A survey on deep learning for patent analysis

Article *in* World Patent Information · June 2021

DOI: 10.1016/j.wpi.2021.102035

4 authors:

Ralf Krestel
ZBW - Leibniz Information Centre for Economics

Renukswamy Chikkamath
Munich University of Applied Sciences

Christoph Hewel

Julian Risch
deepset

A Survey on Deep Learning for Patent Analysis

Ralf Krestel[a], Renukswamy Chikkamath[b], Christoph Hewel[c] and Julian Risch[a]

[a]Hasso Plattner Institute, University of Potsdam, Germany.
email: firstname.lastname@hpi.de

[b]University of Passau, Passau, Germany.
email: chikka.renu@gmail.com

[c]BETTEN und RESCH Patent- und Rechtsanwaelte PartGmbB
email: C.Hewel@bettenpat.com

ARTICLE INFO

Keywords:
Deep learning;
Patent analysis;
Text mining;
Natural language processing;

ABSTRACT

Patent document collections are an immense source of knowledge for research and innovation communities worldwide. The rapid growth of the number of patent documents poses an enormous challenge for retrieving and analyzing information from this source in an effective manner. Based on deep learning methods for natural language processing, novel approaches have been developed in the field of patent analysis. The goal of these approaches is to reduce costs by automating tasks that previously only domain experts could solve. In this article, we provide a comprehensive survey of the application of deep learning for patent analysis. We summarize the state-of-the-art techniques and describe how they are applied to various tasks in the patent domain. In a detailed discussion, we categorize 40 papers based on the dataset, the representation, and the deep learning architecture that were used, as well as the patent analysis task that was targeted. With our survey, we aim to foster future research at the intersection of patent analysis and deep learning and we conclude by listing promising paths for future work.

1. The Patent Life-Cycle and Tasks for Automation

The patent life-cycle requires experts with domain-specific knowledge at all its different stages. First, when companies or individual inventors wish to obtain patent protection for an invention, they need to specify the invention in textual form and as illustrations in a patent application. Typically, such a patent application is drafted by a patent attorney, who has both a technical and legal background. It contains claims that define the desired scope of protection for the invention. A patent will only be granted in case these claims specify a subject-matter that is new and inventive over prior art. Hence, already in this early stage of drafting a patent application, it is beneficial to conduct a cursory prior art search to assess the granting chances and eventually to adapt the claim wording.

However, it can be challenging to retrieve relevant prior art, i.e., previous publications related to the invention in question. Other inventors who had the same idea before might have used different words to describe it. A simple keyword search is therefore not necessarily successful. In particular, computer-implemented inventions as non-tangible products are often described by generic terms such as “system”, “means”, “modules” etc. In order to better structure the entirety of patent applications and granted patents, the patent offices classify newly filed patent applications based on the field of invention. As a result, every published patent application and granted patent is labeled with one or several class codes. Such hierarchical classification schemes, e.g., the International Patent Classification (IPC), may later help to limit the search to the relevant field of invention.

After a patent application has been filed with a patent office, a technically skilled examiner assesses the patentability of the described invention. In particular, the examiner carries out a prior art search and evaluates whether the invention defined in the claims is new and inventive (i.e. not obvious) over the found prior art. In this regard, the examiner mostly cites older published patents or patent publications rather than books or conference papers as prior art. The main reasons for citing mostly other patents and patent applications is the large size of patent data (according to the European Patent Office (EPO) currently 110 million documents[1]), the standardized structure of the patent data in patent classes, and the unquestionable

[1] https://worldwide.espacenet.com

publication date of each patent document.

In case the examiner finds that the claimed invention is not new or not inventive, the applicant has the opportunity to further specify the invention defined in the claims. In this way, the claimed invention may be sufficiently delimited from the relevant prior art. However, the applicant must not add any new information to the originally filed application. In case the patent examiner can be convinced, a patent is granted.

Independent of the outcome of the examination, a patent application is published 18 months after its filing date. Accordingly, the applicant has to disclose its invention to the public even if finally no patent might be granted. In case a patent is granted, said granted patent is published as an additional document. Moreover, since the applicant may obtain patent protection for an invention in several countries, multiple national patent applications and patents with substantially the same content may be published. A granted patent may be enforced against a competitor's products in patent litigation proceedings. As a defence strategy, the competitor may attack the validity of the patent. In particular, the competitor usually tries to find relevant prior art, that has not been revealed during the examination proceedings. In case a novelty destroying prior art document is found, a patent can still be nullified in post-grant proceedings.

1.1. Patent Analysis Tasks

The described patent life-cycle describes a set of tasks that can be (at least) partially automated. These tasks are often summarized under the term "patent analysis". From the literature, we identified the most popular tasks for automatic patent analysis. They can be classified into:

1. Supporting tasks, such as pre-processing, extracting information for further analysis, or translating patents to other languages;
2. Patent classification, where patent documents are categorized hierarchically based on the field of invention;
3. Patent retrieval, which branches into prior art search, automated patent landscaping, infringement search, Freedom-to-Operate search, and passage retrieval;
4. Patent valuation and market value prediction, being an innovative research where content, bibliographic details of the patents are analyzed with certain human made protocols to analyze the quality of patent applications, this analysis is further used to add market value and solved as regression problems;
5. Technology forecasting, where patents are used to asses a technology landscape and which helps researchers to capture new or trendy technologies;
6. Patent text generation, where the structure and styles incorporated in published patent documents are used to automate the process of writing patent claims;
7. Litigation analysis, a legal process where potential patents lead to a dispute or litigation between any two companies by prohibiting the development of business strategies;
8. Computer vision tasks, which work with figures and drawings from patent documents instead of text.

1.2. Patent Offices

The ever growing number of patent applications leaves especially the patent offices in dire need for automation. The potential of deep learning or machine learning in general to automate some of the necessary processing was recognized by major patent offices and AI initiatives where established.[2] Table 1 lists some of the efforts undertaken by EPO, United States Patent and Trademark Office (USPTO), and World Intellectual Property Organization (WIPO). Research on deep learning for the patent domain is thus not only conducted by academia and industry but is actively fueled by the major patent offices by providing benchmark datasets, organizing challenges and workshops.

1.3. Related Surveys

To foster interdisciplinary research, a couple of surveys on machine learning in the patent domain have been published in the past. Joho et al. [43] provide a descriptive analysis of the patent literature and list various requirements of patent search and functionalities of patent users adapted during analysis. Their survey also identifies the demographics of patent specialists and their relation in performing patent search tasks. Based

[2] https://www.wipo.int/about-ip/en/artificial_intelligence/search.jsp

Table 1
Major patent offices have recognized the potential of deep learning and started initiatives to foster its application.

Task	EPO	USPTO	WIPO
Patent classification	Automatic pre-classification, re-classification using CPC scheme	Concept questioning using chat bots to assist automatic claim analysis and classification	Automatic categorization based on IPC main class, sub-class or main group
Prior art search	Gold standard generation, automatic annotation, search, and query generation	Search whole document against corpus of grants and pre-grants	Cognitive or semantic search using AI assisted tools
Data analysis	Develop open source libraries to explore data and technology trend analysis	Browser-based advanced patent analytics	Economic and strategic analysis
Patent examination	Automatic annotation and exclusion detection	Patent quality assurance using text analytics, advance data analytics for statistics description	Advance big data analysis to automate examination workflow
Image analysis	Automatic image and figure search	Image search for patents and trademarks	Image search within global brand database, patent search using images

on the number of patent practitioners and approaches utilized in literature, they outline an ideal patent search system. Abbas et al. [1] summarize text-mining and visualization-based approaches used for patent analysis. Patent collections are rich in structured and unstructured text content which requires intelligent tools to accomplish efficient patent analysis. The study focuses on a taxonomy of tools and techniques to retrieve and analyze the data. The study also discloses the drawbacks associated with approaches that utilize only semantics-based techniques. Data mining techniques for patent analysis were surveyed by Zhang et al. [101], who critically point out technical issues associated with patent mining tasks. For classification, visualization, search, and evaluation tasks, not only technical issues but also solutions proposed in related work are comprehensively discussed. The authors also identified challenges for end-user applications related to patent mining. A recent survey reports on intellectual property analysis as a special type of data science [6]. It refers primarily to non-deep-learning machine learning approaches and introduces four main categories of analysis: knowledge management, technology management, information extraction, and economic value prediction. Another survey [88], which focuses on patent retrieval, quantitatively compares different approaches using benchmark datasets. The retrieval methods discussed are all query reformulation methods and are grouped into: keyword-based, pseudo-relevance feedback, semantics-based, metadata-based, and interactive methods. The retrieval tasks and datasets are described in-depth but deep learning approaches to tackle this task are not presented.

To the best of our knowledge, no survey focuses on deep learning approaches for different patent analysis tasks. Therefore, with this article, we intent to fill this gap and present an overview of datasets, text representation methods, and deep neural network architectures used for various patent analysis tasks.

2. Datasets

Training data are an essential ingredient for machine learning in general and for deep learning in particular. Publicly available datasets facilitate access and thus research in the patent domain. Table 2 gives an overview which papers make use of which data collection. The different patent datasets are listed in Table 3. For a more detailed description of available datasets and benchmark collections, see, e.g., [88] or [66].

