JANUARY 2024

Ph.D. in Industrial Engineering

SELEN YÜCESOY KAHRAMAN

REPUBLIC OF TÜRKİYE

GAZİANTEP UNIVERSITY

GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES

AUTOMATED PATENT CLASSIFICATION FROM THE PERSPECTIVE OF TECHNOLOGY MANAGEMENT: DEEP LEARNING METHODS AND TRANSFORMER MODELS

Ph.D. THESIS

IN

INDUSTRIAL ENGINEERING

BY

SELEN YÜCESOY KAHRAMAN

JANUARY 2024

AUTOMATED PATENT CLASSIFICATION FROM THE PERSPECTIVE OF TECHNOLOGY MANAGEMENT: DEEP LEARNING METHODS AND TRANSFORMER MODELS

Ph.D. Thesis

in

Industrial Engineering
Gaziantep University

Supervisor

Prof. Dr. Türkay DERELİ

Co-Supervisor

Assoc. Prof. Dr. Alptekin DURMUŞOĞLU

by

Selen YÜCESOY KAHRAMAN

January 2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Selen YÜCESOY KAHRAMAN

ABSTRACT

AUTOMATED PATENT CLASSIFICATION FROM THE PERSPECTIVE OF TECHNOLOGY MANAGEMENT: DEEP LEARNING METHODS AND TRANSFORMER MODELS

YÜCESOY KAHRAMAN, Selen
Ph.D. in Industrial Engineering
Supervisor: Prof. Dr. Türkay DERELİ
Co-Supervisor: Assoc. Prof. Dr. Alptekin DURMUŞOĞLU

January 2024

145 pages

Patents have become a widely used source of data for technology development, monitoring, and prediction due to their ability to disclose prior technologies and provide information about new technologies. Examining a patent application and assigning it to the appropriate class is a time-consuming and labor-intensive process for the patent expert. Automatic patent classification approaches can significantly help with this problem. The aim of this thesis is therefore to make a contribution to solving this problem on the basis of various leading studies. The thesis comprises four studies, which can be summarized as follows. The first study includes a bibliometric analysis of automatic patent classification. The second study evaluates patent texts using a hierarchical attention mechanism. The third study compares ensembled deep learning methods with the transformer encoder model. The last study investigates how the pre-trained Bert model addresses the challenge of automatic patent classification. It is worth mentioning that the studies conducted in this thesis aim to provide practical and useful solutions for automatic patent classification.

Key Words: Patent classification, HAN, transformer encoder, BERT, bibliometric analysis

ÖZET

TEKNOLOJİ YÖNETİMİ PERSPEKTİFİNDEN OTOMATİK PATENT SINIFLANDIRMASI: DERİN ÖĞRENME YÖNTEMLERİ VE TRANSFORMATÖR MODELLERİ

YÜCESOY KAHRAMAN, Selen
Doktora Tezi, Endüstri Mühendisliği
Danışman: Prof. Dr. Türkay DERELİ
İkinci Danışman: Doç. Dr. Alptekin DURMUŞOĞLU

Ocak 2024

145 sayfa

Patentler, önceki teknolojiyi açığa çıkarma ve yeni teknoloji hakkında bilgi sağlama yetenekleri nedeniyle teknoloji geliştirme, izleme ve tahmin için yaygın olarak kullanılan bir veri kaynağı haline gelmiştir. Bir patent başvurusunu inceleyerek uygun sınıfa atamak, patent uzmanı için zaman alıcı ve emek yoğun bir süreçtir. Otomatik patent sınıflandırma yaklaşımları bu sorunun çözülmesine önemli ölçüde yardımcı olabilir. Dolayısıyla bu tezin amacı çeşitli yol gösterici çalışmalardan yararlanarak bu sorunun çözümüne katkıda bulunmaktır. Tez aşağıdaki gibi özetlenebilecek dört çalışmadan oluşmaktadır. İlk çalışma, otomatik patent sınıflandırmasının bibliometrik analizini içermektedir. İkinci çalışma ise patent metinlerini hiyerarşik bir dikkat mekanizması kullanarak değerlendirmektedir. Üçüncü çalışma, birleştirilmiş derin öğrenme yöntemlerini transformatör kodlayıcı modeliyle karşılaştırmaktadır. Son çalışma, önceden eğitilmiş Bert modelinin otomatik patent sınıflandırmasının zorluğunu nasıl ele aldığını inceliyor. Bu tez aracılığıyla gerçekleştirilen çalışmaların, otomatik patent sınıflandırma faaliyetlerine yönelik pratik ve faydalı çözümler sunmasının beklendiğini belirtmekte fayda var.

Anahtar Kelimeler: Patent sınıflandırması, HAN, transformatör kodlayıcı, BERT, bibliometrik analiz

"Dedicated to my lovely family"

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof. Dr. Türkay DERELİ and Assoc. Prof Dr. Alptekin DURMUŞOĞLU for their guidance and support throughout the study. I am thankful for their encouragement and motivation.

I should express my special thanks to Assist. Prof. Dr. Yunus EROĞLU, Assist. Prof. Dr. Pınar Kocabey Çiftçi, Res. Asst. Özge VAR for their valuable guidance and support throughout this thesis.

I also would like to thank the other committee members of my thesis Prof. Dr. Serap ULUSAM SEÇKİNER, Prof. Dr. Serkan ALTUNTAŞ, Assist. Prof. Dr. Baki ÜNAL and Assoc. Prof. Dr Zeynep Didem UNUTMAZ DURMUŞOĞLU.

I would like to thank all people in the Department of Industrial Engineering for their support and help during my thesis. I also would like to thank to my friend Döndü EROĞLU for her support.

I also would like to thank my dear lovely family, and my husband's family for their support, prayers and encouragement.

Finally, my genuine, sincere and honest thanks and gratitude are dedicated to my dear husband Cihan KAHRAMAN for his heartfelt support, unfailing patience, and absolute love. I also would like to express my love and gratitude to my son Abdullah Emir KAHRAMAN who gives me moral and motivation with his existence.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page*

LIST OF SYMBOLS

$\mathbf{u_{it}}$	importance of words
$\mathbf{u_w}$	word-level meaning vector
$\boldsymbol{\alpha_{it}}$	normalized weight
$\mathbf{s_i}$	sentence vector
W	weight matrix
b	bias vector
t	time
xt	input vector
i	input gate
c	cell memory
f	forget gate
o	output gate

LIST OF ABBREVIATIONS

APC	Automated Patent Classification
IPC	International Patent Classification
CPC	Cooperative Patent Classification
USPTO	United States Patent and Trademark Office
USPC	United States Patent Classification
WIPO	World Intellectual Property Organization
ECLA	European Patent Classification
EPO	European Patent Office
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
BERT	Bidirectional Encoder Representations from Transformers
HAN	Hierarchical Attention Network
F-Term	File forming term
SVM	Support Vector Machine
kNN	k-Nearest Neighbor
XML	Extensible Markup Language
CLEF	Conference and Labs of the Evaluation Forum
ELECTRA	Pre-training Text Encoders as Discriminators
XLNET	eXtreme Language understanding NETwork,
TRIZ	Theory of Inventive Problem Solving
LDA	Latent Dirichlet Allocation
SBERT	Sentence-Bert
GPT	Generative Pre-trained Transformer
MLM	Masked Language Modeling
NSP	Next-Sentence Prediction
GRU	Gated Recurrent Unit
WOS	Web of Science

CHAPTER 1
INTRODUCTION

1.1 Motivation of Study

Commencing with the issuance of the inaugural patent in 1421, the proliferation of patented innovations has permeated every facet of human existence, encompassing technologies ranging from the printing press to magnetic resonance imaging systems. [1]. A patent may be described as a proprietary privilege conferred upon the inventor for a finite duration, which prohibits any unauthorized entity from manufacturing, utilizing, importing, or commercializing the patented innovation. It contains more than 80% technical information worldwide [2]. Patents are a valuable source of information for predicting high-tech research and development activities and technological progress. They provide electronic accessibility, reliability, information about new technologies, and explanations of the state of the art.

Researchers use the patent classification system when analyzing patents. The patent classification system serves to facilitate access to technological and legal status information contained in patent documents. These systems are also utilized in assessing incoming patent applications based on the criteria of “inventive step", "novelty” and “industrial application". Another purpose of the patent classification system is to speed up access to similar published patent documents and to determine the latest status of the investigated technology under examination. Relaying solely on a keyword-based search to find prior art make it difficult to identify some records. This is mainly due to the fact that patents are written in different languages, the same words may have different meanings in different languages, there are no keywords in English abstracts and words are misspelled. To address this issue, it is essential to utilize the patent classification system.

Today, there are different systems for patent classification. Some of the major classification systems used by patent offices: International Patent Classification (IPC),

the US Patent Classification (USPC) and the European Patent Classification (ECLA). While the IPC system divides technology into a total of 78,682 groups [3], the number of groups in the USPC system is more than 150,000 [4]. ECLA, the extended version of IPC, has a total of 134,000 groups. To avoid the confusion caused by different classification systems, the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO) combined the patent classification systems in 2013and created a new classification system called Cooperative Patent Classification (CPC). This new system brings together more than 250,000 technology groups under one roof[5], [6].

If patent experts cannot register a new invention in existing groups, they may choose to open a new group. Due to the newly added groups, the number of groups in the IPC has increased from 71,437[7] to 78,682 [8] since 2014. Experts constantly monitor these systems to keep their technical knowledge up to date. It can be assumed that the workload of experts will increase due to the regular updates to the systems. The increase in patent applications is thought to be another factor that increases the workload of experts. According to the World Intellectual Property Organization (WIPO), 3,457,400 patents were filed in 2022. This is 56,900 more than the number of applications in 2021 [3]. According to WIPO's statistical data from 2022, 3,982 patent experts working at the EPO were appointed to evaluate 193,610 European patent applications [3]. This means that an expert working at the European Patent Office will be assigned to examine an average of 48.62 patent applications in 2022. In the USA, one of the offices where the most patent applications are made, the average pendency time after demand of investigation made is up to 549 days[8]. Experts at patent office examine and evaluate incoming patent applications based on their own technical knowledge and experience. Evaluating an application by a patent expert, examining the patent classification categories and determining the appropriate patent classes for the application is a time-consuming and labor-intensive process due to the large number of groups in the existing patent classification systems.

1.2 Aim of thesis

This thesis was inspired by the goal of reducing the workload of patent experts. It seeks to develop a solution through various studies aimed at addressing one of the essential

technical problems. Automated classification systems used by experts need to be designed to provide more accurate results. The fact that the classification approaches used by these systems are supported by various analyzes in the scientific literature will increase their trustworthiness. The approaches outlined in this thesis aim to address the gaps in the field of automatic patent classification and contribute to the existing literature.

1.3 Roadmap of the Thesis

In light of the information given in the previous section, this thesis focuses on the solution of automatic patent classification problems. In this context, the roadmap of the thesis is presented to the readers in Figure 1.1.

Chapter 2 examines the technical background of Automated Patent Classification (APC) on the basis of statistical data. This section discusses the structure of patent texts, statistical data on patent applications worldwide, the features of automatic patent classification systems, the workload of patent experts, and the patent dataset used in this study.

Chapter 3 presents published work in the field of APC. The most commonly used dataset in the literature, keyword extraction and classification methods, patent sections utilized by the researchers, and the classification performances achieved are examined. Furthermore, the contributions of this thesis to the literature are listed.

Chapter 4 includes a bibliometric analysis of academic papers and patents focusing on automatic patent classification. A lot of data such as the leading researchers, journals and inventors in the field were examined. An overview of APC is provided, applying keyword analysis to both academic and patent texts.

Chapter 5 includes the study on the analysis of patent texts using the hierarchical attention network method. In this study, using the attention mechanism where the keywords are shown with the embedding vector, patent texts are classified at the subclass level. Attempts were made to increase the classification accuracy with various hyperparameter tests.

Chapter 6 contains a comparison of the classification result obtained by embedding the results of two different deep learning methods (CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory)) with the classification result obtained with the Transformer-Encoder model.

In Chapter 7, the study examines the solution offered by the pre-trained Bert algorithm (Bidirectional Encoder Representations from Transformers) to the problem APC. The study presents the use of different sections of the patent as input for classification and the testing of various hyperparameters.

Chapters 4, 5, 6, and 7 contain various approaches and studies addressing solutions to the problem of APC. These sections can be read separately, but it is recommended to read them in the order given to understand the context.

The results of the studies conducted as part of this thesis, the associated conclusions and suggestions for future studies are summarized in Chapter 8.

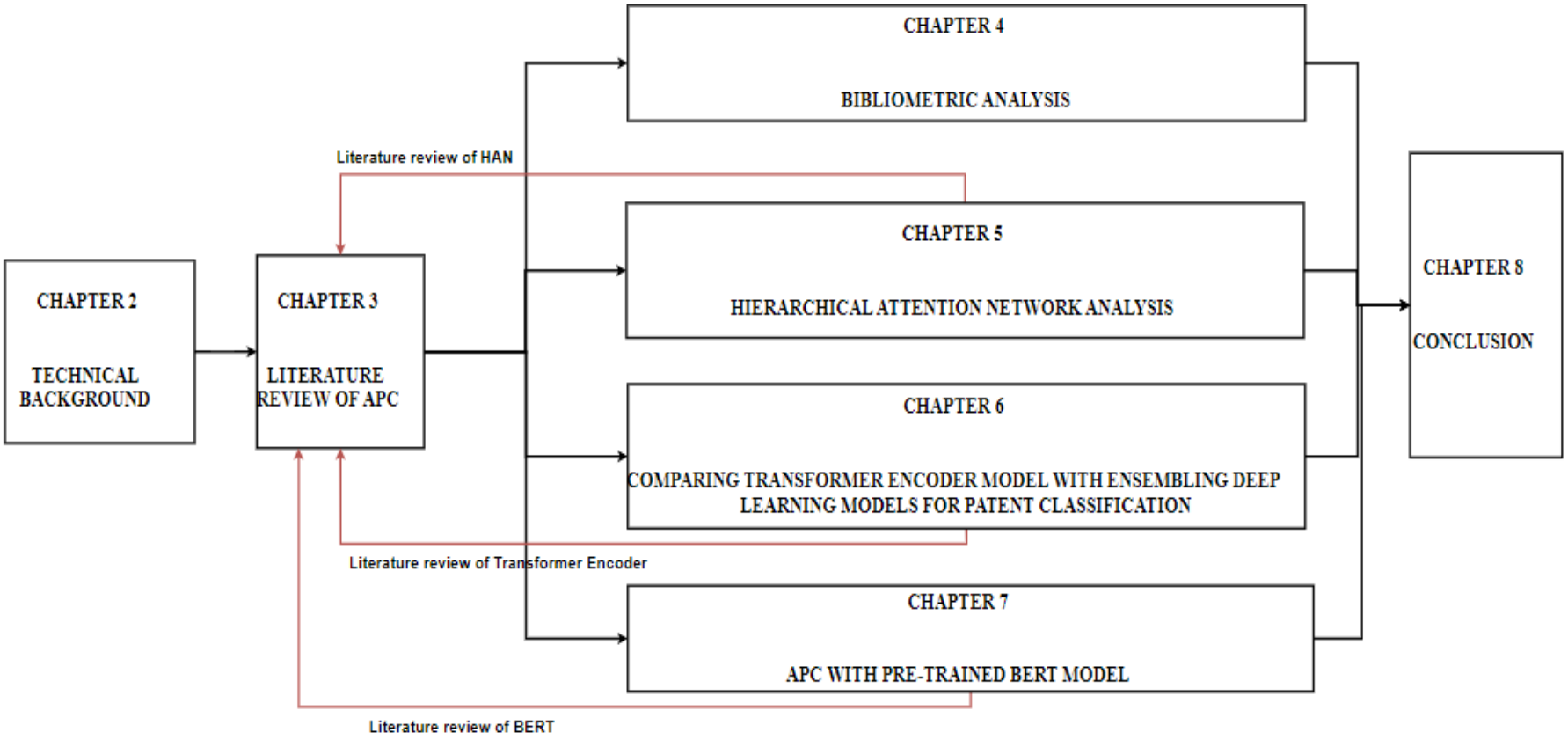


Figure 1.1. Roadmap of the thesis for the readers

1.4 Conclusion

All the studies conducted as part of this thesis are new to the literature. They have been presented at several conferences and have been deemed valuable enough to be published and presented in several reputable scientific journals. These studies contain valuable data that can be of use to anyone conducting technological research, innovators, and especially patent experts. It is expected that investment in new technologies will increase as access to accurate information becomes easier. Some of the considerations and methods presented in this thesis are also likely to inspire other scientists to pursue new fields of study. It is anticipated that the information gathered in this thesis will be valuable to everyone. In the hope that it will benefit both readers and humanity.

