water use efficiency. Transactions Chinese Society for Agricultural Machinery. Trans 2017;48(5):9 <https://doi.org/10.6041/j.issn.1000-1298.2017.05.027>.
 38. Shi BF, Chi GT. Indicator selection model based on information maximization for green industry assessment: Methodology and case study. Systems Engineering: Theory Practice 2014;34(7):1799 – 1810 [https://doi.org/10.12011/1000-6788\(2014\)7-1799](https://doi.org/10.12011/1000-6788(2014)7-1799).
 39. Jolliffe IT. Principal Component Analysis. Springer, 2002.
 40. Dai J, Jing QI, Chi J, Chen S, Yang J, Liping JU, et al. Integrated water resource security evaluation of Beijing based on GRA and TOPSIS. Frontiers of Earth Science in China 2002;4(003):357-362 <https://doi.org/10.1007/s11707-010-0120-7>.
 41. .Trela K, Campbell Y, Dornbusch F, Pohle A. How to find new industry partners for public research: a classification approach. IEEE Transactions on Engineering Management 2002;PP(99):1-18 <https://doi.org/10.1109/TEM.2020.2992060>.
 42. Nair SS, Mathew M, Nag D. Dynamics between patent latent variables and patent price. Technovation 2011;31(12):648-654
<https://doi.org/10.1016/j.technovation.2011.07.002>.

Acknowledgements

The funders have/had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Funding

This work was supported by National Social Science Fund for Young Scholars (25CGL152 to Wenjun Gu).

Author contributions

Wenjun Gu contributed to formal analysis and writing (original draft). Jinhua Wang contributed to formal analysis. Jiyu Li contributed to conceptualization, and funding acquisition. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Data availability statement