Table 2
Datasets and the papers that make us of them. NAT (national: Chinese, Japanese, or Russian patents), COLL (curated collection: NTCIR or CLEF-IP), REP (KISTA or EPO search reports), LEGAL (post-grant documents: lawsuits, litigation)

Dataset	Used in
USPTO	[102, 32, 14, 3, 61, 58, 81, 59, 87, 39, 36, 23, 57, 82, 49, 54, 41, 56, 80, 18, 45, 104, 75, 34, 24, 55]
WIPO	[81, 5, 2, 82, 34]
EPO	[5, 41, 34]
NAT (national)	[47, 48, 103, 70, 67, 60, 34, 105, 63]
COLL (curated collections)	[58, 40, 86, 2, 41]
REP (reports)	[23, 45, 104, 79]
LEGAL (post-grant documents)	[61, 94]

Table 3
Sources of Patent Collections

Dataset	Collection	Source
USPTO	Bulk	`https://bulkdata.uspto.gov/`
	USPTO 2M	`http://mleg.cse.sc.edu/DeepPatent/index.html`
WIPO	Bulk	`https://www.wipo.int/classifications/ipc/en/ITsupport/Categorization/dataset/`
EPO	Bulk	`https://www.epo.org/searching-for-patents/data/bulk-data-sets.html`
NAT	Chinese	`https://oversea.cnki.net`
	Japanese	`https://www.j-platpat.inpit.go.jp/`
	Russian	`https://new.fips.ru/publication-web/?lang=en`
COLL	NTCIR	`http://research.nii.ac.jp/ntcir/data/data-en.html`
	CLEF-IP	`http://www.ifs.tuwien.ac.at/~clef-ip/download-central.shtml`
REP	KISTA	`http://biz.kista.re.kr/patentmap/`
	Search EPO	`https://www.epo.org/searching-for-patents/legal/register.html`
LEGAL	Litigation	`http://lexmachina.com`; `http://www.patexia.com/`; `http://legal.thomsonreuters.com`
	Various USPTO	`https://developer.uspto.gov/data?search_api_fulltext=csv`

2.1. Granted Patents and Patent Applications

The large patent offices, namely USPTO and EPO, are providing various datasets next to the actual granted patents and applications. There are also specific collections serving as benchmark datasets, such as USPTO2M, which comprises the titles, abstracts, and class labels of 2 million granted patents and contains around 235 million tokens, available in JSON format. With 5 million full-text patent publications from 1976 to 2016, USPTO5M is an even larger dataset, which contains around 38 billion tokens. WIPO released excerpts extracted from patent documents in several datasets in XML format. Its most recent dataset, WIPO-delta-en, is from 2019 and comprises an unprecedented set of 55 million excerpts. EPO has patents in English, French, and German and publishes various data products for research. Other national patent offices also publish patent data, e.g., there has been research conducted involving documents from the Japanese, Chinese, and Russian patent offices.

2.2. Curated Collections

To promote computer science research in the patent domain, different initiatives were founded. They provide a forum for researchers by defining shared patent analysis tasks, curating patent collections, and organizing workshops:

CLEF-IP organized shared tasks between 2009 and 2013 and thereby promoted several patent analysis tasks, in particular, cross-language patent analysis [74].[3]

NTCIR was the first initiative focusing on patent retrieval. Patents were a major focus since early on

[3] `http://www.clef-initiative.eu/`

(2001), while in later years, tasks shifted from retrieval to machine translation of patents [64].[4]

TREC-CHEM organized a shared task to identify chemical structures in patent documents from 2009 to 2011 [65].[5]

2.3. Reports

Besides patent collections, there are also a couple of related documents that are explored in the context of patent analysis. Among them are trend and search reports, landscaping reports, or general reports about trending technologies.

KIPO Reports: In addition to the patent offices already mentioned, the Korean Intellectual Property Office (KIPO)[6] is also active in patent analysis research. Along with patent documents, KIPO releases yearly trend reports (KISTA[7]) and landscaping reports (KIPRIS[8]).

Gartner's Hype Cycle: Gartner's Hype Cycle for Emerging Technologies[9] describes technology trends and classifies them into different stages based on maturity. This can be used as ground truth for trend prediction [104].

EPO Search Reports: Search reports published by patent offices are a valuable resource especially for patent retrieval research. In particular, to make sure an invention is novel and non-obvious, search reports help by providing citations to the existing similar inventions [62]. These inventions may be described in patents or pending applications. The search reports can be utilized as a ground truth resource for training machine learning models for prior art search. They play an important role in the patenting process and document how the prior art search was conducted and what relevant documents were found by the examiner.[10] According to the rich format citation standard of EPO and WIPO, the relevant passages are cited and the citation is categorized for each found prior art document. The categories indicate in which way the cited documents are relevant to the patent application's claims, such as technical background (A), novelty-destroying (X), or inventive step (Y).[11]

2.4. Post-Grant Documents

A further interesting data source origins from post-grant proceedings from courts and offices. Such proceedings include office proceedings (e.g., opposition proceedings in Europe or post-grant review and inter partes review (IPR) in the US) and court proceedings at national courts (e.g., the German federal court of justice). Post-grant office proceedings give third parties the possibility to challenge the validity of a granted patent, e.g., to request revocation due to novelty destroying prior art which has not been found by the patent examiner. The documents originating from these proceedings may thus complement the search reports from the pre-grant examination proceedings. One example for these kinds of documents can be found in Rajshekhar et al. [77], who provide a data set of PTAB prior art references extracted from USPTO rulings.[12]

Post-grant court proceedings mostly concern litigation proceedings due to an alleged patent infringement. Accordingly, these documents relate to the question whether a contested product of a third party infringes a patent. Documents originating from court proceedings comprise court decisions (also referred to as case law) and party submissions (including evidence). Meanwhile, decisions are often publicly available (e.g., on a court's website[13]), while party submissions are usually not published and only available under specific conditions (e.g., in Germany due to §299 ZPO). There also exist datasets related to patent litigation proceedings. For example, USPTO provides a dataset with detailed patent litigation data on more than

[4] http://research.nii.ac.jp/ntcir/index-en.html
[5] https://trec.nist.gov/data/chem-ir.html
[6] https://www.kipo.go.kr/
[7] http://biz.kista.re.kr/patentmap
[8] http://www.kipris.or.kr/khome/main.jsp
[9] Understanding Gartner's Hype Cycles https://www.gartner.com/en/documents/3887767
[10] https://www.epo.org/applying/european/Guide-for-applicants/html/e/ga_c5_2_3.html
[11] https://www.epo.org/law-practice/legal-texts/html/epc/2013/e/ar54.html
[12] https://data.world/wzadrozn/ptab-prior-art
[13] https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/list.py?Gericht=bgh&Art=en&Sort=3

81,000 unique district court cases filed between 1963 and 2016.[14] The data include a variety of information, including parties and attorneys, cause of action, location, and litigation history. In addition, commercial services exist that offer litigation data, which is used by researchers (see Table 3).

2.5. Copyright Protection

Machine learning approaches rely on the availability of training data. Complex deep learning models require large amounts of data to learn and thus copyright issues need to be considered. The patent domain seems to be suited very well to deploy deep learning methods. On the one hand, there are huge amounts of publicly available patent documents, that are often annotated by domain experts, e.g., with regard to the international patent classification (IPC) scheme. This saves additional, expensive, manual labeling of the data for training purposes. On the other hand, published patents and patent applications are generally excluded from copyright protection. That is to say, since such patent documents are published by a patent office, they have a public domain status which shall not be subject to any copyright claims. According to this general rule, anyone can reproduce published patent documents, at least as long as the content is not altered and the source is correctly cited, i.e., the patent (application) number.

However, there may exist exceptions to this general rule. Potential copyright claims may concern very different aspects in a patent document. Just to give some examples, ownership might be claimed for drawings of a third party, for specific boilerplate paragraphs used by a law firm, or for third-party text or drawings incorporated in a patent application which itself was protected by copyright. Moreover, even if there exist several multilateral international copyright treaties, copyrights basically constitute national rights. Hence, copyright protection can vary between the national jurisdictions (i.e., between the single countries). For example, any copyright claims with regard to a European patent application would be subject to the individual national laws of the member states of the European Patent Convention. A thorough analysis of the different national laws regarding potential copyright protection of patent documents goes beyond the scope of this survey. For this reason, only a brief summary of the regulations in some selected jurisdictions is given in the following which confirm the above-mentioned general rule.

For example, in Germany and Switzerland patent documents (as published by the patent office) are explicitly excluded from copyright protection (Section 5(2) German UrhG, and Art. 5 d. Swiss CopA). In the UK patent specifications published after 1989 may be reproduced for the purpose of "disseminating information", but other uses are prohibited without a license from the copyright holder.[15] In the USA, the text and drawings of a patent are typically not subject to copyright restrictions according to the USPTO,[16] even if there are limited exceptions reflected in 37 CFR 1.71(d) & (e) and 1.84(s). At least for US patent applications it is thus possible to include a copyright notice or mask work notice (cf. 37 CFR 1.71 (d) and (e)). However, inclusion of such a copyright notice or mask is only allowed if the applicant acknowledges at the same time that facsimile reproduction (i.e., photocopies) are allowed.