CHAPTER 2
TECHNICAL BACKGROUND OF AUTOMATED PATENT CLASSIFICATION

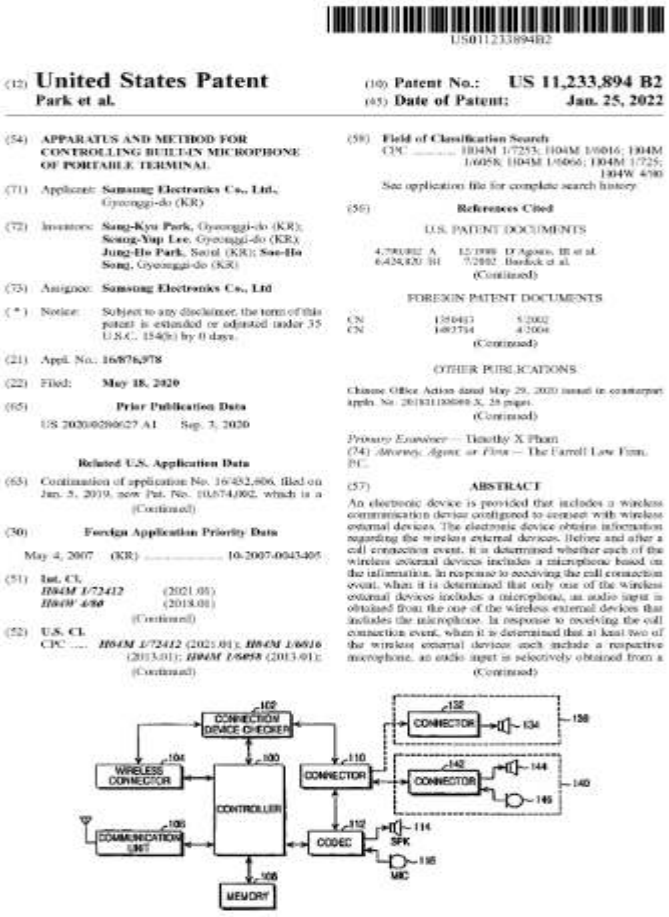
2.1 Introduction

One of the most important tools of the concept of intellectual property, which plays an important role in the concept of sustainable development, are patent documents. A patent is a right granted to the inventor that prevents others from making, using, importing or selling the invention without the inventor's permission. Patent information sources are accessible to any entrepreneur, innovator, or researcher by patent offices. Strong networks can be formed between investors, experts and patent owners who use patent information and innovative solutions can be offered for various technological problems. Patent documents contain quite long texts compared to other documents. For example, in the study by Kim et al., it was determined that the longest patent among the Japanese patents published in 1996 consisted of approximately 60,000 Japanese words [9].

2.2 Patent document structure

There is a standard format for the patent document in many countries. The first page (or cover page) of a patent document contains identifying information. This information includes bibliographic data and textual content. For example, the bibliographic data in a United States patent document includes the type of patent, the name of the applicant, inventor, assignee, application number, fill date, fields of classification search, national classification information, IPC, references cited, principle examiner name and attorney, agent, or firm name (if applicable). Text data includes information on title and abstract. Depending on whether there is space on the first page, a representative illustration of the drawing sheet may also be included. In addition, the other text data in the patent document includes the claims and description sections. An example of United States patent is shown in Figure 2.1.

The title, abstract, claims, and description sections of the patent include essential information for detailing the patent. The title specifies the name of the patent. The abstract is a technical section that summarizes the scope of the patented invention. It has no legal significance. As a result, it is not used when interpreting the scope of patent protection. The claims section explains the legal protection scope of the patent. In patent cases, interpreting the claims is the initial step in challenging the validity of the patent and identifying infringement. The description of the invention is a clear and detailed specification of the invention, presented in such a way that it can be easily understood and applied by those skilled in the art. It is generally recommended that the description should include information such as the technical field of the invention, the prior art, the purpose of the invention, the description of the figures, the explanation of the references in the figures, and the description of the invention[10].



US011233894B2

(12) **United States Patent**
Park et al.

(10) **Patent No.: US 11,233,894 B2**
(45) **Date of Patent: Jan. 25, 2022**

(54) **APPARATUS AND METHOD FOR CONTROLLING BUILT-IN MICROPHONE OF PORTABLE TERMINAL**

(71) Applicant: **Samsung Electronics Co., Ltd.**, Gyeonggi-do (KR)

(72) Inventors: **Sung-Kyu Park**, Gyeonggi-do (KR); **Seung-Yup Lee**, Gyeonggi-do (KR); **Jung-Ho Park**, Seoul (KR); **Soo-Ho Song**, Gyeonggi-do (KR)

(73) Assignee: **Samsung Electronics Co., Ltd**

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **16/876,978**

(22) Filed: **May 18, 2020**

(65) **Prior Publication Data**

US 2020/0280627 A1 Sep. 3, 2020

Related U.S. Application Data

(63) Continuation of application No. 16/452,806, filed on Jun. 5, 2019, now Pat. No. 10,674,002, which is a (Continued)

(30) **Foreign Application Priority Data**

May 4, 2007 (KR) 10-2007-0043405

(51) **Int. Cl.**
H04M 1/72412 (2021.01)
H04W 4/80 (2018.01)
(Continued)

(52) **U.S. Cl.**
CPC ..... ***H04M 1/72412*** (2021.01); ***H04M 1/6016*** (2013.01); ***H04M 1/6058*** (2013.01); (Continued)

(58) **Field of Classification Search**
CPC H04M 1/7253; H04M 1/6016; H04M 1/6058; H04M 1/6066; H04M 1/725; H04W 4/80
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,780,882 A 12/1988 D'Agosto, III et al.
6,424,820 B1 7/2002 Burdick et al.
(Continued)

FOREIGN PATENT DOCUMENTS

CN 1350403 5/2002
CN 1497734 4/2004
(Continued)

OTHER PUBLICATIONS

Chinese Office Action dated May 29, 2020 issued in counterpart Appln. No. 201811100909.X, 25 pages.
(Continued)

Primary Examiner — Timothy X Pham
(74) *Attorney, Agent, or Firm* — The Farrell Law Firm, P.C.

(57) **ABSTRACT**

An electronic device is provided that includes a wireless communication device configured to connect with wireless external devices. The electronic device obtains information regarding the wireless external devices. Before and after a call connection event, it is determined whether each of the wireless external devices includes a microphone based on the information. In response to receiving the call connection event, when it is determined that only one of the wireless external devices includes a microphone, an audio input is obtained from the one of the wireless external devices that includes the microphone. In response to receiving the call connection event, when it is determined that at least two of the wireless external devices each include a respective microphone, an audio input is selectively obtained from a
(Continued)

Figure 2.1. Example of first page of US patent 11,233,894

The patent document occasionally deviates from other documents. Patents are written in a language that differs from everyday language. While the patent abstract describes

the invention in clear and simple terms, the patent claims are written in technical language and contain the legal information necessary for the protection provided by a patent. In fact, these are considered the most essential parts that make up patent documents [11].

Each claim must be expressed in a single sentence. These sentences can be expanded into paragraphs with complex grammatical structures linking several sentences, leading to grammatical parsing issues [12]. In the study [13], the Stanford Natural Language Parser was used to analyze the patent claims and test the small corpus. They found significant variations among independent claims (e.g. Claim 1 in Figure 2.2 is more complex than the dependent claims (e.g., Claims 2, 3, 4, and 5 in Figure 2.2), which explain and expand the other claims. Parsing dependent claims is easier and generates more successful results than independent ones because dependent claims contain no more than 35 words. On the other hand, independent claims contained more than 120 words. Thus, the incorrect parse increased by 20-50% according to the heap size[11].

Another aspect that sets patent documents apart from other types of documents is the use of specialized language by assignees to protect their exclusive rights for a specific period of time. Common terms are frequently eliminated from patents and replaced with more specific words, making it challenging for researchers to identify relevant patents. On the contrary, many companies use more generic words to expand the reach of their patents as much as possible. Additionally, the patent aims to protect novel inventions. Thus, new technologies bring novel terms with them, and scientists researching similar new technologies at the same time may use various terms for those technologies. After the development of technology, it has its own jargon. Scientists begin to use standard terminology, but until then, they will generate different vocabularies. For those reason, making patent search by using all relevant keywords is time consuming and needs large investment of effort.

When reviewing patent documents, it is important to consider the unique language structures used in patents. Patent examiners typically do not rely solely on keywords, but instead use patent classification systems.

What is claimed is:

1. A fluorescent lamp system for lighting a space comprising:

a discharge lamp for producing ultraviolet light;

a phosphor coating external of the discharge lamp for converting ultraviolet light to visible light;

an ultraviolet reflective/visible light transmissive layer between the phosphor coating and the space to be lit; and

an outer jacket having an inner and an outer surface, the outer jacket surrounding the discharge lamp, the inner surface defined as the surface closer to the discharge lamp and the outer surface defined as the surface further from the discharge lamp.

2. The fluorescent lamp system of claim 1 wherein the phosphor coating is on the inner surface of the outer jacket.

3. The fluorescent lamp system of claim 1 wherein the phosphor coating is on the outer surface of the discharge lamp.

4. The fluorescent lamp system of claim 1 wherein the ultraviolet reflective/visible light transmissive layer is on the outer surface of the outer jacket.

5. The fluorescent lamp system of claim 1 wherein the outer jacket comprises:

an ultraviolet reflective/visible light transmissive layer.

Figure 2.2. Beginning of claims section from US patent 6348763, Fluorescent lamp luminaire system ".

2.3 Statistics of patent

Innovators field 3.45 million patent applications worldwide in 2022. This corresponds to an increase in the number of patent applications, 56900 more than in 2021. In 2019, there was a 3% decrease in the number of patent applications due to Covid19. But in the following years, the number of applications continued to rise [14].

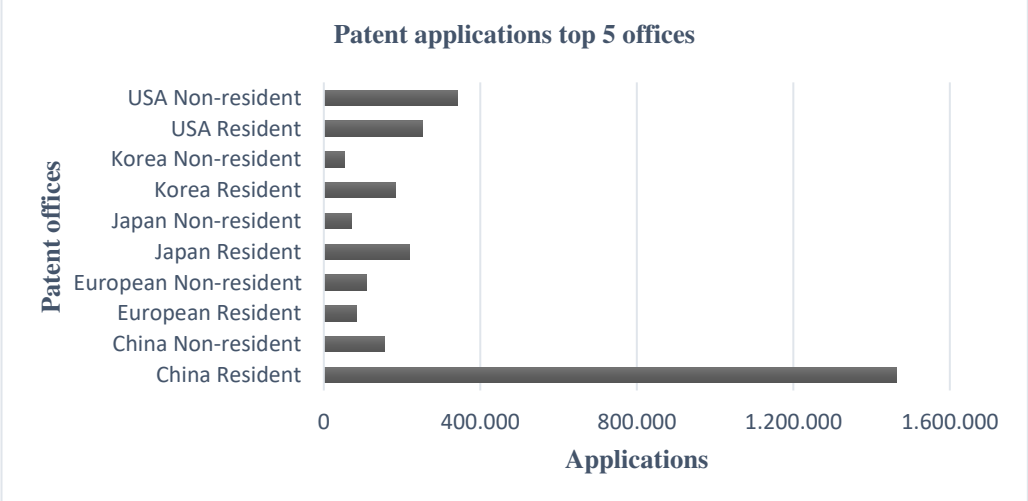


Figure 2.3. The top five offices with the most patent applications in 2022, Source: WIPO

Figure 2.3 shows that China accounts for almost half of the world's patent applications. China has more than twice as many applications as the USA and Japan combined. China received more than twice as many applications as the USA and Japan combined. It was followed by the USA, Japan, the Republic of Korea, and the EPO [14].

Patent offices are responsible for evaluating whether the claims in a patent application satisfy the patentability criteria for novelty, non-obviousness, and industrial applicability. Therefore, the patenting process, from application to registration, consumes both resources and time. However, the offices cannot conduct this examination without the applicant filling out an examination request form. This condition affects the duration of time that patent applications remain in the system. In 2022, there will be approximately 6.9 million pending patent applications worldwide (pending applications include all applications awaiting a decision from the patent office, regardless of the stage of the patenting process (including applications that do not yet have an examination request). This prediction is based on information gathered from 140 patent offices worldwide. For example, China had nearly 2.6 million pending applications in 2022 [14]. Figure 2.4 shows the average waiting time from the request for examination to the first office transaction for the 5 patent offices that receive the most applications. According to the 2020 data, the Chinese patent office employs 13704 patent experts. According to the 2020 data, the average number of patent applications examined by the patent expert working in the China office, which has the longest waiting period, is 109.2 days.

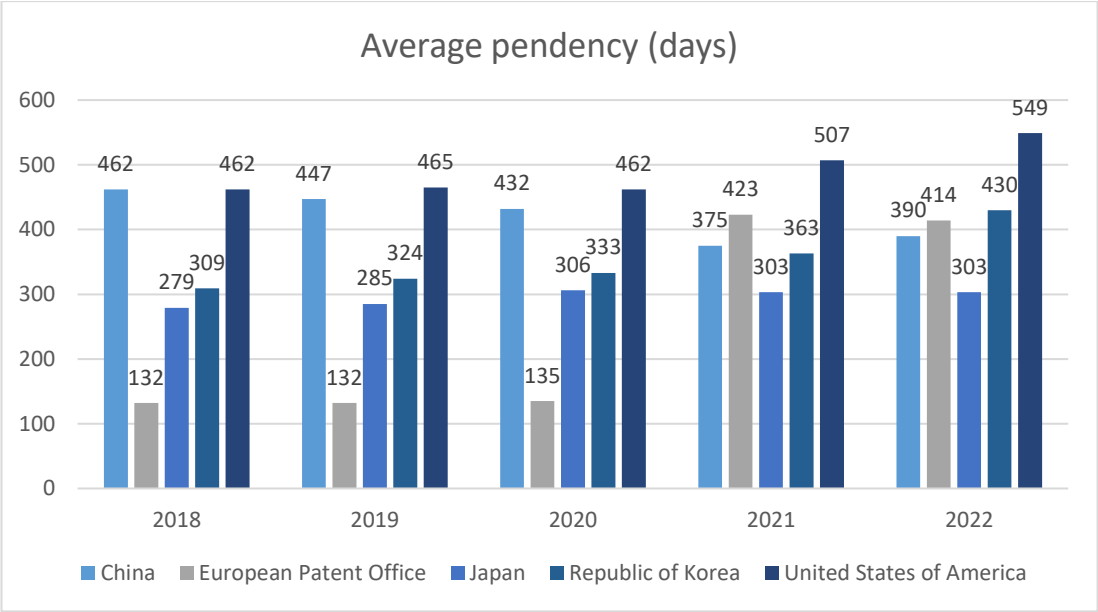


Figure 2.4. Average pendency of patent application from request for examination to the first office action (days)

2.4 International Patent Classification

Patents are classified into hierarchical systems of categories by patent office's according to the type of invention they represent. This thesis is concentrated on the IPC since it is the best-known and most widely spread classification system, it is used by over 140 patent-issuing bodies worldwide and contains around 78,000 classes [14]. In addition to that, most other important systems are directly based on it; they take the IPC hierarchy and add additional subclasses. Table 2.1 shows the major patent classification systems. Most systems contain between 110,000 and 330,000 classes and are relying on the IPC.

Table 2.1. Major patent classification systems

System abbreviation	Full Name	Number of Groups
IPC	International Patent Classification	78,682
USPC	United States Patent Classification	150,000
CPC	Cooperative Patent Classification	250,000
ECLA	European Classification	134,000
DPK	Deutsche Patentklassifikation	110,000
F-Term	File forming term	330,000
FI	File Index	190,000
DWPI	Derwent classification system	26,000

The IPC hierarchy for all versions is divided into eight sections that correspond to very general categories such as “Performing operations; transporting" (Section A) or “Textiles; paper" (Section D). Table 2.2 shows an overview of all eight sections of the IPC. Each section is made up of numerous classes (e.g., G06), and each class contains multiple subclasses such as G06F. Each subclass is again divided into main groups (e.g., G06F 16/00), and for most main groups there are additional subgroups such as G06F 16/35. Figure 2.5 presents an illustrative hierarchical representation of the IPC codes.

