Predicting Patent Quality Based on Machine Learning Approach

Zulfiye Erdogan, Serkan Altuntas, and Turkay Dereli

***Abstract*—The investment budget allocated by companies in R&D activities has increased due to increased competition in the market. Applications for industrial property rights by countries, investors, companies, and universities to protect inventions obtained as an outcome of investments have also increased. The selection of the patent to be invested becomes more difficult with the increasing number of applications. Therefore, predicting patent quality is quite significant for companies to be successful in the future. The level to which a patent meets the expectations of decision makers is referred to as patent quality. Patent indices represent decision makers' expectations. In this study, an approach is proposed to predict patent quality in practice. The proposed approach uses supervised learning algorithms and analytic hierarchy process (AHP) method. The proposed approach is applied to patents related to personal digital assistant technologies. The performances of individual and ensemble machine learning methods have been also analyzed to establish the prediction model. In addition, 75% split ratio and the five-fold cross-validation methods have been used to verify the prediction model. The multilayer perceptron algorithm has 76% accuracy value. The proposed prediction model is essential in directing R&D studies to the right technology areas and transferring the incentives to patent applications with a high quality rate.**

***Index Terms*—Analytic hierarchy process (AHP), machine learning, multilayer perceptron, patent indices, supervised learning algorithms.**

I. Introduction

Meeting customer needs quickly and effectively is very critical for companies and investors to maintain their continuity in the marketplace. Firms allocate a large amount of budget to R&D studies to gain an advantage over their competitors. Patents protect the inventions obtained from R&D studies. The number of patent applications is increased through the increase in technological developments. National and international patent applications are time-consuming and costly processes. In addition, patent documents may not consistently achieve economic benefit. Many inventions have been patented but cannot be commercialized. Therefore, determining promising technologies and directing R&D activities to these technologies has a critical role to obtain high-quality patents. Measuring the quality of patents using their subtechnology fields provides advantages in terms of both time and cost. Predicting the patent quality is important in terms of directing R&D studies to the profit technology fields and directing incentives for patent applications to patents in technology fields with high quality rates. Various studies have been conducted in the literature to evaluate the quality of patents and technologies [1], [2]. Various indices, such as number of claims [3], [4], number of forward citations [5], [6], and patent family size [2], [3], have been used in the literature. However, these indices are calculated using data that can be obtained after the patent registration certificate is obtained. In this study, a supervised learning algorithm is proposed to predict patent quality. The importance of this prediction model is that data to be used for training the supervised learning algorithm consists of the information that can be obtained before the patent is obtained.

In this study, patent quality indicates the level of meeting the expectations of decision makers from a patent. The index values used in the literature are taken into account to measure these expectations while predicting the patent quality. The indices used to analyze the patent quality were selected based on the literature. Nine patent indices including the number of claims, the number of annual forward citations, the number of back citations [3], [5], pendency [7], the remaining lifetime of the patent [2], [5], the interaction power, the technology cycle time (TCT) [8], technology score [9], and patent family size [6] are discussed to determine the patent quality. These indices have different levels of effect on patent quality. These indices show the quality level of patents in various fields, such as their reliability, commercialization potential, and ability to affect the market. The effect levels of these areas are determined to calculate the overall quality of the patent. Therefore, the weights of these indices are calculated based on analytic hierarchy process (AHP) to determine these effect levels in this study. However, there has not been a comprehensive study relating to the weighting of these patent indices in the literature. To fill this gap, the opinions of academicians working in the field of technology management and technology transfer office (TTO) experts responsible for the patent registration process are taken into account to determine the weights of patent indices in this study. Information about the participants is given in Table XI. Participants include mechanical, electrical and electronic, computer, industrial, mechatronic engineers, and economists. Individuals with at least 5 years of

Manuscript received 27 November 2021; revised 9 April 2022; accepted 8 September 2022. Date of publication 27 September 2022; date of current version 5 January 2024. Review of this manuscript was arranged by Department Editor F. Tietze. *(Corresponding author: Zulfiye Erdogan.)*

Zulfiye Erdogan is with the Department of Industrial Engineering, İskenderun Technical University, 31200 Hatay, Turkey (e-mail: zulfiye.erdogan@iste.edu.tr).

Serkan Altuntas is with the Department of Industrial Engineering, Yildiz Technical University, 34349 Istanbul, Turkey (e-mail: serkan@yildiz.edu.tr).

Turkay Dereli is with the Office of President, Hasan Kalyoncu University, 27000 Gaziantep, Turkey (e-mail: turkay.dereli@hku.edu.tr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TEM.2022.3207376.

Digital Object Identifier 10.1109/TEM.2022.3207376

TABLE I
PATENT INDICES IN THE LITERATURE

Authors	Claim	Forward Citation	Backward Citation	TCT	Patent Family Size	Remaining lifetime of the patent	Patent Application Date	Patent Acceptance date	Novelty	Patent Classes
Abraham and Moitra [1]	✓						✓	✓		✓
Harhoff et al. [15]		✓	✓		✓		✓		✓	✓
Haupt et al. [7]	✓	✓	✓				✓	✓		
Lai and Che [3]	✓	✓	✓		✓				✓	✓
Shih et al. [5]		✓	✓	✓		✓				✓
Tseng et al. [2]	✓	✓	✓		✓	✓				✓
Gao et al. [16]	✓		✓				✓			✓
Choi et al. [17]		✓			✓					
Grimaldi et al. [18]	✓	✓								✓
Ni et al. [19]	✓	✓	✓		✓	✓				
Verhoeven et al. [20]		✓	✓							✓
Chang and Fan [21]	✓	✓	✓							✓
Kim and Bae [6]	✓	✓			✓					✓
Kyebambe et al. [8]	✓	✓		✓					✓	✓
Song et al. [22]	✓	✓	✓	✓					✓	✓
Kwon and Geum [23]		✓	✓	✓						✓
Hur and Oh [24]	✓	✓	✓							✓

experience are included in the analysis Trademark attorneys are not included in the study. More than half of the surveyed experts have also knowledge on project management and entrepreneurship. After receiving expert opinions, the weights of patent indices are calculated using the AHP method. Patent indices are normalized using the min—max method to calculate patent scores. Patent scores are calculated using normalized patent indices and their weights. Patent score is determined by the weighted sum method. After determining the weights and scores of the patent indices, quality values of the patents are calculated. Patents are labeled at four levels in terms of their quality level using clustering algorithm. In the literature, various factors are analyzed while selecting and evaluating the performance of clustering algorithms. These factors are the data' size, dimension, and noise level, etc. [10]. Algorithms, such as BIRCH [11] and CLARANS [12], can be preferred in the analysis of large data sets. Algorithms, such as k-medoids [13] and CURE [14], are effective in handling noisy data. In this study, the self-organizing maps (SOM) algorithm is preferred, since the dataset used in the study is not large. The SOM algorithm does not request parameter values from the user. The algorithm calculates the optimum parameter values itself. Due to this advantage, the SOM algorithm is used for analysis. Patent quality values are clustered using the SOM algorithm. These clusters are patent score clusters. Patent score clusters are the output values of the prediction model. Patent score clusters and corporate patent classification (CPC) codes are data to be used to create a prediction model. These patent score clusters refer to patent quality levels. Patents with high-quality levels should be preferred for R&D investments. Various supervised learning algorithms have been tested to determine the method to be used to predict patent quality. Here, the multilayer perceptron algorithm with the best performance results is selected as the prediction method. The patent quality value calculated using patent indices is based on CPC codes. CPC codes represent the technology field of invention. Therefore, the technology field of the invention has an impact on patent quality. This relationship is used to predict patent quality.

The rest of this study is organized as follows. In Section II, the literature on patent indices, prediction, and assessment of technologies are given. In Section III, the proposed approach is explained. In Section IV, application of the proposed approach is given. Finally, Section V concludes this article.

II. LITERATURE REVIEW

A. Patent Indices

Patents are the most frequently used documents to assess technologies. Various patent indices have been used to evaluate patents and technologies. The effects of patents on technologies have been analyzed using these indices. Studies using patent indices in the literature are given in Table I.

When the literature is reviewed, there is no a study that deals with all of the indices given in Table I. In addition, there is no a comprehensive study relating to the weighting of these patent indices. In this study, nine indices are considered and these indices are weighted. Most of the indices in Table I cannot be obtained before the patent application. Therefore, the use of these indices may not be appropriate to predict quality of patent. In this study, these indices are used as label values, not attribute values.