3. Representations

Patent documents consist of text data, metadata, such as citation information, and sometimes image data. These data need to be pre-processed into a suitable (vector) representation to allow deep learning models to use the data as input. Table 4 lists the various representations, which can be divided into representations learned by deep learning methods, such as word2vec [69], and "raw" input representations, as used primarily in traditional machine learning models.

Deep learning models can use raw input directly to learn a semantically meaningful representation given enough training data. The imprecise, labor-intensive step of extracting features is then unnecessary. These learned representations already contain valuable information so that they can be directly used for certain tasks and don't require further processing by deep learning methods. Examples are clustering of patents based on their learned representations, or visualizing patent similarity by using dimensionality reduction [68] of the embedding space. Especially embeddings of words, i.e., the representation of words as dense vectors, can also be used as input for traditional machine learning methods [102, 103, 39, 5, 36, 49, 94, 34]

[14] https://www.uspto.gov/about-us/news-updates/patent-litigation-data-through-2016-now-available
[15] https://webarchive.nationalarchives.gov.uk/20140603113132/http://www.ipo.gov.uk/types/copy/c-other/c-other-faq/c-other-faq-type/c-other-faq-type-patspec.htm
[16] https://www.uspto.gov/terms-use-uspto-websites

Table 4
Representations and the papers that make use of them

Representation			Used in
NUM (numeric features)			[3, 61, 59, 104]
CIT (citation networks)			[59, 61, 70, 75]
IMG (image data)			[48, 60, 41]
WE (word embeddings)	domain-specific	word2vec	[58, 81, 39, 5, 57, 86, 18]
		fastText	[81, 82, 18]
		integrated	[47, 40]
	general-purpose	word2vec	[102, 32, 14, 3, 61, 59, 103, 5, 67, 36, 23, 2, 49, 94, 75, 34, 24, 63]
		fastText	[23, 2, 49]
		GloVe	[2, 18]
DE (document embeddings)			[14, 87, 36, 23, 45, 63]
GE (graph embeddings)			[61, 59, 23, 75]
CTX) (contextual word embeddings)			[54, 23, 56, 55, 79]

Before deep learning revolutionized machine learning in many areas, extracting and selecting appropriate *features* was a major task and crucial processing step for traditional machine learning approaches. Especially text and image data needed to be transformed into features that could be used by methods such as support vector machines or decision trees. This type of features is still valuable when available, e.g., as metadata or citation information. However, in the context of deep learning, the raw input text or image data is typically represented by learned representations, usually called embeddings, which automates the cumbersome feature extraction and selection process. It is also possible to combine traditional features with embedding representations, either by combining the different input data or by combining the representations. In the deep learning for patent analysis literature, there are three types of traditional features used: numerical, citation, and raw image features. Alternatively and more common for deep learning approaches are different embeddings to represent the input data.

3.1. Numerical Features

These are quantities that can be extracted either manually or automatically from patent data. Examples are citation counts [61, 104], dates [61], number of claims [59, 104] or other numeric quantities derived from metadata [61, 59, 104], but also categorical features represented as one-hot encoded vectors, e.g., references [3] or class codes [3, 61].

3.2. Citation Networks

Another valuable data source for patent analysis are the references. They can be extracted and a citation network can be constructed. This network contains a lot of interesting information which can be combined with other data, such as texts or images, within deep learning models. The citation network can be used directly as input for further processing, e.g., to detect communities [70], or as input to learn a representation using graph embedding methods [61, 59, 75].

3.3. Image Data

Deep learning models do not require the extraction of features beforehand. This makes deep learning very successful when it comes to image data. Here, the extraction of salient features is especially cumbersome and making use of deep learning to learn suitable representations is extremely beneficial. As input serve the raw pixels from images of sizes between 100 x 100 and 300 x 300 [48, 60, 41].

3.4. Word Embeddings

Representing words in a meaningful way, in contrast to one-hot encoded bag-of-words vectors, is one of the reason for the success of deep learning for natural language processing. Word embeddings [13] are dense vectors that are learned from large text collections by looking at the context of each word. The most employed embedding method (not only) in the patent domain is word2vec [69]. It learns vector representations of words in an unsupervised fashion, i.e., by training on a large collection of documents, words with a similar meaning get assigned a similar vector. There are two other word embedding methods used for patents, namely fastText [9] and GloVe [71]. Instead of using pre-trained embeddings, it is also possible to integrate this step into a larger deep learning framework and learn embeddings implicitly [47, 40]. Further, one can distinguish the approaches based on the document collection that was used to learn the word vectors. Domain-specific embeddings were trained on a collection of patents or related documents, while general embeddings were trained on huge amounts of general text, e.g., on Wikipedia and news articles. These pre-trained general word vectors can be downloaded from the Web.[17] Domain-specific word vectors have the advantage, that they contain also representations of highly domain-specific words. This is especially important in the patent domain. Word vectors trained on millions of patent documents are also available for download[18] and can be used directly to represent words in the patent domain for downstream tasks. Alternatively, one can learn domain-specific representations by training these methods on selected document collections. Table 4 lists the different methods used in the corresponding papers.

3.5. Document (Sentence/Paragraph) Embeddings

When it comes to representing phrases, sentences, paragraphs, or whole documents, various methods have been proposed. The simplest approach takes the average of the word embedding vectors of a text, ignoring the ordering of the words in the text. Given that the meaning of a sentence is more than the sum of the meaning of its words, this simple approach cannot capture subtle semantic differences on a sentence level. More sophisticated approaches, such as doc2vec [51], try to capture not only the semantics of individual words, but of longer text parts (sentences, paragraphs, or documents). With the upcoming of more complex encoder neural network architectures, a variety of sentence embeddings [25, 15, 78] has been developed. In general, sentence embeddings outperform word embeddings on natural language processing downstream tasks [15]. Their disadvantage is the complexity of the underlying models and thus their need for larger training datasets and more computational power.

3.6. Graph Embeddings

The idea of representing text data as semantically meaningful vectors has also been adopted to represent graph data [72, 33]. Besides these general approaches to embed graphs [12], researchers in the patent domain came up with their own methods to embed graph data prevalent in patent documents, e.g., in the shape of citation networks [61, 59, 75], or adopted other approaches [84, 23].

3.7. Contextual Word Embeddings

The most recent deep learning methods to represent textual data are context-dependent word embeddings [28]. The most popular one of these methods is BERT [27], which is based on a transformer architecture [95]. It learns individual word vectors for each word in a sentence dependent on the other words in the sentence. Alternative approaches are ULMFiT [38] or ELMo [73], and, especially for text generation, the different versions of GPT [11]. On the one hand, these contextual word embedding models have billions of parameters and need a lot of training data. On the other hand, they can be trained in parallel and they represent textual data very accurately and thus make it easy for downstream tasks, e.g., patent classification, to produce great results by just adding one more (fully connected, dense) layer on top of the representation-generating deep neural network. This step is called fine-tuning and has been applied to many NLP-tasks very successfully: you take a pre-trained (general) BERT model and fine-tune it by training the exchanged last layer on task-specific data. Google has recently released[19] a BERT model pre-trained on over 100 million patent publications from the U.S. and other countries [83]. Close to the patent domain is LEGAL-BERT [16], which was pre-trained on court cases and legislation documents. Besides BERT, which becomes more and more popular in the patent domain [23, 54, 56, 79], a recent approach tries to use GPT to generate patent claims [55].

[17] word2vec: `https://code.google.com/archive/p/word2vec/`; fastText: `https://fasttext.cc/docs/en/crawl-vectors.html`; GloVe: `https://nlp.stanford.edu/projects/glove/`

[18] `https://hpi.de/naumann/projects/web-science/deep-learning-for-text/patent-classification.html`

[19] `https://cloud.google.com/blog/products/ai-machine-learning/how-ai-improves-patent-analysis`

Table 5
Deep learning architectures classified after the primary layer type employed and the papers that make use of them

DL Architecture	**Used in**
FC (fully connected network)	[3, 61, 23, 104, 75, 24, 63]
CNN (convolutional neural network)	[48, 61, 58, 59, 67, 57, 60, 2, 41, 75, 105, 24, 63]
RNN ((simple) recurrent neural network)	[40]
LSTM (long short term memory network)	[32, 47, 3, 87, 70, 86, 60, 80, 18, 24]
GRU (gated recurrent unit network)	[81, 82, 67, 105, 18, 63]
SEQ2SEQ (sequence-to-sequence network)	[40, 60, 80]
GAN (generative adversarial network)	[104]
AE (autoencoder network)	[45]
TRANS (transformer-based network)	[23, 54, 56, 55, 79]

4. Deep Learning Architectures

Before we dive into the different architecture types, we give a brief introduction to the basic concepts of deep learning. For a much more detailed introduction, we recommend Goodfellow et al. [30], Alom et al. [4] for a general introduction or Young et al. [99] for a more specific introduction to deep learning for NLP.

4.1. Deep Learning Basics

Deep learning can be seen as learning a transformation of the input data to the output data.[30] This is a major difference in comparison to traditional machine learning. Apart from that, deep learning shares the basic building blocks with the supervised machine learning process.