Table 2.2. Overview of all eight sections of the IPC

IPC Sections	Definition of Patent Sections
A	human necessities
B	performing operations; transporting
C	chemistry; metallurgy
D	textiles; papers
E	fixed constructions
F	mechanical engineering; lighting; heating; weapons; blasting
G	Physics
H	Electricity

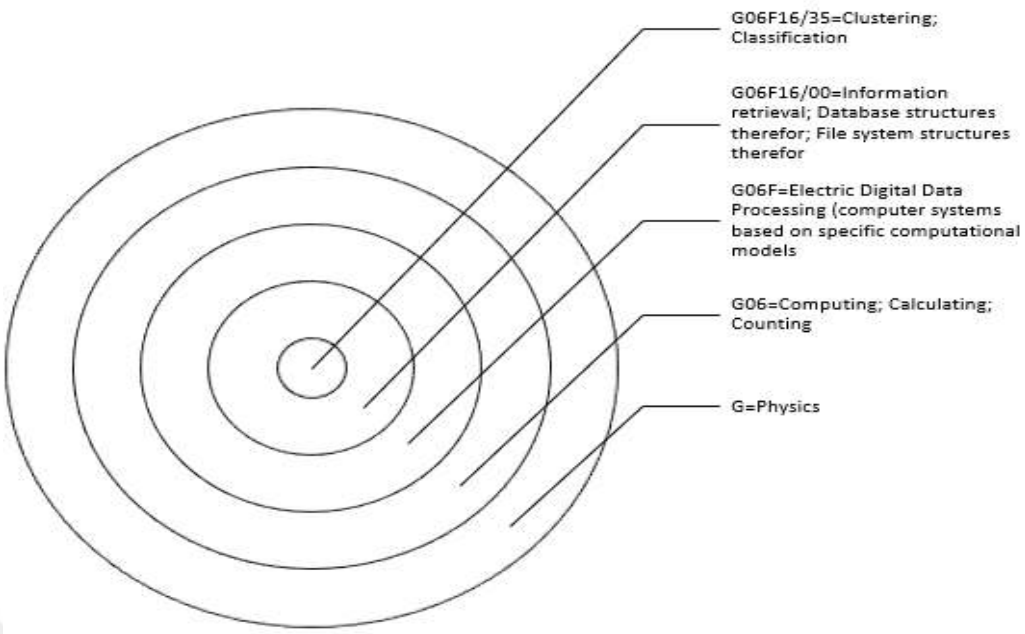


Figure 2.5. IPC codes description

2.5 Dataset used in this thesis

Wipo-alfa [15] is the first patent collection published in the field of automatic classification. This dataset, published in 2002, consists of approximately 75000 records containing more than 5000 categories down to the main group level. Wipo-alfa's 46324 records are from the training document and 28926 records are from the test document. WIPO-alpha documents are distributed in 8 sections, 114 classes, 451 subclasses and 4427 main groups. Each patent document in the dataset is saved in XML (Extensible Markup Language) format. An example XML document from the WIPO-alpha collection is given in Figure 2.6. In this record, the country information is recorded as "cy" in the record-tag (If it is a PCT document, the country information is entered as "WO"). The document's reference number (dnum), publication type (kind), application number and publication number are indicated as "an" and "pn", respectively. The prs-tag shows patent publication numbers and dates in different countries. The "title, abstract, claims and description" sections, which constitute the textual sections of the document, are shown with the tags "tis, abs, cls and txts", respectively. The main IPC code is indicated by ic under the "ipc" tag. IPC version information is indicated with "ed".

The training corpus consists of documents that are fairly evenly distributed across the main IPC groups, with each subclass limited to between 20 and 2,000 records. The test

collection consists of records that are approximately distributed based on the number of patent applications in a typical year (2001 was chosen for this purpose), with the restriction that each subclass contains between 10 and 1,000 records [21]. The training collection contains 46,324 patents, while the test documents contain 28,926 patents. There are a total of 75,250 documents.

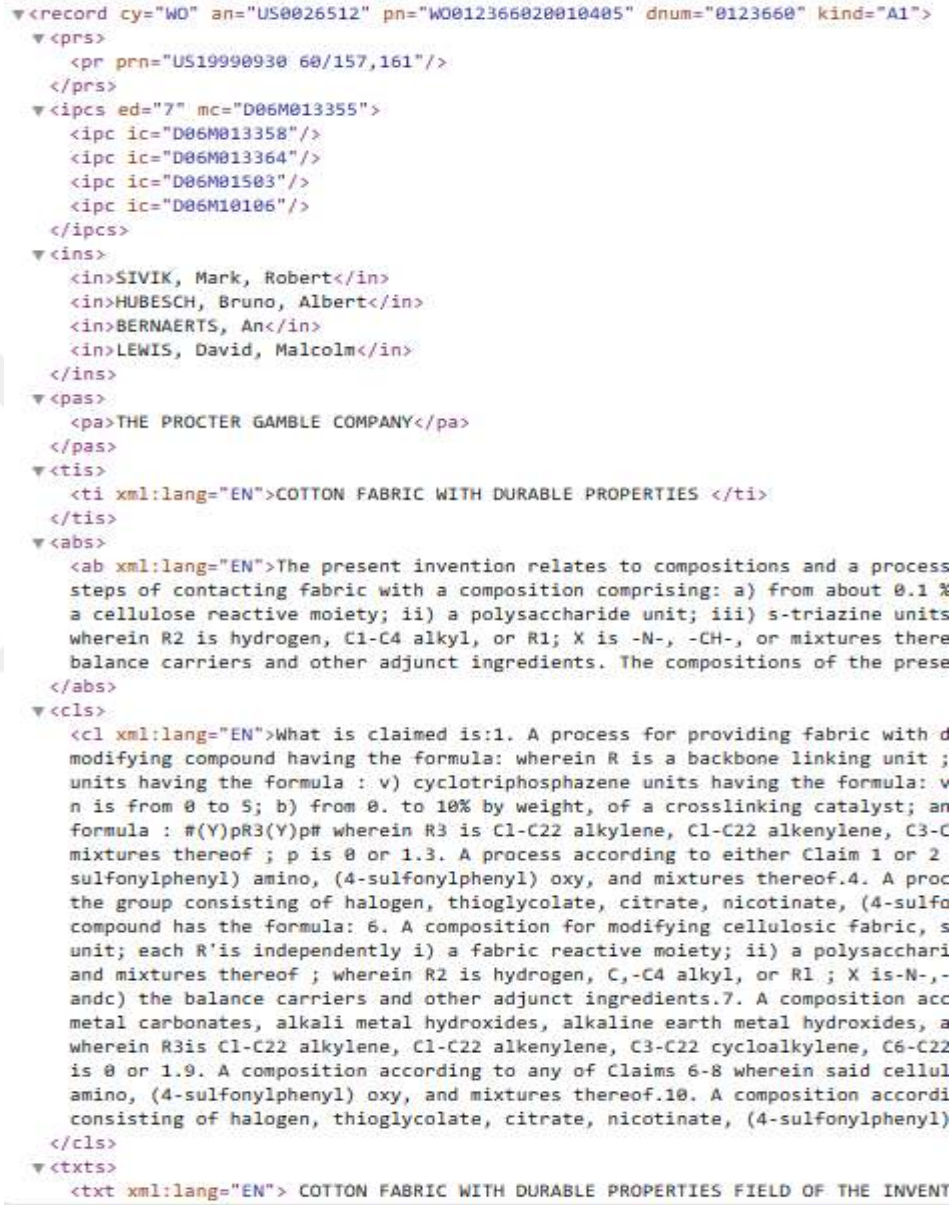


Figure 2.6. An XML example of WIPO-alpha dataset

2.6 Concluding Remark

This section provides general information and statistical data about patent texts. The semi-structural features of patent texts, the complexity of the technical information

they contain, and the difficulties in examining this information are explained in detail. Statistical information regarding patent classification systems and related numerical data regarding the workload of patent experts are provided. All this data aims to reveal the necessity of establishing automatic patent classification systems and the complexity that the hierarchical structure of the system brings to the classification system. Thus, based on this information, it is clear that the approaches, algorithms and various parameters included in the classification method used by automatic classification systems need to be adjusted precisely.

CHAPTER 3
LITERATURE REVIEW ON AUTOMATED PATENT CLASSIFICATION

3.1 Introduction

Over the past four decades, researchers in patent classification have experienced with different approaches. These studies initially used conventional classification algorithms but have evolved with advancements in information technology. Nowadays, research on deep learning algorithms underway. Due to the continuous updates of these technologies, algorithms can quickly become outdated after their introduction. If we consider the algorithm as a product, it can quickly reach its saturation in a very short period of time. Although this process may have taken longer in the past, technological advancements in information have begun to reduce the average time. Figure 3.1 illustrates the evolution of automated patent classification studies over time.

Figure 3.1 shows that the earlier two decades of patent classification research produced significantly fewer progress than the last twenty years. Handling vast amounts of data was a challenge due to advances in technology, and computer capacity was inadequate for the task. Machine learning methods have always been popular in studies on automatic patent classification. Although many methodologies have been developed, it is possible to see that these methods continue to be used even in studies published today[16], [17]. One of the most cited studies in this field is the article by Chakrabarti et al. (1998). In this study, a computerized system that begins with a small corpus sample where topics are manually assigned. It then updates the database as the corpus grows, quickly and accurately assigning topics to new records [18]. The earliest investigations in this area employed methods from machine learning. Fall et al. set up the first dataset (Wipo-alpha) that was analyzed for the automatic classification of patents in 2003[15].

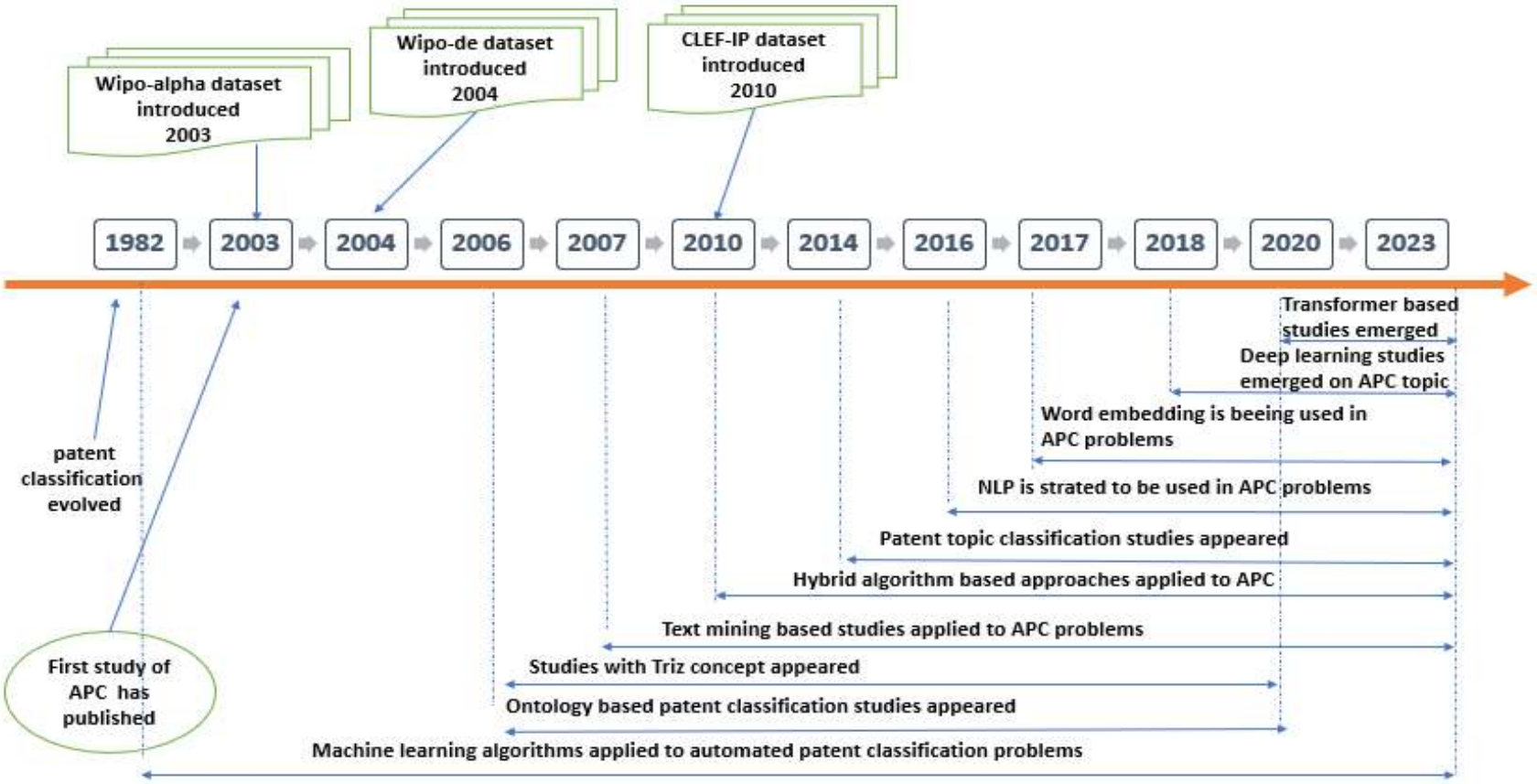


Figure 3.1. Evolution of APC studies over time

Ontology has been the focus of many studies on patent analysis subject. According to Figure 3.1, the studies using the ontology concept started in 2006. In the study by Trappey et al., an ontology-based semantic-concept approach was proposed. The keywords were generated using the TFIDF method, and the patents were trained and evaluated by using back-propagation artificial neural network (BPANN)[19]. One other study which used ontology concept is produced by Trappey et al in 2009. In this study, a hybrid strategy that combines ontology-based and TF-IDF concept categorization was used to analyze patent documents. The concept of ontology represents the comprehensive knowledge as well as vital contents of patent documents in a specific field. The proposed approach gathers, categorizes, and integrates patent data to generate a concise summary and a cluster tree graphic for appropriate phrases that involve [20].

Just as the ontology concept was tested on patent documents, the Triz (Theory of Inventive Problem Solving) method also started to be tested in the same time period. For example, Loh et al. propose an automatic patent classification system for the use of Triz users using the Triz method [21]. They updated their study in their next study, in which they included only 6 of the 40 Contradictions in the Triz method. In Cong and Tong's (2008) updated study, this time all 40 Contradictions were included and they tried to design an expert system for Triz practitioners by removing the single label and balance assumption found in the previous study [22].

Data mining has long been an approach used in various fields. Text mining has evolved as studies in this field have focused on texts. Afterward, studies were also conducted on patent texts. Tseng et al. (2007) proposed an approach that utilizes various text mining algorithms to automatically classify patents based on the content of the patent texts[10]. During text mining, researchers have utilized established word extraction algorithms (e.g., decision tree[23], naïve Bayes[24], SVM[25], [26], [27], Hidden Markov Model (HMM)[28], and etc.) for supervised approaches, Term Frequency-Inverse Document Frequency (Tfidf) [29], [30], TextRank [31], and Rake [32], [33], etc. for unsupervised approaches, while also developing new techniques[34]. For example, in [35] researchers offer a patent keyword extraction algorithm (PKEA) for patent classification relying on the distributed Skip-gram model in this study. They

also introduce a set of performance metrics for word extraction analysis according to information gain and cross-validation. They evaluated the performance of PKEA on a benchmark dataset. Another example is a study that evaluates unsupervised word harvesting methods which do not make use of a corpora. It compares the outcome of Position Rank, which takes into account the location of each word in the document, with TextRank and RAKE [36].