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Figures

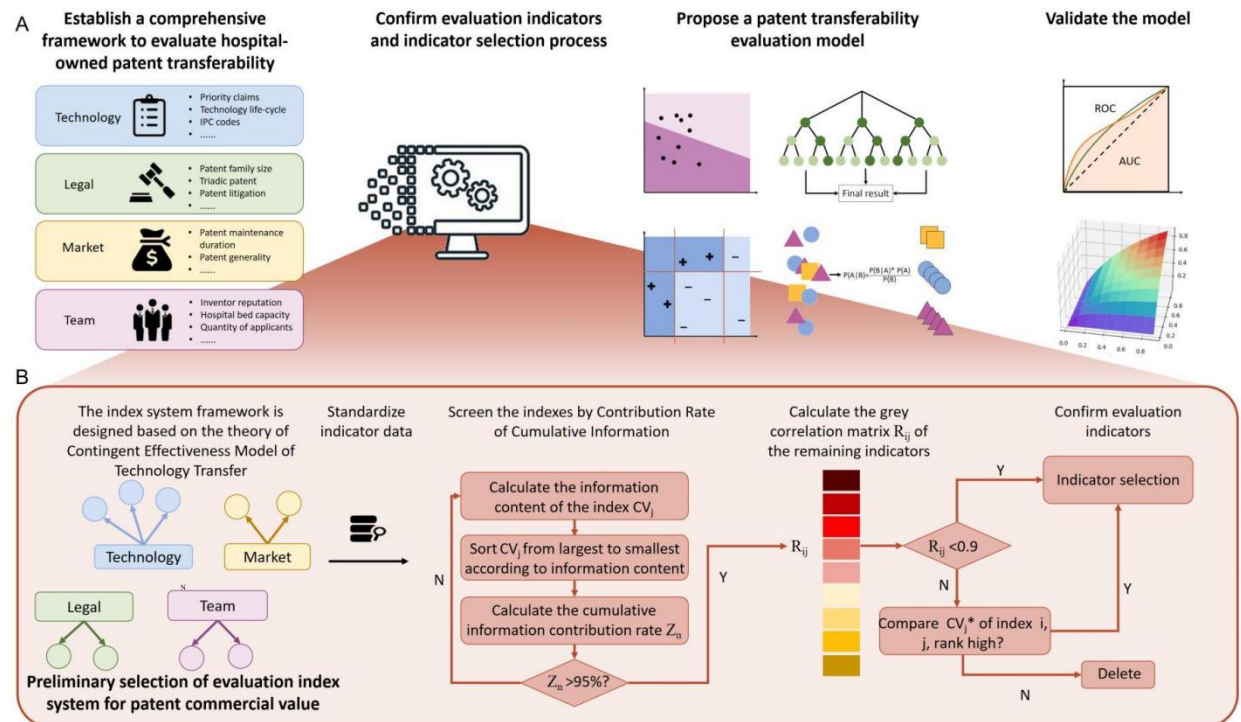


Figure 1. Workflow of machine learning-based evaluation of hospital-owned patent transferability
(A) the workflow of this study. (B) The steps of establishment of a metrics system for performance assessment.

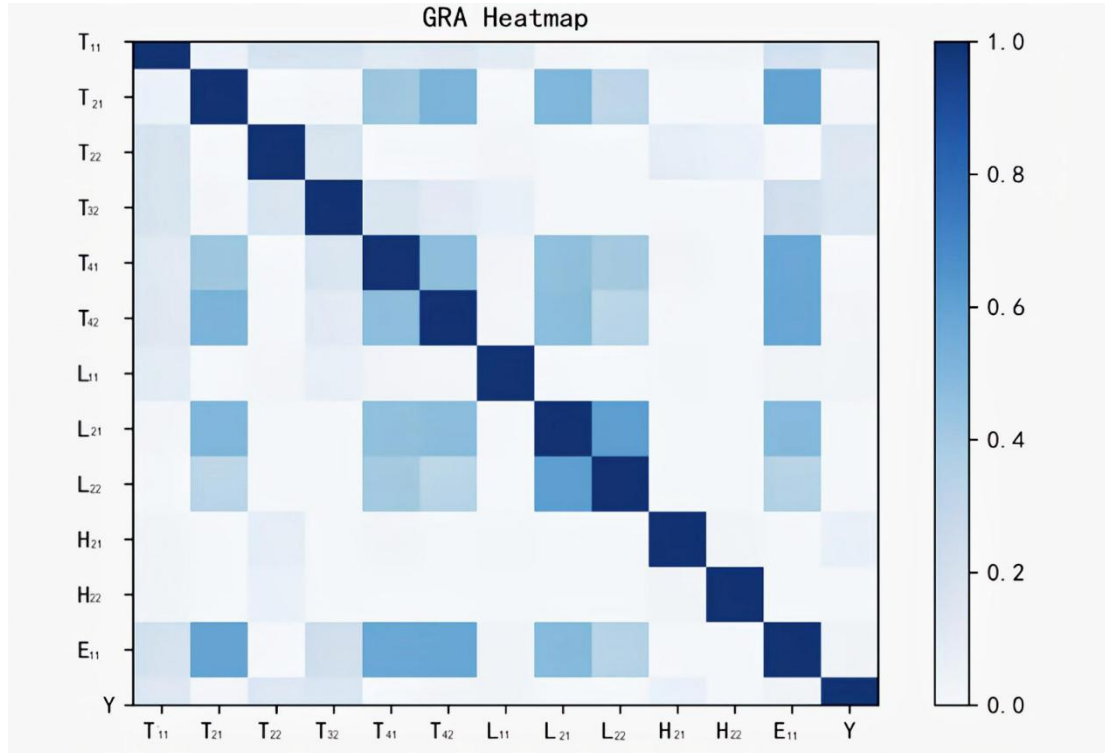


Figure 2. Feature index correlation matrix heat-map

Note: In the heat-map visualization, correlation strength is encoded in color saturation, with darker hues denoting more significant correlations. Here: T₁₁: Standardized citation counts; T₂₁: Priority claims; T₂₂: Technology life-cycle; T₃₂: Non-patent literature references; T₄₁: IPC code; T₄₂: Claim quantity; L₁₁: Patent litigation frequency; L₂₁: Patent family size; L₂₂: Triadic patent counts; H₂₁: inventor reputation; H₂₂: hospital bed capacity; E₁₁: patent maintenance duration.