The CPC codes are used to identify technical fields which patents are applicable. It is critical to determine whether these technology fields are emerging technology field. The power of CPC codes to influence the market is also important. The combination of some CPC codes or the number of CPC codes referring the patent determines the value of the patent. As the number of technology codes increases, the number of different

TABLE II
STUDIES RELATED TO TECHNOLOGY ANALYSIS

Authors	Method	Aim
Wang and Hsu [34]	Gray theory and genetic algorithm	Technology Prediction
Karsak and Ahiska [35]	Data envelopment analysis	
Durmusoglu and Dereli [36]	Data envelopment analysis	
Lim and Anderson [37]	Data envelopment analysis and regression model	
Yoon et al. [38]	Data envelopment analysis	
Altuntas et al. [39]	Data fusion	
Obrecht and Denac [40]	Delphi method	
Dahooie et al. [41]	SWARA and fuzzy MULTIMOORA methods	
Kim et al. [42]	Random forest	
Altuntas and Aba [43]	S curves	
Golkar et al. [44]	Pareto frontier forecasting and game-theoretic planning	
Yuan and Cai [45]	S-curve simulation for growth curves and entropy weight	
Sartori et al. [46]	The semi-structured interview and non-participant observation, qualitative and descriptive approach	
Adamuthe and Thampi [47]	Technology life cycle and S-curve, trend projection methods and growth curve methods	
San Cristóbal [48]	AHP-VIKOR	Technology Assessment and Selection
Ren and Lützen [49]	AHP-VIKOR	
Chu et al. [50]	Common weighted data envelopment analysis	
Pakdil et al. [51]	Kemeny median indicator row fit	
Mohammed [52]	Fuzzy AHP-TOPSIS	
Mondragon et al. [53]	AHP and fuzzy AHP methods	
Zhou et al. [54]	Sentiment analysis, SAO analysis, machine learning, and expert judgments	
Han et al. [55]	SAO Networks and Link Prediction	
Torkayesh et al. [56]	Stratified multi-criteria decision making	
Mall and Anbanandam [57]	Fuzzy AHP, VIKOR	
Park and Jun [58]	Text mining, Bayesian additive regression trees	
Ranaei et al. [59]	Term counting technique, the emergence score, and LDA	
Woo et al. [60]	Bayesian structural equation model	Patent Analyze
Han et al. [61]	Text mining, LDA topic model	
Trappey et al. [62]	Deep learning and PCA	
Jeon et al. [63]	A doc2vec and local outlier factor approach	
Sinigaglia *et al.* [64]	Logistic growth model, patent indicators	
Sharma and Sharma [65]	Citation analysis	
Song et al. [66]	Natural language processing and machine learning, K-means, AdaBoost algorithm	
Chao et al. [67]	Document term frequency analysis, clustering method, and LDA	
You et al. [68]	Word2vec, Bert, LSTM, CRF, and TextCNN	
Kim and Geum [69]	Logistic regression, naive Bayes classifier, support vector machine, random forest, and deep neural network	
Ma et al. [70]	Link prediction	

technologies originating the patent also increases. The ratio of cumulative technology codes to number of patents shows how the economy's creative efforts are being driven [25]. The prevalence of technology class code combinations of a patent shows whether the patent is different or radical [26]. Active or passive CPC and IPC codes also determine the value of the patent. Therefore, these codes are critical, while the patent quality or success is determined [27]. Choi et al. [28] proposed an approach to obtain emerging technology combinations with high success ratio. This approach analyzes the co-occurrence of CPCs in all patents to create a universal CPC network containing technological combination registries with high potential for success. Altuntas and Gok [29] proposed an approach considering frequency of each technology class and different importance levels among technologies based on patent citations. The proposed approach uses the importance of CPC codes and the number of repetitions of each CPC code. The significance value of CPC codes is calculated by multiplying the average number of citations per code by the number of CPC codes in the patent data. Oh et al. [30] proposed a systematic approach based on link prediction to forecast possible development aspects of a product. In this approach, a target product is represented as a set of CPCs included in product-related patents. Shim [31] analyzed relationships among CPC codes to forecast future patent trends. Jeon et al. [32] proposed a model to analyze applicants' technical inventions based on CPC codes. The model demonstrates that applicants have different invention patterns and tendencies to invent technology. Chae and Gim [33] showed that patent classification systems can be used to infer trends in applicants' technological inventions and monitor changes in their innovative models.

In this study, the various patent indices are used to analyze the effect of CPC codes on the prediction of the quality of patents. The reliability of using CPC codes to predict the score obtained by using patent indices has been tested. The activity level or difference of technologies or patents within a technology field significantly affects the importance of technologies. The effect levels of patent index values on the evaluation of technologies or patents are determined in this study. CPC codes and patent indices, which have been utilized in the literature to assess technologies, are approached from a different angle in order to present a model and dataset. The lack of data is a significant issue in the field of technology management. A different perspective is offered to researchers in order to fill this deficiency. The suggested model and dataset are open to improvement for future research.

TABLE III
STUDIES USING DATA MINING METHODS FOR TECHNOLOGY PREDICTION

Authors	Method
Jun et al. [71]	Matrix map and k medoid- support vector clustering
Lee and Lee [72]	Portfolio analysis, association mining, and collaborative filtering
Madvar et al. [73]	Text mining
Jeon and Suh [74]	Association rule mining, brokerage network, and centrality analysis
Yu et al. [75]	K-medoid clustering method and time series
Ampornphan and Tongngam [76]	K-means clustering, text mining, and association rule mining
Feng et al. [77]	Latent Dirichlet allocation and association rule mining
Naik et al. [78]	Decision tree algorithm and Naive Bayes algorithm
Tang et al. [79]	Association rule mining and information entropy
Lee et al. [80]	Logistic regression
Sasaki and Sakata [81]	Random forest algorithm
Daim et al. [82]	Text mining, expert judgment
Li et al. [83]	Text mining, expert judgment
Ozcan et al. [84]	Text mining and an expert review
Gui et al. [85]	Deep learning
Liu et al. [86]	Topic Analysis and social Network Analysis
Xu et al. [87]	Topic analysis
San Kim and Sohn [88]	SVM, NN, decision tree, random forest, deep semantic analysis
Zhou et al. [89]	Gartner's hype cycle, generative adversarial network, deep learning
Kim et al. [90]	Text mining, patent analysis

B. Prediction and Assessment of Technologies

In the literature, various approaches have been proposed to predict, assess, and selection of technologies. Studies related to technology analysis are given in Table II.

Various methods, such as multicriteria decision making, data mining, and qualitative methods are used to evaluate technologies. The studies using these methods are given in Table II. The main problem of studies on the evaluation of technologies is the lack of data. Patent data has significantly overcome this problem. Patents are the most frequently used data sources to predict technologies. Data mining and machine learning techniques are generally used for patent analysis. Studies using data mining methods for technology prediction are given in Table III.

Studies provided in Table III use patent data to predict technologies. When the literature is reviewed, it is found that patent analysis is the most frequently used method for identifying and predicting promising technologies. Data mining functions such as classification, clustering, and association rule mining are used to analyze data quickly and accurately. The classification algorithms are used methods for generating prediction models in different study fields. However, the data used in the technology management field are generally analyzed using unsupervised learning algorithms, such as clustering and association rule mining due to the lack of label values. In the literature, studies using classification algorithms usually use patent index values as attribute values. However, these indices are calculated using data that can be obtained after the patent registration certificate is obtained. In the literature, technology clusters are created by using clustering algorithms for the evaluation or prediction of technologies. However, all of the patents in the same technology cluster have not the same quality rate. The patents with the different quality level may available in the same cluster. Although the technology cluster obtained using various clustering algorithms has a low quality rate, there may be patents with high quality rates in this technology cluster. A different perspective is offered to researchers in order to fill this deficiency. The data to be used for training of the proposed model is consists of the information that can be obtained before the patent is obtained. A dataset suitable for the use of supervised learning algorithms is created in this study.

III. PROPOSED APPROACH

The proposed approach is introduced in this section. The flowchart of the proposed approach is given in Fig. 1. The data used to create this model consists of two parts, namely attribute and label values. Patent data are obtained from USPTO database. In Step 1, a patent-class matrix is created using CPC codes of patents. Patents that remaining lifetime is higher than 0 are included in the patent-class matrix.

Patent-class matrix forms the attribute values of the data that the machine learning algorithm will use to create predict model. Nine patent indices are taken into account to evaluate patents in the patent class matrix. In Step 2, the weights of these patent indices are determined using the AHP method. In Step 3, the scores of patent indices are calculated for each patent. Then, the patent scores are determined using weights and scores of patent indices. In Step 4, the patent scores are clustered using the SOM algorithm. Each patent is assigned to a score cluster. These score clusters form the label values of the data. In Step 5, label values representing the quality of patents are classified using individual and ensemble machine learning algorithms. In Step 6, the predict performances of these algorithms are tested using Kappa statistics, mean absolute error, accuracy values. In Step 7, the MLP algorithm with the highest performance values is preferred to create the prediction model. In Step 8, the quality of patents is predicted using the MLP algorithm.