- Input data: In the patent domain, input data can be in the form of texts, images, or references from patent documents, but also from *external* sources, such as court documents or citation counts.
- Expected output data: The output is task specific: one or multiple, discrete or continuous variables. E.g., for a classification task, the outputs are different class labels, e.g., the classes that should be assigned to a patent.
- Model: A model, such as a complex neural network or a naive Bayes classifier, whose parameters are adjusted based on the input data and expected output data, is the final result of the learning process.
- Metric: A way to measure the progress of the model is necessary based on task-specific loss functions, e.g., the fraction of correct classifications.

Deep learning is a special kind of machine learning that uses multi-layer, artificial neural networks. *Learning* means that the parameters of the network, i.e., the weights of the layers, are modified in a way that a loss function is minimized for a set of training samples — the training dataset. The loss function needs to be differentiable so that stochastic gradient descent can be used to find a local minimum of the function. If the training data comprises labeled ground truth pairs of inputs and expected outputs, it is called supervised learning in contrast to unsupervised learning. The word *deep* in deep learning is related to the number of stacked layers that learn a hierarchical representation mainly in computer vision tasks. To summarize, the core principle adopted in neural networks is to learn representations from the given input data in a layered transformation which provides a mapping from input data to output data.

Deep learning methods evolved from more basic neural networks. While the story of artificial neural networks goes back more than 70 years, we focus on the most recent development, starting with convolutional neural networks. The basic artificial neural networks are similar to other traditional machine learning approaches, such as support vector machines or decision trees, since they require numerical (hand-crafted) features as input and they show similar performance and can be applied to similar tasks. Nevertheless, these basic neural networks set the ground for modern deep learning and there are several research papers making use of neural networks to solve patent analysis tasks [91, 19, 93, 50, 44, 53, 92, 104]. We briefly introduce the main deep neural network architectures in the following. For a more detailed introduction, we refer to the original papers for the individual architectures referenced in the corresponding subsections.

4.2. Deep Neural Network Architectures

Although the breakthrough of deep learning, especially in computer vision and in natural language processing, happened rather recently, the ideas and basic concepts of deep learning are older [85]. Today, the terms *deep learning* and *deep neural networks* typically refer to convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The difference of these two architectures is that they are tailored to recognizing different kinds of patterns in the data.

Deep neural networks can be used either to compute representations of the data, e.g., word embeddings or contextual word embeddings (BERT, etc.) as discussed earlier, or to solve a specific task. In this section, we focus on the architectures used to solve a task, i.e., to classify, predict, cluster, etc. A variety of deep learning architectures have been developed, which are tailored to different problem settings. Table 5 lists the most popular neural network architectures used in patent analysis.

Fully connected network: The most basic deep neural network architecture consists of dense layers (also called fully connected layers), i.e., each neuron in one layer is connected to each neuron in the previous layer. In the context of patent analysis, networks consisting only of dense layers are rarely used. More often, these layers are used to combine or bring together the output of other networks [61, 3, 23, 75, 24, 63].

Convolutional neural network: CNNs [52] are predominantly used for image classification because their architecture allows recognizing local, spatial patterns, e.g., for a group of neighbored pixels. But they can also be employed for one-dimensional data, e.g., text. CNNs have less parameters to train than more complex architectures such as RNNs, which makes them attractive also for textual data. In the context of patent analysis, CNNs were deployed to solve different tasks related to image data [48, 60, 41] as well as text data [61, 58, 59, 67, 57, 2, 75, 105, 24, 63].

Recurrent neural network: RNNs were developed to work with sequences of data (time series, text sequences). They can memorize earlier parts of a sequence for the classification of later parts. They are typically either based on long short-term memory (LSTM) cells or gated recurrent units (GRUs), but there is also a simple version lacking these complex extensions. Simple RNNs are used, e.g., for machine translation of patents [40].

Long short term memory network: Neural networks based on LSTMs [37] are the most popular RNNs and are frequently used for patent analysis tasks [32, 47, 3, 87, 70, 86, 60, 80, 18, 24].

Gated recurrent unit network: Compared to LSTMs, networks using gated recurrent units have a smaller number of parameters. From an application point of view, the difference to LSTMs is often negligible, with comparable performance. This results in GRUs being also often used for patent analysis [81, 67, 82, 18, 105, 63].

Sequence-to-sequence network: SEQ2SEQ network architectures were developed in the context of machine translation [90, 21]. These architectures can learn a "translation" from generic input to output data and are therefore not limited to machine translation. A sequence-to-sequence network consists of two components: an encoder and a decoder. In the context of patent analysis, SEQ2SEQ architectures are used in various contexts [40, 80, 45]. Some of the architectures allow to choose suitable networks for the encoder and the decoder parts. One approach uses, e.g., a CNN to encode image data and then an LSTM to decode it to the output text data [60].

Generative adversarial network: GANs consist of two sub-networks: a generative network, which generates candidates and a discriminative network, which aims to distinguish generated candidate samples from real data [31]. Roughly speaking, the two sub-networks compete during the training process and thereby improve each other's parameters. One major area of application for GANs is the generation of authentic looking, but artificially generated data. In the context of patent data, GANs have not been used for this task so far. One reason might be the low quality of the resulting texts. Even when solving the problem of the non-differentiable selection of the next token to generate a sentence with an RNN [100], the output of these GAN models is far away from genuine-looking texts. Besides

Table 6
Patent analysis tasks and the papers that addressed them

Analysis Task	Used in
SUP (supporting tasks)	[102, 47, 14, 103, 40, 39, 86, 18, 34]
CLASS (classification)	[32, 58, 81, 87, 67, 57, 2, 82, 56, 80, 105, 63]
RETR (retrieval)	[3, 5, 36, 23, 57, 49, 45, 75, 63, 79]
QUAL (quality analysis and market valuation)	[59, 24]
TECH (technology forecasting)	[70, 104]
GEN (data generation)	[54, 55]
LIT (litigation analysis)	[61, 94]
CV (computer vision)	[48, 60, 41]

generating texts, GANs can also be used to generate other types of data. There is an approach using GANs to generate the features of artificial samples to create more training data for standard machine learning approaches in the patent domain [104].

Autoencoder network: The main idea behind AEs [96] are to learn a condensed, lower-dimensional representation of the input data in an unsupervised fashion. To this end, the autoencoder tries to reconstruct the encoded data in a decoding step. It learns to represent the data in a way that keeps the reconstruction error minimal. In the patent domain, AEs have not been employed very frequently [45]. An extension to plain autoencoders are variational AEs [46] which allow to generate data, but haven't been employed for patent data yet.

Transformer-based network: Transformer models [95] have been developed in the context of machine translation and consist of an encoder and a decoder. As such, they form the foundation of contextual word embedding models, such as BERT [27], based on a transformer's encoder, or GPT [11], based on a transformer's decoder. These models learn powerful representations (therefore categorized as "representation" in this survey). By exchanging the last layer of BERT with a task-specific layer, these transformer-based network architectures can be trained on different tasks. This step is called fine-tuning. Since the underlying representations are so powerful, one fully connected layer on top is sufficient to get very good results. In the patent domain, researchers have fine-tuned BERT for different tasks [23, 54, 56, 79]. Some experiments have been conducted to use GPT to generate patent texts [55].

Besides these main concepts, there are a couple of further improvements and specializations. An important one is the attention mechanisms [7], which allows neural networks to learn which part of the input is most relevant for the desired output. A side effect of the attention mechanism is that the words on which the attention is placed can serve as an explanation for the network's output, e.g., to explain a classification decision. Attention can be used on top of convolutional or recurrent layers. It is especially popular in combination with sequence-to-sequence network architectures. The training of architectures based only on attention without an underlying LSTM or GRU layer can be better parallelized and thus enables the processing of more data in shorter time. Transformer-based networks make heavy use of attention allowing to learn huge models.

5. Patent Analysis Tasks

There are different patent analysis tasks that have been automated at least partially in the past. Table 6 lists different tasks together with the publications that propose deep learning methods to automate them.

5.1. Supporting Tasks

There are a couple of tasks that can be considered supporting tasks, in the sense that they produce results that can be used to analyze patents in a later step. Among these supporting tasks, there are three broader areas that were investigated in the context of patent data employing deep learning methods: *extraction* of information from patents, *segmentation* of patent documents into semantically meaningful smaller parts, and *translation* of patents from a foreign language.

Extraction: Named entity extraction is a very important task in various domains, e.g., from news articles or tweets. In the patent domain, automatically extracting chemical named entities [102] or biomedical named entities [86] is particularly interesting. Not only entity mentions can be extracted, but also the relations between a pair of entities [18]. Besides entities, extracting general keywords from patent texts can also be very useful, e.g., for classification [39].

Segmentation: Patents or patent applications are semi-structured documents consisting of different sections. If patents are not available in electronic form, this structure might get lost. Deep learning based representations can be used to segment large OCRed text into predefined sections [14]. But even within a large section, such as the "description" section, text can often be further segmented, primarily into the part describing the invention and the part describing experiments [34].

Translation: If patents are only available in a certain language, translating these patents is the first step to further analyze them. Given the huge success of deep learning methods in the area of machine translation, it comes with no surprise that there has been research focusing on patent texts in the context of translation [47, 40].

These supporting tasks are useful as a preparation step to then further analyze the results or use the results in subsequent steps, e.g., for classification.