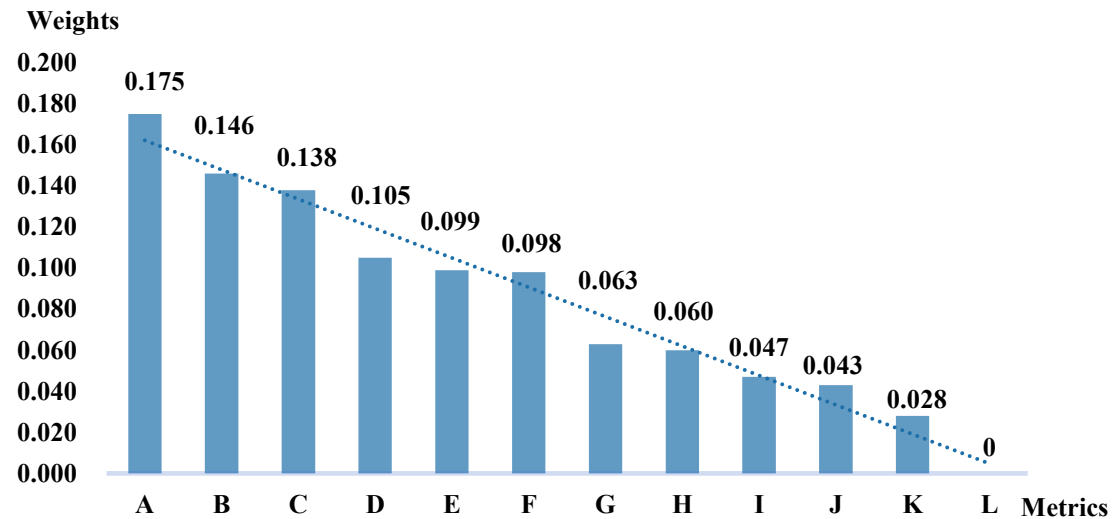


Figure 3. Feature Importance Analysis of the Random Forest Model for Assessing Hospital-owned Patent Transferability

Note: A: Claim quantity; B: patent maintenance duration; C: inventor reputation; D: hospital bed capacity; E: Technology life-cycle; F: Standardized citations; G: Non-patent literature references; H: IPC codes; I: Priority claims; J: Patent family size; K: Triadic patents; L: Patent litigation frequency.

