Valuation of University-Originated Technologies: A Predictive Analytics Approach

Young-Choon Kim, Joon Mo Ahn, Ohjin Kwon, and Changyong Lee

***Abstract*—Experts have difficulty assessing the economic value of university-originated technologies due to the high level of uncertainty associated with the commercialization of early stage and basic technologies. This article proposes a random forest approach to the valuation of university-originated technologies that integrates monetary value and patent value models for technology valuation. First, a technological characteristics-value matrix was constructed after defining a total of 23 indicators from the U.S. Patent and Trademark and Scopus databases and extracting the value of university-originated technologies from technology transaction databases. Second, a random forest model, an ensemble machine learning model based on a multitude of decision trees, was employed to assess the economic value of university-originated technologies. Finally, the performance of our approach was assessed using quantitative metrics. A case study of the technologies registered in the Office of Technology Licensing of Stanford University confirms, with statistically significant outcomes, that our method is valuable as a complementary tool for the valuation of university-originated technologies.**

***Index Terms*—Patent databases, publication databases, random forest, technology transaction databases, technology valuation, university-originated technologies.**

I. Introduction

TECHNOLOGY valuation is an intractable task since the economic value of technologies is affected by various factors and is realized only after its commercialization [60]. This is especially true for the university-originated technologies that are characterized as being early stage and basic (rather than mature and applied) inventions.[1] While there is a growing body of literature on university-originated technologies and university licensing activities—such as the performance of technology licensing officers [10], [24], license payment policies and structures [8], [32], business models of technology licensing officers [2], and patterns of licensing outcomes [19], [52], [59], [65]—a major question remains in the literature regarding how to use quantitative data and scientific methods to assess the economic value of university-originated technologies.

A variety of technology valuation methods—ranging from intuitive judgments to complex real options theory—have been presented. The review of previous methods suggests that the use of a single model is not appropriate for the valuation of university-originated technologies, while the integration of existing valuation methods can create synergies and overcome the current limitations in each. Four main issues are central to this problem and need to be addressed. First, any approach that is proposed should measure the monetary worth of the technologies, rather than the relative preference or technological characteristics, to give practical assistance [45]. Second, distinct technological characteristics should be incorporated into the valuation since the technological factors are the most important consideration in the commercialization of university-originated technologies [53]. However, there is a lack of theoretical understanding about the relationships between the technological characteristics and the economic value of technologies, and moreover, the complexity and nonlinearity associated with technology valuation make the design of certain functions impractical [38]. Hence, previous deductive approaches should be extended to context-specific inductive approaches. Third, validation has usually been omitted in the literature, and the validation that has been performed was qualitative and domain specific. The performance and utility of the approach should be assessed through extensive application to ensure external validity. Finally, any approach that is proposed needs to enable a quick analysis of a wide range of technologies and support decision making with acceptable levels of time and cost [1], [39].

Considering these issues, we propose a random forest approach to the valuation of university-originated technologies that integrates the monetary value and patent value models for technology valuation. The premise of this research is that the technological characteristics or superiority captured by quantitative indicators can provide valuable information on the economic value of university-oriented technologies. With a historical technology transaction database integrated with patent and publication databases, we have defined a computational problem for the valuation of university-originated technologies: given the many quantitative indicators for university-originated

Manuscript received October 22, 2018; revised April 10, 2019; accepted August 19, 2019. This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIP) under Grant 2017R1C1B2011434. Review of this manuscript as arranged by Department Editor Y. Zhou. *(Corresponding author: Changyong Lee.)*

Y.-C. Kim is with the School of Business Administration, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea (e-mail: yckim@unist.ac.kr).

J. M. Ahn is with the Graduate School of Management of Technology, Sogang University, Seoul 04107, Republic of Korea (e-mail: jmahn@sogang.ac.kr).

O. Kwon is with the Korea Institute of Science and Technology Information, Seoul 130-741, Republic of Korea (e-mail: dbajin@kisti.re.kr).

C. Lee is with the School of Management Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea (e-mail: changyong@unist.ac.kr).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TEM.2019.2938182

[1]Existing literature presents that only a few (12%) of university-originated technologies are ready for practical or commercial use, while the overwhelming majority are either a proof of concept (45%) or a lab scale prototype (37%) [59].

technologies, can we classify them according to their economic value? To answer this, first, a technological characteristics-value matrix was constructed after defining a total of 23 indicators from patent and publication databases and extracting the value of university-originated technologies from technology transaction databases. Second, a random forest model—that is, an ensemble machine learning model based on a multitude of decision trees—was employed to assess the economic value of university-originated technologies. The primary advantage of this method for the valuation of university-originated technologies lies in its ability to infer the complex and nonlinear relationships among the technological characteristics and economic value of the technologies through empirical and inductive learning processes [9]. Moreover, this method has the ability to ignore irrelevant input variables [57] and is thereby appropriate for exploratory problems where there is little theoretical understanding about the relationships among variables (i.e., the technological characteristics and economic value of technologies). Finally, the performance and utility of the proposed approach were assessed using quantitative metrics.

The proposed approach was applied to the valuation of the university-originated technologies that are registered in the Office of Technology Licensing (OTL) of Stanford University. We employed the U.S. Patent and Trademark Office (USPTO) and Scopus databases as our data sources due to their scope and reliability. Our approach obtained statistically significant results at acceptable levels of time and cost. The case study also identified a way to improve the proposed approach, which we expect to be a useful tool to support those in charge of technology licensing but have little domain-specific knowledge. The systematic process and quantitative outcomes offered by the proposed approach are expected to serve as a starting point for developing more general models.

II. Background

A. Characteristics of University-Originated Technologies

We define "university-originated technologies" as the inventions[2] that are created by university researchers through their research work and disclosed to the university's OTL for commercialization. University-originated technologies, although coming from a broad spectrum of technological areas, including computer sciences, engineering, and life sciences, represent a type of technology characterized by inherent uncertainties about commercial viability. These technologies are typically in the embryonic stage of development, located somewhere between theory and prototypes, often serving as source technology for further applications. According to a university survey [27], a majority of university inventions are either in a proof of concept stage (45%) or a lab-scale prototype stage (37%). Only a small fraction of university inventions is known for the manufacturing feasibility of the technologies (15%) or for being ready for practical or commercial use (12%).

[2]The inventions include many types of discoveries and technical innovations, including processes, methods, machines, article of manufacture, devices, chemicals, and compositions of matter. An invention, if novel, useful, and nonobvious, may be protected by a patent.

This embryonic nature of these technologies makes it difficult to fully anticipate their properties and potential applications. Prior research documented that significant applications from an early stage technology may arise in contexts that its inventors may not anticipate; furthermore, the focal technology's potential is often dependent on improvements in other technologies, which makes it even harder to foresee the viability of a particular invention [34], [49]. Adding to such uncertainty, due to the early stage nature of the technology, it is difficult to evaluate how much demand a commercialized technology will command, what specific market segment may take interest, and how much revenue it will generate through commercialization.

In the U.S., university inventions, once created in research laboratories, are disclosed to the university's OTL. Then, the OTL files for patent applications for the disclosed university inventions and licenses it to commercial partners through licensing arrangements. The process of commercialization requires mutual collaboration between the inventors and licensees for commercial success. Among the disclosed inventions, usually less than 20% are licensed [59]. Licensing requires financial negotiations (such as upfront payments, royalties, and sometimes equity stakes) and specifies the coverage of the intellectual property rights and development schedules. Successful commercialization usually takes several years. Prominent examples of successful commercialization of university inventions include Recombinant-DNA, jointly developed by UC Berkeley and Stanford faculty members, the cotransformation process, developed at Columbia University, the three-dimensional (3-D) printing process, developed by MIT researchers, and PageRank, an Internet search algorithm developed at Stanford and licensed to the inventor's own start-up, Google.