5.2. Classification

Patent classification is the most prominent patent analysis task, where one needs to assign a classification code to a patent document based on IPC or CPC classification schemes. In practice, patent documents are analyzed manually and then the classification codes are assigned by the applicant and patent officers. These manual labeling tasks require domain expertise and are time-consuming. Besides traditional machine learning, deep learning techniques can be used for automatic patent classification. Since the classification schemes are hierarchical, different variants for the setting exists, e.g., only predicting the top-level class. This is the simplest version and in practice not very useful. More interesting settings require to prediction of the subclass up to a certain level. This setting was also used for large shared-task competitions (e.g., CLEF-IP 2010 [8]). Regarding the evaluation, there are also different variants possible. Given that a patent can have multiple subclasses assigned to it, three measures can be deduced [29]: The straightforward measure compares the top prediction with the main subclass assigned to the patent. Another measure compares the top three predictions with the main subclass. And the third measure compares the top prediction with the main class and the incidental subclasses assigned to the patent.

In the context of machine learning for patent analysis, classification is by far the most popular task. One reason for this is the availability of large quantities of training data, i.e., patent documents with assigned class labels. Another reason is the straightforward setting: given a document, predict the subclass codes. The deep learning approaches differ therefore only slightly, e.g., with respect to the data they use as input (abstract, claims, metadata, etc.) or the network architecture (GRU, LSTM, etc.). Some approaches also explicitly model the hierarchy of the classification codes [80]. Besides predicting classes based on a classification scheme, other classification tasks are possible when analyzing patent data. One approach tries to classify citations into applicant-provided or examiner-provided [63]. Another approach uses classification to train representations to improve clustering later on [75]. In general, predefined class labels can be used to learn semantically meaningful representations, since the class labels work as a very short summary of the patent itself [57].

5.3. Retrieval

Finding patents is important for a variety of reasons and with different intentions. In our description of the retrieval subtasks we mostly follow Shalaby and Zadrozny [88], who provide a good overview of patent retrieval tasks. In addition, we include passage retrieval and clustering as further subtasks, since there are a couple of papers dedicated to them.

The most obvious subtask is finding *prior art* for a given patent (application). But also finding patents related to a specific area or dealing with a specific topic, often called *landscaping*, is an important retrieval task. Finding not patents as such but particular sentences or paragraphs within patent documents is called *passage retrieval*. And finally, a more general, indirect retrieval task is *clustering* patents: for each patent, the most similar patents need to be identified.

Prior art search: Prior art (in other words state of the art or background art) is composed of all publicly available information that has been made accessible in any readable form prior to a given date that might be similar to a patent's claims. If prior art already describes the same invention as a newly filed application, then another patent on that respective invention cannot be granted. Patents mistakenly granted after the publication of such prior art can be revoked. There are different reasons for conducting prior art search: *Related work* search needs to be done by the patent applicant to find and list related patents. *Novelty detection* or patentability search is carried out by patent applicants, patent examiners, patent attorneys, patent agent professionals. They search in patents and patent application databases as well as in other scientific literature to identify the novelty of an invention. Novelty detection takes place before and after an inventor files a patent application. *Validity detection* tries to discover a prior art overlooked by the patent examiner in order to invalidate a patent. A validity check is carried out by patent infringement entities or patent owners. They search in patents and patent application databases, other scientific literature, technical society websites, and archives, usually after a patent was issue. *Infringement search* or *freedom-to-operate* are a special form of prior art search, where the purpose is to discover whether claims of patent applications and patents are infringed by any process or product. It is carried out by patent attorneys and professional patent searchers (often directed by attorneys), both in patents and patent application databases. This form of search is done before and after an inventor gets a grant.

Traditionally, the whole process of prior art search was carried out through expert-generated queries or term-based search methods. These approaches consume a lot of human labour, require domain expertise, and are also often associated with sub-optimal results. To alleviate these problems, several deep learning approaches were proposed [5, 36, 49].

Landscaping: Related to prior art search is automated patent landscaping [3]. Patent landscaping helps in finding technology related patents to avoid infringement issues and also to asses the trends in technology. The change in technology may lead to several implications towards business, economy, and policies. It has been complex and time consuming process to conduct a careful technology survey.

Abood and Feltenberger [3] proposed an approach to patent landscaping using 5.9 million USPTO patent abstracts, citations, and CPC codes. Patent documents are further used to generate seed sets by human experts, perform feature extraction, and create embeddings. Choi et al. [23] presented benchmark datasets from KISTA trend reports. They propose the use of graph embeddings based on metadata, such as USPC, IPC, or CPC codes. Another approach making use of citation information finds similar technology patents [63].

Passage Retrieval: The workshops of CLEP-IP 2013 and NTCIR 2005 provide datasets for passage retrieval. In this task, relevant passages (paragraphs) need to be retrieved from a patent document. For example, an instance of this task might consist of a patent application and a prior art patent. Only those passages from the patent shall be retrieved that are relevant to judge the novelty of the application's claims.

Since this task is very complex and even for experts very hard, not many have tried to automate this task. Only one very recent approach using EPO search reports to learn to match claims and paragraphs [79] has been proposed. It is based on contextual word embeddings and trained with positive and negative examples of matching paragraphs and claims.

Clustering: Clustering is the unsupervised grouping of patents based on a similarity measure. In contrast to classification, where class labels exist that can be learned, clustering does not need any labels. In the context of patent analysis, clustering is often used in combination with visualization methods, grouping similar patents close together in a vector space and then visualizing a 2-dimensional projection of this space.

It is often possible to use the semantically loaded embeddings directly to compute similarity between words, sentences, or documents [57]. The evaluation of clustering is then more difficult, since no ground-truth labels exist. Reports from patent offices can be used to this end, e.g., from KIPO [45]. Simple algorithms, such as k-means, can be used to cluster the patents based on similarity and the results can be visualized using dimensionality reduction methods [75], such as t-SNE [68].

5.4. Quality Analysis and Market Valuation

Analysing the quality of patents plays a major role in determining the economic value of a patent portfolio. A quality analysis and evaluation typically relies on domain expertise, technical knowledge and other factors, such as market and finance strategies. Such an analysis is of great interest for patent applicants, venture capitalist, policymakers, and business organisations. Although there are several metrics that act as consensus to measure the quality of an invention, having global metrics from major patent offices is challenging. Major offices, such as USPTO and EPO, approach the problem with custom metrics. For instance, USPTO proposed indicators with respect to a product, process and perception.[20] Others [35] identified forward citations in patents as a reliable indicator to detect the value of a patent. Other indicators include quality of the claims, family size of the patent, and the validity of the patent.

Approaches using deep learning are still very rare. One approach focuses on the citation network of a patent to assess its value [59]. Another approach tries to predict the number of forward citations as an indicator of patent value using abstract and claim text as well as hand-crafted features [24].

5.5. Technology Forecasting

Technology forecasting provides an opportunity for both, public and private enterprises, to predict upcoming technologies and make sure about their capital investment. Patent documents are a major source to base such predictions on. Technology forecasting started with unsupervised approaches, where text and data mining techniques were employed [17, 20]. Several approaches [22, 10] considered citation networks and Bayesian models for clustering to provide technology clusters. However, these unsupervised methods lack external domain knowledge and hence must incorporate domain experts interpretations in the end, which are time-consuming and costly.

One deep learning approach analyses citation networks and then tries to predict the number of future citations within a community using LSTM [70]. Zhou et al. [104] proposed an approach to augment training data using GANs to make up for a lack of annotated data. To this end, features are extracted using classical methods, then, synthetic data is generated resembling the extracted feature combinations, and finally, classical machine learning methods are trained on the augmented data.

5.6. Data Generation

The most progress of employing deep learning in the patent domain could be gained by automating the patent writing process itself. Advancements in language modeling and also GANs make it at least theoretically possible to generate patents or patent claims automatically given some seed information. Systems, such as GPT-3 [11] have demonstrated their capabilities in generating text. The question remains, how to improve the quality of the generated texts to make it interesting for the patent domain.

In preliminary work, Lee [54] proposed a transformer model for generating claims. He proposes to fine-tuned a GPT-2 model by personalizing the training data. In follow-up work [55], the proposed model was realized and patent claims actually generated. The reported results are still very poor, leaving a lot of room for improvement.

5.7. Litigation Analysis

Patent litigation is a legal process where potential patents lead to a dispute or litigation between any two companies by prohibiting the development of business strategies. This process often helps the companies to protect their profits and other proprietary values. The identification of patents that might induce litigation between companies is a tedious, costly and time-consuming process, which is carried out manually. It often

[20] https://www.uspto.gov/patent/initiatives/quality-metrics-1

comes with various intentions such as protection of market shares and product features but also fighting competitors (patent war). In the patent domain, several feature engineering-based approaches were proposed to automate the detection of litigation risk, e.g., using collaborative filtering [42]. Litigation risk is highly related to patent quality and is therefore sometimes considered to be one facet of it. We treat it as a separate task, reflecting the different type of input data that is necessary to asses litigation risk, namely legal documents in addition to patents.