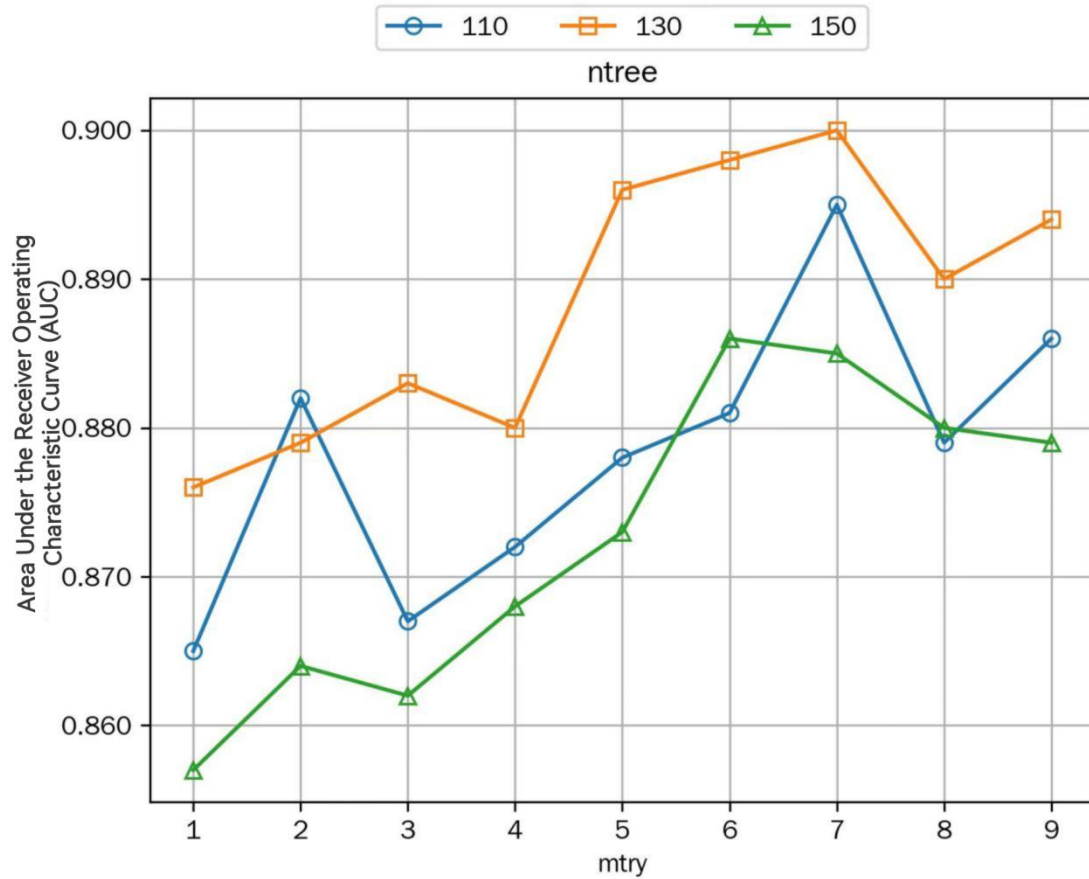


Figure 4. Parameter tuning plot of random forest based on ten - fold cross - validation

Note: In this study, we focused on two key parameters: *mtry* (the number of variables randomly sampled as candidates at each split) and the number of trees (represented by 110, 130, and 150 in the figure). As shown in Figure_4, for all three tree numbers, the AUC values exhibited a trend of first increasing and then decreasing as *mtry* increased. When the number of trees was set to 130, the RF model achieved the highest AUC value (approaching 0.900) at *mtry*=7. For the cases with 110 and 150 trees, the peak AUC values were lower than that of the 130 - tree model, and they also reached their maxima around *mtry*=7 before declining.

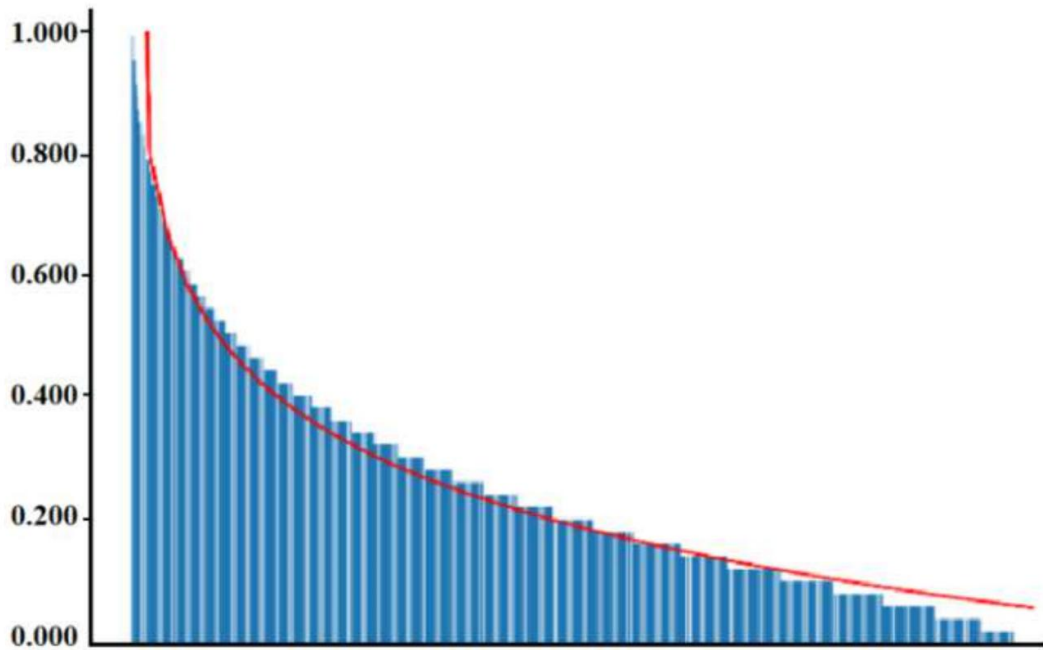


Figure 5. Distribution trend of transferability for non-commercialized domestic patents samples