B. Models and Methods for Technology Valuation

1) Monetary Value Model: Monetary value models estimate the monetary worth of technologies based on capital budgeting methods such as net present value and internal rate of return and can broadly be classified into four approaches according to valuation criteria and procedures, i.e., cost approach, market approach, income approach, and real options approach. First, the cost approach determines the value of a technology based on the principle of substitution that postulates that a buyer would pay no more and a seller could command no more than the cost to create an intellectual asset of equal desirability and utility [54]. Although the concept of the approach makes sense to some extent, the drawback here is that this method relies on accurate cost and depreciation data, which is unlikely to exist for early stage and basic technologies. Equally, the cost approach is unable to consider future profits. Second, the market approach estimates the value of technology by comparing the value of comparable technologies based on historical transaction data [47]. This approach is likely to assign a realistic value that buyers and sellers can agree on and yet has not been deployed in practice due to the lack of such data [45]. Third, the income approach assesses the value of a technology based on the expected economic benefits generated by the technology [6]. This approach is the most desirable in theory because it results in a real meaning of value, but theoretical complexities make it difficult to estimate the relevant parameters [1], [64]. Moreover, the embryonic nature of university-originated technologies makes parameter estimation more difficult than that of mature technologies. Finally, the real options approach is rooted in the income approach, but what distinguishes the real options approach from the general income approach is flexibility [46]. The introduction of real options theory provides a robust means of dealing with uncertain future circumstances and offers managers flexibility in decision making [7], [66]. However, this approach suffers from the same shortcomings as the income approach.

Given that university-originated technologies are early stage and basic inventions and that there is high level of uncertainty associated with the commercialization processes, the market approach is considered the most practical and promising solution since, if applicable transaction data does exist, it assigns realistic values that buyers and sellers can agree on and incorporates the economic benefits generated by similar technologies into the valuation. Moreover, it should also be noted that the technological factors are more important than other factors for the valuation of university-originated technologies, although non-technological factors become more important in the later phases of technology commercialization [29]. Considering these issues, this article adopts a market approach and employs patents and publications to incorporate distinct technological characteristics into the valuation of university-originated technologies.

2) Patent Value Model: Patent value models measure the quality of patents as a proxy for the value of the relevant technologies [11], [26], [35], [37], [53]. The economic and innovation literature has presented a wide range of patent indicators, such as patent forward citation, family, and originality, that may be indicative of the quality of patents and further, the relevant technology's economic value [22]. In line of these efforts, recent studies have employed patent value models for different objectives of technology and innovation management research, for instance, to identify promising technologies [36], [56], to examine emerging technologies [38], to investigate a technology's progression through its life cycle [23], [37], and to assess the cross-country innovation gaps [31].

Patent value models are appropriate for the valuation of university-originated technologies for the following reasons: First, whether a university-originated technology is protected by patents is critical to its valuation, since universities often apply for patents when the commercial viability of a technology is known [59]. In many cases, it is difficult to convince the patent office of the "utility" of university-originated technologies that are only at a proof of concept stage. Second, the technological characteristics measured by the quantitative patent indicators, if integrated appropriately, make technology valuation more objective and reliable. Many empirical studies have found significant relationships among the patent indicators and the quality and economic value of patents [15], [44]. Finally, this approach enables the valuation of a wide range of university-originated technologies at acceptable levels of time and cost. The operational efficiency can be enhanced by software systems, giving specific practical help to the staff in charge of technology valuation.

However, despite their potential utility, patent value models have rarely been employed for the valuation of university-originated technologies. This is mainly because, as with scoring models, the results of this method indicate the technological characteristics or patent quality, rather than the economic value of a technology. Moreover, the complexity and nonlinearity associated with technology valuation make the design of certain functional relationships among the patent indicators and the value of technologies impractical [38]. With recent rapid advances in machine learning technologies, we employ a random forest model to develop a new technology valuation model through empirical, inductive, and repeated learning processes.

III. Methodology

The overall process of the proposed approach is shown in Fig. 1. The proposed approach is designed to be executed in three discrete steps: data collection and preprocessing; construction of a technological characteristics-value matrix; and finally, assessment of the value of university-originated technologies.

A. Data Collection and Preprocessing

The relevant patents for the technology set in the transaction database (Set 1) are collected based on the patent number that is the identifier that distinguishes issued patents by using the search formula, *PN/patent number*. The patent documents collected at this stage are a mixture of both structured and unstructured data in either HTML or XML formats. The documents are thus parsed according to the type of information (e.g., patent number, assignee, and class) and stored in a structured database. The patents and journal papers that are published by the inventors of Set 1 patents (Set 2 and Set 3) are also collected and preprocessed to analyze the development efforts and capabilities and prior academic achievements of the inventors. Moreover, the patents that cite the Set 1 and Set 2 patents (Set 4) are collected and preprocessed to analyze the patent citations made by later patents.

B. Construction of a Technological Characteristics-Value Matrix

Our approach incorporates the technological characteristics that are measured by quantitative indicators into technology valuation. We form a technological characteristics-value matrix that is divided into three parts: the identifier of the technologies, 23 indicators extracted from the database constructed in the preceding step, and the economic value of a technology that is transacted in the market.

Previous studies have presented numerous patent indicators that may be indicative of the quality of a patent and further, a relevant technology's economic value. However, not all of them can be employed for the valuation of university-originated technologies. For example, forward citations, a crucial variable in patent value models, cannot be used for the valuation of the latest university-originated technologies, since sufficient time is required for the relevant patents to be cited or fail to be cited [23]. Considering *what* the indicators measure and *when* they can be

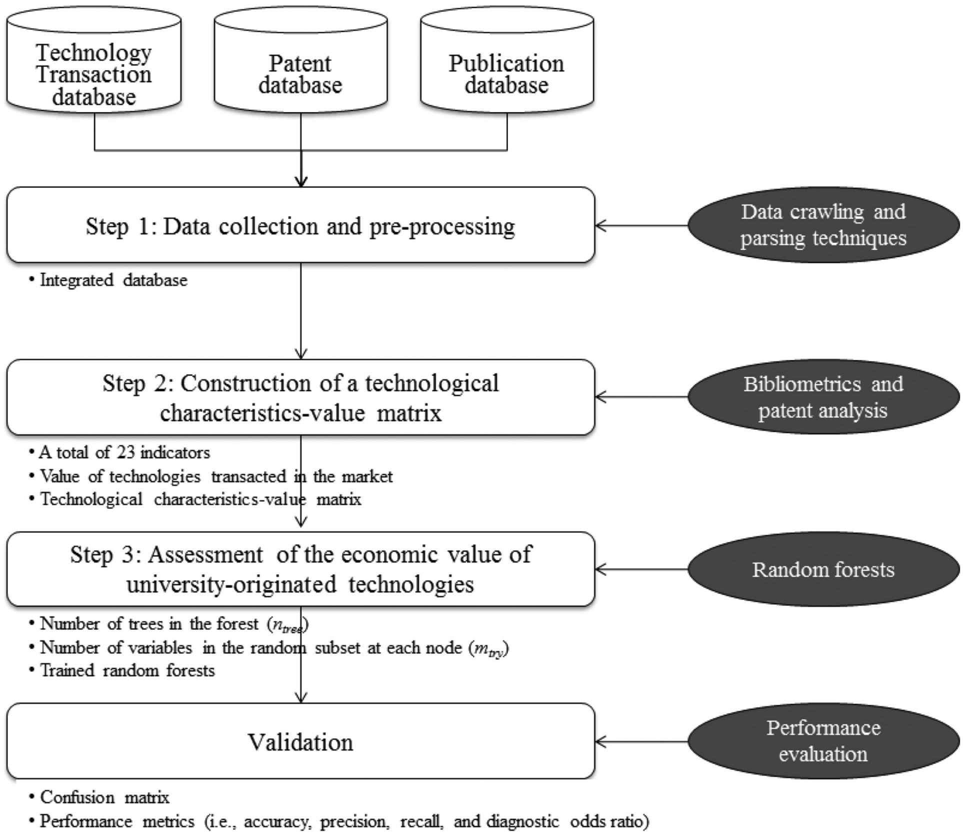

Fig. 1. Overall process of the proposed approach.