Various hybrid studies in patent classification analysis have been conducted, including hybrid keyword extraction algorithms and classification with hybrid approaches. In a study by Wu et al., a hybrid result was obtained by integrating genetic algorithm and SVM methods. The SVM parameters are optimized using a genetic algorithm [37]. In patent classification problems, hybrid techniques, which are the integration or merging of various approaches in order to eliminate individual limitations of an algorithm or method and improve the accuracy of classification, have been encountered. Xue et al present one of these methods as an example. They developed a new IG & k-NN based patent classification method. While performing the feature selection process with the new IC, they also aimed to improve the classification performance with the new kNN [38]. The authors in [39] conducted research using patent citation data to evaluate the use of graph kernels for classifying patent documents. They employed seven hybrid graph kernels, trained the classification algorithm with the kernel matrix of the training dataset's input, and then applied the model to the testing set. They applied SVM as the kernel machine in this study due to its well-established performance.

According to Figure 3.1, the number of studies concerning the topic of modeling has increased since 2014. In the study by Venugopalan and Rai, linguistic features of topic modeling are utilized to categorize patent texts into subcategories using four different classifiers (SVM, second-order discriminant analysis (QDA), Linear discriminant analysis (LDA), and neural networks) with a hierarchical method based on natural language processing. The primary contribution of the study is the use of topic distributions generated by the topic model as features to train classifiers [40]. Two different methods were used in another topic modeling study on the APC. Yun and Geum [41] offer a topic model for APC problem using SVM estimation. The study focuses on two main challenges in patent categorization: text representation and

category prediction. For text representation, the topic modeling approach and latent Dirichlet allocation (LDA) are used. The output of the LDA is then used for category estimation. SVM prediction is employed to perform automatic patent categorization. Some studies perform the classification process by extracting important information from the existing data with rule-ensemble methods [42].

In some studies, NLP was used to classify multi-label data. In a study testing natural language processing tools on patent texts, various methods were combined for different parts of the analysis. LDA was used for topic modeling. In the multi-label classification phase, machine learning methods such as KNN, decision trees, and random forest were utilized. It has also been reported that the results are promising [43].

Due to Figure 3.1, new techniques in automated patent classification have been implemented in the past 5 years. In recent years, there has been a surge in research using deep learning, and the use of the word embedding method has become seen in APC problems. With advancements in computer technology, new techniques started being utilized to address the APC problem. Some studies have made classifications by adding word embedding to methods such as LSTM and CNN. For example, in [44], the study presents a patent classification system that uses word embedding and a long short-term memory (LSTM) network to categorize patents up to the IPC subclass level. The method described in the paper differs from others because it utilizes a semantic approach, representing semantics with Word2Vec-generated word vectors [44]. The study examining the CLEF-IP (Conference and Labs of the Evaluation Forum) and USPTO-2M data sets using the CNN deep learning model is one of the exemplary ones in this field [45]. In[46], the authors propose a new classification approach that depicts documents as Fixed Hierarchy Vectors (FHV), that represent the content structure. FHVs represent a document on multiple levels, each of which reflects the whole document but with a distinctive local setting. Researchers additionally utilize LSTM-based models to design and categorize data. The research shows that FHVs give better representation of patent documents, and sequential categorization improves the accuracy of classifying patents within the IPC hierarchy. Since its introduction in 2017, the Transformer model has been used to classify patent texts. In [47], the researchers

suggested a deep patent landscaping approach which employs a transformer and Diff2Vec architectures to solve the challenges of patent landscaping. The researchers provide a novel dataset, and their model achieves strong overall classification performance in patent landscaping. They also conduct experiments to investigate the impact of technical codes and text data on patent classification models. In the study [48], the authors created a Chinese patent dataset and used RBT3, BERT, and RoBERTa models to train and testing the multi-label patent classification

3.2 Automated Patent Classification Literature Details

Krestel et al. categorized patent analysis research into eight tasks. These include supporting tasks, patent classification, retrieval, valuation, market value prediction, technology forecasting, text generation, litigation analysis, and computer vision tasks [49]. The main focus of this thesis is the automatic classification of patent texts. Table 3.1 provides the information, which includes details such as achieved research accuracy F1 or precision values (based on the information provided by the authors), analysis techniques, studied data sets, utilized sections of patent texts, and so on.

The main challenge for researchers on APC is achieving satisfactory classification performance in subgroup level analyzes. Various studies have been conducted in this area. Some researchers have adopted hierarchical classification approaches [15], [50] while others use different keyword acquisition methods to achieve accurate classification. Researchers have analyzed different patent datasets for comparative study. For instance, WIPO-alpha [51], [52], [53], WIPO-de [54], CLEF-IP series [45], [55], [56], [57] USPTO-2M (dataset includes 2,679,443 utility patent documents) [45], [58], [59], and USPTO-3M (3,050,615 patents) [60], [61], [62] are patent datasets.

The researchers analyzed various combinations of patent text attributes to determine which part had the greatest impact on patent examiners' classification decisions. For example, [63] achieved 52.98% accuracy by integrating the first three hundred keywords from a patent document's description, title, and abstract. The developers of the WIPO-alpha collection [15] compared Naive Bayes (NB), a variant of the Winnow algorithm, SVM, and k-NN [64]. The study results indicated that both NB and SVM had the highest classification rates at the class level (55%), while SVM demonstrated better accuracy at the subclass level (41%). Utilizing only the first 300 keywords from

the descriptions, titles, inventors, abstracts, and applicants sections improved classification accuracy compared with utilizing all parts, according to their results. With the establishment of the WIPO-alpha collection, numerous scientists utilized it in their research. In study [54], the authors analyzed all sections of the patent paper and proposed a multiclass SVM for categorizing the WIPO-alpha dataset. The main group level of IPC achieved a maximum accuracy of 42.9%. The hierarchical online classifier (HITEC) algorithm, presented by [64] achieved 0.5325 accuracy at the subclass level and 0.3689 at the main group level, demonstrating approximately 12% higher accuracy than single-level classifiers. Some studies [65] utilized the CLEF-IP 2010 dataset and received 77% accuracy at the class level and 68% accuracy at the subclass level using the Balanced Winnow approach. The authors of [66] used a back-propagation network model as a classifier to categorize at the main group and subgroup levels. The dataset only included one WIPO-alpha component (class B25). They only used the top 67 words for their classification. The best results are 92% at the main group level and 90% at the subgroup level. Although the accuracy results improved compared to prior studies, this study only involved 124 patents, so it should be tested with more complex dataset in order to be conclusive.

The published studies contain valuable data on keyword extraction approaches. They assume that carefully chosen keywords accurately describe the document's content. The issue of how many terms and which section of a document should be used for classification is fundamental to these APC studies. If the analyst selects too many words, there will be noisy data, but choosing fewer words will result in inaccurate classification outcomes. In this thesis, various cases are examined to identify the sections of patent documents from which consistent and specific keywords can be extracted. Most earlier studies have focused their examination on the abstracts or titles of patent documents [67]. Yücesoy et al. added extra stop words and phrases to the abstract, title and claim sections of the patent and reached an accuracy value of 0.69 at the Wipo-alpha subclass level in their classification. The main purpose of the authors is to investigate which sections should be used in patent classification studies. Contrary to studies in the literature on this subject where only abstract and title are used, they argue that this is not sufficient and other sections should also be added [68]. This thesis aims to analyze different textual content of patent document through various case

studies. The thesis seeks to find out if including only an abstract or title can provide brief and unspecific details about a patent. It is expected that adding additional sections will result in more consistent terms. Furthermore, due to the description section of a patent encompasses a background a summary and a detail of the intended usage of the invention, and claims include legal information, it is assumed that if they are used in a patent classification of text research, they will make a significant contribution to the patent classification accuracy.

With the introduction of Latent Dirichlet Allocation (LDA), researchers' interest in classifying patents by splitting them into topics has increased. In study [40] researchers divide the patent according to their technological subject. In their study, where they proposed a two-stage technique, they used an SVM-based classifier to separate relevant patents from irrelevant ones. In the second stage, they used topic modeling to divide technology into subclasses.

With the deep learning approach, new methods have been used in the analysis of patent texts. In the studies, the authors use methods that emerge with deep learning such as CNN, LSTM, BILSTM, and also create their own structures by making various additions to these methods. For example, in study Hu et al.[69] authors established a hybrid hierarchical feature extraction model for multi-label mechanical patent classification called HFEM. They compared this model with BILSTM, LSTM and CNN algorithms. They used the entire text of the patent at the subclass level. They worked with the CLEF-IP data set and the maximum F1 value they obtained was 63.97%, giving better results than other models. Li et al.[45] tried the deep learning approach on the subclass level using datasets called CLEF-IP and USPTO-2M. The work, known as DeepPatent, is based on CNN and word vector embedding. It achieved a precision value of 83.98% on the CLEF-IP dataset, while it achieved a precision of 73.88% on the USPTO-2M dataset. In his research, Hepburn used the Fine-Tuned Universal Language Model (ULMFiT) and SVM algorithms to categorize Australian patents at the section level, achieving an F1 score of 78.4% 1[70].

Transformers are deep learning architectures that include both encoder and decoder structures, as well as the attention mechanism. The BERT model was developed using frameworks that include an encoder from the transformers. GPT (Generative Pre-

trained Transformer) models are models that include the decoder. BERT can be fine-tuned for various text analysis tasks. Some studies utilize models and work with patent documents.

Pujari et al. combined transformer and hierarchical algorithms and used Wipo alpha and a dataset of 70,000 sample patents from USPTO. They developed the Transformer-based Multi-Task Model (TMM) and its hierarchical version. The researchers proposed a model to improve classification performance by adding hierarchical layers between classification links. When comparing Transformer-based Hierarchical Multi-Task Model THMM-CLS with TMM-CLS on both datasets, the former outperforms in terms of macro-F1, while the latter has slightly higher micro-F1, indicating that adding links is particularly beneficial for less frequent tags [71].

Table 3.1. Patent classifications studies details

Study	*Year*	*Classification Method*	*Dataset*	*Level of Classification*	*Section used*	*Result*
[66]	2006	NN	Class B25 from WIPO-alpha (124 patents for testing)	Main group and subgroup	DF (the 67 most frequent words are selected)	0.92 accuracy at main group level, 0.9 accuracy at subgroup level
[55]	2012	SVM and kNN	A subset of WIPO-alpha (21,104 patents, 12,042 for training and 9,062 for testing)	Subgroup	Title and claims (combined)	0.36 accuracy
[15]	2003	SVM or NB or kNN or SNoW	WIPO-alpha	Subclass and group	(a) Title or (b) claims (separate) (c) 300 first words of titles, inventors, applicants, abstracts and descriptions (combined) (d) titles, inventors, applicants, and abstracts combined)	0.55 accuracy
[72]	2006	H-M3 (Maximum Margin Hierarchical Multilabel Classifier)	Section D of WIPO-alpha	Subclass	Not defined	0.76 F1
[73]	2003	UFEX	WIPO-alpha, WIPO-de	Class, subclass and main group	Title, inventor, applicant, abstract, claims (combined)	Acc Top=0.66, for WIPO-alpha at class level/ Acc Top=0.55, for WIPO-alpha at subclass level/ Acc Top=0.38, for WIPO-alpha at main group level/ Acc Top=0.65, WIPO-de at class level/ Acc Top=0.55, for WIPO-de at subclass level/ Acc Top=0.38, for WIPO-de at main group level

Table 3.1. Patent classifications studies details*(continued)*

Study	*Year*	*Classification Method*	*Dataset*	*Level of Classification*	*Section used*	*Result*
[54]	2007	hSVM (hierarchical SVM)	WIPO-alpha using 3-fold cross validation over the whole dataset	Main group	Title and claims (combined)	0.38 accuracy
[63]	2009	Kernel classification model	WIPO-alpha	Subclass	Title, abstract and first 300 words of Description	0.52 F1
[65]	2010	Balanced Winnow	CLEF-IP 2010	Class and Subclass	Title or abstract or names or description (each section separated)	Best performance (using abstracts, titles, names, addresses in combination 0.77 F1 value at the class level and (using all 3 languages words) 0.68 accuracy at the subclass level
[57]	2013	Balanced Winnow	CLEF-IP 2010	Class	Abstract	0.7951 Precision
[40]	2015	LDA and SVM	USPTO 10,201 data	Subclass	Claims and Abstract	0.98 accuracy separate relevant patent from irrelevant patent
[44]	2017	Word2Vec and LSTM	USPTO Bulk Data	Subclass	Description	0.63 Accuracy
[68]	2018	kNN and SVM	Section D of WIPO-alpha	Subclass	Title, abstract and claim	0.69 Accuracy
[45]	2018	CNN and word vector embedding	USPTO-2M, CLEF-IP	Subclass	Title abstract and IPC	0.7388 Precision

Table 3.1. Patent classifications studies details*(continued)*

Study	***Year***	***Classification Method***	***Dataset***	***Level of Classification***	***Section used***	***Result***
[69]	2018	HFEM CNN, LSTM and BİLSTM	CLEF-IP	Subclass	Title, abstract claim and description	HFEM outperformed other models with entire text 0.6397 F1 value
[70]	2018	ULMFIT	Australian patents	Subclass	Section	0.784 F1 score
[62]	2020	BERT CNN with word embeddings	USPTO-3M	Subclass	Claims	0.6683 F1 score with CPC + Claim, 0.8175 Precision with IPC+ Title + Abstract
[71]	2021	THMM	Wipo-alpha, USPTO	Subclass	Title + abstract	THMM-CLS received 0.698 micro F1 result
[74]	2021	LSTM	2 datasets ((1) 430k and (2) 1925k)	Subclass	Dataset1: All text + Inventor + Assignee Dataset 2: All text	Dataset1 67% Acc. Dataset :74% Acc.
[58]	2022	BERT, XLNet, RoBERTa, and ELECTRA	USPTO-2M and M-Patent	Subclass	Description	XLNET achieved 0.8272 precision on USPTO-2M dataset and 0.8229 on M-Patent dataset

3.3 Literature related with Hierarchical Attention Network algorithm

There are studies in the literature that consider the text document hierarchically and take the attention situation into account in the analysis. Some attention-based methods are hierarchical attention, self-attention sentence embedding, knowledge powered attention and others [75]. In 2016, the HAN algorithm was created in order to extract more detailed results from the text [76]. The method, which includes 2 words and 2 sentence encoders, can produce specific results for inputs. This method collects the words it deems important into the vector of sentences, then applies the same operation to the document vector, in other words, gathers the sentences it deems important. The bidirectional and two-level attention mechanism of the method benefits the classification results. The researchers who developed the method tested this method in 6 different data sets. The data sets used in the study were: Imdb review, Amazon review, Yahoo answer and Yelp datasets (Yelp corpus of 2013, 2014 and 2015). The model produced as a result of the study outperformed previous studies in all data sets and higher accuracy was obtained [76]. A method called Hierarchical Attention-based Recurrent Neural Network (HARNN) has been developed in which the attention mechanism was applied to education dataset and patent dataset (USPTO dataset) [77]. The researchers applied a different attention mechanism than the HAN model. This new method has two layers. In the first stage, it creates the representation and hierarchical structure of the texts with the documentation representation layer. In the second stage, the hierarchical attention-based recurrent layer repeats the hierarchical structure from top to bottom and models it by revealing the dependencies between the levels [77]. The method reached 74.2 % precision; according to paper they achieved highest value in accordance with all evaluation matrices.

3.4 Literature related with Transformer Encoder Model

After the emergence of the transformer model, many variations of this technique have occurred. BERT and GPT models used today are actually transformer-based techniques. Many applications of BERT can be found in patent classification. One of them is the PatentBERT article. With their new dataset, called USPTO-3M, they implemented BERT in a way that includes patent claims. They state that the results they found are better than the results in the DeepPatent article[78]. Among the studies using the Attention method, they are also frequently encountered in the literature. A

method combining label embedding and self-interaction attention was proposed by Dong et al. The Bert method was used in the model, but the training of the Bert method was thought to be hard [79]. Zeng et al. combined CNN and LSTM using the Automatic Content Extraction (ACE) 2005 dataset. They tried to find lexical-level and sentence-level features. Trigger labeling and argument role labeling results were competitive with other studies[80]. In the study conducted with a pre-trained BERT model using more than two million patent claims, analyses were made with CNN with word embeddings and better results were obtained in the analyses with BERT than CNN [81]. In the study [82], in which the automated patent landscaping model was proposed for the analysis of patent texts, patent texts were analyzed with a modified transformer encoder method and patent metadata with a graphic placement method called Diff2Vec, and they benefited from the collaboration of these two methods. They also exemplified their work with four new reference datasets.