A. SOM Algorithm

SOM algorithm is an unsupervised algorithm proposed by Kohonen [91]. The method aims to maximize intragroup similarity and intergroup difference [92]. The weight values of the neurons

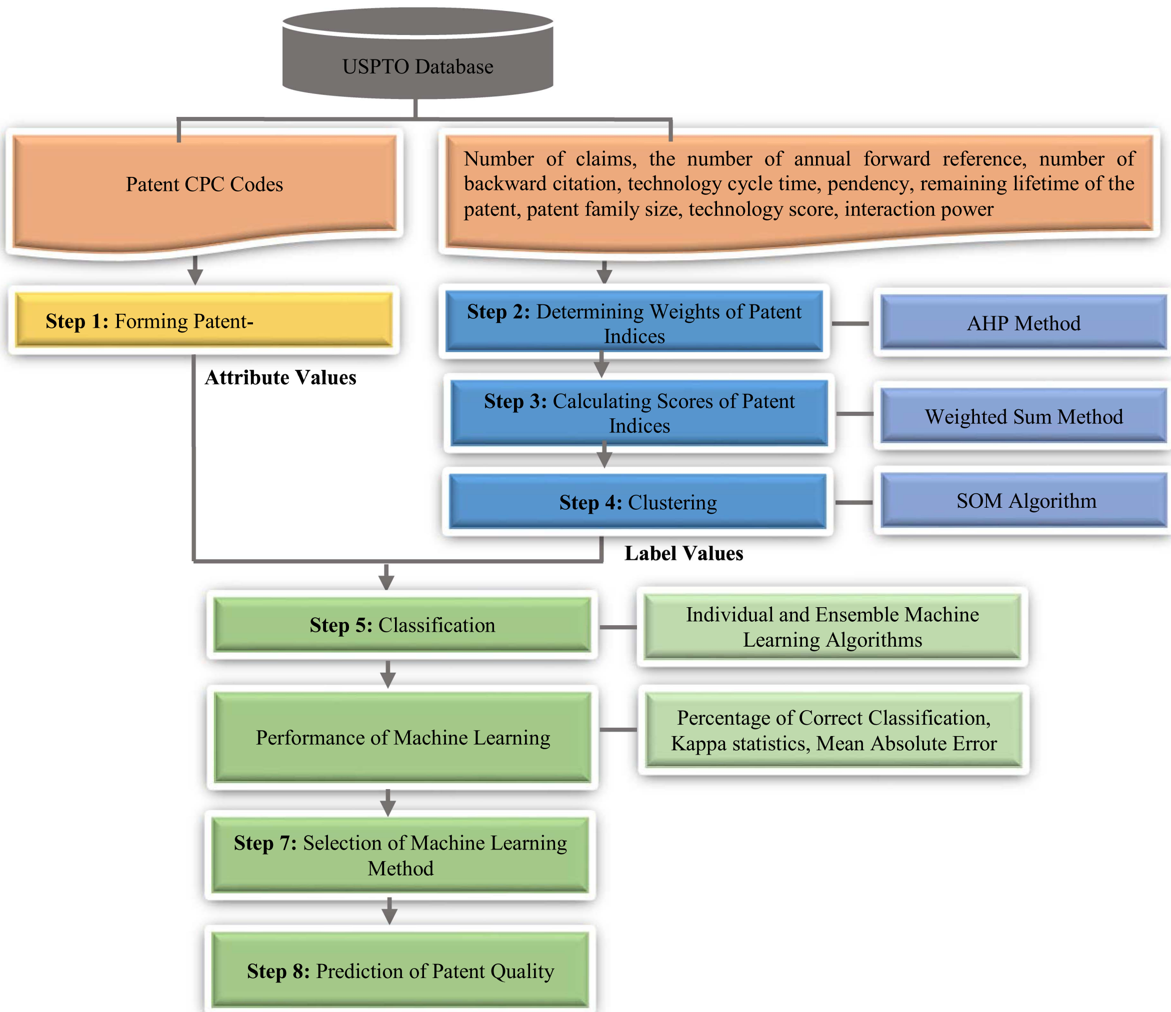

Fig. 1. Proposed approach.

in the network are assigned randomly by the algorithm. The input values are taken. Subsequently, all the values on the map are navigated. The distances between the values are calculated and the node with the shortest distance among these values is taken. This process continues until all values have been assigned to a cluster. The SOM algorithm is applied to create technology clusters. These clusters are used to calculate TCT and interaction power in Step 3. In addition, the SOM algorithm is used to obtain patent score clusters in Step 4. The purpose of clustering patent scores is to determine the correct limit values for converting patent score values to categorical data.

B. Patent Indices

In the literature, various indices are used to calculate the quality of patents. In this study, nine patent indices are considered: number of claims, number of annual forward citations, number of back citations, TCT, pendency, remaining lifetime of the patent, patent family size, technology score, and interaction power.

The number of claims: Claims are the part in which the technical features of the invention requested to be protected are specified. The high number of claims positively affects the patent score [1].

The number of annual forward citations: The number of forward citations is the number of cited to patent document by other patents or academic studies [6]. Patent citation information is useful resources for measuring the economic value of technologies [93], [94]. The number of forward citations of a patent is an important representative for the value of the patent and related technology [95]. The number of forward citations of the patent is a critical indicator to predict and evaluate value of technologies [96]. In this study, the number of annual forward citations is obtained by dividing the number of forward citations by the age of the patent. The number of annual forward citations of a patent is calculated using the following equation:

$$\text{The number of annual forward citations}_i = \frac{\text{Total number of forward citations}_i}{\text{age of patent}_i}. \quad (1)$$

The number of backward citations: Back citation is the reference used by the inventor or the person who wrote the description when writing the patent [21]. The higher backward number of patents means that this technology field is based on a more robust foundation [3]. Some studies in the literature argue that the number of citations to nonpatent literature is closely related to patent originality and scientific indices [8]. The number of references to the patent literature has a significant positive coefficient for all technical fields [15]. The backward citations and forward citations contribute positively to R&D activities [97]. Therefore, the high number of back citations positively affects the patent score.

Pendency: Pendency is the period from the patent application to the patent acceptance [3]. More controversial claims lead to slower grants and faster withdrawals, but well-documented applications are approved faster and withdrawn more slowly [98]. Highly cited applications are approved by the EPO more quickly than less important ones, but are less quickly withdrawn by the applicant [99]. The pendency value of the *i*th patent is calculated using the following equation. The low pendency value positively affects the patent score

$$\text{Pendency}_i = \text{Patent acceptance date}_i - \text{Patent application date}_i. \quad (2)$$

The remaining lifetime of the patent: Patent age and year of application have negatively related to the patent value [100]. The older the patent, the more likely the underlying technology is outdated. Patent age lowers patent's salability. Therefore, the age of the patent also determines the commercial value of the patent. Also, the age of the patent has a negative effect with the auction price [101]. The protection period of a patent is generally 20 years. The remaining lifetime of a patent is determined by subtracting the age of the patent from the total protection period [19]. The remaining lifetime of an *i*th patent is calculated using the following equation. The high remaining lifetime value positively affects the patent score

$$\text{Remaining lifetime}_i = 20 - \text{age of patent}_i. \quad (3)$$

Patent family size: Patent family size is the number of countries that a registration certificate is obtained [3]. The high patent family size value positively affects the patent score.

Technology score: The technology score, or T-Score, grades the CPC codes. Statistics are used to determine whether these codes are strong or weak. The technology score is calculated by considering the growth rate, maintenance rate, transaction rate, and allowance rate of a patent's CPC codes. This score is obtained from the AcclaimIP software. This value has a positive effect on the patent score. Growth of class is the number of patents received in that class. The number of patents refers the rate of technological progress in that field [27]. Maintenance rate of class is the ratio of the number of maintained patents the total number of patents in that class. Technological impact of patents has a positive effect on maintenance decisions [102]. Maintenance rates of patents are used to measure the commercial quality of patents [103], [104]. Transaction rate of class is the ratio of the number of transacted patents to the total number of patents in that class. Patent transaction can be used to assess patent quality. Patents licensed or sold are considered high-quality patents [101], [105]. Allowance rate of class is the ratio of the number of patents with an allowance to the total number of patents in that class. USPTO issues a notice of allowance when it thinks an invention qualifies for a patent. Patent allowance has a positive effect on the value bargaining of the invention [106], [107].

TCT: In this study, TCT is calculated based on subtechnology clusters. The subtechnology clusters are obtained using the SOM algorithm. Patents involving emerging technologies are expected to have relatively smaller TCT values than other patents in a rapidly growing technology cluster [5], [8]. A lower TCT indicates that companies can acquire a competitive edge by innovating more quickly [2]. In this study, technology clusters are created by clustering patents with similar technology fields. Then, cycle time of each technology cluster is calculated. Since the TCT is preferred to be low, it is included in the calculation with a negative effect on the patent score. TCT is calculated using the following equation:

$$\text{TCT}_k = \text{median}_k \ \{|T_y - T_x|\} \quad (4)$$

where TCT_k is the TCT of the *k*th technology cluster. T_x is application date of the *x*th patent. Patent *x* and Patent *y* are members of the same technology cluster.

Interaction power: Interaction power refers to the effect level of CPC codes belonging to the current technology field. The most cited patents or technology fields that can cooperate with the most technology in technology opportunity analysis studies are considered promising technologies in the literature. For the calculation of the interaction power of a patent, the interaction power of the CPC codes of that patent is calculated first. Subtechnology clusters of the selected technology are determined to calculate the interaction power of CPC codes using SOM algorithm. Thus, the number of subtechnology clusters in which a CPC code operates are determined. The interaction power of a patent increases as the relation level of CPC codes to which it is associated with subtechnology clusters increases. The minimum and maximum importance level of a CPC code are 1 and the total number of clusters, respectively. After determining the importance levels of CPC codes, their interaction powers are calculated. The interaction power of a CPC code is denoted by $\text{Interaction power}_l$

$\text{Interaction power}_l$ is calculated using the following equation:

$$\text{Interaction power}_l = a_l + \sum_{m=1}^{n} a_m \quad (5)$$

where a_l is importance level of the lth CPC code. a_m is importance level of the *m*th CPC code that has an association with the *l*th CPC code and $M = \{m = 1, \ldots n\}$.

The determination of codes together in patents has been taken into account when determining the codes to which a CPC code is related.

TABLE IV
SCALE OF BINARY COMPARISON

Degree of Importance	9	7	5	3	1	8, 6, 4, 2
Definition	Absolute importance	Demonstrated importance	Strong importance	Week importance	Equal importance	Values between 9-7, 7-5, 5-3, 3-1

The interaction power of *i*th patent is calculated using the following equation:

$$\text{Interaction power}_i = \sum_{i=1}^{n} \text{Interaction power}_l \ . \tag{6}$$

Where Interaction power$_i$ is interaction power of the *i*th patent and interaction power$_l$ is interaction power of the *l*th CPC code.