extracted from the database, we define a total of 23 indicators, which can be divided into five categories according to their characteristics and implications: (1) technological field and scope, (2) novelty and originality, (3) protection and market coverage, (4) development efforts and capabilities, and (5) prior academic achievements. The indicators that belong to the same category measure similar concepts (e.g., core area strength and peripheral area strength) or measure the same concept at different levels (e.g., class-level originality and mainline subclass-level originality) to enable a fine-grained characterization. Originating from bibliometrics, the first four categories rely on patent databases, while the fifth uses publication databases. Importantly, all the indicators can be extracted from the databases immediately after the relevant technologies are registered in the OTL, enabling an early assessment of the value of university-originated technologies. Table I summarizes the operational definitions of the indicators with the references that can serve as a starting point for obtaining more details.

C. Assessment of the Economic Value of University-Originated Technologies

The economic value of university-originated technologies is affected by many factors and cannot be ascertained fully. Likewise, industrial practitioners in the OTLs assess the economic value of university-originated technologies using ordinal scales (e.g., A, B, and C). As such, we employ a random forest model to classify the technologies according to their economic value which is estimated based on the indicators defined in the preceding section.

The random forest algorithm, as proposed by Leo Breiman [9], is an ensemble machine learning model consisting of a collection of decision trees $\{T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \ldots, T_{n_{\mathrm{tree}}}(\boldsymbol{X})\}$, where $\boldsymbol{X} = \{x_1, x_2, \ldots, x_p\}$ is a p-dimensional input vector (i.e., the indicators defined in the preceding step). Compared to standard decision trees, the generalization error is reduced significantly by introducing random perturbations into the learning procedure and combining the predictions of several different models produced from a single learning set [13]. Specifically, this method produces n_{tree} outputs from the individual decision trees $\{\hat{Y}_1 = T_1(\boldsymbol{X}), \hat{Y}_2 = T_2(\boldsymbol{X}), \ldots, \hat{Y}_{n_{\mathrm{tree}}} = T_{n_{\mathrm{tree}}}(\boldsymbol{X})\}$, which are aggregated to provide the final prediction, $\hat{Y}$. For classification problems, the final prediction $\hat{Y}$ is determined by using majority rules.

The operation of the random forest algorithm involves three steps, as follows. The first step draws n_{tree} bootstrap samples from the original data (i.e., randomly sampled with replacement). The second step constructs a fully grown and unpruned decision tree for each of the bootstrap samples. Here, in contrast to standard decision trees where the best split among all of the input variables is chosen, in the random forest algorithm, the best split from among m_{try} randomly selected input variables is chosen. Finally, the new data is classified by aggregating the predictions of the n_{tree} decision trees $(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_{n_{\mathrm{tree}}})$ (i.e.,

TABLE I
SUMMARY OF THE INDICATORS EMPLOYED IN THIS RESEARCH

Category	Subcategory	Indicator	Operational definition	References
Technological field and scope	Technological field	Main field (MF)	Primary class of a patent	Lee et al. (2009)
	Technological scope	Patent count (PC)	Number of patents for a technology	Ernst (2003)
		Class-level scope (CS)	Number of classes to which a patent belongs	Lerner, 1994; Lee et al. (2018)
		Mainline subclass-level scope (MS)	Number of mainline subclasses to which a patent belongs	
Novelty and originality	Novelty	Publication year (PY)	Year when a patent was issued	Lee et al. (2015)
		Prior knowledge (PK)	Number of backward citations of a patent	Harhoff et al. (2003); Lee et al. (2016)
		Scientific knowledge (SK)	Number of non-patent literature references of a patent	Trajtenberg (1990)
		Technology cycle time (TCT)	Median age of cited patents	Bierly and Chakrabarti (1996)
	Originality	Class-level originality (CO)	Herfindahl index on classes of cited patents	Bessen (2008); Jaffe and Trajtenberg (2002); Lee et al. (2016)
		Mainline subclass-level originality (MO)	Herfindahl index on mainline subclasses of cited patents	
Protection and market coverage	Protection coverage	Independent claims (IC)	Number of independent claims of a patent	Reitzig (2004); Trappey et al. (2012)
		Dependent claims (DC)	Number of dependent claims of a patent	
	Market coverage	Patent family (PF)	Number of patents registered in multiple countries with the coverage of the same invention	Gullect and Potterie (2000)
Development efforts and capabilities	Development efforts	Collaboration (Col)	1 if a patent has more than one assignee, otherwise 0	Ma and Lee (2008)
		Human resources (HR)	Number of inventors of a patent	
	Development capabilities	Total know-how (TKH)	Number of patents issued by an assignee	Meyer (2006)
		Core area know-how (CKH)	Number of patents in a technology field of interest issued by an assignee	
		Peripheral area know-how (PKH)	Number of patents in other technology fields issued by an assignee	
		Total technological strength (TTS)	Number of forward citations of patents issued by an assignee	Ernst (2003)
		Core area technological strength (CTS)	Number of forward citations of patents in a technology field of interest issued by an assignee	
		Peripheral area technological strength (PTS)	Number of forward citations of patents in other technology fields issued by an assignee	
Prior academic achievements	Research performance	Research quantity (RQual)	Number of citations of the papers published by a lead inventor	Van Looy et al. (2006)
		Research quality (RQuan)	Number of papers published by a lead inventor	

using averaging processes for regression and majority rules for classification).

The use of bootstrapping techniques enables a random forest model to offer internal estimates of important statistics by using the left-out samples, which are called out-of-bag (OOB) samples (D^{OOB}). On average, each decision tree is grown using approximately $1 - e^{-1} \approx \frac{2}{3}$ of the samples, leaving $e^{-1} \approx \frac{1}{3}$ as OOB samples [57]. In the case of the generalization errors, the data in the OOB samples (i.e., input variables $\boldsymbol{X_i}$ and the corresponding class label Y_i) is predicted by the individual decision trees, $T_{\theta_m}(\cdot)$, whose bootstrap samples (θ_m) did not include $\boldsymbol{X_i}$ and Y_i, and the error rate can be calculated by aggregating the error rates for the OOB predictions, as defined

$$\mathrm{Err}_{\mathrm{OOB}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m} I(T_{\theta_j}(X_i) \neq Y_i) \tag{1}$$

where $I(\cdot)$ is the indicator function, *m* is the number of individual decision trees, $T_{\theta_m}(\cdot)$, whose bootstrap samples (θ_m) did not include $\boldsymbol{X_i}$ and Y_i, and *n* is the number of samples. Moreover, this method assesses the importance of an input variable x_j by investigating how much the prediction error increases when the values of x_j are randomly permuted in the OOB samples, as shown below

$$VI(x_j) = \frac{1}{n_{\mathrm{tree}}}\sum_{i=1}^{n_{\mathrm{tree}}}\left(\mathrm{Err}^{j}_{\mathrm{OOB},i} - \mathrm{Err}_{\mathrm{OOB},i}\right) \tag{2}$$

where $\mathrm{Err}_{\mathrm{OOB},i}$ represents the error of a single decision tree $T_{\theta_j}(\boldsymbol{X})$ on its OOB samples and $\mathrm{Err}^{j}_{\mathrm{OOB},i}$ indicates the error of the decision tree $T_{\theta_j}(\boldsymbol{X})$ on the perturbed sample where the values of x_j in its OOB are permuted.