3.5 Literature related with BERT Model

BERT offers advantages such as capturing bi-directional contextual information, enabling fine-tuning without the need for architectural changes, and accurately understanding and classifying patent documents [83], [84]. BERT algorithms, known for their superior language understanding and learning capabilities, are frequently used in automatic patent classification studies. Some of these studies combined various variations of word embedding and Bert algorithms. For example, Roudsari et al. In their study [85], they took the USPTO-2M and M-patent data sets as input data in the abstract and title sections and performed a multi-label patent classification process using pre-trained XLNet, BERT, ELECTRA and RoBERTa algorithms at the IPC subclass level. Analysis results were compared with deep learning models and revealed that pre-trained language models performed better and fastText embedding was better than word2vec embedding. The XLNet algorithm outperformed other pre-trained models with an F1 value of 72.08 on the M-patent dataset and an F1 value of 63.33 on the USPTO-2M dataset. The study differs from the PatentNet study [85], in terms of the type of data set it uses (USPTO-2M and M-patent data sets), the type of classification made (multi-label patent classification) and the methods it uses (BERT, XLNET and ELECTRA). In another study where the Sentence-Bert (SBERT) algorithm was applied to the CPC data set, 54% accuracy and 66% F1 score were

obtained at the subclass level in 1,492,294 patents. The analysis of 10000 sentences with the SBERT algorithm took 5 seconds, and the same process took 65 hours when done with BERT and RoBERTa [86]. In their study, Lee et al. developed a new and pre-trained BERT-based transformer model for patent classification under the name PatentBERT. Patents at the CPC subclass level of the USPTO-3M (3,050,615 patents) dataset were used for analysis. Patent title, abstract and claims were used as input data and 44.75% F1 value was reached [87]. In the patent mapping study conducted by Wang et al., Bert and Triz methods were used. Patent data was collected from the Derwent database. A vector space was established by collecting similar technology elements under the same headings. While Triz helped find existing technologies, Bert came into play at the classification stage [88]. Zhipeng and Zheng [89] combined the output feature vectors of a deep learning model with the physical meaning of quotes to solve some questions in designing new technologies. They accepted the references to patent texts as vectors and added them to their analysis. Using the Bert model in analysis, the study developed an approach that will help inventors design new technologies through prospective citations of patents whose citation relationships and vectors have great similarities with the initial vectors. In their study [90], Choi et al. proposed a deep patent mapping model. This model consists of 2 parts. The first is a transformer encoder [91], and the second is Diff2Vec, which is a graph-embedding process. An attempt was made to find a solution to the classification problem with a deep patent mapping model by making use of textual information and citation-graph information data. The patent dataset of the study was obtained from Google BigQuery. Analysis results were compared with APL and PatentBERT [90].

This study is different from DeepPatent and PatentBert studies, which are frequently encountered in the field of automatic patent classification. In the DeepPatent article, an algorithm based on ESA and word embedding was developed using USPTO 2M and CLEF-IP datasets. In the study, only the title and abstract parts of the patent texts were used and an accuracy of 73.88% and an F1 score of 55.09% were obtained. In the PatentBert article[60], in addition to the USPTO 3M data set, the USPTO-2M data set used by DeepPatent [45] was used and the analyzes were examined comparatively. It has been argued that only the claim parts of the patent texts are sufficient. In both studies, they include patent tags in the analysis as well as text data. Looking at other

studies using the Wipo-alpha dataset, Gomez et al. [92] Wipo-alpha and Wipo-de patent data sets were used in the study. In their study, both the label and the text part of the patent were taken as input. NB, SVM, LR and KNN methods were used and the best result achieved at the Wipo-alpha subclass level (accuracy = 0.533) was obtained with logistic regression and description input. In the study of Abdelgawad et al.[93], different word embedding methods (glove, fasttext) were used together with the ESA method. An accuracy of 55.02% was achieved at the subclass level. This thesis's study on Bert differs from the studies of PatentBert and DeepPatent in many aspects. These can be listed as 1) using the Wipo-alpha data set for patent classification, 2) trying to obtain classification performance by using only the textual part of the patent document, 3) analyzing the hyper parameters of the Bert model and discussing their effects on classification performance.

The contributions of this study to the literature can be listed as follows. First of all, the idea that is widely used in the literature and that only the abstract or claim section of the patent is sufficient as input data has become invalid according to the classification performance result achieved by this study. Secondly, the results obtained from experiments carried out with many hyperparameter analyzes will create a starting point for researchers who want to analyze in this field and will save time by preventing unnecessary use of data such as computer time and GPU. Thirdly, the accuracy obtained by the classification analyzes at the subclass level exceeded the accuracy achieved by similar studies in the literature.

3.6 Conclusion of Literature Review Analysis

A few gaps in the existing literature on automatic patent classification have been attempted to be solved in this PhD thesis. Four independent studies performed different analyses on automatic patent classification problems. First study performs the bibliometric analysis of articles and patents related to automatic patent classification. In this study, all studies related to APC in the literature (matching the search criteria) were brought together and examined with a broad bibliometric approach. Studies in the literature on this subject are old and only include academic articles, do not include patent data, and do not provide such a comprehensive analysis.

The second study, focuses on patent classification using the hierarchical attention network (HAN) technique. This study evaluates patent texts on both a word and sentence basis. This study contributes to the literature by analyzing various hyperparameter analyzes and different input data.

The third study, compares the Transformer encoder model to ensembled-deep learning algorithms for automatic classification of patent documents. The study not only focused on deep learning methods and their static and dynamic versions, but also examined the parameters that influence changes in performance through various experimental analyses. The study involved conducting various experimental analyses on the transformer model, aiming to contribute to the literature by comparing the combined outcomes of deep learning methods with the transformer model. The fourth study, employs the pre-trained BERT approach to solve automatic patent classification problem. It was used to classify the Wipo-alpha data set at the subclass level. In this study, the special structure of patent texts was taken into account and the study was examined with hyperparameter analyses. These four studies are comprehensively discussed in their respective chapters.

CHAPTER 4
A BIBLIOMETRIC ANALYSIS ON AUTOMATED PATENT CLASSIFICATION

4.1 Introduction

One of the most important indicators of the development of technology is the existence of patent documents [94]. These documents reflect all kinds of innovations in an original, impartial and unique language. Patent classification systems such as IPC and CPC have been developed for the purpose of facilitating access to patent documents and grouping these documents under the same technological categories. These systems have a hierarchical structure with tens of thousands of subclasses [95]. Automatic patent classification systems help patent experts in the process of assigning new patents to the suitable technology group. In addition, these systems are also used by inventors, experts and researchers in their research, listing similar patents and reducing the possibility of patent infringement. A classification system with a large number of misclassified patents wastes time, money and effort. The issues encountered in classification systems have led to the emergence of the problem of automatic patent classification in the literature [54].

In general, a basic text classification procedure includes four steps [96]. The first of these is the preprocessing step. At this point, the text content is edited with processes such as tokenization, stop word elimination, and stemming. The second step is feature selection, where the text vector with the smallest size is determined in order to minimize classification errors. Techniques used in the text representation phase include TFIDF, bag of words, and word2vec. Lastly, in the classification step, one of the classical or model classification approaches is used to complete the process. Researches in the field of patent classification also utilizes these basic steps.

Academic studies published in the field of APC have been conducted for more than 40 years. Studies looking for a solution to the APC problem using small-sized data and traditional clustering methods (such as SVM, naive Bayes and k-NN) are among the first examples of research in this field [54], [97], [98], [99], [100]. In addition, hybrid studies in which heuristic algorithms perform the classification process are also among the studies carried out in this field [98].

Traditional classification methods have been the most widely used in APC's 40-year history. In recent years, with the advent of deep learning methods (such as CNN, RNN, and BiLSTM), patent texts have begun to be analyzed using algorithms that employ pre-trained word embeddings and can analyze large-scale data [93], [101], [102], [103], [104] With the addition of attention situations being prioritized, algorithms such as BERT, ELMo, GPT, and Transformer encoders and named entity recognition started to be used [78], [83], [105], [106], [107], [108]. Some studies classify patent documents by using both textual and bibliometric data. For instance, in [109] bibliometric data were analyzed with the modified transformer structure, patent textual part, and the diffusion graph Diff2Vec. Since patent texts have a multi-label structure, the classification of these labels is another field of study in APC. While there are studies that focus on the estimation of a single label by considering patent texts as a flat model, there are also studies that make a multi-label classification [110], [111], [112]. Researchers have also sought an answer to the question of which part of the patent text (title, abstract, claims, and descriptions) is used for more effective classification [68], [113]. There are also differences in the studies on how to represent the patent vector. While keyword frequency [114] is used in some studies, TFIDF [115], [116] or binary frequency [113] is used in others, word embedding [117], [118], [119] has been tried to be included in recent patent classification studies.

Some published literature review studies in the field of patent analysis are as follows. Abbas et al. [120] conducted a literature analysis in which they examined studies using text mining and visualization-based methods. In the study, a comprehensive literature review is provided highlighting the potential limitations of relying solely on that may be encountered when Sao-based methods. They also noted that patent and academic journal articles should be used in combination for patent retrieval studies. In study

[121], 40 different studies that conducted patent analysis using deep learning were examined. The study compiles a wealth of information, covering analysis methods such as retrieval, computer vision and classification as well as datasets. It also discusses details of the deep learning trends and potential future research directions. The potential for conducting previously unexplored studies, such as patent-related adversarial attacks, is also mentioned in this field. In study [122], a combined keyword and bibliometric analysis was conducted according to examine the evolution of patent mining. In the study, of 143 different articles, it was found that words such as "innovation", "science", and "nanotechnology" were among the most frequently used keywords. The change in patent mining was revealed with 13 clusters created from the abstracts of examined papers. In the study, only academic publications in the field of patent mining between 1992 and 2013 were included, and no analysis was included for patent documents published in this field. Furthermore, this study does not include a distinct title for patent classification. In their literature review on automatic hierarchical patent classification, Gomez and Moenz [123] examined the studies published up to 2014 using the WIPO and CLEF-IP datasets.

Bibliometric analysis is a valuable statistical tool that allows for mapping of the literature related to a scientific research field and provides insight into the current state of technology. The output of this method gives quantitative information, including annual publications, authors, citations, journals, and leading countries, as well as qualitative information such as expected future study topics and research methods [124], [125]. This method has been applied in many fields, such as APC [126], multi-criteria decision- making [127], environmentally friendly products [128], ChatGPT [129], and COVID-19 research [130].

This chapter includes all patents and articles published in the field of automatic patent classification, sorted based on search parameters. It also incorporates bibliometric, keyword, and citation analysis. According to our knowledge, other than Yücesoy et al [126] no research has thoroughly integrated bibliometric, citation, and keyword analysis of both patents and scientific papers in the published works on automatic patent classification. While some research has focused on patent mining or analysis [122], others have conducted analysis based solely on scholarly publications [123] and

have disregarded the role that patents play in this field. This chapter intends to point out current research and the most recent advancements in this gap in the academic literature on automatic patent classification which Yücesoy et al. endeavored to fill.

The bibliometric perspective in this study is expected to contribute to addressing the APC problem in three ways. Firstly, with the increasing number of patent applications worldwide, the APC is an active research area, and the main structure of this field is presented in detail. Secondly, trends and research methods in the APC field were examined in detail, providing insight into the future of research in the APC field. Finally, there are steps that facilitate decision making process for scholars or researchers who are continuing or beginning their studies in this field, helping them determine in which direction of their work.

The rest of this chapter is organized as follows: First, the methods used in the bibliometric analysis process are detailed. Bibliometric and keyword analyzes of academic studies on APC and published patents in this field were conducted. In the conclusion, the study discusses its findings and identifies potential research topics resulting from gaps in the literature.

4.2 Data Sources and Methodology

It is evident that the most commonly used databases in the literature are Scopus, Web of Knowledge, Google Scholar, and Pubmed[131], [132], [133]. This study utilized the Scopus database platform to select articles related to APC. While CiteSpace only accepts WoK (Web of Knowledge) as the primary source of input [134], the Scopus database is preferred over WoK for a more comprehensive paper analysis.

It is important to include patents and papers on APC for general inclusivity. When searching the patent text in the Scopus database, around 16,000 patent records were found. The same research was conducted in WOK, leading to 71 patents being listed. The study utilized the WOK database for patent analysis.

During the research phase of this study, all subject areas and document types were included. There are no restrictions on the starting year for the search. The analysis was conducted on December 6, 2023, to avoid bias resulting from any updates to Scopus.

The analysis utilizes the "topic term search" option. This Scopus feature enables research within a record's "Title," "Abstract," and "Keywords" fields. After reviewing the extracted papers, irrelevant data is removed. Various searches created from different combinations of keywords used are listed in Table 4.1. The initial search found 463 records.

Table 4.1. Different combinations of keywords search query

Search	Keywords	Count	Document h-index
S.1	"Document Class*" "Automat*" "Patent"	22	7
S.2	"Document Categor*" "Patent" "Automat*"	10	6
S.3	"Document Cluster*" "Patent" "Automat*"	10	3
S.4	"Text Cluster*" "Patent" "Automat*"	4	2
S.5	"Text Categor*" "Patent" "Automat*"	14	5
S.6	"Text Class*" "Automat*" "Patent"	37	9
S.7	"Automat* Class*" "Patent"	70	16
S.8	"Patent Mining" "Automat*""Categor*"	4	2
S.9	"Patent Mining" "Automat*""Class*"	14	5
S.10	"Patent Mining" "Automat*""Cluster*"	5	4
S.11	"Patent Analysis" "Automat*""Class*"	49	12
S.12	"Patent Analysis" "Automat*""Cluster*"	27	11
S.13	"Patent Analysis" "Automat*""Categor*"	17	6
S.14	"Patent Class*" "Automat*"	158	26
S.15	"Patent Cluster*" "Automat*"	7	5
S.16	"Patent Categor*" "Automat*"	15	7
	Total	463	

The individual searches are merged to remove duplicate and redundant information. The search query is created from this combination: (TITLE-ABS-KEY("Patent Analysis" "Automat*""Cluster*") OR TITLE-ABS-KEY("Patent Analysis" "Automat*""Categor*") OR TITLE-ABS-KEY("Patent Analysis" "Automat*""Class*") OR TITLE-ABS-KEY("Patent Mining" "Automat*""Cluster*") OR TITLE-ABS-KEY("Patent Mining" "Automat*""Class*") OR TITLE-ABS-KEY("Patent Mining" "Automat*""Categor*") OR TITLE-ABS-KEY("Patent Classification" "Automat*") OR TITLE-ABS-KEY("Patent Categor*" "Automat*") OR TITLE-ABS-KEY("Patent Cluster*" "Automat*") OR TITLE-ABS-KEY("Document Classification" "Automat*" "Patent") OR TITLE-ABS-KEY("Document Cluster*" "Patent" "Automat*") OR TITLE-ABS-KEY("Document

Categorization" "Automat*" "Patent") OR TITLE-ABS-KEY("Text Cluster*" "Automat*" "Patent") OR TITLE-ABS-KEY("Text Categorization" "Patent" "Automat*") OR TITLE-ABS-KEY("Text Classification" "Automat*" "Patent") OR TITLE-ABS-KEY("Automat* Classification" "Patent"))AND PUBYEAR < 2025

285 records were found in the search results. {Next, 28 data points were excluded from the records by removing articles with unknown authors, as shown in Figure 1 of the flowchart.} Articles on "APC" were written in five languages, mostly in English.In the patent data search phase, the same terms are used as in the paper search. In total, 78 patents are listed. With manual elimination, 7 patents not related to APC were excluded from the list. The patent search process is illustrated in Figure 4.1.