Note: The fitted logarithmic function for domestic patents was: $y = -0.187\ln x + 1.976$ ($R^2 = 0.967$)

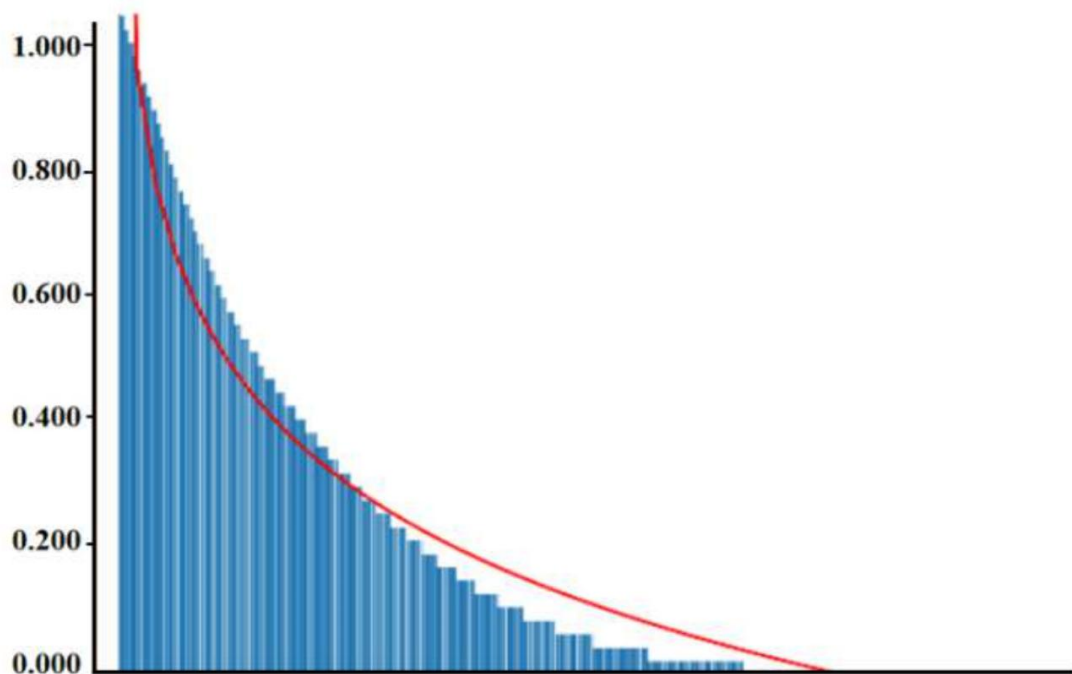


Figure_6 Distribution trend of transferability for non-commercialized international patents samples

Note: The fitted logarithmic function for international patents was: $y = -0.248\ln x + 2.312$ ($R^2 = 0.938$)

Table 1. Multidimensional evaluation framework for hospital-owned patent transferability