This method has been widely employed in scientific research such as QSAR modeling and compound classification [57] as well as land cover classification [17]. However, the application of this method has rarely been considered in technology and innovation management research. Guns and Rousseau [21] proposed a method using a random forest approach for recommending high-potential future collaborators. Kwakkel and Cunningham [33] developed an improved version of the patient rule induction method, which was inspired by the random forest algorithm to improve scenario discovery.

A random forest model is considered appropriate for the valuation of university-originated technologies for the following reasons. First, unlike conventional statistical models such as multiple linear regression and logistic regression analysis, this method can deal with the complexity and nonlinearity associated with technology valuation without any assumptions about predetermined curves and probability distributions, since this method develops nonlinear mathematical models through empirical, inductive, and repeated learning processes [58]. Second, this method is useful for exploratory problems where there is little theoretical understanding about the relationships among the variables (i.e., the technological characteristics and the value of technologies), since random forest models have the ability to ignore irrelevant input variables [57] and give reliable estimates of generalization errors and variable importance [16]. Finally, random forest models are found to have consistently lower generalization errors than other methods thus avoiding overfitting problems and handling missing data effectively and thereby enhancing applicability in practice [5].

D. Performance Evaluation

The performance and utility of our approach are ultimately related to the ability to classify university-originated technologies according to their economic value, which corresponds to a multiclass classification problem. Considering this, we scrutinize the accuracy and reliability of the proposed method by using several performance metrics [55]. First, we measure the accuracy per class and the overall accuracy of the proposed approach, as defined in (3) and (4)

$$\text{Accuracy}_i = \frac{tp_i + tn_i}{fp_i + fn_i + tp_i + tn_i} \tag{3}$$

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{fp_i + fn_i + tp_i + tn_i}}{l}. \tag{4}$$

Here, a true positive (tp_i), true negative (tn_i), false positive (fp_i), and false negative (fn_i) for class i represents the number of positive examples correctly classified, the number of negative examples correctly classified, the number of negative examples wrongly classified as positive, and the number of positive examples wrongly classified as negative, respectively, while l is the number of classes. Second, to compensate for the classes having different weights in our case study and because of possible imbalances in the dataset [30], we add the precision (positive predictive value), recall (true positive rate or sensitivity), and diagnostic odds ratio (DOR) [18]. Precision is the number of true positive results divided by the number of all positive results, whereas recall is the number of true positive results divided by the number of positive results that should have been returned. The DOR is a measure of the overall effectiveness of a classifier and is defined as the ratio of the probability of a classification being positive if the subject is actually positive to the probability of a classification being positive if the subject is actually negative, as shown in (5)

$$\text{DOR}_i = \frac{\text{sensitivity} \times \text{specificity}}{(1 - \text{sensitivity}) \times (1 - \text{specificity})}. \tag{5}$$

Here, sensitivity and recall are equivalent, whereas specificity is the ratio of accurately identified negatives. The DOR is independent of prevalence or balanced sets and ranges from zero to infinity. A DOR of one means that the test can be expected to predict a positive outcome regardless of the true condition, thus giving no information, whereas a higher DOR indicates better performance.

IV. Empirical Analysis and Results

A. Data

Our data came from a full sample of inventions that were disclosed to the OTL of Stanford University from January 1970 to July 2014. Starting its operation in 1970, the Stanford OTL has been one of the most active offices in the field of university technology transfer. Cumulatively, the OTL has over 11 000 invention disclosures and has completed over 3500 license agreements and generated over \$1.8 billion in revenue to the University and Stanford inventors (the OTL website).

Stanford inventions, once generated during research work, are disclosed to the OTL and assigned to a licensing team in the relevant technological domain. At the OTL, the inventions are evaluated for manufacturing feasibility, patentability, novelty, potential applications, and possible markets and then are processed for patent protection. The OTL develops a licensing strategy for the particular technology to find industry partners to further commercialize the invention. Once the industry partners are identified as target licensees, the OTL negotiates the licensing contracts, maintains and amends the agreements, monitors the development and commercialization process, and tracks the royalty payments. Among all the disclosed inventions, approximately 50% have patent applications filed; approximately 25% of them were licensed. When royalty payments are generated from the licensing contracts, they are equally distributed to the inventor(s), the inventor's department and to the inventor's school after deductions are made to support the OTL's operation and reimburse any direct expenses (such as patent costs that are not paid by the licensees).

The Stanford data include various types of intellectual property, including emblematic licenses, options, material transfer agreements, and university inventions. From the data, we only selected the university inventions that were issued a U.S. patent up to July 2014. We used the patent numbers from the patents that were issued to university inventions and matched them to the patent information in the USPTO database. When an invention is issued with multiple patents (including continuations and modifications), we matched the invention with all the patents that belong to the invention.

TABLE II
THREE CATEGORIES OF TECHNOLOGIES ACCORDING TO THE (EXPECTED) ECONOMIC VALUE

Category	(Expected) economic value	Number of technologies
L1	Above $500 000	83 (3.8%)
L2	$50 000 - $500 000	356 (16.1%)
L3	0 - $50 000	1772 (80.1%)
Sum		2211 (100%)

TABLE III
PART OF THE TECHNOLOGICAL CHARACTERISTICS-VALUE MATRIX

Technology ID	Patent Number	MF	CS	PY	…	PTS	RQual	RQuan	Value
00-003	6 808 925	435	1	3	…	1	8	1	L2
	7 141 426	435	1	2	…	1	8	1	
	7 732 585	536	2	2	…	98	8	1	
00-009	7 070 681	204	2	3	…	211	404	7	L3
00-010	6 420 119	435	2	3	…	0	126	7	L2
…	…	…	…	…	…	…	...	…	…
03-274	7 766 903	606	1	2	…	52	1877	34	L1
	8 409 180	606	2	3	…	52	1877	34	
	8 616 216	128	2	3	…	82	1877	34	
03-278	7 943 306	435	2	2	…	0	0	0	L3
03-282	7 265 093	514	2	2	…	930	9899	36	L2
…	…	…	…	…	…	…	…	…	…
12-461	8 388 519	600	1	1	…	0	0	0	L3
14-276	6 506 554	435	3	5	…	0	0	0	L3
	6 841 657	530	3	3	…	0	0	0	

Since 1970, Stanford inventions have generated $1.8 billion in licensing income, but only 95 have generated over $1 million. Some of the successful inventions from Stanford include functional antibodies ($613 M), improved hypertext searching ($342 M), recombinant DNA ($255 M), fibre optic amplifier ($48.4 M), discrete multitone technologies for DSL ($29.6 M), FM sound synthesis ($22.9 M), data visualisation software ($14.8 M), and code error detection software ($11.1 M).