C. AHP Method

AHP is a decision-making method developed by Saaty [108]. Decision-makers make binary comparisons between criteria. Thus, the importance values of the criteria relative to each other are determined. In Step 2, the AHP method is used to calculated weights of the patent indices used in the calculation of patent scores. Decision-makers make binary comparisons of criteria using the scale proposed by Saaty. This scale is given in Table IV.

In AHP applications, the group comparison matrix is obtained by taking the geometric mean of the pair-wise comparison matrices. Weights of patent indices are calculated using the group comparison matrix. The AHP method is a method that calculates consistency. The consistency of the answers given by the experts is also tested using the consistency rate (CR) calculated in this method. Consistency Index (CI) and CR values are calculated to test the usability of the calculated weights. If the CR value is lower than 0.1, the criterion weights are used. The label values are calculated after testing the consistency of criterion weights.

D. Patent Score Clusters

Patent score clusters constitute the label values of the data. Patent score clusters are created by clustering patent scores using the SOM algorithm. Patent scores are calculated using scores of patent indices and weights of these indices. The scores of patent indices are normalized using the min–max normalization method. Patent scores are calculated by taking the weighted arithmetic sum of patent indices. Patent scores are calculated using the following equation:

$$\text{Patent Score}_i = \sum_{j=1}^{9} X_j \times A_j \tag{7}$$

Where *j* is the patent index, X_j is the score of the patent index, A_j is the weight of the patent index. The scores of the pendency, remaining lifetime, and TCT indices, which have a negative effect on the patent score, are included in the calculation by converting. After the patent scores are calculated in Step 3, patents are assigned to the patent score clusters using the SOM algorithm in Step 4. Label values of data are created.

E. Selection of the Prediction Method

The quality of patents is classified using the individual and ensemble machine learning algorithms in Step 5. In Step 6, the performances of these algorithms are tested to select predict method. Cross-validation and split ratio methods can be used to verify the prediction models. In the cross-validation method, $n-1$ of the n observations is used as the training dataset to create the prediction model. The remaining observations are used as test data. Iteration is continued until all of the observations are predicted [109]. In the split ratio method, the data are divided into two datasets as training and test. Training dataset is used to establish predict model. Test dataset is used to test the performance of this model. In this study, five-fold cross-validation and 76% split ratio methods are applied. In Step 6, Kappa statistics, accuracy, and mean absolute error values are used to evaluate performances of established predict models. The mean absolute error is the absolute value of the difference between the predicted value and the actual value. Accuracy is calculated by dividing the number of correctly predicted samples by the total number of samples. Kappa statistics are used to measure the consistency between predicted and observed categorizations of a dataset [110]. The algorithm with the lowest mean absolute error value and the highest accuracy and Kappa statistics values has the best prediction performance. The results of the first four algorithms with the best prediction performance are given in this study. It should be noted that the MLP algorithm has the best prediction performance among these algorithms.

F. MLP Algorithm

MLP algorithm is a variant of the basic perceptron model proposed by Rosenblatt [111]. Neurons are arranged as layers. The model has an input layer and an output layer, and one or more hidden layers [112]. The output signal value of a neuron is calculated using the following equation:

$$y_i = f\left(\sum_{j=1}^{J} w_{ij} x_i\right) \tag{8}$$

Where w_{ij} is connecting weight of neurons, x_i is the input signal of neuron, and *f* is the activation function. There are various activation functions, such as a simple lower bound, hyperbolic tangent, and sigmoid function. In this study, the sigmoid function is used. MLP algorithm is the most widely used feed-forward neural networks due to its ease of application, speed of operation, and the need for a smaller training set [113]. In Step 7, MLP algorithm has the best prediction performance among these algorithms is selected to create the prediction model. MLP algorithm is used for the prediction of patent quality in Step 8.

TABLE V
GROUP COMPARISON MATRIX

	Number of claims	Number of annual forward citations	Number of back citations	TCT	Pendency	Remaining lifetime	Patent family size	Technology score	Interaction power
Number of claims	1	0.63	0.98	0.36	0.67	0.27	0.30	0.31	0.30
Number of annual forward citations	1.5939	1	0.63	0.34	0.74	0.30	0.33	0.33	0.27
Number of back citations	1.0246	1.5892	1	0.25	0.44	0.20	0.22	0.26	0.24
TCT time	2.7831	2.9665	4.0636	1	0.70	0.35	0.34	0.44	0.30
Pendency	1.4927	1.3426	2.2716	1.4363	1	0.22	0.25	0.27	0.27
Remaining lifetime	3.7257	3.3242	4.8938	2.8957	4.6061	1	0.35	0.47	0.36
Patent family size	3.34022	3.0068	4.5369	2.9069	3.9463	2.8329	1	0.43	0.34
Technology score	3.1898	3.0027	3.9157	2.2509	3.6985	2.1309	2.3050	1	0.38
Interaction power	3.3563	3.7046	4.0861	3.3854	3.6693	2.7806	2.9151	2.6628	1

The precision, recall, and F-measurement values of the predicted patent score clusters are calculated using the following equation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

where TP is the number of correctly predicted positive samples, and FP is the number of incorrectly predicted positive samples

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

where FN is the number of incorrectly predicted negative samples

$$F - \text{measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{11}$$

F-measurement value is the harmonic mean of precision and recall value. If these values are high, the predict performance is good. The performance values of the algorithm are given based on technology score clusters in this study.

IV. Application of the Proposed Approach

The proposed approach is applied to patents on personal digital assistant (PDA) technologies in this section. The patent documents are gathered using keywords (PDA) [114]. Patents to be remaining lifetime is higher than 0 are considered in this study. CPC codes of the patents are used to create the patent-class matrix. In Step1, patent-class matrix forms the attribute values of the data that the machine learning algorithm will use to establish predict model. Nine patent indices are considered to calculate scores of the patents. These indices are number of citations, number of annual forward citations, number of back citations, TCT, pendency, remaining lifetime of the patent, patent family size, technology score, and interaction power.

In Step 2, expert opinion is received to determine the weights of these indices. In this stage, the AHP method is applied. The view of patent attorneys, academics working on technology management, and techno park, and TTO experts responsible for the commercialization of patents are received to determine the weights of patent indices. The pair-wise comparison matrices are obtained for 22 experts. The group comparison matrix is given in Table V. After the matrix is normalized, the weight values of the patent indices are calculated. The weights of patent indices are given in Table VI. As can be seen in Table VI, the most effective indices on patent quality are interaction power, technology score, patent family size, and remaining life of the patent. The calculated CI value is 0.0975 and the CR value is 0.067. Since the CR value is lower than 0.1, the inconsistency of comparison is within acceptable limits. The scores of the patent indices are calculated for each patent. The minimum, maximum, and mean values of the patent indices are given in Table VII.

In Step 3, the patent scores are calculated using the scores and weights of the patent indices. The scores of patent indices are normalized using the min–max normalization method. In Step 4, the patent scores are assigned to 4 patent score clusters using the SOM algorithm. The descriptive statistics of normalized patent score clusters are given in Table VIII.

As can be seen in Table VIII, the patents in Cluster-4 have the highest patent score value. The quality levels of clusters from high to low are Cluster-4, Cluster-3, Cluster-2, and Cluster 1, respectively. After patent score clusters are calculated, the distribution of these patent scores within the technology clusters is also analyzed. These technology clusters are obtained using the SOM algorithm. Patents related to personal digital assistant technologies are grouped in four technology clusters CPC codes. Therefore, patents with similar CPC codes are assigned to the same technology cluster. However, all of the patents in a technology cluster are not of the same quality level. Patents with different quality values can exist in the same cluster. The distribution of patent scores within technology clusters is given in Fig. 2.

As can be seen in Fig. 2, there are patents with different patent scores in the same technology cluster. The patents in Clusters 2, 3, and 4 have the highest patent scores. There are patents with different quality values for all of the technology clusters. For this reason, considering patents individually to determine the patent quality rate can give more accurate and more precise results. The proposed prediction model considers the quality of patents individually. The obtained patent score clusters are the

TABLE VI
WEIGHT VALUES OF THE PATENT INDICES

Patent Indices	Number of claims	Number of annual forward citations	Number of back citations	TCT	Pendency	Remaining lifetime	Patent family size	Technology score	Interaction power
Weights	0.0415	0.0457	0.0403	0.085	0.0605	0.1417	0.1608	0.1726	**0.2519**

TABLE VII
VALUES OF PATENT INDICES

Patent indices	Number of claims	Number of annual forward citations	Number of back citations	TCT	Pendency	Remaining lifetime	Patent family size	Technology score	Interaction power
Min. value	1	0	0	13	127	147	1	9	0
Max. value	89	50.9108	104	63	7228	7147	65	98	99
Mean value	18.288	2.664	13.937	30.781	1572.604	2269.246	8.459	63.639	28.1805

TABLE VIII
NORMALIZED SCORES OF CLUSTERS

Clusters	Number of Patents	Min Patent Score	Max Patent Score	Mean Patent Score
Cluster-1	54	0	0.2819	0.1795
Cluster-2	**168**	0.2878	0.4776	0.3934
Cluster-3	120	0.4786	0.6820	0.5584
Cluster-4	57	**0.6871**	**1**	**0.7984**

The bold entities indicate the best result of the comparison methods.