Type	Indicator	Definition	Significance	Calculation Method	Reference
Techni- cal	Standardized citation count	Normalized number of patent citations	The number of citations a patent receives correlates strongly with its publication date, given the cumulative nature of citation behavior. This temporal bias necessitates citation count normalization. The indicator measures technological influence.	Citations/ (Patent expiration date Publication date or database lock date)	Federico Caviggioli et al.(2020); ¹⁵ Alberto Méndez-Morales et al.(2022); ¹² Roger Svensson(2022) ²⁷
	Priority claims	Number of priority claims in patent documents	More claims suggest higher innovation level.	Directly extracted from incoPat database	Grimaldi, M. L. Cricelli.(2020) ¹⁷
	Technology lifecycle	Lifecycle stage of patented technology	Longer lifecycle indicates slower innovation speed.	Patent grant year minus Average grant year of cited patents in the patent document	David H.Hsu et al.(2021) ²⁸
	Non-patent literature references	Number of non-patent literature citations	The indicator reflects research background depth.	The reference lists in patent texts were extracted directly from the incoPat database followed by manual screening	You, Xindong et al.(2021) ²⁹ ; Alberto Méndez-Morales et al.(2022) ¹²
	Number of backward citations	Number of the prior art cited in a patent	Represent the technological foundations upon which the current invention is built	Directly extracted from incoPat database	S.Yang et al(2017) ³⁰
	IPC codes count	Number of IPC classifications	More IPC codes indicate broader technical scope.	Directly extracted from incoPat database	Federico Caviggioli et al.(2020) ¹⁵ ; Michele Grimaldi et al.(2020) ¹⁷
	Claim quantity	Count of claims in patent text	More claims indicate broader protection scope.	Directly extracted from incoPat database	Sergio Gómez Arroyave et al.(2019) ³¹ ; Weidong Liu et al.(2020) ³² ; Shu-Hao Chang et al.(2018) ¹⁶

	Word counts of independent claim	The number of words contained in the standalone claims	Define the broadest scope of patent protection	Directly extracted from incoPat database	New added in this study
Legal	Patent litigation frequency	Number of patent lawsuits	Higher litigation frequency suggests greater patent market value.	Directly extracted from incoPat database	Cheng-Chin Tsao et al.(2015) ²¹
	Patent family size	Number of countries/regions in patent family	Larger families indicate wider legal protection.	Directly extracted from incoPat database	Anthony F.J. van Raan(2017) ³³
	Triadic patent counts	Patents granted in US, Europe, and Japan	These regions represent global economic leaders and the indicator reflects technological impact.	Directly extracted from incoPat database	Tahmooresnejad, L.& C. Beaudry.(2019) ³⁴
	Publication-to-grant period	The time elapsed between a patent's initial publication and its final authorization	Reflect the efficiency of the patent examination process	Directly extracted from incoPat database	You, Xindong et al.(2021) ²⁹
Team	Inventor reputation	The technical invention experience of the first inventor in the specialized field	More experienced inventors may create higher-value patents.	The number of the first inventor's patent output in the relevant technical domain/the number of all patents in the relevant technical domain	Wei Du et al.(2021) ²⁴
	Hospital bed capacity	Number of beds in patent-owning medical institutions	More beds suggest stronger clinical capabilities.	Collected from hospital websites or literatures	New added in this study
	Quantity of applicants	The number of applicants	Reflect the collaborative nature of the invention	Directly extracted from incoPat database	Grimaldi, M.L. Cricelli.(2020) ¹⁷
	Applicant category	The applicants were classified into three main categories: enterprises (e.g, corporations and startups), academic institutions (e.g, universities and research centers), and individual applicants.	The indicator were employed to characterize the level of technological innovation and market applicability in this study.	Directly extracted from incoPat database	Xin X.et al(2019) ³⁵
	Inventor team size	The total count of individuals credited with contributing to a patented invention	A proxy for research team size and collaborative intensity	Directly extracted from incoPat database	Yan Y.(2014) ³⁶

Market	Patent maintenance duration	Years of patent validity maintenance	Longer maintenance indicates higher expected economic returns.	Expiration date or Database lock date - Granted date.	Cheng-Chin Tsao et al.(2015) ²¹ ; Jurrie'n Bakker(2017) ¹⁸
	Filing-to-transfer period	The period from patent application submission to the first legally recorded licensing or assignment agreemen	There exists a positive relationship between filing-to-assignment/license duration and total commercialization costs	Directly extracted from incoPat database	Federico Caviggioli et al.(2020) ¹⁵
	Patent generality	The number of subsequent citations received from inventions within the same IPC code	General-purpose technologies enable broader applications	Executed under Python 3.11.8	Sonmenz,Z.(2018) ²⁰