B. *Random Forest Approach to the Valuation of University-Originated Technologies*

1) Data Collection and Preprocessing: For the patent database, the USPTO[3] serves as a data source, since the U.S. is the world's largest patent market—the majority of quality patents are submitted to the USPTO—and its database is also well organized and holds the historical information back to 1976 [38]. There are a total of 3157 patents (Set 1) for the 2211 technologies registered in the OTL of Stanford University. Here, a crawling system was developed to download the relevant patents automatically, since the number is sufficiently large that we could not collect them all manually. These documents were then parsed based on their structures, distinguishing each document by its content. The details about patent numbers, assignees, citations, claims, classes, and other information were stored in a patent database that we constructed using Microsoft Office Access. In addition, the patents that are published by the inventors of Set 1 patents (Set 2) and the patents that cite the Set 1 and Set 2 patents (Set 4) were collected, preprocessed, and transformed in the same manner. Finally, we also collected information on the academic achievements (i.e., number of papers and number of total citations) of the inventors of Set 1 patents (Set 3) from the Scopus database.[4]

2) Construction of a Technological Characteristics-Value Matrix: As previously noted, the technologies are classified according to their (expected) economic value. As such, we grouped the technologies into three categories according to the (expected) economic value, as summarized in Table II. Here, L1 technologies are the most valuable, while L3 technologies are the least valuable. It is noteworthy that only 83 technologies

[3][Online]. Available: https://www.uspto.gov

[4][Online]. Available: https://www.scopus.com

TABLE IV
PART OF THE RESULTS OF VALUATION OF UNIVERSITY-ORIGINATED TECHNOLOGIES

Technology ID	Patent Number	MF	CS	MS	…	PTS	RQual	RQuan	Expected value
00-003	6 808 925	435	1	3	…	1	8	1	L2
	7 141 426	435	1	2	…	1	8	1	
	7 732 585	536	2	2	…	98	8	1	
00-009	7 070 681	204	2	3	…	211	404	7	L2
00-010	6 420 119	435	2	3	…	0	126	7	L2
…	…	…	…	…	…	…	...	…	…
03-274	7 766 903	606	1	2	…	52	1877	34	L1
	8 409 180	606	2	3	…	52	1877	34	
	8 616 216	128	2	3	…	82	1877	34	
03-278	7 943 306	435	2	2	…	0	0	0	L3
03-282	7 265 093	514	2	2	…	930	9899	36	L2
…	…	…	…	…	…	…	…	…	…
12-461	8 388 519	600	1	1	…	0	0	0	L3
14-276	6 506 554	435	3	5	…	0	0	0	L3
	6 841 657	530	3	3	…	0	0	0	

(3.8%) belong to L1 technologies, accounting for 58% of the total revenue.

Moreover, we extracted a total of 23 indicators from the integrated database and constructed a technological characteristics-value matrix. The resulting matrix was a 2212 by 25 matrix, which was used to train the random forest model. The matrix is not reported here in its entirety owing to lack of space, but a part of the matrix is shown in Table III. In the table, the first column represents the identifier of technology followed by the 23 input indicators, and the last column represents the categorized value of technologies. For the technologies with multiple patents (e.g., 03-274), their technological characteristics are captured by calculating the average of the patent indicators or by identifying the majority. For instance, technology 03-274 has three patents which were published in 2010, 2013, and 2013 and belong to U.S. patent class 606, 606, and 218 (as a primary class), respectively; the publication year and the main field are thus 2012 and U.S. patent class 606. This technology has high values for many indicators such as originality, patent family, academic papers and number of citations and was classified as an L1 technology, whereas technology 03-278, that has one patent published in 2011 and has low values for the main indicators such as technology cycle time, originality, and patent family, was classified as an L3 technology.

3) Assessment of the Economic Value of University-Originated Technologies: We used the *RandomForest* module provided by the R software to develop a random forest model for the valuation of university-originated technologies. First, the number of trees in the forest (n_{tree}) and the number of variables in the random subset at each node (m_{try}) should be carefully determined. For this, we constructed the random forests with sizes ranging from 1 to 500 trees in increments of ten trees. We observed that building trees beyond 100 did not result in significant additional performance but did increase the run time considerably. Given the trade-off between execution time and classification performance, the number of trees was set to 100.

Under these conditions, a random forest model was developed to classify the technologies according to the predicted economic value. In addition to the OOB predictions, we used a five-fold stratified sampling technique for a thorough performance evaluation, since the number of technologies for each category is imbalanced. Specifically, the random sampling was applied within each stratum with sampling after the technologies were divided into homogeneous strata according to their patent citation counts (i.e., L1, L2, and L3). A total of 80% of the technologies were used as a training dataset, while the remaining 20% of the technologies were employed as a test dataset.

The results are not reported here in their entirety owing to lack of space, but parts of the results are given in Table IV. The last column represents the predicted economic values when the corresponding technologies were employed in a test dataset. For instance, technology 03-274, which has three patents published in 2010 and has high values for many indicators such as originality, patent family, academic papers and number of citations, was classified as an L1 technology, whereas technology 03-278, that has one patent published in 2011 and has low values for the main indicators such as technology cycle time, originality, and patent family, was classified as an L3 technology.

4) Performance Evaluation: Several metrics using five-fold cross-validation techniques[5] were examined to assess the performance of our approach after a confusion matrix was constructed,

[5] *k*-fold cross-validation is a statistical technique for assessing how the results of an analysis will generalize to an independent dataset and how accurately a predictive model will perform in practice. This technique partitions the data into *k* nearly equally sized folds. Subsequently *k* iterations of training and validation are performed such that, in each iteration, a different fold of the data is held-out for validation while the remaining *k-1* folds are used for training the model. Upon completion, *k* samples of the performance metric are available and they are combined to derive a more accurate estimate of model performance.

TABLE V
RESULTS OF PERFORMANCE EVALUATION

(a) Confusion matrix for the proposed approach

Category	L1	L2	L3
L1	4	7	2
L2	32	75	49
L3	47	274	1721

(b) Performance evaluation metrics for the proposed approach

Category	Overall accuracy	Accuracy	Precision	Recall	Specificity	DOR
L1	0.81	0.96	0.31	0.05	1.00	11.92
L2		0.84	0.48	0.21	0.96	5.85
L3		0.83	0.84	0.97	0.27	12.40

(c) Confusion matrix for manual appraisal

Category	L1	L2	L3
L1	33	39	59
L2	23	134	380
L3	23	159	1038

(d) Performance evaluation metrics for manual appraisal

Category	Overall accuracy	Accuracy	Precision	Recall	Specificity	DOR
L1	0.64	0.92	0.25	0.42	0.95	12.53
L2		0.68	0.25	0.40	0.74	1.94
L3		0.67	0.85	0.70	0.56	2.98

(e) Performance evaluation metrics for two major classification models

Model	Category	Overall accuracy	Accuracy	Precision	Recall	Specificity	DOR
Neural network	L1	0.78	0.96	0.32	0.14	0.99	14.22
	L2		0.80	0.35	0.24	0.91	3.36
	L3		0.79	0.84	0.91	0.30	4.57
Support vector machine	L1	0.81	0.96	0.50	0.01	1.00	25.94
	L2		0.84	0.49	0.05	0.99	5.12
	L3		0.81	0.81	1.00	0.07	22.36

as shown in Table V(a). The overall accuracy of the proposed approach is 81% and is much higher than that of the naïve rule, 64% [the accuracy for each class of the naïve rule: 93% (L1), 72% (L2), and 65% (L3)], as reported in Table V(b). Although there are differences in the degree of accuracy across different classes, the proposed approach is found to be effective in assessing the economic value of university-originated technologies given the limited number of indicators that can be defined and extracted immediately after technologies are registered in the OTL. The precision and recall measures also show that the proposed approach is accurate and reliable in classifying the university-originated technologies. In particular, it is noteworthy that the proposed approach is effective in identifying the least valuable technologies that are very unlikely to be licensed. The DOR also shows that the proposed approach is effective in classifying the university-originated technologies, although there are differences in the degree of classification effectiveness across different classes. It also confirms that our method is the most effective in screening the least valuable technologies.