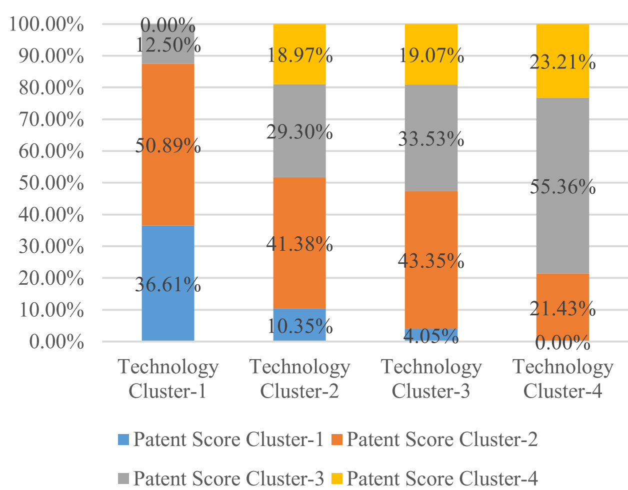

Fig. 2. Distribution of patent scores within technology clusters.

label value of the data used by the machine learning algorithm. Attribute values are CPC codes.

In Step 5, these label values are classified using individual and ensemble classification algorithms. In Step 6, these algorithms are tested to establish a prediction model. Ensemble machine learning algorithm detects a weak classifier in each part of the data. Combines each of these weak classifiers into a global classifier. Each of these weak classifiers is combined into a global classifier [115]. Voting [116], Stacking [117], and AdaBoost algorithm [118] have the best performance values among the ensemble machine learning methods. The SMO [119] and MLP algorithms [111] are used as subclassifier by the Voting algorithm. The Stacking algorithm are used the logistic model tree [120] and the J48 [121] algorithm as subclassifier and MLP algorithm as meta classifier. The MLP algorithm is used as subclassifier by AdaBoost algorithm.

The MLP algorithm has the best performance values among the individual machine learning methods. In Step 6, algorithms with the best classification performance among the tested machine learning algorithms are analyzed in detail. The performance values of these algorithms are given in Fig. 3.

The accuracy, MAE, and Kappa statistics values of the machine learning algorithms are given in Fig. 3. Prediction models with a high accuracy value and Kappa statistics are preferred in this study. The algorithm with the highest Kappa statistics and accuracy value is the MLP (Kappa statistics = 65.66 and Accurate value = 76). The MAE value is expected to be small. The machine learning algorithm with the lowest MAE value is AdaBoost (MAE = 14.93), using the MLP algorithm as a subclassifier. The MAE value of the MLP algorithm is also lower than other algorithms. For this reason, the MLP algorithm is selected for the establishment of the prediction model in Step 7.

Parameter values of the MLP algorithm are given in Table IX.

The algorithm provides the best prediction performance values when the number of hidden layers is 8.

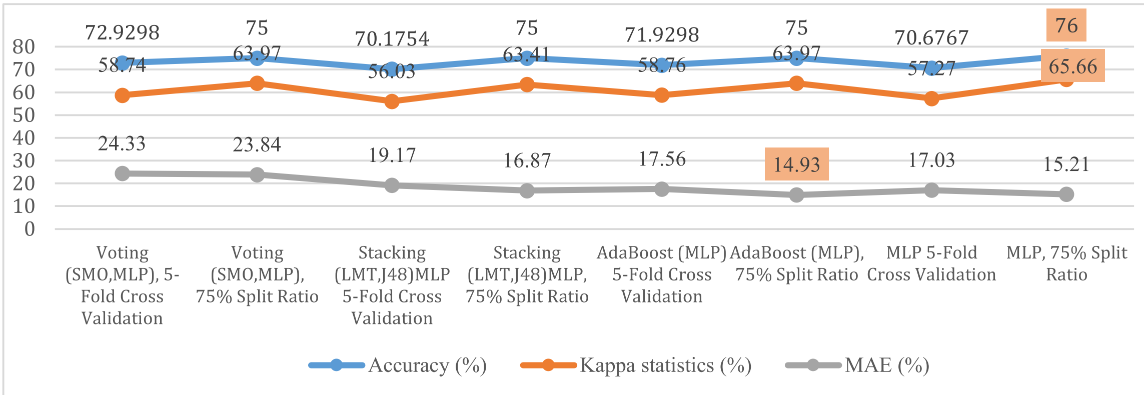

Fig. 3. Performance values of machine learning algorithms.

TABLE IX
PARAMETER SETTINGS OF MLP ALGORITHM

Parameter	Settings
Hidden layer	a
Learn rate	0.3
Momentum	0.2
Training time	370
Validity threshold	20
Activation function	Sigmoid

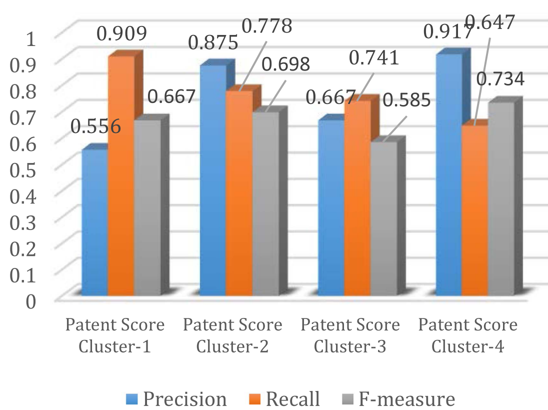

Fig. 4. Performance values for the 75% split ratio method.

In Step 8, patent quality is predicted using MLP algorithm. Performance values of the MLP algorithm for the 75% split ratio method are given in Fig. 4. Additionally, performance values of the MLP algorithm for the five-fold cross-validation method are given in Fig. 5.

Figs. 4 and 5 show the performance values of the prediction models for each cluster. High precision, recall, and *F*-measurement values mean good prediction performance. These values are required to give a high value for all output values. Therefore, these values are given for all output values in Figs. 4 and 5. The accuracy value of MLP is 76% with the 75% split ratio method. However, when the prediction results for all label values are analyzed, the performance values obtained using the five-fold cross-validation method are better.

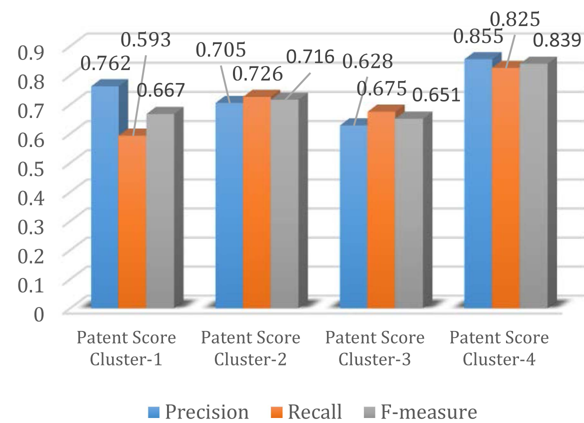

Fig. 5. Performance values for the five-fold cross validation method.

TABLE X
RANK VALUES OF ATTRIBUTES

Attribute No	Rank Value
G06F	0.1392
B41J	0.1040
Y02D	0.1023
H04W	0.0924
G06K	0.0866
G02B	0.0692
G11C	0.0507
H04M	0.0441
B25G	0.0192
B25D	0.0192
B25F	0.0192
G11B	0.0175

The performance of the prediction model for Cluster-4 is higher than other clusters. The codes with the highest info gain value among the CPC codes taken as input variables and info gain values are given in Table X.

The CPC code having the highest impact on patent quality is G06F, which refers to the processing of digital electrical data.

V. Conclusion

In this study, an approach is proposed to predicting patent quality based on machine learning algorithm. First, nine patent indices used in the literature are considered to measure the quality of patents. These indices are the number of claims, the number of annual forward citations, the number of back citations, the patent family size, pendency, the remaining lifetime, the TCT, the technology score, and the interaction power. In the literature, there is no study for weighting patent indices. However, these indices have different effect levels on patent quality. Expert opinion is asked to fill this gap in the literature. The AHP method is used for the calculation of weights in the proposed approach. As a result of expert evaluations, the most effective indices for assessing the patents are determined as the interaction power, technology score, the size of the patent family, and the remaining lifetime. The weighted sum of the indices represents the patent score. These patent scores are clustered using the SOM algorithm to obtain the label values of the data to be used. The score clusters obtained for each patent are predicted using CPC codes of the patents. MLP algorithm provided the highest accuracy value. This algorithm has a 76% accuracy value.

A. Theoretical Implications

In the literature, technology clusters are generally used by patent-based technology evaluation approaches. All of the patents in the same technology cluster have not the same quality rate. The patents with the different quality level may available in the same cluster. Although the technology cluster obtained using various clustering algorithms has a low quality rate, there may be patents with high quality rates in this technology cluster. The proposed approach fills this gap in the literature considering patents individually. The activity level or difference of technologies or patents within a technology field significantly affects the value of technologies. In this study, the effect levels of patent index values on the evaluation of technologies or patents are determined. CPC codes and patent indices, which have been utilized in the literature to assess technologies, are approached from a different angle in order to present a model and data set. The lack of data is a significant issue in the field of technology management. Monitoring patents in the database is essential to establish a prediction model. This study differs from the studies in the literature in terms of the proposed data set and methodology. Every stage of the methodology is open to improvement for future studies. Each stage of the proposed methodology is a separate research topic in the literature. Therefore, this study provides a new perspective to researchers in different fields of study.