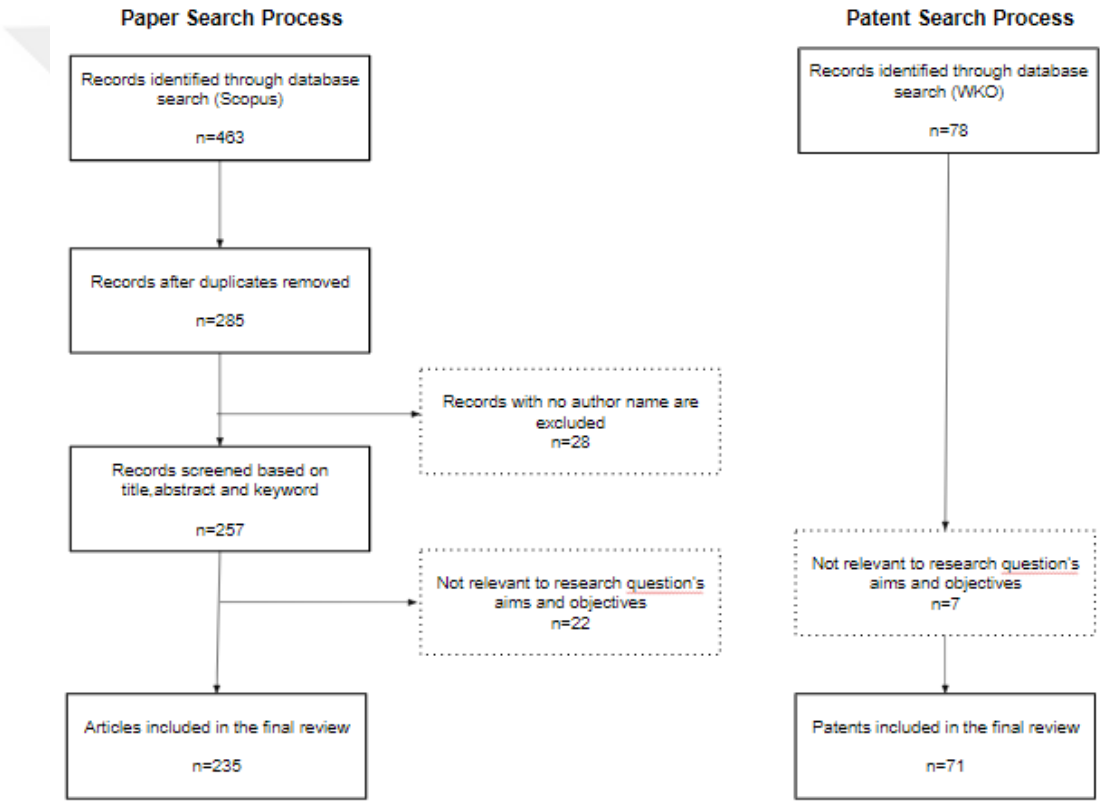


Figure 4.1. Paper and patent selection process

The following inquiries are being considered with respect to the extracted documents: What kinds of publications are available? What are the annual trends in APC publications? What are the most productive publications, associations, authors, and countries in terms of productivity? What is the current citation status of the APC subject? What are the primary areas for keyword analysis in research on APC? Within the APC, which assignee demonstrates higher productivity? Which patent subjects are

most productive in terms of APCs? To find the answer to this question, Yücesoy et al. conducted a detailed quantitative analysis to see the trends in automatic patent classification research and how researchers around the world contribute to this field [126]. This section of the thesis concentrates on Yücesoy et al. [126] research in the field of APC and provides an expanded version of their study.

4.3 Results and Discussion of Bibliometric Analysis

This study encompasses the steps of analyzing scientific articles and patent documents within the domain of APC. The general framework of the analyzes is depicted in Figure 4.2.

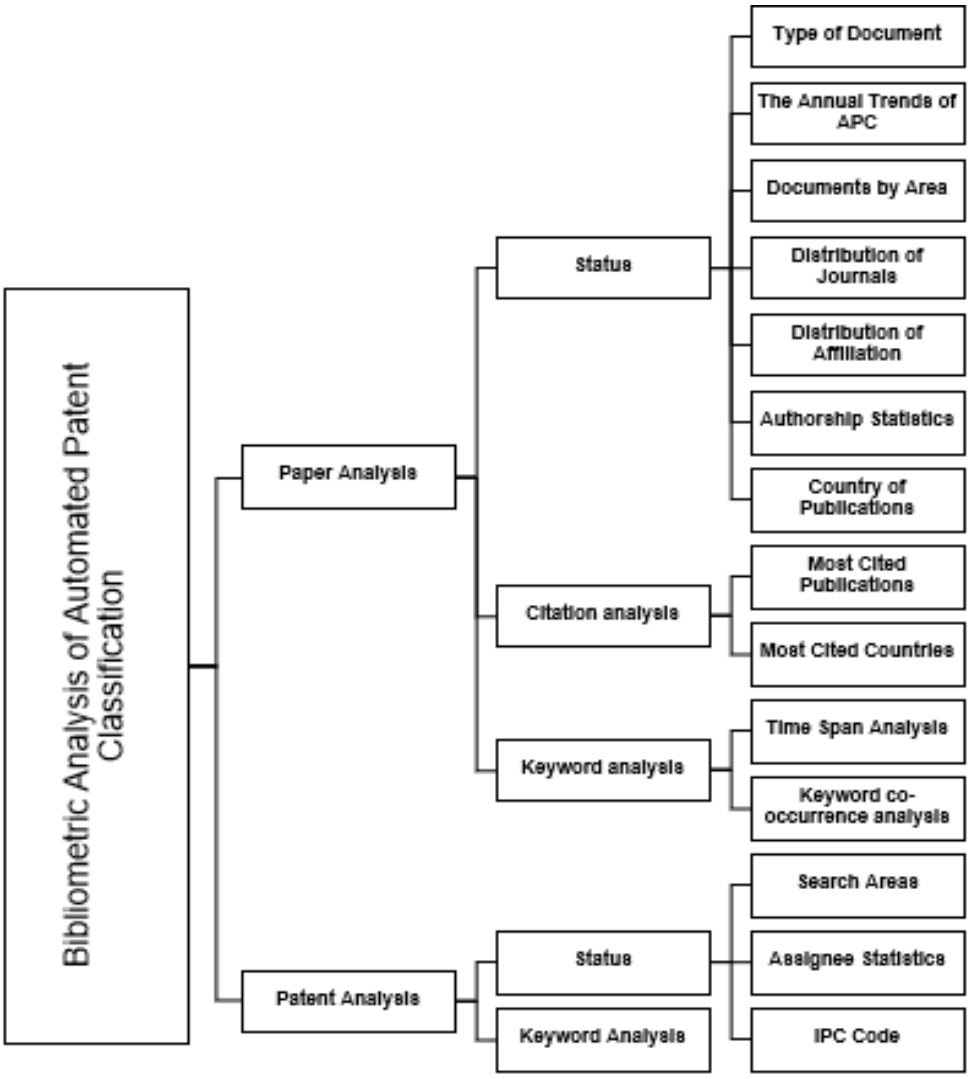


Figure 4.2. The general overview of the major steps of the Bibliometric analysis

4.3.1 Current APC-Related Scientific Publications Status

This section provides an overview of the present status of scientific publications related to APC. The study made use of the Scopus database, and the visualization process was facilitated using R-studio's Bibliometrics tool [135]. Before commencing the assessment, it is useful to identify a particular situation that will only be discerned by attentive readers. Taiwan is excluded from certain bibliometric analysis studies. For instance, Taiwan is not included in the list presented in Figure4.8, which displays corresponding authors countries. The primary reason for this is that Biblioshiny does not include Taiwan in its list of countries. The Biblioshiny application represents Taiwan data as "NA" in certain charts, despite the fact that Taiwan should be included. This application is necessary for some analyses. As a result, it appeared imperative to issue this problem prior to starting the analysis.

4.3.2 Type of Document

The extracted documents encompass five categories of publications, including journal articles, conference proceedings, book chapters, reviews, and short surveys. Upon analysis of the studies with regard to document types, it is apparent that more than half of these studies (134) consist of journal articles. The conference proceedings are ranked second (93). The significance of book chapters (5), review studies (1), letter (1), and short surveys (1) was determined to be notably lower in comparison to journal articles and conference papers. The graph depicting the distribution of document types is shown in Figure 4.3.

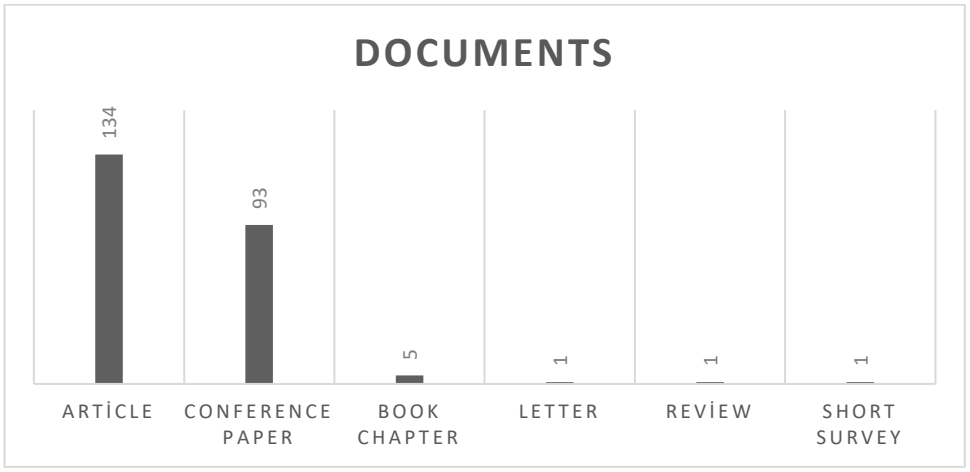


Figure 4.3. Various types of documents exist within the field of "APC."

4.3.3 The Annual Trends of APC- Related Publications

The annual trends of APC-related publications are shown in Figure 4.4. The main title of the researched subject is determined as "APC" and the time interval is accepted as "all years". 235 publications related to the APC subject have been identified over an average period of 40 years. The number of articles published during the 20 years after the first article was published in 1982 corresponds to 5% (12 papers) of the articles published on APC. In the third decade, it is seen that there were 5 times more publications (67 papers) compared to the first 20 years. Articles published between 2007-2023 (212 papers) make up 90% of the total records. In the last decade, the number of publications in this field has not fallen below 12, except for 7 publications in 2015. Advances in big data have also accelerated research in the field of APC. It is seen that the volume of publications has been increasing since 2017. One of the main factors contributing to this growth is the advancement of computer technology. There will be fewer studies in the field of APC in 2022 compared to previous years. This decrease is replaced by an increase in academic studies conducted in 2023.

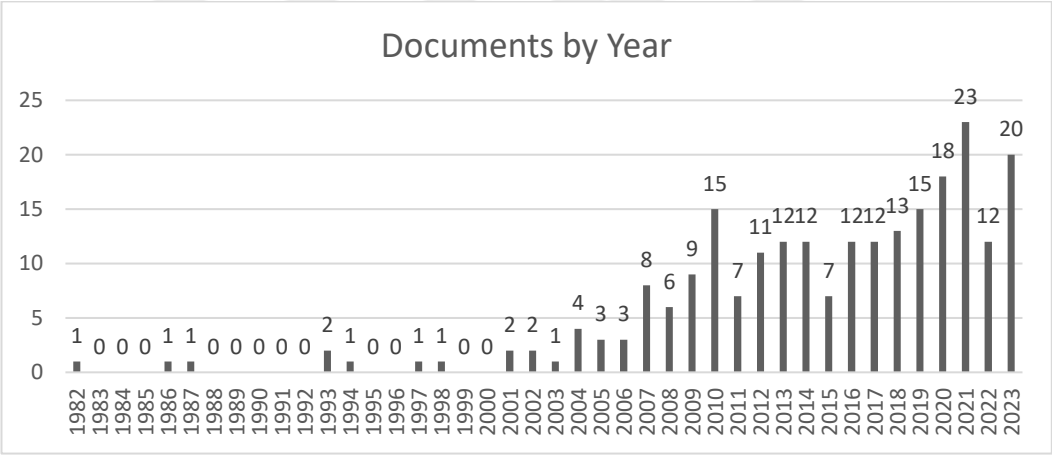


Figure 4.4. The annual trends of automated patent classification papers

4.3.4 Documents by Area

According to Figure 4.5, over 33% of the articles are published in the computer science field, with engineering and social sciences following closely behind. There is not much scholarly literature on APC in other fields. This suggests that APC is a concept that has connections to a wide range of fields, including energy, materials science, medicine, decision science, mathematics, social sciences, and chemical engineering.

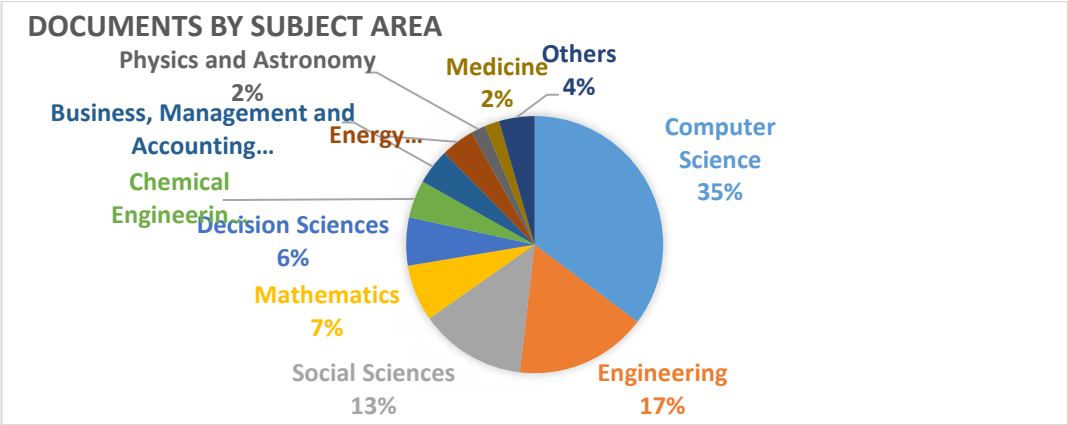


Figure 4.5 The subject of papers published between 1982–2023 on APC topic.

4.3.5 Journal Distribution in the APC Study

Table 4.2 shows the top 10 journals that feature APC the most. These journals published over one fifth of the publications listed in this field according to the search criteria. In terms of article publications, "World Patent Information," "Scientometrics," and "Expert Systems with Applications" were the top contributors to APC research. The impact factor statement refers to the number of citations of full-text and review papers published in an academic journal in subsequent years, per full-text or review paper[136]. The journals having the highest impact factor are "*Technological Forecasting and Social Change"* and "*Computers in Industry*". Having a high impact factor is a crucial factor for researchers when selecting a journal. If a journal lacks an impact factor, it does not necessarily mean that researchers do not prefer it. One of the most significant examples of this situation is the "*World Patent Information*" journal. This journal does not have an impact factor until 2022. Despite this, the most preferred journal in the field of APC is World Patent Information. Since the journal covers various patent-related issues, it is favoured by researchers in the field of APC.

Jorge Hirsch developed the "Hirsch index" or "h-index," a performance indicator that evaluates the quantity and visibility of publications [25]. According to the journal index scores (Table 4.2), Expert Systems with Applications and Applied Soft Computing had the highest h-index ratings. Scopus developed SNIP and SJR characteristics to measure the impact of articles. These two indicators are calculated for three-year periods and apply only to articles. While SNIP softens a journal's distinctive features, SJR evaluates the prestige of the citing journal and adjusts it based

on its high or low [137]. The CiteScore of a journal reflects the average number of citations to recent articles in that journal. It is an alternative to JCR (Journal Citation Reports). It examines the citations preserved in the Scopus database and uses articles from the last four years [49]. Table 4.2 displays the journals with the highest SJR values, including "Computers in Industry" (2.646) and "Technological Forecasting and Social Change" (2.644). "Information Processing and Management" and "Computers in Industry" had the highest SNIP levels. "Computers in Industry" has the greatest CiteScore value (21.1).

Journal Q scores provide information about a journal's citation performance as well as its position within the group of journals in the relevant scientific area. Researchers can use Q scores to assess a journal's status when selecting a journal to submit an article to[136]. The Q rankings of the journals in Table 4.2 correspond to Q1 and Q2, placing them in the top 50% of the groups to which they are associated, according to JCR category ratings.