The performance of the proposed approach is compared to that of the expert appraisal approach that is used in the OTL of Stanford University, as summarized in Table V(c) and (d). Here, the number of technologies in the confusion matrix for expert appraisal is different from that of the proposed approach since some technologies do not have expert appraisal results in our dataset. According to the DOR, the expert appraisals tend to be different from the valuations from our method in that they identify the most valuable technologies well but they underperform relative to the proposed approach in terms of accuracy and reliability. Although the recall measure shows that the expert appraisals assess L1 and L2 technologies better than the proposed approach, it should be noted that the precision of the proposed approach is

TABLE VI
RESULTS OF ROBUSTNESS TESTS

(a) OTL-specific input variables

Indicator	Description
CntSponsors	Number of sponsors for technology development
CntRecipients	Number of marketing recipients
FieldOfUse	Number of potential application areas
Federal	1 if technology development is funded by federal governments, otherwise 0
Bioflag	1 if a technology is related to bio science, otherwise 0.

(b) Five categories of technologies according to the (expected) economic value

Category	(Expected) economic value	Number of technologies
L1	Above $500 000	83 (3.8%)
L2	$150 000 - $500 000	153 (6.9%)
L3	$50 000 - $150 000	203 (9.5%)
L4	0 - $50 000	627 (28.3)
L5	0	1145 (51.8)
Sum		2211 (100%)

(c) Performance evaluation metrics for Model 1

	Overall accuracy	Accuracy	Precision	Recall	Specificity	DOR
L1	0.82	0.96	0.11	0.01	1.00	3.23
L2		0.84	0.51	0.36	0.93	7.91
L3		0.85	0.87	0.96	0.41	15.38

(d) Performance evaluation metrics for Model 2

	Overall accuracy	Accuracy	Precision	Recall	Specificity	DOR
L1	0.59	0.96	0.29	0.10	0.99	11.24
L2		0.92	0.30	0.14	0.98	6.66
L3		0.92	0.36	0.06	0.98	3.76
L4		0.74	0.57	0.40	0.88	4.93
L5		0.66	0.62	0.89	0.43	5.76

higher than that of the expert appraisals, which means that our method provides more conservative forecasts.

Moreover, considering that many other machine learning models can be employed for this purpose, the performance of the proposed approach is compared to that of two major classification models (i.e., neural networks and support vector machines), as summarized in Table V(e). The neural network is an adaptive system that changes its structures based on the information that flows through the network [5]. The support vector machine has its roots in statistical learning theory and uses the concept of maximum-margin hyperplanes that have the largest distance to the nearest training data point of any class [50]. As shown in Table V(e), all the models outperformed the expert appraisals, although there are slight differences in the performance among the models. While the random forest shows better performance than the other classification models in our dataset, industrial practitioners need to select and customize appropriate methods according to the context of analysis and characteristics of data available.

Putting these things together, the results support our contention that predictive analytics can assess the value of university-originated technologies. Moreover, given the different behaviors of the two approaches (i.e., manual appraisal and predictive analytics approaches), we are confident that predictive analytics is a useful complementary tool for the valuation of university-originated technologies, creating synergies with manual appraisals.

V. DISCUSSION

A. Robustness Tests

Although the proposed approach is found to be accurate and reliable, especially for identifying the least valuable technologies, the performance of the proposed approach may vary according to the analysis context. For this reason, we conducted two additional analyses using different input and output variables to assess the robustness of the proposed approach. Specifically, model 1 employed five OTL-specific input variables, such as cntSponsors (the number of research sponsors) and federal (i.e., whether the invention is funded by the federal government), as summarized in Table VI(a). Model 2 used five categories to classify the economic value of the university-originated technologies, as reported in Table VI(b). The robustness tests using models 1 and 2 were conducted in the experimental setup that is identical to that of the experiment in Section IV-B. As summarized in Table VI(c) and (d), although there are slight differences in performance among the models, all the models show reliable and significant performance. Specifically, a slight performance improvement is found in Model 1: although the effectiveness in identifying L1 technologies drops slightly, the ability to discern L2 and L3 technologies increases. The performance metrics for Model 2 again confirm that our approach is very effective in identifying the least valuable technologies that are very unlikely to be licensed. Therefore, the proposed approach is considered robust across different analysis contexts and can be deployed in the different OTL contexts although it should be customized according to the analysis context.

B. Implementation and Customization of the Proposed Approach

A number of considerations should be made before applying a novel method. First, the valuation of the university-originated technologies relied in many cases solely on the expert appraisals in the relevant domain. However, such methods have become extremely time-consuming and labor-intensive as the number of technologies and the complexity of the inventions mount. As such, there is a huge increase in the intensive use of intelligent computerized approaches which apply expert feedback to the results derived [39]. In this respect, our method can be used as a real-time decision support system for the valuation of university-originated technologies once it is trained and developed. Moreover, the data and throughput are reusable, and new data can be added and analyzed easily. If the OTL has registered a new technology, it only remains to extract the relevant input indicators for the patents and publications. Second, we implemented *classification* models to assess the value of university-originated technologies. However, *regression* models can also be used for the valuation and can include performance metrics such as the mean absolute error and mean absolute percentage error (MAPE). Third, the proposed approach should be customized according to organizational contexts. In particular, different organizations have different strategies for data collection and management, and thus implementing a parsimonious model that accomplishes a desired level of explanation or prediction with as few predictor variables as possible is an important issue in practice. For this reason, we measured the importance of the input variables for the valuation of university-originated technologies to provide guidelines for feature selection, as reported in Online Appendix A. The following variables were found to be important in the valuation of university-originated technologies: (1) market coverage as measured by the number of patents registered in multiple countries with the coverage of the same invention, (2) development capabilities as measured by the core area technological strengths, (3) technological scope as measured by the number of patents, (4) novelty as measured by the number of nonpatent literature references of the patents, publication year, and the technology cycle time, and (5) technological field as measured by the primary class of the patents. Moreover, such OTL-specific variables such as FieldOfUse and CntRecipeints are also important in the valuation of university-originated technologies. In particular, FieldOfUse is recognized as the most important variable, and thus the communication between inventors and those in charge of technology licensing remains critical so as to identify the potential application areas of inventions. These variables should be given a high priority to implement and deploy the proposed approach in practice. Fourth, organizations should use as many different approaches as practical within the resource limitations. Although predictive analytics is effective for the valuation of university-originated technologies that integrates monetary value and patent value models, it cannot completely replace expert-centric approaches. Other quantitative approaches should be incorporated into the valuation of university-originated technologies. In this respect, sophisticated monetary value models based on discounted cash flows [64] and real options theory [1] and patent value models based on data envelopment analysis [51] can be helpful for this purpose. Multiple criteria decision making methods, such as analytic hierarchy processes and analytic network processes, can also improve the credibility and objectivity of the expert appraisal results [12]. Moreover, simulation techniques such as system dynamics and Monte Carlo simulation can be useful in modeling the complex interactions among the factors affecting the value of university-originated technologies and the uncertainties associated with the commercialization process [28]. Finally, our model should be periodically adjusted and updated. Systematic processes for doing so need to be established, although they may differ across organizational contexts (i.e., the speed and rate of technology registration in the OTL).