B. Managerial Implications

In this study, patent quality is predicted using a machine learning-based approach. Patent quality indicates the level of meeting the expectations of decision makers from a patent. Patent indices represent decision makers' expectations. The higher a patent's potential in areas, such as acceptance, commercialization, and attribution, the higher the satisfaction level of the decision maker will be. Therefore, the index values used in the literature are taken into account to measure these expectations while predicting the patent quality. The choice of technology or patent to invest in is critical for decision makers. This approach is a practical and useful tool for decision makers. Decision makers can direct the budget allocated to R&D studies to promising technology fields through the proposed approach. In addition, the incentives given for patent applications can be transferred to the patents in the technology field that has a high quality rate. The proposed approach has been applied to patents about PDA. The result shows that G06F CPC code (processing of digital electrical data) has the highest information gain value for predicting the patent quality. B41J (typewriters, selective printing mechanisms), Y02D (Information and communication technologies used to minimize the use of energy), H04W (wireless communication systems), and G06K (recognition and presentation of data) CPC codes have high information gain values. These codes refer to technologies that provide the highest information to predict the quality of the patents related to personal digital assistant technologies. The patents with these codes have higher quality rates.

C. Limitations and Further Research

For future studies, monitoring the quality of patents and recording these patent quality data as label values in the database will increase the accuracy of the prediction. Since promising technology fields may change over time, monitoring patents will be useful to identify changes reflect on the prediction model. In addition, a detailed analysis can be performed by adding patent data of interconnected technologies to the dataset. Indices used to calculate the output value can be updated. Different indicators measuring commercial value can also be included in the output value. Technology-based differentiations can be eliminated such as citation or review time. In addition, methods such as social network analysis and structural hole analysis can be used to assess technologies. In this study, all patents about PDA technologies are included in the analysis. However, the patents of the last 4 years should be removed from the analysis for more accurate results in the future research. A fuzzy approach can be used for the weighting of the index. Additionally, fuzzy clustering algorithms can be applied to create patent score clusters. Thus, more precise results can be obtained. Different MCDM methods can be used for weighting the indices. In this study, each patent is evaluated individually. However, the subtechnology to which a patent is linked is the most crucial factor impacting its quality. In this study, only the 4 digits of the CPC code are considered while examining the substudy fields. Future studies may consider all digits of CPC codes for a more comprehensive analysis. The SOM algorithm is preferred, since the dataset used in this study is not large. Future studies working with large datasets can use different clustering algorithms to get faster results. Optimization can also be made for parameter values of clustering algorithms.

TABLE XI
INFORMATION OF PARTICIPANTS INCLUDED IN THE AHP STUDY

Title	Bachelor degree	Experience (year)
Academician (Prof. Dr.)	Mechanical engineer	17
Academician (Assoc. Prof.)	Industrial engineer	11
Academician (Lecturer)	Industrial engineer	12
Academician (Lecturer)	Industrial engineer	9
Academician (Res. Assist.), TTO-IPRs specialist	Industrial engineer	5
Academician (Res. Assist.)	Industrial engineer	5
Academician (Res. Assist.), TTO- Entrepreneurship and business development specialist	Industrial engineer	5
Academician (Asst. Prof.), TTO assistant director	Electrical electronics engineer	6
Patent attorney	Mechanical engineer	20
Patent attorney	Mechatronics engineer	5
Patent attorney	Electrical electronics engineer	6
Patent attorney	Mechanical engineer	17
TTO-IPRs specialist, Patent attorney	Mechatronics engineer	11
Trademark and patent attorney, TTO-IPRs specialist	Economist, Lawyer	25
TTO-IPRs specialist	Industrial engineer	8
TTO-IPRs specialist, Technology transfer specialist	Economist	8
TTO-IPRs specialist, project specialist	Computer engineering	6
TTO- Technology transfer specialist	Economist	7
TTO-IPRs specialist	Computer engineering	7
TTO Coordinator, Technology transfer specialist	Economist	10
TTO-Technology transfer specialist	Electrical electronics engineer	16
TTO-IPRs specialist	Electrical electronics engineer	5

*Technology Transfer Office (TTO)
*Intellectual Property Rights (IPRs)

APPENDIX

See Tables XI.

ACKNOWLEDGMENT

The authors would like to acknowledge that this article is submitted in partial fulfilment of the requirements for Ph.D. degree at Yildiz Technical University.

REFERENCES

[1] B. P. Abraham and S. D. Moitra, "Innovation assessment through patent analysis," *Technovation*, vol. 21, no. 4, pp. 245–252, 2001.

[2] F.-M. Tseng, C.-H. Hsieh, Y.-N. Peng, and Y.-W. W. Chu, "Using patent data to analyze trends and the technological strategies of the amorphous silicon thin-film solar cell industry," *Technological Forecasting Social Change*, vol. 78, no. 2, pp. 332–345, 2011.

[3] Y.-H. Lai and H.-C. Che, "Modeling patent legal value by extension neural network," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10520–10528, 2009.

[4] J. Yoon, W. Seo, B. Coh, I. Song, and J.-M. Lee, "Identifying product opportunities using collaborative filtering-based patent analysis," *Comput. Ind. Eng.*, vol. 107, pp. 376–387, 2017.

[5] M.-J. Shih, D.-R. Liu, and M.-L. Hsu, "Discovering competitive intelligence by mining changes in patent trends," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 2882–2890, 2010.

[6] G. Kim and J. Bae, "A novel approach to forecast promising technology through patent analysis," *Technological Forecasting Social Change*, vol. 117, pp. 228–237, 2017.

[7] 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.

[8] M. N. Kyebambe, G. Cheng, Y. Huang, C. He, and Z. Zhang, "Forecasting emerging technologies: A supervised learning approach through patent analysis," *Technol. Forecast. Soc. Change*, vol. 125, pp. 236–244, 2017.

[9] S. Altuntas, Z. Erdogan, and T. Dereli, "A clustering-based approach for the evaluation of candidate emerging technologies," *Scientometrics*, vol. 124, no. 2, pp. 1157–1177, 2020, doi: 10.1007/s11192-020-03535-0.

[10] A. Nayyar and V. Puri, "Comprehensive analysis & performance comparison of clustering algorithms for big data," *Rev. Comput. Eng. Res.*, vol. 4, no. 2, pp. 54–80, 2017.

[11] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," *ACM Sigmod Rec,* vol. 25, no. 2, pp. 103–114, 1996.

[12] R. T. Ng and J. Han, "CLARANS: A method for clustering objects for spatial data mining," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 5, pp. 1003–1016, Sep./Oct. 2002.

[13] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009.

[14] S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," *ACM Sigmod Rec.*, vol. 27, no. 2, pp. 73–84, 1998.

[15] 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.

[16] L. Gao et al., "Technology life cycle analysis method based on patent documents," *Technological Forecasting Social Change*, vol. 80, no. 3, pp. 398–407, 2013.

[17] S.-W. Choi, Y.-Y. You, and K.-S. Na, "Forecasting promising technology using analysis of patent information: Focused on next generation mobile communications," *J. Central South Univ.*, vol. 21, no. 11, pp. 4303–4310, 2014.

[18] M. Grimaldi, L. Cricelli, M. Di Giovanni, and F. Rogo, "The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning," *Technological Forecasting Social Change*, vol. 94, pp. 286–302, 2015.

[19] J. Ni, R. Shao, C. O. L. Ung, Y. Wang, Y. Hu, and Y. Cai, "Valuation of pharmaceutical patents: A comprehensive analytical framework based on technological, commercial, and legal factors," *J. Pharmaceut. Innov.*, vol. 10, no. 3, pp. 281–285, 2015.

[20] D. Verhoeven, J. Bakker, and R. Veugelers, "Measuring technological novelty with patent-based indicators," *Res. Policy*, vol. 45, no. 3, pp. 707–723, 2016.

[21] S. H. Chang and C. Y. Fan, "Identification of the technology life cycle of telematics: A patent-based analytical perspective," *Technological Forecasting Social Change*, vol. 105, pp. 1–10, 2016.

[22] K. Song, K. Kim, and S. Lee, "Identifying promising technologies using patents: A retrospective feature analysis and a prospective needs analysis on outlier patents," *Technological Forecasting Social Change*, vol. 128, pp. 118–132, 2018.

[23] U. Kwon and Y. Geum, "Identification of promising inventions considering the quality of knowledge accumulation: A machine learning approach," *Scientometrics*, vol. 125, no. 3, pp. 1877–1897, 2020.