Table 4.2. The leading scientific journal on "APC" (1982-2023)

	Journal	Index	Cite Score 2022	SJR 2022	SNIP 2022	Publisher	Impact Factor	5-year I.F	JCR Category	h	TC	NP	PY
1	World Patent Information	ESCI	3.5	0.637	0.914	Elsevier	2.7	2.4	Information Science & Library Science 65/163 - Q2	34	289	17	1987
2	Scientometrics	SCIE/ SSCI	6.0	1.019	1.520	Springer	3.9	4.1	Computer Science, Interdısciplinary Applications 52/110 -Q2	133	304	8	1978
3	Expert Systems with Applications	SCIE	12.6	1.873	2.582	Elsevier	8.5	8.3	Computer Science, Artificial Intelligence 22/145 -Q1	249	355	6	1990
4	Information Processing and Management	SCIE/ SSCI	14.8	2.106	3.032	Elsevier	8.6	8.2	Computer Science, Information Systems 11/158-Q1	114	849	6	1974
5	Technological Forecasting and Social Change	SSCI	17.2	2.644	3.008	Elsevier	12	12	Business 11/154-Q1	155	362	4	1970
6	Advanced Engineering Informatics	SCIE	11.8	1.709	2.432	Elsevier	8.8	8.7	Computer Science, Artificial Intelligence 19/145-Q1	99	26	4	2002
7	Applied Sciences	SCIE	4.5	0.492	0.974	MDPI	2.7	4.5	Engineering, Multidisciplinary-Q2	101	2	2	2011
8	Applied Soft Computing	SCIE	14.3	1.882	2.314	Elsevier	8.7	7.9	Computer Science, Artificial Intellıgence 21/145-Q1	171	140	2	2001
9	Computers in Industry	SCIE	21.1	2.646	3.283	Elsevier	10	10.4	Computer Science, Interdisciplinary Applications 8/110-Q1	117	29	2	1979
10	Journal of Informetrics	SCIE/ SSCI	6.9	1.269	1.756	Elsevier	3.7	7.5	Computer Science, Interdisciplinary Applications 54/110-Q2	83	8	2	2007

JCR=Journal Citation Reports SJR=SCImago Journal Rank, SNIP=Source Normalized Impact per Paper, PY=publication year, TC=Total Citations, NP=Number of Publication

4.3.6 Distribution of Affiliation on APC Study Statistics

The top 20 affiliations that made the most significant contributions to the APC topic are depicted in Figure 4.6. Universities from four different continents, each with unique origins, contribute to the field of APC. Taiwan, China, Singapore, and South Korea are among the top 20 Asian countries. However, in the Asian continent, universities in Taiwan and China have the highest number of publications in this field. Out of the top 20 organizations, 5 are Taiwanese and 7 are Chinese. Only two universities from European countries are in the top 20, with each having three registrations. On the American continent, Brazil has six records, while the United States has three records. Australia is home to four registered universities on the Oceania continent. Gaziantep University, situated in Turkey, spans both Asia and Europe and is individually represented in the top 20 with three records. Out of the 161 different institutions on the list, only 76 are universities. The remainder includes research institutes, companies, and laboratories. The presence of non-university-affiliated researchers conducting studies in the field of APC highlights that APC is not solely studied for academic purposes.

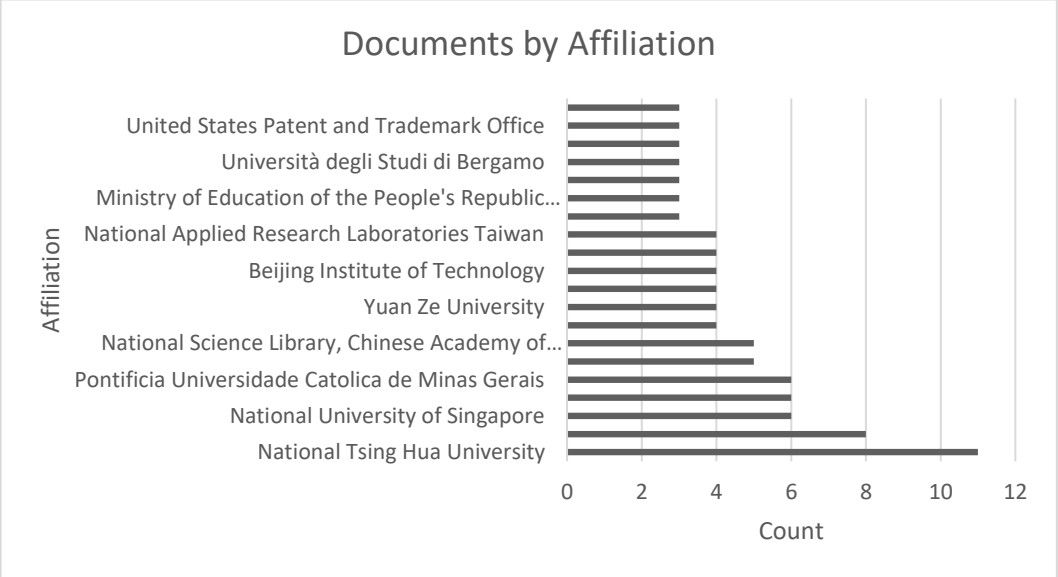


Figure 4.6. The top 20 affiliations on the basis of the number of articles published on APC areas in 1982 – 2023

4.3.7 Most productive authors

The list of the 20 most productive authors is presented in Table 4.3. Automatic patent classification and patent mining are closely related topics. Researchers from South

Korea are ranked among the top five authors in the study [122] that examines the evolution of patent mining. In addition, Porter, A.L. He is one of the top three writers. The author is not ranked in the top 20 on the APC list. In the same article, Trappey, A.J.C. is the most prolific author in the APC area, despite being listed as the last author.

Six Taiwanese authors, three Chinese authors, and three Spanish writers contribute to the top 20. Eight authors are affiliated with European university programs. Of the authors mentioned above, sixteen are affiliated with different academic institutions. Authors from the same country are known to have collaborated on works together. It appears that group work at APC follows a structure. Upon closer inspection, Trappey, A.J.C. and Trappey, C.V. appear as the most notable authors on the list. It seems that the majority of these authors' publications were co-written. Among those who prefer publishing jointly is Meireles M.R.G. Almeida, P.E.M. Considering the authors' h-index values, it is evident that they have high scores, which is indicative of their pioneering and highly cited publications in their respective fields.

Table 4.3. The number of publications among the most prolific 20 contributing authors

Author	Count	Institute	Country	h-index
Trappey, A.J.C.	11	National Tsing Hua University	Taiwan	43
Trappey, C.V.	10	National Chiao Tung University Taiwan	Taiwan	38
Hu, J.	8	University of South Carolina	USA	42
Salampasis, M.	5	International Hellenic University	Greece	17
Chang, P.C	4	Yuan Ze University	Taiwan	60
Fan, Chinyan	4	National Applied Research Laboratories	Taiwan	23
Li, Y.	4	Sichuan University	China	60
Meireles, M.R.G.	4	Pontificia Universidade Catolica de Minas Gerais	Brazil	8
Tsao, C.C.	4	Yuan Ze University	Taiwan	4
Wu, C.Y.	4	National Applied Research Laboratories	Taiwan	8
Almeida, P.E.M.	3	Federal Center for Technological Education of Minas Gerais	Brazil	10
Bermudez-Edo, M.	3	Universidad de Granada	Spain	14
Chen, E.	3	University of Science and Technology of China	China	60
Cuxac, P.	3	INIST-CNRS	France	6
Dereli, T.	3	Hasan Kalyoncu University	Turkey	37
Ferraro, G.	3	The Australian National University	Australia	11
Geva, S.	3	Queensland University of Technology	Australia	19
Giachanou, A.	3	Utrecht University	Netherlands	16
Hajlaoui, K.	3	INIST-CNRS	Spain	16
Hurtado-Torres, N.	3	Facultad de Ciencias Económicas y Empresariales	Spain	6

4.3.8 Country of publications

Table 4.4 shows that Taiwan, China, and the United States have the highest number of academic papers in the field of APC. Four of the East Asian countries (China, Taiwan, South Korea, and Japan) are included in the top 20 list. There are 9 European countries with a total of 54 publications in the top 20. The countries in which China broadcasts the most are the USA, Taiwan, and France. The countries with which the USA cooperates more are China, South Korea, and Singapore. Germany has produced joint work with Brazil, Spain, and Japan. This situation reveals that geography is not important in joint work and that collaborative studies are produced even between continents.

Table 4.4 The lists of APC corresponding authors' countries

Rank	Country	Documents	Rank	Country	Documents
1	China	60	11	Italy	9
2	USA	36	12	Greece	8
3	Taiwan	28	13	Singapore	8
4	Germany	16	14	Spain	7
5	South Korea	16	15	India	5
6	Japan	10	16	Australia	4
7	Switzerland	10	17	Belgium	4
8	UK	10	18	Netherlands	4
9	Brazil	9	19	Russia	4
10	France	9	20	Turkey	3

Figure 4.7 shows the geographical distribution of the publications by country, based on the total number of authors for each publication. For example, Turkey has three publications. All eight authors of these publications are from Turkey. Therefore, Turkey's total production is 8, while China's production reaches 292 in this case. This chart displays 32 different countries. When it comes to the number of publications produced by a country, China ranks highest. China possesses three times as many records as its nearest competitor, the US. South Korea, Germany, and Japan follow the United States. While the United States and Asian countries have shown the highest productivity in the field of APC, it is noteworthy that there is a lack of African scholars (with the exception of Morocco and Saudi Arabia) who have conducted research on this topic. This phenomenon may be attributed to the close proximity of major patent offices, collectively known as the IP5 group, which comprises the patent offices of Europe (EPO), the United States (USPTO), China (SIPO), Japan (JPO), and Korea

(KIPRIS). APC research seems to be carried out in both developed and developing countries. Upon examination of the roster of nations that have submitted patent applications [138]. similar results emerge. This emphasizes the importance of the APC issue once again.

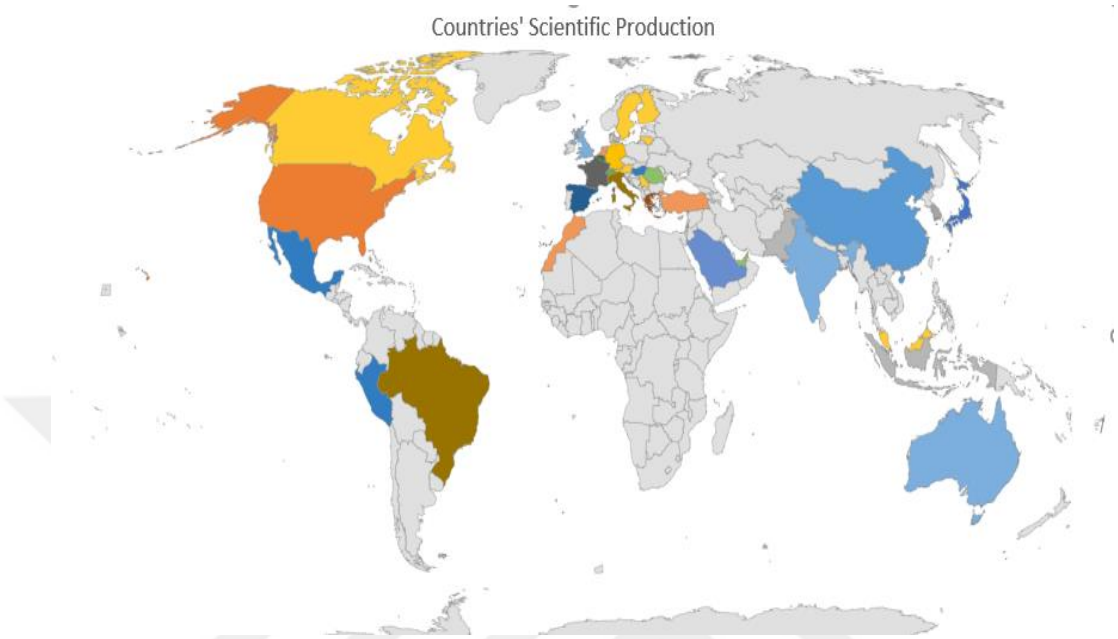


Figure 4.7. Geographical distribution of papers based on all authors, 1982-2023

Figure 4.8 illustrates those publications resulting from various collaborations are less favored in studies on APC. Upon examination of publications produced through collaboration between local and foreign researchers in various countries, it is evident that the method of inter-country study is significantly more prevalent than intra-country study. Out of the top 20 countries, seven opted not to engage in inter-country cooperation. It conducts intranational studies, including those focused on China, the United States, and Korea. Out of the publications conducted in China, 11 are global studies and 52 are local studies. With the exception of three studies, all research conducted in the USA is domestic in nature. Likewise, all except one of the South Korean studies are local studies.

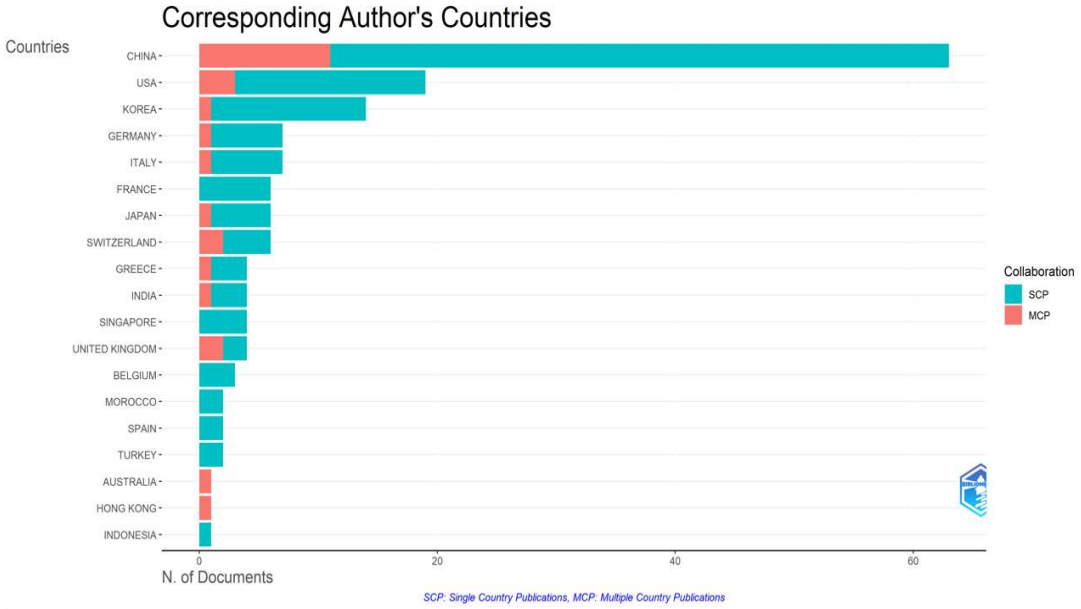


Figure 4.8. The countries of the corresponding authors are presented based on the number of papers they have published on the topic of APC between 1982 and 2023 (SCP: the intra-country, MCP: Inter-country collaboration)

4.3.9 Citation analysis

The quantification of the quality of a publication is commonly achieved through the assessment of the number of citations it receives, which is a popular bibliometric method [139]. Figure 4.9 illustrates the annual citation count for articles published in the field of APC. It is noteworthy that all APC papers have collectively received 3585 citations.

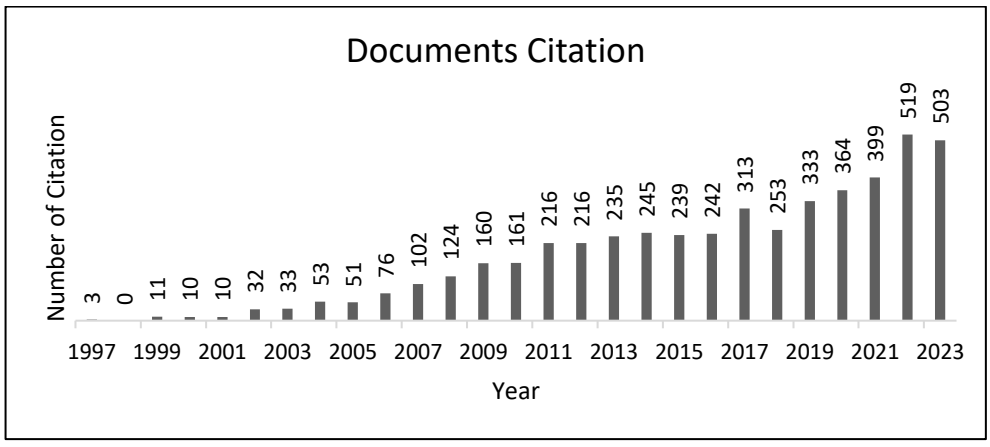


Figure 4.9. The number of citations of the documents published in the field of APC by years

4.3.10 Most cited publications

Table 4.5 presents the top 10 most cited publications on APC worldwide. The term "local citation" refers to the frequency with which a document has been referenced by other publications within the 235-document. The term "global citation" refers to the overall Scopus citation count for the article. There is a disparity between the counts of citations at the local and global levels. This indicates that scholars from different fields also demonstrate an interest in APC. This finding reveals the presence of citations from articles that are not displayed in the research results.