VI. Conclusion

This article proposed an approach to the valuation of university-originated technologies that integrated monetary value and patent value models for technology valuation via a random forest model. The premise of this article was that the technological characteristics that were captured by using quantitative indicators could provide valuable information on the economic value of university-originated technologies. To this end, we constructed a technological characteristics-value matrix that includes a total of 23 indicators and the value of technologies transacted in the market and employed a random forest model to classify new technologies according to their expected economic value. The specific case of the technologies registered in the OTL of Stanford University was used to verify the performance and potential utility of the proposed approach.

From an academic perspective, the primary contribution of this research was to develop a new approach to the valuation of university-originated technologies by integrating monetary value models and patent value models. The random forest model made this integration possible by developing a nonlinear model to link the technological characteristics and the economic value of the university-originated technologies. From a practical standpoint, the proposed approach was of assistance, especially for screening the least valuable technologies. This approach does not require any assumptions for parameter estimation, which are difficult to identify, especially for early stage and basic inventions. Moreover, our approach enabled the quick analysis of wide-ranging technologies and supports decision making at acceptable levels of time and cost. We expected that the proposed approach and software system we developed could be useful as a complementary tool for supporting those in charge of technology marketing and negotiation in a wide array of technological fields but who have little domain-specific knowledge.

Despite the confirmed validation, this article has limitations that should be complemented by future research. First, only technological factors were incorporated into valuation, although technological factors are more important than other factors in the early stages of commercialization of university-originated technologies. As nontechnological factors become more important in later phases of technology commercialization, the proposed approach could be further elaborated by including other types of indicators such as sponsor types, market size, marketing channels, and relationships between licensors (or inventors) and licensees. Second, given technology's multiple patents, the characteristics of the technology are captured by the average or majority of individual patent indicator values. Different indexes and further algorithms need to be devised according to the analysis context. Third, it is difficult to understand the role of the input indicators and the detailed relationships among the indicators (i.e., which input indicators influence the value of the university-originated technologies and how much) from the proposed approach. Methods such as Bayesian networks could be helpful for investigating the relationships among the indicators. Fourth, our approach is only applicable to the university-originated technologies that are registered in patent offices. Although whether a university-originated technology is protected by patents is critical to its valuation, the proposed approach needs to be extended to more generic models that can examine different types of university-originated technologies. More advanced analysis techniques, such as text mining and novelty detection methods, could be helpful for this purpose. Finally, our case study has been limited to the technologies registered in the OTL of Stanford University. Further testing on diverse databases would help establish the external validity of our method. Nevertheless, we argue that the analytical power our approach offers makes a substantial contribution, both to current research and to future practice.

References

[1] D. H. Baek, W. Sul, and K. P. Hong, H. Kim, "A technology valuation model to support technology transfer negotiations," *R&D Manage*, vol. 37, no. 2, pp. 123–138, 2007.

[2] D. Baglieri, F. Baldi, and C. L. Tucci, "University technology transfer office business models: One size does not fit all," *Technovation*, vol. 76, pp. 51–63, 2018.

[3] J. Bessen, "The value of U.S. patents by owner and patent characteristics," *Res. Policy*, vol. 37, no. 5, pp. 932–945, 2008.

[4] P. Bierly and A. Chakrabarti, "Determinants of technology cycle time in the U.S. pharmaceutical industry," *R&D Manage*, vol. 26, no. 2, pp. 115–126, 1996.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[6] F. P. Boer, *The Valuation of Technology*. New York, NY, USA: Wiley, 1999.

[7] F. P. Boer, "Valuation of technology using "real options," *Res. Technol. Manage.*, vol. 43, no. 4, pp. 26–30, 2000.

[8] M. J. Bray and J. N. Lee, "University revenues from technology transfer: Licensing fees vs. equity positions," *J. Bus. Venturing*, vol. 15, no. 5, pp. 385–392, 2000.

[9] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[10] W. Chapple, A. Lockett, D. Siegel, and M. Wright, "Assessing the relative performance of UK university technology transfer offices: Parametric and non-parametric evidence," *Res. Policy*, vol. 34, no. 3, pp. 369–384, 2005.

[11] Y. S. Chen and K. C. Chang, "The relationship between a firm's patent quality and its market value—The case of U.S. pharmaceutical industry," *Technol. Forecast. Soc.*, vol. 77, no. 1, pp. 20–33, 2010.

[12] Y. J. Chiu and Y. W. Chen, "Using AHP in patent valuation," *Math. Comput. Model.*, vol. 46, no. 7–8, pp. 1054–1062, 2007.

[13] D. R. Cutler *et al.*, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.

[14] H. Ernst, "Patent information for strategic technology management," *World Patent Inf.*, vol. 25, no. 3, pp. 233–242, 2003.

[15] T. Fischer and J. Leidinger, "Testing patent value indicators on directly observed patent value—An empirical analysis of Ocean Tomo patent auctions," *Res. Policy*, vol. 43, no. 3, pp. 519–529, 2014.

[16] R. Genuer, J. M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, 2010.

[17] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[18] A. Glas, J. G. Lijmer, M. H. Prins, G. J. Bonsel, and P. M. Bossuyt, "The diagnostic odds ratio: A single indicator of test performance," *J. Clin. Epidemiol.*, vol. 56, no. 11, pp. 1129–1135, 2003.

[19] C. Grimpe and H. Fier, "Informal university technology transfer: A comparison between the United States and Germany," *J. Technol. Transfer*, vol. 35, no. 6, pp. 637–650, 2010.

[20] D. Guellec and B. V. P. de la Potterie, "Applications, grants and the value of patent," *Econ. Lett.*, vol. 69, no. 1, pp. 109–114, 2000.

[21] R. Guns and R. Rousseau, "Recommending research collaborations using link prediction and random forest classifiers," *Scientometrics*, vol. 101, no. 2, pp. 1461–1473, 2014.

[22] D. Harhoff, F. M. Scherer, and K. Vopel, "Citations, family size, opposition and the value of patent rights," *Res. Policy*, vol. 32, no. 8, pp. 1343–1363, 2003.

[23] R. Haupt, M. Kloyer, and M. Lange, "Patent indicators for the technology life cycle development," *Res. Policy*, vol. 36, no. 3, pp. 387–398, 2007.

[24] D. W. Hsu, Y. C. Shen, B. J. Yuan, and C. J. Chou, "Toward successful commercialization of university technology: Performance drivers of university technology transfer in Taiwan," *Technol. Forecast. Soc.*, vol. 92, pp. 25–39, 2015.

[25] A. B. Jaffe and M. Trajtenberg, *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA, USA: MIT Press, 2002.

[26] H. J. Jang, H. Woo, and C. Lee, "Hawkes process-based technology impact analysis," *J. Informetr.*, vol. 11, no. 2, pp. 511–529, 2017.

[27] R. Jensen and M. C. Thursby, "Proofs and prototypes for Sale: The tale of university licensing," *Am. Econ. Rev.*, vol. 91, pp. 240–259, 2001.