Combining legal documents and patents can be done using deep learning methods. Liu et al. [61] made use of USPTO patents and Patexia lawsuits[21]) to train a model to predict the risk of litigation, which influences the value of a patent. To this end, the authors proposed to combine network embeddings learned from hand-crafted features and word embeddings followed by a CNN to learn a representation for patents. Tensor factorization is used to predict the probability of litigation based on this learned representation. Closely related to patent litigation is trademark litigation. One approach uses word embeddings to represent trademark case judgments [94]. Exploring the learned space can help to find relevant precedents. Clustering using k-means is further proposed to facilitate this search process.

5.8. Computer Vision

Inventors use figures, flowcharts or work-flows to depict their invention. Analysing such image data is a great challenge and specific tasks include *classification*, *image captioning*, and *image-based retrieval*.

Classification: A first step in analyzing images is the classification of what is depicted in the figure into more specific classes, such as technical drawing, chemical structure, sequence of gene, flowchart etc. The results can be used to improve search or enable faceted search for particular figure types.

One deep learning approach that tackles this task specifically for patents uses CNNs to classify patent images into different categories [48].

Captioning: Semantically understanding what is depicted in an image is a very active research field. One application is to generate captions automatically describing what can be seen on an image. If successful, these generated captions can be used for classification or retrieval, e.g., to find semantically similar images. Compared to standard image captioning, which is applied to photos, images in patents are much more difficult to describe meaningfully.

One team of researchers proposed an image captioning model using a combination of CNN and an LSTM [60]. The authors considered design patents to train an encoder-decoder model. The CNN is pre-trained on ImageNet [26] and is used to encode design patent images in 300-dimensional vectors. Further, the text descriptions of the images are encoded with word embeddings of the same dimensionality and a mapping from image features to word embeddings is learned.

Image-Based Retrieval: Image data are also used for patent retrieval based on the content of images. Especially when dealing with design patents, images are the best cue to find relevant patents. This field is called content-based image retrieval [97].

Deep learning models can be utilised to improve over classical approaches. One approach [41] uses a dual VGG network [89] to learn the representation of two images by minimizing the cosine distance of similar images as defined by the IPC class labels.

6. Literature Discussion

In this section, we summarize our findings on the literature discussed above. Table 7 gives an overview in matrix-style, highlighting the main ingredients of each identified related work article. The timeline on the left emphasizes the growing number of research papers using deep learning in the field of patent analysis.

Before the year 2016, automated patent analysis was conducted using traditional information retrieval and machine learning methods. While there are still some approaches using these traditional methods today, the vast majority of research in the patent domain happens in the field of deep learning. Deep learning models consistently outperform traditional approaches on *perceptive* tasks, i.e., tasks where semantic information,

[21] https://www.patexia.com/

Table 7
Survey Summary. NAT (national: Chinese, Japanese, or Russian patents), COLL (curated collections: NTCIR, CLEF-IP, or TREC-CHEM), REP (reports: Gartner, KISTA or EPO), LEGAL (post-grant documents), NUM (numeric features), CIT (citation networks), IMG (image data), WE (word embeddings), DE (document/paragraph embeddings), GE (graph/network embeddings), CTX (contextual word embeddings), FC (fully connected network), SEQ2SEQ (sequence-to-sequence network), AE (autoencoder network), TRANS (transformer-based network), SUP (supporting task: extraction, segmentation, or translation), CLASS (classification), RETR (retrieval: prior art search, landscaping, passage retrieval, or clustering), QUAL (quality and market valuation), TECH (technology forecasting), GEN (data generation), LIT (litigation analysis), CV (computer vision: captioning, classification, or image-based retrieval)

		Dataset							Representation							Architecture									Analysis Task							
		USPTO	WIPO	EPO	NAT	COLL	REP	LEGAL	NUM	CIT	IMG	WE	DE	GE	CTX	FC	CNN	RNN	LSTM	GRU	SEQ2SEQ	GAN	AE	TRANS	SUP	CLASS	RETR	QUAL	TECH	GEN	LIT	CV
	Zhang et al. [102]	x										x													x							
2017	Grawe et al. [32]	x										x							x							x						
	Kinoshita et al. [47]				x							x							x						x							
	Kravets et al. [48]				x						x						x															x
	de Carvalho and Nguyen [14]	x										x	x												x							
2018	Abood and Feltenberger [3]	x							x			x				x			x								x					
	Liu et al. [61]	x						x	x	x		x		x		x	x														x	
	Li et al. [58]	x				x						x					x									x						
	Risch and Krestel [81]	x	x									x								x						x						
	Lin et al. [59]	x							x	x		x		x			x											x				
	Zhong et al. [103]				x							x													x							
	Jehl and Riezler [40]					x						x						x			x				x							
	Shalaby et al. [87]	x											x						x							x						
	Nakai et al. [70]				x					x									x										x			
	Hu et al. [39]	x										x													x							
	Aras et al. [5]		x	x								x															x					
2019	Lyu and Han [67]				x							x					x			x						x						
	Saad [86]					x						x							x						x							
	Helmers et al. [36]	x										x	x														x					
	Choi et al. [23]	x					x					x	x	x	x	x								x			x					
	Liu et al. [60]				x						x						x		x		x											x
	Lei et al. [57]	x										x					x									x	x					
	Abdelgawad et al. [2]		x			x						x					x									x						
	Risch and Krestel [82]	x	x									x								x						x						
	Krishna et al. [49]	x										x															x					
	Lee [54]	x													x									x						x		
2020	Jiang et al. [41]	x		x		x					x						x															x
	Lee and Hsiang [56]	x													x									x		x						
	Risch et al. [80]	x										x							x		x					x						
	Chen et al. [18]	x										x							x	x					x							
	Kim et al. [45]	x					x						x										x				x					
	Trappey et al. [94]							x				x																			x	
	Zhou et al. [104]	x					x		x							x						x							x			
	Qi et al. [75]	x								x		x		x		x	x										x					
	Habibi et al. [34]	x	x	x	x							x													x							
	Zhu et al. [105]				x							x					x			x						x						
	Chung and Sohn [24]	x							x			x				x	x		x									x				
	Lu et al. [63]				x							x	x			x	x			x						x	x					
	Lee and Hsiang [55]	x													x									x						x		
	Risch et al. [79]			x			x								x									x			x					

either from natural language or image data, plays an important role. Given the complexity of patents, especially compared to other textual data, such as news articles, product reviews, or tweets, this advantage becomes even larger. Deep learning methods are able to capture semantics much better and allow for a much more fine-grained analysis, e.g., in prior art search or classification. This comes with the cost of typically requiring much more *annotated* training data. This is especially true for the large contextual word embedding models having billions of parameters that need to be learned. One way to cope with this is *transfer learning*: Instead of training solely on the in-domain training data, these complex models can be trained on general text to capture common semantics, and then only require to be fine-tuned on domain-specific training data. This concept is responsible for the large success of contextual word embeddings, such as BERT, ELMo, and ULMFiT, for domain-specific tasks.

We already discussed the different ingredients of the considered papers. Nevertheless, we want to briefly summarize the insights with regard to Table 7. While most research was and still is considering USPTO patents for their experiments, patents from other countries are not ignored. The advantage of English language patents, apart from their commercial international importance, is that there are already pre-trained models available for English and it is easier to compare with other approaches when they report results on similar data. Further, the new possibilities of deep learning fosters the combination of patent documents and non-patent literature, such as court documents or reports. We expect to see more of this especially for technology forecasting and quality and market valuation.

The vast majority of approaches uses word embeddings as a deep learning method to represent patent documents. Some approaches combine word embeddings with other representations, such as graph embeddings or document embeddings. Only a very small number of papers deal with non-textual input. This is either citation information or image data. Numeric or discrete input data, e.g., metadata of patents, is rarely used and if, then in combination with embeddings. This is consistent with the promise of deep learning not requiring feature engineering and being able to extract the crucial information from the raw (textual) input data.

The employed deep learning architectures are more diverse. The classic network architectures, CNN and LSTM, are the most popular for conducting patent analysis research, but the more complex architectures are gaining momentum. ENDEC and GAN architectures are more specialized architectures and therefore not suited for all tasks. Nevertheless, we expect to see more of those architectures, especially for difficult tasks. The large contextual word embedding models represent data extremely accurately and therefore only require simple dense neural network layers to fulfill their tasks. Often, these kind of layers are also used to combine two architectures and different input data.

Classification is the most popular patent analysis task. In its basic form, it is also the easiest task and the task with the most annotated training data, given that all published patents have classes assigned to them. In addition, evaluating classifiers is rather simple since there are plentiful ground truth datasets available. Patent retrieval with all its subtask is also very popular. The remaining identified tasks have been investigated only sparely so far. One reason for this is the more difficult nature of these tasks. When even human experts do not agree on a solution, or the task requires a lot of common sense and background knowledge, then automatic methods still have a hard time. However, we expect that more complex deep learning methods will be able to handle these difficult tasks better in the future.

7. Trends and Conclusions

Currently, there is a trend towards training more complex neural network architectures with larger number of parameters and thus a need for larger training datasets. After bi-directional encoder representations with transformers (BERT) [27] and XLNet [98] with 340 million parameters, GPT-3 [11] pushed the limit to 175 billion parameters. Many architectures build on the underlying attention mechanism using transformer models [95]. These large models reveal their full potential for few-shot and zero-shot learning [76], e.g., with label embeddings [80]. It is a challenge to access and handle the required amounts of training data, e.g., more than 181 billion English words [11]. Thus, we see another trend to reduce the amount of training data with the help of transfer learning and training on auxiliary tasks to reduce the need for labeled task-specific training data. This development shows a direction from supervised to self-supervised/semi-supervised training.