[24] W. Hur and J. Oh, "A man is known by the company he keeps?: A structural relationship between backward citation and forward citation of patents," *Res. Policy*, vol. 50, no. 1, pp. 1–15, 2021.
[25] D. Strumsky, J. Lobo, and S. Van der Leeuw, "Using patent technology codes to study technological change," *Econ. Innov. New Technol.*, vol. 21, no. 3, pp. 267–286, 2012.
[26] K. R. Harrigan, M. C. Di Guardo, E. Marku, and B. N. Velez, "Using a distance measure to operationalise patent originality," *Technol. Anal. Strategic Manage.*, vol. 29, no. 9, pp. 988–1001, 2017.
[27] R. M. Margolis and D. M. Kammen, "Underinvestment: The energy technology and R&D policy challenge," *Science*, vol. 285, no. 5428, pp. 690–692, 1999.
[28] J. Choi, B. Jeong, J. Yoon, B.-Y. Coh, and J.-M. Lee, "A novel approach to evaluating the business potential of intellectual properties: A machine learning-based predictive analysis of patent lifetime," *Comput. Ind. Eng.*, vol. 145, 2020, Art. no. 106544.
[29] F. Altuntas and M. Ş. Gök, "Analysis of patent documents with utility mining: A case study of wind energy technology," *Kybernetes*, vol. 50, no. 9, pp. 2548–2582, 2020.
[30] S. Oh, J. Choi, N. Ko, and J. Yoon, "Predicting product development directions for new product planning using patent classification-based link prediction.," *Scientometrics*, vol. 125, no. 3, pp. 1833–1876, 2020.
[31] J. Shim, "Analysis of technology association rules between CPC codes of the Internet of Things (IoT)'patent," *J. Korea Inst. Inf., Electron. Commun. Technol.*, vol. 12, no. 5, pp. 493–498, 2019.
[32] J. Jeon, S. Chae, and J. Gim, "Technical invention trend analysis of applicants based on CPC classification," in *Proc. Int. Conf. Web Eng.*, 2019, pp. 10–17.
[33] S. Chae and J. Gim, "A study on trend analysis of applicants based on patent classification systems," *Information*, vol. 10, no. 12, pp. 1–12, 2019.
[34] C.-H. Wang and L.-C. Hsu, "Using genetic algorithms grey theory to forecast high technology industrial output," *Appl. Math. Comput.*, vol. 195, no. 1, pp. 256–263, 2008.
[35] E. E. Karsak and S. S. Ahiska, "Practical common weight multi-criteria decision-making approach with an improved discriminating power for technology selection," *Int. J. Prod. Res.*, vol. 43, no. 8, pp. 1537–1554, 2005.
[36] A. Durmusoglu and T. Dereli, "On the technology forecasting using data envelopment analysis (TFDEA)," in *Proc. Technol. Manage. Energy Smart World*, 2011, pp. 1–6.
[37] D.-J. Lim and T. R. Anderson, "Incorporating value judgment in technology forecasting using data envelopment analysis," in *Proc. Technol. Manage. IT-Driven Serv.*, 2013, pp. 2093–2097.
[38] B. S. Yoon, A. Charoensupyan, N. Hu, R. Koosawangsri, M. Abdulai, and X. Wang, "Technology forecasting of digital single-lens reflex camera market: The impact of segmentation in TFDEA," in *Proc. Technol. Manage. IT-Driven Serv.*, 2013, pp. 2116–2129.
[39] S. Altuntas, T. Dereli, and A. Kusiak, "Forecasting technology success based on patent data," *Technological Forecasting Social Change*, vol. 96, pp. 202–214, 2015.
[40] M. Obrecht and M. Denac, "Technology forecast of sustainable energy development prospects," *Futures*, vol. 84, pp. 12–22, 2016.
[41] J. H. Dahooie, N. Mohammadi, T. Daim, A. S. Vanaki, and E. K. Zavadskas, "Matching of technological forecasting technique to a technology using fuzzy multi-attribute decision-making methods: Case study from the aerospace industry," *Technol. Soc.*, vol. 67, pp. 1–11, 2021.
[42] Y. C. Kim, J. M. Ahn, O. Kwon, and C. Lee, "Valuation of university-originated technologies: A predictive analytics approach," *IEEE Trans. Eng. Manage.*, vol. 68, no. 6, pp. 1813–1825, Dec. 2021.
[43] S. Altuntas and S. Aba, "Technology forecasting of unmanned aerial vehicle technologies through hierarchical s curves," *Defence Sci. J.*, vol. 72, no. 1, pp. 18–29, 2022.
[44] A. Golkar, I. Yuskevich, K. Smirnova, and R. Vingerhoeds, "Model-based approaches for technology planning and roadmapping: Technology forecasting and game-theoretic modeling," *Technological Forecasting Social Change*, vol. 173, pp. 1–12, 2021.
[45] X. Yuan and Y. Cai, "Forecasting the development trend of low emission vehicle technologies: Based on patent data," *Technological Forecasting Social Change*, vol. 166, pp. 1–9, 2021.
[46] R. V. Sartori, D. R. dos Reis, M. Bronzeri, and A. Q. Silva, "Technology forecast: A case study in Daiken company," *Int. J. Innov. Sci.*, pp. 203–217, 2020.
[47] A. C. Adamuthe and G. T. Thampi, "Technology forecasting: A case study of computational technologies," *Technological Forecasting Social Change*, vol. 143, pp. 181–189, 2019.
[48] J. R. San Cristóbal, "Multi-criteria decision-making in the selection of a renewable energy project in Spain: The Vikor method," *Renew. Energy*, vol. 36, no. 2, pp. 498–502, 2011.
[49] J. Ren and M. Lützen, "Fuzzy multi-criteria decision-making method for technology selection for emissions reduction from shipping under uncertainties," *Transp. Res. Part D: Transport Environ.*, vol. 40, pp. 43–60, 2015.
[50] J. Chu, J. Wu, C. Chu, and M. Liu, "A new DEA common-weight multi-criteria decision-making approach for technology selection," *Int. J. Prod. Res.*, vol. 58, no. 12, pp. 3686–3700, 2020.
[51] F. Pakdil, P. Toktaş, and G. F. Can, "Six sigma project prioritization and selection: A multi-criteria decision making approach in healthcare industry," *Int. J. Lean Six Sigma*, vol. 12, no. 3, pp. 553–578, 2020.
[52] H. J. Mohammed, "The optimal project selection in portfolio management using fuzzy multi-criteria decision-making methodology," *J. Sustain. Finance Investment*, pp. 1–17, 2021, doi: 10.1080/20430795.2021.1886551.
[53] A. E. C. Mondragon, E. Mastroncinque, J.-F. Tsai, and P. J. Hogg, "An AHP and fuzzy AHP multifactor decision making approach for technology and supplier selection in the high-functionality textile industry," *IEEE Trans. Eng. Manage.*, vol. 68, no. 4, pp. 1112–1125, Aug. 2021.
[54] X. Zhou et al., "Identifying and assessing innovation pathways for emerging technologies: A hybrid approach based on text mining and altmetrics," *IEEE Trans. Eng. Manage.*, vol. 68, no. 5, pp. 1360–1371, Sep. 2020.
[55] X. Han, D. Zhu, X. Wang, J. Li, and Y. Qiao, "Technology opportunity analysis: Combining SAO networks and link prediction," *IEEE Trans. Eng. Manage.*, vol. 68, no. 5, pp. 1288–1298, Oct. 2019.
[56] A. E. Torkayesh, B. Malmir, and M. R. Asadabadi, "Sustainable waste disposal technology selection: The stratified best-worst multi-criteria decision-making method," *Waste Manage.*, vol. 122, pp. 100–112, 2021.
[57] S. Mall and R. Anbanandam, "A fuzzy analytic hierarchy process and VIKOR framework for evaluation and selection of electric vehicle charging technology for India," *Transp. Developing Economies*, vol. 8, no. 1, pp. 1–11, 2022.
[58] S. Park and S. Jun, "Patent analysis using bayesian data analysis and network modeling," *Appl. Sci.*, vol. 12, no. 3, pp. 1–20, 2022.
[59] S. Ranaei, A. Suominen, A. Porter, and S. Carley, "Evaluating technological emergence using text analytics: Two case technologies and three approaches," *Scientometrics*, vol. 122, no. 1, pp. 215–247, 2020.
[60] H.-Y. Woo, J. Kwak, and C. Lim, "A study on patent evaluation model based on Bayesian approach of the structural equation model," *Korean J. Appl. Statist.*, vol. 30, no. 6, pp. 901–916, 2017.
[61] X. Han, D. Zhu, M. Lei, and T. Daim, "R&D trend analysis based on patent mining: An integrated use of patent applications and invalidation data," *Technological Forecasting Social Change*, vol. 167, pp. 1–11, 2021.
[62] A. J. C. Trappey, C. V. Trappey, U. H. Govindarajan, and J. J. H. Sun, "Patent value analysis using deep learning models—The case of IoT technology mining for the manufacturing industry," *IEEE Trans. Eng. Manage.*, vol. 68, no. 5, pp. 1334–1346, Oct. 2021.
[63] D. Jeon, J. M. Ahn, J. Kim, and C. Lee, "A doc2vec and local outlier factor approach to measuring the novelty of patents," *Technological Forecasting Social Change*, vol. 174, pp. 1–11, 2022.
[64] T. Sinigaglia, M. E. S. Martins, and J. C. M. Siluk, "Technological evolution of internal combustion engine vehicle: A patent data analysis," *Appl. Energy*, vol. 306, 2022, Art. no. 118003.
[65] A. Sharma and R. Sharma, "Patent citing patterns of foreign firms: An analysis of Indian IT services industry," *Asian J. Technol. Innov.*, pp. 1–27, 2022, doi: 10.1080/19761597.2022.2026234.
[66] K. Song, C. Ran, and L. Yang, "A digital analysis system of patents integrating natural language processing and machine learning," *Technol. Anal. Strateg. Manage.*, pp. 1–17, 2022, doi: 10.1080/09537325.2022.2035349.
[67] M.-H. Chao, A. J. C. Trappey, and C.-T. Wu, "Emerging technologies of natural language-enabled chatbots: A review and trend forecast using intelligent ontology extraction and patent analytics," *Complexity*, vol. 2021, pp. 1–26, 2021.
[68] X. You, J. Liu, Z. Wang, Z. Liu, X. Lv, and J. Y. Kim, "Applying deep learning technologies to evaluate the patent quality with the collaborative training," *Wireless Commun. Mobile Comput.*, pp. 1–23, 2021, doi: 10.1155/2021/5006974.
[69] M. Kim and Y. Geum, "Predicting patent transactions using patent-based machine learning techniques," *IEEE Access*, vol. 8, pp. 188833–188843, 2020.