[28] C. Jeon and J. Shin, "Long-term renewable energy technology valuation using system dynamics and Monte Carlo simulation: Photovoltaic technology case," *Energy*, vol. 66, pp. 447–457, 2014.

[29] V. K. Jolly, *Commercialising New Technologies: Getting from Mind to Market*. Boston, MA, USA: Harvard Business Press, 1997.

[30] J. Kim, S. Kim, and C. Lee, "Anticipating technological convergence: Link prediction using Wikipedia hyperlinks," *Technovation*, vol. 79, pp. 25–34, 2019.

[31] D. Kong, Z. Yuan, Y. Liu, and L. Xue, "Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country," *Technol. Forecast. Soc.*, vol. 119, pp. 80–97, 2017.

[32] R. Kotha, P. Crama, and P. H. Kim, "Experience and signaling value in technology licensing contract payment structures," *Acad. Manage. J.*, vol. 61, no. 4, pp. 1307–1342, 2018.

[33] J. H. Kwakkel and S. C. Cunningham, "Improving scenario discovery by bagging random boxes," *Technol. Forecast. Soc.*, vol. 111, pp. 124–134, 2016.

[34] D. A. Lane and R. R. Maxfield, "Ontological uncertainty and innovation," *J. Evol. Econ.*, vol. 15, no. 1, pp. 3–50, 2005.

[35] C. Lee, Y. Cho, H. Seol, and Y. Park, "A stochastic patent citation analysis approach to assessing future technological impacts," *Technol. Forecast. Soc.*, vol. 79, no. 1, pp. 16–29, 2012.
[36] C. Lee, B. Kang, and J. Shin, "Novelty-focused patent mapping for technology opportunity analysis," *Technol. Forecast. Soc.*, vol. 90, pp. 355–365, 2015.
[37] C. Lee, J. Kim, O. Kwon, and H. G. Woo, "Stochastic technology life cycle analysis using multiple patent indicators," *Technol. Forecast. Soc.*, vol. 106, pp. 53–64, 2016.
[38] C. Lee, O. Kwon, M. Kim, and D. Kwon, "Early identification of emerging technologies: A machine learning approach using multiple patent indicators," *Technol. Forecast. Soc.*, vol. 127, pp. 291–303, 2018.
[39] C. Lee, B. Song, and Y. Park, "How to assess patent infringement risks: A semantic patent claim analysis using dependency relationships," *Technol. Anal. Strategic Manage.*, vol. 25, no. 1, pp. 23–38, 2013.
[40] S. Lee, B. Yoon, C. Lee, and J. Park, "Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping," *Technol. Forecast. Soc.*, vol. 76, no. 6, pp. 769–786, 2009.
[41] J. Lerner, "The importance of patent scope: An empirical analysis," *RAND J. Econ.*, pp. 319–333, 1994.
[42] Z. Ma and Y. Lee, "Patent application and technological collaboration in inventive activities: 1980–2005," *Technovation*, vol. 28, no. 6, pp. 379–390, 2008.
[43] M. Meyer, "Are patenting scientists the better scholars?: An exploratory comparison of inventor-authors with their non-inventing peers in nanoscience and technology," *Res. Policy*, vol. 35, no. 10, pp. 1646–1662, 2006.
[44] S. S. Nair, M. Mathew, and D. Nag, "Dynamics between patent latent variables and patent price," *Technovation*, vol. 31, no. 12, pp. 648–654, 2011.
[45] Y. Park and G. Park, "A new method for technology valuation in monetary value: procedure and application," *Technovation*, vol. 24, no. 5, pp. 387–394, 2004.
[46] M. Perlitz, T. Peske, and R. Schrank, "Real options valuation: The new frontier in R&D project evaluation?," *R&D Manage.*, vol. 29, no. 3, pp. 255–270, 1999.
[47] R. F. Reilly and R. P. Schweihs, *Valuing Intangible Assets*. New York, NY, USA: McGraw-Hill, 1998.
[48] M. Reitzig, "Improving patent valuations for management purposes–Validating new indicators by analyzing application rationales," *Res. Policy*, vol. 33, no. 6–7, pp. 939–957, 2004.
[49] N. Rosenberg, "Uncertainty and technological change," In R., Landau, T., Talyer, G. Wright (Eds.), *The Mosaic of Economic Growth*. Stanford, CA, USA: Stanford Univ. Press, 1996.
[50] B. Schölkopf *et al.*, "Comparing support vector machines with Gaussian kernels to radial basis function classifers," *IEEE Trans. Signal Proces.*, vol. 45, no. 11, pp. 2758–2765, Nov. 1997.
[51] H. Seol, S. Lee, and C. Kim, "Identifying new business areas using patent information: A DEA and text mining approach," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 2933–2941, 2011.
[52] S. Shane, "Selling university technology: Patterns from MIT," *Manage. Sci.*, vol. 48, no. 1, pp. 122–137, 2002.
[53] J. Shin, B. Y. Coh, and C. Lee, "Robust future-oriented technology portfolios: Black–Litterman approach," *R&D Manage.*, vol. 43, no. 5, pp. 409–419, 2013.
[54] G. V. Smith and R. L. Parr, *Valuation of Intellectual Property and Intangible Assets*. New York, NY, USA: Wiley, 2000.
[55] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Comm. Com. Inf. Sc.*, vol. 45, no. 4, pp. 427–437, 2009.
[56] K. Song, K. Kim, and S. Lee, "Identifying promising technologies using patents: A retrospective feature analysis and a prospective needs analysis on outlier patents," *Technol. Forecast. Soc.*, vol. 128, pp. 118–132, 2018.
[57] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *J. Chem. Inf. Comp. Sci.*, vol. 43, no. 6, pp. 1947–1958, 2003.
[58] P. N. Tan, *Introduction to Data Mining*. Chennai, India: Pearson Educ. India, 2006.
[59] J. G. Thursby, R. Jensen, and M. C. Thursby, "Objectives, characteristics and outcomes of university licensing: A survey of major U.S. universities," *J. Technol. Transf.*, vol. 26, no. 1–2, pp. 59–72, 2001.
[60] J. Tipping, E. Zeffren, and A. Fusfeld, "Assessing the value of your technology," *Res. Technol. Manage.*, vol. 38, no. 5, pp. 22–39, 1995.
[61] M. Trajtenberg, *Economic Analysis of Product Innovation: The case of CT scanners*. Cambridge, MA, USA: Harvard Univ. Press, 1990.
[62] A. J. Trappey, C. V. Trappey, C.-Y. Wu, and C.-W. Lin, "A patent quality analysis for innovative technology and product development," *Adv. Eng. Informat.*, vol. 26, no. 1, pp. 26–34, 2012.
[63] B. Van Looy, J. Callaert, and K. Debackere, "Publication and patent behavior of academic researchers: Conflicting, reinforcing or merely co-existing?," *Res. Policy*, vol. 35, no. 4, pp. 596–608, 2006.
[64] S. Van Triest and W. Vis, "Valuing patents on cost-reducing technology: A case study," *Int. J. Prod. Econ.*, vol. 105, no. 1, pp. 282–292, 2007.
[65] Y. Wu, E. W. Welch, and W. L. Huang, "Commercialization of university inventions: Individual and institutional factors affecting licensing of university patents," *Technovation*, vol. 36, pp. 12–25, 2015.
[66] M. Zettl, "Valuing exploration and production projects by means of option pricing theory," *Int. J. Prod. Econ.*, vol. 78, no. 1, pp. 109–116, 2002.