For the specific research direction of patent analysis with deep learning, we envision new tasks. Patent text generation is a rather new task and there are also almost no deep learning approaches for passage retrieval. The reason for this is the complexity of the task [79], which makes it much harder in comparison to standard document classification and requires other evaluation scenarios, e.g., a ranking of documents. Further, instead of full documents, subsections of documents need to be matched. Finally, there are no standard labels or annotations available. The extraction of information from search reports that can be used as labels is challenging.

Another new task that we expect to become more relevant is litigation analysis. While the vision of an artificial intelligence that can handle the entire patent life-cycle (AI patent lawyer) is emerging on the horizon, today's machine-learned models can still be easily fooled if targeted. To the best of our knowledge, there is no work on adversarial attacks on, e.g., image classification or text classification, in the patent domain — yet. We are confident that more semi-automated applications will be developed in research and will eventually find their way into industry in the near future.

With this survey, we summarized existing approaches which make use of deep learning for a variety of patent analysis tasks. While research in this area is still in its early stages, we outlined current trends of using various deep learning methods. We anticipate a shift in automated patent analysis away from classical machine learning to more and more deep learning. Further, we gave an overview of the available datasets for supervised learning needed by these methods. We hope that our work fosters interest in deep learning for patent analysis and serves as a comprehensive survey for researchers and practitioners from academia and industry.

References

[1] Abbas, A., Zhang, L., Khan, S.U., 2014. A literature review on the state-of-the-art in patent analysis. World Patent Information 37, 3–13.

[2] Abdelgawad, L., Kluegl, P., Genc, E., Falkner, S., Hutter, F., 2019. Optimizing neural networks for patent classification, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 688–703.

[3] Abood, A., Feltenberger, D., 2018. Automated patent landscaping. Artificial Intelligence and Law 26, 103–125.

[4] Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A., Asari, V.K., 2019. A state-of-the-art survey on deep learning theory and architectures. Electronics 8, 1–66.

[5] Aras, H., Türker, R., Geiss, D., Milbradt, M., Sack, H., 2018. Get your hands dirty: Evaluating word2vec models for patent data, in: Proceedings of the Posters and Demos Track of the International Conference on Semantic Systems (SEMPDF), pp. 1–4.

[6] Aristodemou, L., Tietze, F., 2018. The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. World Patent Information (WPI) 55, 37–51.

[7] Bahdanau, D., Cho, K., Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate, in: Proceedings of the International Conference on Learning Representations (ICLR), pp. 1–11.

[8] Beney, J., 2010. Lci-insa linguistic experiment for clef-ip classification track, in: Proceedings of the Conference and Labs of the Evaluation Forum (CLEF), pp. 1–11.

[9] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics (TACL) 5, 135–146.

[10] Breitzman, A., Thomas, P., 2015. The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems. Research policy (RP) 44, 195–205.

[11] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language models are few-shot learners. ArXiv e-prints arXiv:2005.14165.

[12] Cai, H., Zheng, V.W., Chang, K.C.C., 2018. A comprehensive survey of graph embedding: Problems, techniques, and applications. Transactions on Knowledge and Data Engineering 30, 1616–1637.

[13] Camacho-Collados, J., Pilehvar, M.T., 2018. From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research 63, 743–788.

[14] de Carvalho, D.S., Nguyen, M.L., 2017. Efficient neural-based patent document segmentation with term order probabilities, in: Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), pp. 171–176.

[15] Cer, D., Yang, Y., Kong, S.y., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., et al., 2018. Universal sentence encoder. arXiv preprint arXiv:1803.11175 .

[16] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I., 2020. LEGAL-BERT: The muppets straight out of law school, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing: Findings (EMNLP), pp. 2898–2904.

[17] Chang, C.K., Breitzman, A., 2009. Using patents prospectively to identify emerging, high-impact technological clusters. Research evaluation 18, 357–364.
[18] Chen, L., Xu, S., Zhu, L., Zhang, J., Lei, X., Yang, G., 2020. A deep learning based method for extracting semantic information from patent documents. Scientometrics 125, 289–312.
[19] Chen, Y.S., Chang, K.C., 2010. Exploring the nonlinear effects of patent citations, patent share and relative patent position on market value in the US pharmaceutical industry. Technology Analysis & Strategic Management 22, 153–169.
[20] Chiavetta, D., Porter, A., 2013. Tech mining for innovation management. Technology Analysis & Strategic Management 25, 617–618.
[21] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734.
[22] Choi, S., Jun, S., 2014. Vacant technology forecasting using new bayesian patent clustering. Technology Analysis & Strategic Management (TA&SM) 26, 241–251.
[23] Choi, S., Lee, H., Park, E., Choi, S., 2019. Deep patent landscaping model using the transformer and graph embedding. ArXiv e-prints `arXiv:1903.05823`.
[24] Chung, P., Sohn, S.Y., 2020. Early detection of valuable patents using a deep learning model: Case of semiconductor industry. Technological Forecasting and Social Change 158, 120–146.
[25] Conneau, A., Kiela, D., Schwenk, H., Barrault, L., Bordes, A., 2017. Supervised learning of universal sentence representations from natural language inference data, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 670–680.
[26] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255.
[27] Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 4171–4186.
[28] Ethayarajh, K., 2019. How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 55–65.
[29] Fall, C.J., Törcsvári, A., Benzineb, K., Karetka, G., 2003. Automated categorization in the international patent classification, in: Proceedings of the Special Interest Group on Information Retrieval (SIGIR), pp. 10–25.
[30] Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press.
[31] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), pp. 2672–2680.
[32] Grawe, M.F., Martins, C.A., Bonfante, A.G., 2017. Automated patent classification using word embedding, in: Proceedings of the International Conference on Machine Learning and Applications (ICMLA), pp. 408–411.
[33] Grover, A., Leskovec, J., 2016. node2vec: Scalable feature learning for networks, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), pp. 855–864.
[34] Habibi, M., Rheinlaender, A., Thielemann, W., Adams, R., Fischer, P., Krolkiewicz, S., Wiegandt, D.L., Leser, U., 2020. Patseg: a sequential patent segmentation approach. Big Data Research 19-20, 100–133.
[35] Harhoff, D., Scherer, F.M., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Research policy 32, 1343–1363.
[36] Helmers, L., Horn, F., Biegler, F., Oppermann, T., Müller, K.R., 2019. Automating the search for a patent's prior art with a full text similarity search. PLOS ONE 14.
[37] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.
[38] Howard, J., Ruder, S., 2018. Universal language model fine-tuning for text classification, in: Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), pp. 328–339.
[39] Hu, J., Li, S., Yao, Y., Yu, L., Yang, G., Hu, J., 2018. Patent keyword extraction algorithm based on distributed representation for patent classification. Entropy 20, 104–123.
[40] Jehl, L., Riezler, S., 2018. Document-level information as side constraints for improved neural patent translation, in: Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), pp. 1–12.
[41] Jiang, S., Luo, J., Pava, G.R., Hu, J., Magee, C.L., 2020. A CNN-based patent image retrieval method for design ideation. ArXiv e-prints `arXiv:2003.08741`.
[42] Jin, B., Che, C., Yu, K., Qu, Y., Guo, L., Yao, C., Yu, R., Zhang, Q., 2016. Minimizing legal exposure of high-tech companies through collaborative filtering methods, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), pp. 127–136.
[43] Joho, H., Azzopardi, L.A., Vanderbauwhede, W., 2010. A survey of patent users: An analysis of tasks, behavior, search functionality and system requirements, in: Proceedings of the Symposium on Information Interaction in Context (IIiX), p. 13–24.
[44] Kim, J., Lee, S., 2017. Forecasting and identifying multi-technology convergence based on patent data: the case of IT and BT industries in 2020. Scientometrics 111, 47–65.
[45] Kim, J., Yoon, J., Park, E., Choi, S., 2020. Patent document clustering with deep embeddings. Scientometrics 123, 1–15.
[46] Kingma, D.P., Welling, M., 2014. Auto-encoding variational bayes, in: Proceedings of the International Conference on