Table 2. Descriptive analysis between domestic and international granted hospital-owned patents

[illegible]

Table 2. Descriptive analysis between domestic and international granted hospital-owned patents(continued table)

[illegible]

Table 3. CRCI-based indicator screening for hospital-owned granted invention patent transferability value assessment (International)

Metric	Information content CV_j	Sort by Indicator	The sorted CV_j^*	Contribution Rate of Cumulative Information(%)	Variable Selection
T ₁₁	3.811	L ₁₁	29.45	53.874	Retain
T ₂₁	1.234	T ₁₁	3.811	60.845	Retain
T ₂₂	2.072	T ₃₂	3.639	67.502	Retain
T ₃₁	0.099	H ₂₁	3.259	73.463	Retain
T ₃₂	3.639	L ₂₂	2.639	78.291	Retain
T ₄₁	1.300	T ₂₂	2.072	82.081	Retain
T ₄₂	1.230	L ₂₁	1.301	84.461	Retain
T ₄₃	0.064	T ₄₁	1.300	86.838	Retain
L ₁₁	29.450	H ₂₂	1.245	89.116	Retain
L ₁₂	0.145	T ₂₁	1.234	91.374	Retain
L ₂₁	1.301	T ₄₂	1.230	93.623	Retain
L ₂₂	2.639	E ₁₁	0.986	95.427	Retain
H ₁₁	0.695	H ₁₂	0.836	96.956	Elimination
H ₁₂	0.836	H ₁₁	0.695	98.227	Elimination
H ₁₃	0.432	H ₁₃	0.432	99.018	Elimination
H ₂₁	3.259	E ₂₁	0.173	99.334	Elimination
H ₂₂	1.245	L ₁₂	0.145	99.600	Elimination
E ₁₁	0.986	T ₃₁	0.099	99.782	Elimination
E ₁₂	0.055	T ₄₃	0.064	99.899	Elimination
E ₂₁	0.173	E ₁₂	0.055	100.000	Elimination

Note: : T₁₁: standardized citation counts; T₂₁: priority claims; T₂₂: technology life-cycle; T₃₁: number of patent citations; T₃₂: non-patent literature references; T₄₁: IPC code; T₄₂: claim quantity; T₄₃: word counts of independent claim; L₁₁: patent litigation frequency; L₁₂: publication-to-grant period; L₂₁: patent family size; L₂₂: triadic patent counts; H₁₁: quantity of applicants; H₁₂: inventor team size; H₁₃: applicant category; H₂₁: inventor reputation; H₂₂: hospital bed capacity; E₁₁: patent maintenance duration; E₁₂: filing-to-transfer period; E₂₁: patent generality.

Table 4 CRCI-based indicator screening for hospital-owned granted invention patent transferability value assessment (Domestic)

Metric	Information content CV_j	Sort by indicator	The sorted CV_j^*	Contribution Rate of Cumulative Information(%)	Variable Selection
T ₁₁	1.896	L ₁₁	48.791	63.441	Retain
T ₂₁	5.523	T ₁₁	9.313	75.551	Retain
T ₂₂	0.880	T ₃₂	5.523	82.731	Retain
T ₃₁	0.013	H ₂₁	2.703	86.246	Retain
T ₃₂	1.574	L ₂₂	1.896	88.712	Retain
T ₄₁	0.751	T ₂₂	1.574	90.758	Retain
T ₄₂	0.801	L ₂₁	1.068	92.147	Retain
T ₄₃	0.028	T ₄₁	0.880	93.290	Retain
L ₁₁	48.791	H ₂₂	0.824	94.362	Retain
L ₁₂	0.055	T ₂₁	0.801	95.404	Retain
L ₂₁	0.788	T ₄₂	0.788	96.428	Elimination
L ₂₂	9.313	E ₁₁	0.751	97.405	Elimination
H ₁₁	0.416	H ₁₂	0.689	98.301	Elimination
H ₁₂	0.689	H ₁₁	0.518	98.974	Elimination
H ₁₃	0.518	H ₁₃	0.416	99.516	Elimination
H ₂₁	2.703	E ₂₁	0.265	99.860	Elimination
H ₂₂	0.824	L ₁₂	0.055	99.931	Elimination
E ₁₁	1.068	T ₃₁	0.028	99.968	Elimination
E ₁₂	0.012	T ₄₃	0.013	99.985	Elimination
E ₂₁	0.265	E ₁₂	0.012	100.000	Elimination