[70] J. Ma, Y. Pan, and C. Y. Su, "Organization-oriented technology opportunities analysis based on predicting patent networks: A case of alzheimer's disease," *Scientometrics*, vol. 127, pp. 5497–5517, 2022.

[71] S. Jun, S. S. Park, and D. S. Jang, "Technology forecasting using matrix map and patent clustering," *Ind. Manage. Data Syst.*, vol. 112, pp. 786–807, 2012.

[72] M. Lee and S. Lee, "Identifying new business opportunities from competitor intelligence: An integrated use of patent and trademark databases," *Technological Forecasting Social Change*, vol. 119, pp. 170–183, 2017.

[73] M. D. Madvar, F. Ahmadi, R. Shirmohammadi, and A. Aslani, "Forecasting of wind energy technology domains based on the technology life cycle approach," *Energy Rep.*, vol. 5, pp. 1236–1248, 2019.

[74] J. Jeon and Y. Suh, "Multiple patent network analysis for identifying safety technology convergence," *Data Technol. Appl.*, vol. 53, pp. 269–285, 2019.

[75] J. Yu et al., "Identification of vacant and emerging technologies in smart mobility through the GTM-Based patent map development," *Sustainability*, vol. 12, no. 22, 2020, Art. no. 9310.

[76] P. Ampornphan and S. Tongngam, "Exploring technology influencers from patent data using association rule mining and social network analysis," *Information*, vol. 11, no. 6, 2020, Art. no. 333.

[77] L. Feng, Y. Li, Z. Liu, and J. Wang, "Idea generation and new direction for exploitation technologies of coal-seam gas through recombinative innovation and patent analysis," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020, Art. no. 2928.

[78] D. A. Naik, C. J. Brunda, and S. Seema, "A feasible dashboard to predict patent mining using classification algorithms," *Procedia Comput. Sci.*, vol. 167, pp. 2011–2021, 2020.

[79] Y. Tang, X. Lou, Z. Chen, and C. Zhang, "A study on dynamic patterns of technology convergence with IPC co-occurrence-based analysis: The case of 3D printing," *Sustainability*, vol. 12, no. 7, 2020, Art. no. 2655.

[80] C. Lee, S. Hong, and J. Kim, "Anticipating multi-technology convergence: A machine learning approach using patent information," *Scientometrics*, vol. 126, pp. 1867–1896, 2021.

[81] H. Sasaki and I. Sakata, "Identifying potential technological spin-offs using hierarchical information in international patent classification," *Technovation*, vol. 100, 2021, Art. no. 102192.

[82] T. Daim, E. Bukhari, D. Bakry, J. VanHuis, H. Yalcin, and X. Wang, "Forecasting technology trends through the gap between science and technology: The Case of software as an E-Commerce service," *Форсайт*, vol. 15, no. 2, pp. 12–24, 2021.

[83] X. Li, Q. Xie, T. Daim, and L. Huang, "Forecasting technology trends using text mining of the gaps between science and technology: The case of perovskite solar cell technology," *Technological Forecasting Social Change*, vol. 146, pp. 432–449, 2019.

[84] S. Ozcan, A. Homayounfard, C. Simms, and J. Wasim, "Technology roadmapping using text mining: A foresight study for the retail industry," *IEEE Trans. Eng. Manage.*, vol. 69, no. 1, pp. 228–244, Feb. 2022.

[85] M. Gui and X. Xu, "Technology forecasting using deep learning neural network: Taking the case of robotics," *IEEE Access*, vol. 9, pp. 53306–53316, 2021.

[86] J. Liu, J. Wei, and Y. Liu, "Technology forecasting based on topic analysis and social network analysis: A case study focusing on gene editing patents," *J. Sci. Ind. Res. (India)*, vol. 80, no. 5, pp. 428–437, 2021.

[87] S. Xu, L. Hao, G. Yang, K. Lu, and X. An, "A topic models based framework for detecting and forecasting emerging technologies," *Technological Forecasting Social Change*, vol. 162, 2021, Art. no. 120366.

[88] T. San Kim and S. Y. Sohn, "Machine-learning-based deep semantic analysis approach for forecasting new technology convergence," *Technological Forecasting Social Change*, vol. 157, pp. 1–10, 2020.

[89] Y. Zhou, F. Dong, Y. Liu, Z. Li, J. Du, and L. Zhang, "Forecasting emerging technologies using data augmentation and deep learning," *Scientometrics*, vol. 123, no. 1, pp. 1–29, 2020.

[90] K. H. Kim, Y. J. Han, S. Lee, S. W. Cho, and C. Lee, "Text mining for patent analysis to forecast emerging technologies in wireless power transfer," *Sustainability*, vol. 11, no. 22, 2019, Art. no. 6240.

[91] T. Kohonen, "Analysis of a simple self-organizing process," *Biol. Cybern.*, vol. 44, no. 2, pp. 135–140, 1982, doi: 10.1007/BF00317973.

[92] T. W. Liao, "Clustering of time series data—A survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, 2005.

[93] J. Lerner, "The importance of patent scope: An empirical analysis," *RAND J. Econ.*, vol. 25, no. 2, pp. 319–333, 1994.

[94] F. Narin, E. Noma, and R. Perry, "Patents as indicators of corporate technological strength," *Res. Policy*, vol. 16, no. 2–4, pp. 143–155, 1987.

[95] C. Lee, O. Kwon, M. Kim, and D. Kwon, "Early identification of emerging technologies: A machine learning approach using multiple patent indicators," *Technological Forecasting Social Change*, vol. 127, pp. 291–303, 2017.

[96] C. Lee, Y. Cho, H. Seol, and Y. Park, "A stochastic patent citation analysis approach to assessing future technological impacts," *Technological Forecasting Social Change*, vol. 79, no. 1, pp. 16–29, 2012.

[97] I. von Wartburg, T. Teichert, and K. Rost, "Inventive progress measured by multi-stage patent citation analysis," *Res. Policy*, vol. 34, no. 10, pp. 1591–1607, 2005.

[98] D. Harhoff and S. Wagner, "The duration of patent examination at the European patent office," *Manage. Sci.*, vol. 55, no. 12, pp. 1969–1984, 2009.

[99] D. Harhoff and S. Wagner, "Modeling the duration of patent examination at the European patent office," *Discuss. Pap. Bus. Adm.*, pp. 1–35, 2006. [Online]. Available: http://epub.ub.uni-muenchen.de

[100] J. Y. Og, K. Pawelec, B.-K. Kim, R. Paprocki, and E. Jeong, "Measuring patent value indicators with patent renewal information," *J. Open Innov.: Technol. Market Complexity*, vol. 6, no. 1, pp. 1–16, 2020.

[101] 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.

[102] H. Yamada, "Identification methods and indicators of important patents," *Library Hi Tech*, vol. 40, pp. 750–785, 2021.

[103] M. R. Taylor, "The influence of government actions on innovative activities in the development of environmental technologies to control sulfur dioxide emissions from stationary sources," Ph.D. Dissertation, Carnegie Mellon Univ., Pittsburgh, PA, USA, 2001.

[104] A. Pakes and M. Schankerman, "The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources," in *R&D, Patents, and Productivity*. Chicago, IL, USA: Univ. of Chicago Press, 1984, pp. 73–88.

[105] A. J. C. 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.

[106] E. F. Sherry and D. J. Teece, "Royalties, evolving patent rights, and the value of innovation," *Res. Policy*, vol. 33, no. 2, pp. 179–191, 2004.

[107] K.-B. Min, C. Lee, and Y.-C. Kim, "The impact of the timing of patent allowance on technology licensing performance: Evidence from university invention commercialization," *R&D Manage.*, vol. 52, pp. 633–649, 2021.

[108] T. L. Saaty, "The analytic hierarchy process," in *Agricultural Economics Review*. Berlin, Germany: Springer, 1980, pp. 49–68.

[109] G. Fernandez, *Statistical Data Mining Using SAS Applications*. 2nd ed. Boca Raton, FL, USA: CRC Press, 2010.

[110] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques Third Edition*, 2nd ed., Burlington, NJ, USA: Morgan Kaufmann, 2011.

[111] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[112] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, no. 14/15, pp. 2627–2636, 1998.

[113] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, 2011.

[114] S. Lee, B. Yoon, and Y. Park, "An approach to discovering new technology opportunities: Keyword-based patent map approach," *Technovation*, vol. 29, no. 6/7, pp. 481–497, 2009.

[115] H. Kargupta, J. Han, P. S. Yu, and F. Group, *Next Generation of Data Mining*. Boca Raton, FL, USA: CRC Press, 2008.

[116] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Anal. Appl.*, vol. 1, no. 1, pp. 18–27, 1998.

[117] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[118] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Int. Conf. Mach. Learn.*, 1996, pp. 148–156.

[119] M. Cottrell, J. C. Fort, and G. Pagès, "Theoretical aspects of the SOM algorithm," *Neurocomputing*, vol. 21, no. 1–3, pp. 119–138, 1998.

[120] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[121] J. R. Quinlan, "Improved use of continuous attributes in C4. 5," *J. Artif. Intell. Res.*, vol. 4, pp. 77–90, 1996, doi: 10.1613/jair.279.