Learning Representations (ICLR), pp. 1–14.
[47] Kinoshita, S., Oshio, T., Mitsuhashi, T., 2017. Comparison of smt and nmt trained with large patent corpora: Japio at wat2017, in: Proceedings of the Workshop on Asian Translation (WAT), pp. 140–145.
[48] Kravets, A., Lebedev, N., Legenchenko, M., 2017. Patents images retrieval and convolutional neural network training dataset quality improvement, in: Proceedings of the International Research Conference on Information Technologies in Science, Management, Social Sphere and Medicine (ITSMSSM), pp. 287–293.
[49] Krishna, A., Jin, Y., Foster, C., Gabel, G., Hanley, B., Youssef, A., 2019. Query expansion for patent searching using word embedding and professional crowdsourcing. ArXiv e-prints arXiv:1911.11069.
[50] Kyebambe, M.N., Cheng, G., Huang, Y., He, C., Zhang, Z., 2017. Forecasting emerging technologies: A supervised learning approach through patent analysis. Technological Forecasting and Social Change (TF&SC) 125, 236–244.
[51] Le, Q., Mikolov, T., 2014. Distributed representations of sentences and documents, in: Proceedings of the International Conference on Machine Learning (ICML), pp. 1188–1196.
[52] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. IEEE 86, 2278–2324.
[53] Lee, C., Kwon, O., Kim, M., Kwon, D., 2018. Early identification of emerging technologies: A machine learning approach using multiple patent indicators. Technological Forecasting and Social Change (TF&SC) 127, 291–303.
[54] Lee, J., 2019. Patenttransformer: A framework for personalized patent claim generation, in: Proceedings of the JURIX Doctoral Consortium, pp. 1–13.
[55] Lee, J.S., Hsiang, J., 2020a. Patent claim generation by fine-tuning openai gpt-2. World Patent Information (WPI) 62, 101983.
[56] Lee, J.S., Hsiang, J., 2020b. Patent classification by fine-tuning bert language model. World Patent Information (WPI) 61, 101965.
[57] Lei, L., Qi, J., Zheng, K., 2019. Patent analytics based on feature vector space model: A case of iot. IEEE Access 7, 45705–45715.
[58] Li, S., Hu, J., Cui, Y., Hu, J., 2018. Deeppatent: patent classification with convolutional neural networks and word embedding. Scientometrics 117, 721–744.
[59] Lin, H., Wang, H., Du, D., Wu, H., Chang, B., Chen, E., 2018. Patent quality valuation with deep learning models, in: Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA), pp. 474–490.
[60] Liu, H., Dai, Q., Li, Y., Zhang, C., Yi, S., Yuan, T., 2019. The design patent images classification based on image caption model, in: Proceedings of the International Conference on Brain Inspired Cognitive Systems (BICS), pp. 353–362.
[61] Liu, Q., Wu, H., Ye, Y., Zhao, H., Liu, C., Du, D., 2018. Patent Litigation Prediction: A Convolutional Tensor Factorization Approach., in: Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI), pp. 5052–5059.
[62] Loveniers, K., 2018. How to interpret EPO search reports. World Patent Information (WPI) 54, 23–28.
[63] Lu, Y., Xiong, X., Zhang, W., Liu, J., Zhao, R., 2020. Research on classification and similarity of patent citation based on deep learning. Scientometrics , 1–27.
[64] Lupu, M., Fujii, A., Oard, D.W., Iwayama, M., Kando, N., 2017. Patent-related tasks at ntcir, in: Current Challenges in Patent Information Retrieval, pp. 77–111.
[65] Lupu, M., Huang, J., Zhu, J., Tait, J., 2009. Trec-chem: large scale chemical information retrieval evaluation at trec, in: ACM SIGIR Forum, pp. 63–70.
[66] Lupu, M., Piroi, F., Hanbury, A., 2010. Aspects and analysis of patent test collections, in: Proceedings of the International Workshop on Patent Information Retrieval, pp. 17–22.
[67] Lyu, L., Han, T., 2019. A comparative study of Chinese patent literature automatic classification based on deep learning, in: Proceedings of the Joint Conference on Digital Libraries (JCDL), pp. 345–346.
[68] Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-sne. Journal of machine learning research (JMLR) 9, 2579–2605.
[69] Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. ArXiv e-prints arXiv:1301.3781.
[70] Nakai, K., Nonaka, H., Hentona, A., Kanai, Y., Sakumoto, T., Kataoka, S., Carreón, E.C.A., Hiraoka, T., 2018. Community detection and growth potential prediction using the stochastic block model and the long short-term memory from patent citation networks, in: Proceedings of the International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1884–1888.
[71] Pennington, J., Socher, R., Manning, C., 2014. Glove: Global vectors for word representation, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543.
[72] Perozzi, B., Al-Rfou, R., Skiena, S., 2014. Deepwalk: Online learning of social representations, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), pp. 701–710.
[73] Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations, in: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 2227–2237.
[74] Piroi, F., Hanbury, A., 2017. Evaluating information retrieval systems on european patent data: The clef-ip campaign, in: Current Challenges in Patent Information Retrieval, pp. 113–142.
[75] Qi, J., Lei, L., Zheng, K., Wang, X., 2020. Patent analytic citation-based vsm: Challenges and applications. IEEE Access 8, 17464–17476.

[76] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., 2019. Language models are unsupervised multitask learners. OpenAI Blog 1, 9.

[77] Rajshekhar, K., Zadrozny, W., Garapati, S.S., 2017. Analytics of patent case rulings: Empirical evaluation of models for legal relevance, in: Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL), pp. 1–9.

[78] Reimers, N., Gurevych, I., 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3982–3992.

[79] Risch, J., Alder, N., Hewel, C., Krestel, R., 2020a. PatentMatch: A dataset for matching patent claims with prior art. ArXiv e-prints `arXiv:2012.13919`.

[80] Risch, J., Garda, S., Krestel, R., 2020b. Hierarchical document classification as a sequence generation task, in: Proceedings of the Joint Conference on Digital Libraries (JCDL), p. 147–155.

[81] Risch, J., Krestel, R., 2018. Learning patent speak: Investigating domain-specific word embeddings, in: Proceedings of the Thirteenth International Conference on Digital Information Management (ICDIM), pp. 63–68.

[82] Risch, J., Krestel, R., 2019. Domain-specific word embeddings for patent classification. Data Technologies and Applications (DTA) 53, 108–122.

[83] Rob Srebrovic, J.Y., 2020. Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery. Technical Report. Google. URL: `https://services.google.com/fh/files/blogs/bert_for_patents_white_paper.pdf`.

[84] Rozemberczki, B., Sarkar, R., 2018. Fast sequence-based embedding with diffusion graphs, in: Proceedings of the International Workshop on Complex Networks, pp. 99–107.

[85] Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

[86] Saad, F., 2019. Named entity recognition for biomedical patent text using bi-lstm variants, in: Proceedings of the International Conference on Information Integration and Web-based Applications & Services (iiWAS), pp. 617–621.

[87] Shalaby, M., Stutzki, J., Schubert, M., Günnemann, S., 2018. An LSTM approach to patent classification based on fixed hierarchy vectors, in: Proceedings of the SIAM International Conference on Data Mining (SDM), pp. 495–503.

[88] Shalaby, W., Zadrozny, W., 2019. Patent retrieval: a literature review. Knowledge and Information Systems (KAIS) 61, 631–660.

[89] Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: Proceedings of the International Conference on Learning Representations (ICLR), pp. 1–14.

[90] Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), pp. 3104–3112.

[91] Trappey, A.J., Hsu, F.C., Trappey, C.V., Lin, C.I., 2006. Development of a patent document classification and search platform using a back-propagation network. Expert Systems with Applications 31, 755–765.

[92] Trappey, A.J., Trappey, C.V., Govindarajan, U.H., Sun, J.J., 2019. Patent value analysis using deep learning models—the case of IoT technology mining for the manufacturing industry. Transactions on Engineering Management , 1–13.

[93] Trappey, A.J., Trappey, C.V., Wu, C.Y., Lin, C.W., 2012. A patent quality analysis for innovative technology and product development. Advanced Engineering Informatics 26, 26–34.

[94] Trappey, C.V., Trappey, A.J., Liu, B.H., 2020. Identify trademark legal case precedents—using machine learning to enable semantic analysis of judgments. World Patent Information (WPI) 62, 101980.

[95] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008.

[96] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A., Bottou, L., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research 11.

[97] Vrochidis, S., Papadopoulos, S., Moumtzidou, A., Sidiropoulos, P., Pianta, E., Kompatsiaris, I., 2010. Towards content-based patent image retrieval: A framework perspective. World Patent Information (WPI) 32, 94–106.

[98] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R., Le, Q.V., 2019. XLNet: Generalized autoregressive pretraining for language understanding, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), pp. 5753–5763.

[99] Young, T., Hazarika, D., Poria, S., Cambria, E., 2018. Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine 13, 55–75.

[100] Yu, L., Zhang, W., Wang, J., Yu, Y., 2017. Seqgan: Sequence generative adversarial nets with policy gradient, in: Proceddings of the Conference on Artificial Intelligence (AAAI), pp. 2852–2858.

[101] Zhang, L., Li, L., Li, T., 2015. Patent mining: A survey. SIGKDD Explorations Newsletter 16, 1–19.

[102] Zhang, Y., Xu, J., Chen, H., Wang, J., Wu, Y., Prakasam, M., Xu, H., 2016. Chemical named entity recognition in patents by domain knowledge and unsupervised feature learning. Database 2016.

[103] Zhong, Q., Qiao, X., Zhang, Y., 2018. Automatic indexing of patent right-claiming document based on deep learning, in: Proceedings of the International Conference on Applied Mathematics, Modelling and Statistics Application (AMMSA), pp. 135–139.

[104] Zhou, Y., Dong, F., Liu, Y., Li, Z., Du, J., Zhang, L., 2020. Forecasting emerging technologies using data augmentation and deep learning. Scientometrics , 1–29.

[105] Zhu, H., He, C., Fang, Y., Ge, B., Xing, M., Xiao, W., 2020. Patent automatic classification based on symmetric hierarchical convolution neural network. Symmetry 12, 1–12.