Note: : T₁₁: standardized citation counts; T₂₁: priority claims; T₂₂: technology life-cycle; T₃₁: number of patent citations; T₃₂: non-patent literature references; T₄₁: IPC code; T₄₂: claim quantity; T₄₃: word counts of independent claim; L₁₁: patent litigation frequency; L₁₂: publication-to-grant period; L₂₁: patent family size; L₂₂: triadic patent counts; H₁₁: quantity of applicants; H₁₂: inventor team size; H₁₃: applicant category; H₂₁: inventor reputation; H₂₂: hospital bed capacity; E₁₁: patent maintenance duration; E₁₂: filing-to-transfer period; E₂₁: patent generality.

Table 5. Performance metrics of six machine learning classifiers

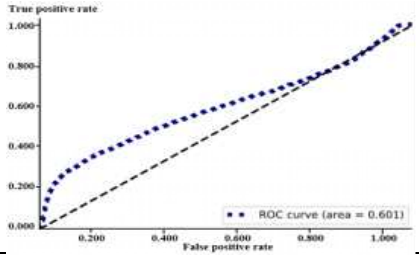
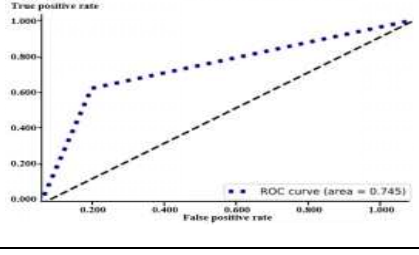
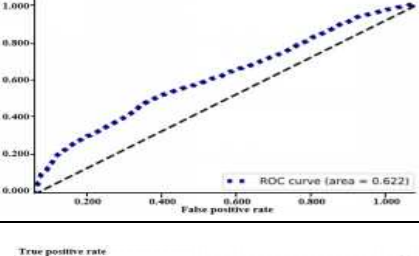
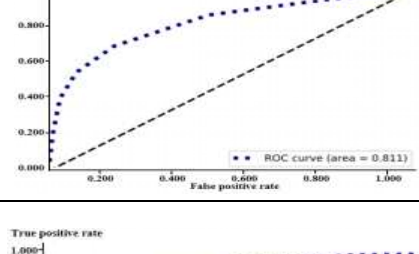
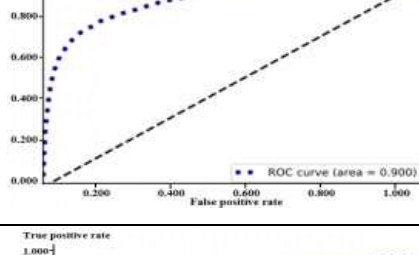
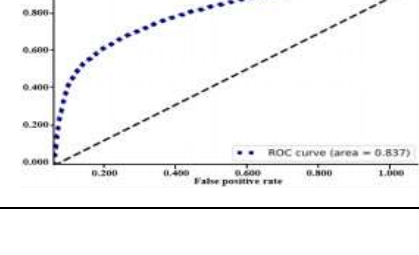
Algorithm name	Accuracy	<i>F1-score</i>	ROC Curve	AUC Area
Support Vector Machine	0.720	0.665		0.601
Decision Tree	0.808	0.806		0.745
Naive Bayes	0.718	0.705		0.622
K-nearest neighbors	0.821	0.822		0.811
Random Forest	0.871	0.869		0.900
AdaBoost	0.826	0.819		0.837

Table S1. Feasibility assessment results of classification algorithms

[illegible]