Tseng et al. [140], receive first place in global and local citations in the top ten list of most cited articles in the field of APC. The scope of this article provides an in-depth analysis of a text mining approach. The main goal of the authors is to automate the entire process by creating an analysis based on patent mining. The study of Chakrabarti et al. [18] comes in second place. Despite the fact that the paper received 561 global citations, the total number of local citations remained at 6. An iterative relaxation algorithm is employed in this study to develop a novel method for automatically classifying hypertext.

One of the most locally cited articles in this study is [66]. However, it holds the 7th position in global citation rankings. In this study, the classification of patents is automated through the utilization of pre-trained neural network models. It is possible to conduct two-stage research with the prototype system created. Ranked third with 294 global and 6 local citations, the study by, [54] introduces a novel hierarchical classification method that extends the scope of SVM learning. This method deals with task-specific loss functions. The effectiveness of the proposed method has been tested through experimental approaches.

Table 4.5 shows that four publications with fewer than five local citations are in the top ten list due to the high number of global citations. The automatic classification of texts and the listing of publications linked to APC when searched under the primary topic heading reflect the worldwide popularity of the topic. Figure 4.10 illustrates the top ten articles based on citations at both the local and worldwide levels. Figure 4.11 depicts the entire article plotted against the global citation rate using the VOSviewer tool.

Table 4.5. The most global and local cited documents ranked according to global citation

Document	Year	Authors	GC	LC
Text mining techniques for patent analysis[140]	2007	Tseng, Lin & Lin	627	23
Enhanced hypertext categorization using hyperlinks [18]	1998	Chakrabarti, Dom & Indyk	561	6
Hierarchical document categorization with Support Vector Machines [54]	2004	Cai & Hofmann	294	6
Text analysis and knowledge mining system [141]	2001	Nasukawa & Nagano	147	1
"Term clumping" for technical intelligence: A case study on dye-sensitized solar cells [142]	2014	Zhang, Y., Porter, A.L., Hu, Z., Guo, Y., Newman, N.C.	136	2
Monitoring emerging technologies for technology planning using technical keyword-based analysis from patent data[143]	2017	Joung, J., Kim, K	119	4
Development of a patent document classification and search platform using a back-propagation network [66]	2006	Trappey, Hsu, Trappey, & Lin	110	22
Patents and nanomedicine [144]	2007	Bawa	100	1
DeepPatent: patent classification with convolutional neural networks and word embedding[45]	2018	Li, S., Hu, J., Cui, Y., Hu, J	97	23
Grouping of TRIZ Inventive Principles to facilitate automatic patent classification [22]	2008	Cong & Tong	96	9

GC: Global citation, LC: Local Citation

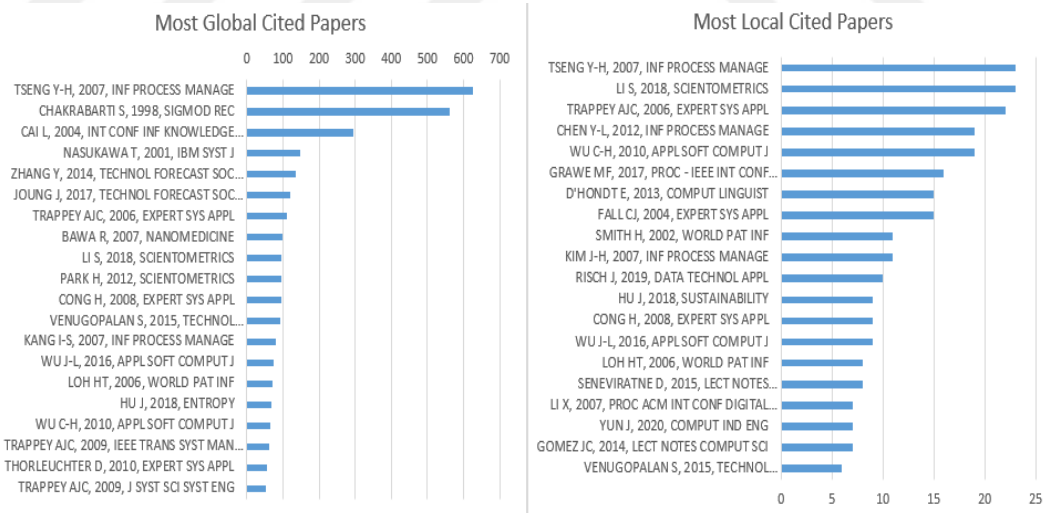


Figure 4.10. The most global and local cited papers from 1982 to 2023

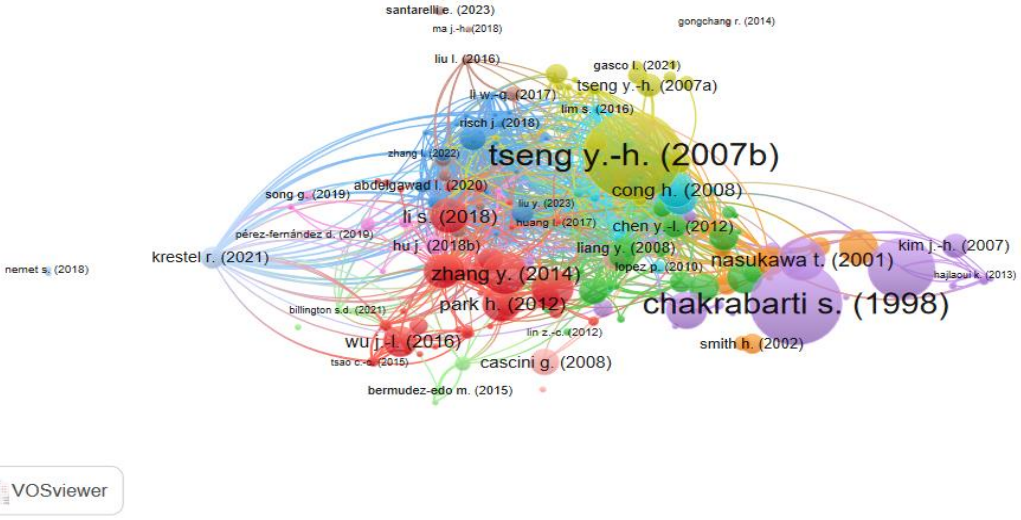


Figure 4.11. Document global citation

4.3.11 Most Cited Countries

The most significant thing following the release of a publication is its citation in other studies. The impact of the findings presented in the paper on the advancement of science may be limited if they are not referenced. The absence of citations in a publication may not allow for the introduction of a new trend or a different perspective. The availability of citations enhances the visibility of the article. Figure 4.12 depicts the most frequently cited countries. The top three countries are USA, Taiwan and China Similarly, these countries also have a high number of publications. This phenomenon may be attributed to the presence of publications with global collaborations. The perspectives and visions of authors from different nations vary from those expressed by authors from the same country. This exerts a positive impact on the quality of the publication, as evidenced by its citation count. China has shown a growing interest in the field of APC over the past decade, with over half of the publications emerging in the last five years. China has been the subject of 25 publications within the last 5 years, resulting in 745 citations across a total of 61 documents. Over the past five years, America has been cited 1698 times in 37 publications, with 11 publications specifically focusing on this topic. Taiwan lacks a publication on this topic post-2021. Nevertheless, this does not present a challenge in terms of attribution. Taiwan has 28 of the 235 publications published in the APC field. The publications have received 1303 citations. Given this circumstance, publications on APC from Taiwan are deemed significant in the global academic community. The

prevalence of citations from Taiwan could be attributed to the relatively older publication dates compared to those from other countries. One possible explanation for the high citation numbers of countries like Germany, Japan, and Switzerland is the publications resulting from international collaboration.

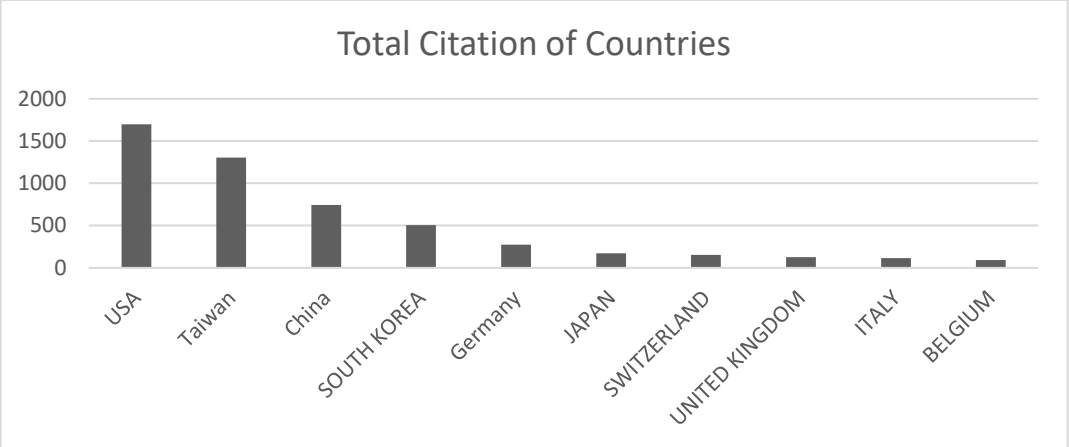


Figure 4.12. Most cited 10 Countries on APC topic (1982-2023)

4.3.12 Keywords Analysis of Research on APC Study

In this section, keyword analysis will be done in two stages. Firstly, the words of the authors will be analyzed. Secondly, the evolution of the APC method will be analyzed with the help of word analysis over a period of 15 years. Based on the words, this section will be finalized with the trend analysis of APC publications. The co-occurrence network map of the keywords (refer to Figure 4.13) and the timeline view of the keywords (refer to Figure. 4.14) will be presented. In a total of 235 publications related to APC, a combined total of 5375 keywords were obtained. In the co-occurrence analysis of the authors' keywords, a minimum threshold of five words is accepted. A total of 266 out of 5375 keywords meet this threshold. Vosviewer identifies 60% of the most pertinent terms. A total of 160 terms have been chosen for analysis. Five keyword clusters were formed.

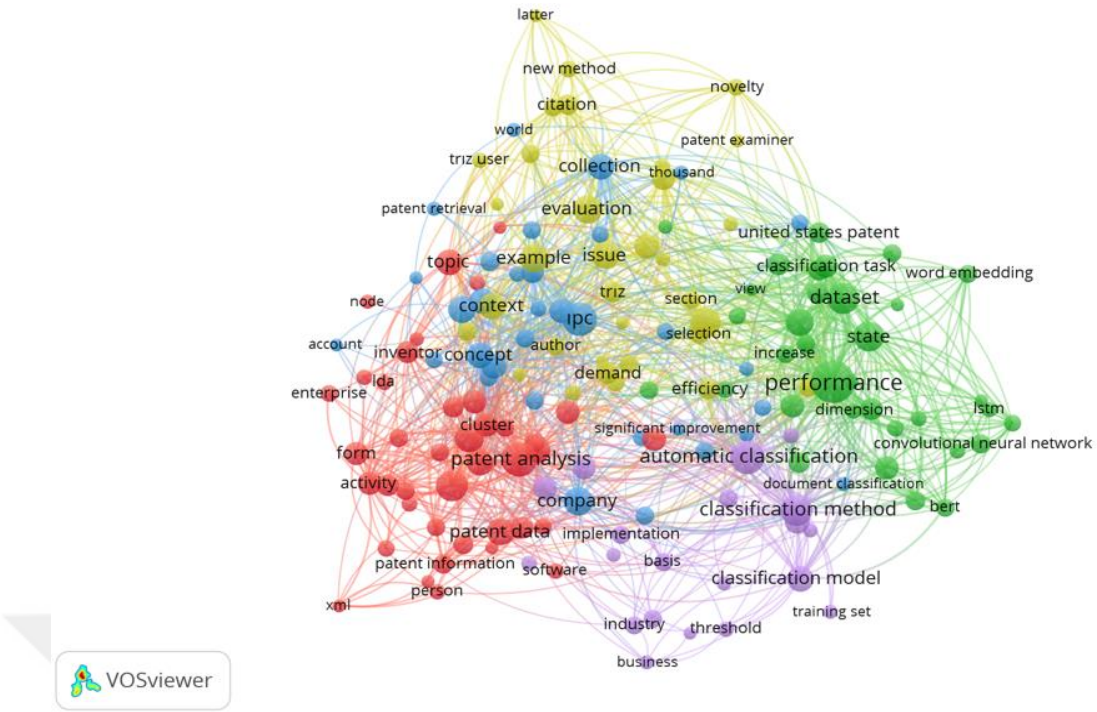


Figure 4.13. Keywords co-occurrence network of APC-related publications

In Cluster 1 (red color), keywords such as database, latent dirichlet allocation, ontology, support vector machine, method and patent analysis are clearly related to the topic “patent classification methods”. In automatic patent classification, a variety of methods, including ontology-based networks and support vector machines (SVM), are employed to enhance the system's quality and ensure its sustainability. The keywords in Cluster 2 (green color) include studies on researching the superiority of the methods to be used in patent classification over each other, which involves studies on deep learning, Bert, LSTM, deep learning, CNN, attention mechanisms, etc., concentrated on the aspect of ‘‘Deep patent classification". In Cluster 3 (indicated by the blue color), the keywords such as information retrieval, big data, patent document classification, large number, are focused on the aspect of "Patent big data analysis in information retrieval". Cluster 4 (yellow color) comprises keywords such as new method, novelty, patent domain, decision maker, TRIZ, automation etc., are center focused on advancement development of new methodologies and automation systems within patent classification. Meanwhile, the last Cluster 5 (purple color) focuses on patent text classification techniques and applications, which is related to difficulty, classification model, text mining, industry, function, classification method, etc.

4.3.13 Time span analysis

Two distinct illustrations are utilized in the time span analysis section. The first figure is Figure 4.14 from VOSviewer, and the second is Figure 4.15 from Bibliometrics. The time interval is restricted to a 20-year period from 2003 to 2023. Figure 4.14 is utilized due to its ability to synthesize contemporary issues. Furthermore, Figure 4.15 is preferred due to its superior representation of temporal transitions. When generating both figures, a minimum word frequency value of 5 was utilized. In Figure 4.14, color is noted in the words as time progresses. Words that were popular in previous eras are depicted in darker shades of blue and green, whereas words in contemporary articles are represented in yellow tones.

The field of text mining began to gain attention in the early 2000s, primarily through the analysis of papers related to patent categorization. However, its significance in the patent field increased notably towards the end of the 2000s [141], [145]. Classical classification methods have been increasingly utilized in computerized patent classification procedures, incorporating the concept of hierarchical categorization of documents [54], [146]. Classification algorithms such as Support Vector Machines (SVM) and k-Nearest Neighbours (kNN) were widely utilized during this period [68], [99]. For an extended period, scholars have opted for the automated classification of patents using TRIZ [21], [22], [147].There are publications with the idea of spreading the concept of fuzzy and applying it as a solution to APC problems [148]. The application of neural networks in patent classification has emerged with advancements in the field of neural network studies. Subsequently, researchers have employed the artificial neural network method in conjunction with a variety of heuristic and/or fuzzy methods [98], [149], [150]. Processes of classification based on words and phrases have emerged. During this period, there was a rise in ontology-based studies [19], [151]. Several topic classification methods, including the LDA method, have been utilized in the automated classification of patents based on the concept of topic classification [152], [153], [154]. As the concept of big data expands, the word embedding method has been used in patent papers [119], [155]. Consequently, there was a surge in research on deep learning. Numerous studies have been published in this field, employing techniques including word embedding, CNN, LSTM, and RNN used [104], [117], [121], [156]. Subsequent advancements in transformer techniques

were later developed [82], [157]. In light of the progress made in this field, the predominant approach employed is BERT [78], [106], [107], [158], [159].The strategies and studies utilized in addressing the APC problem remain unsatisfactory, primarily because of the hierarchical nature of patent classification systems. The combination of textual content and semantic information can enhance the accuracy of categorization results in analysis. Classical methodologies were unable to effectively handle big data, while deep learning techniques continue to struggle to attain the necessary level of accuracy and consume excessive time. Staying abreast of advancements in computer science is crucial for facilitating the rapid adoption of new technologies.

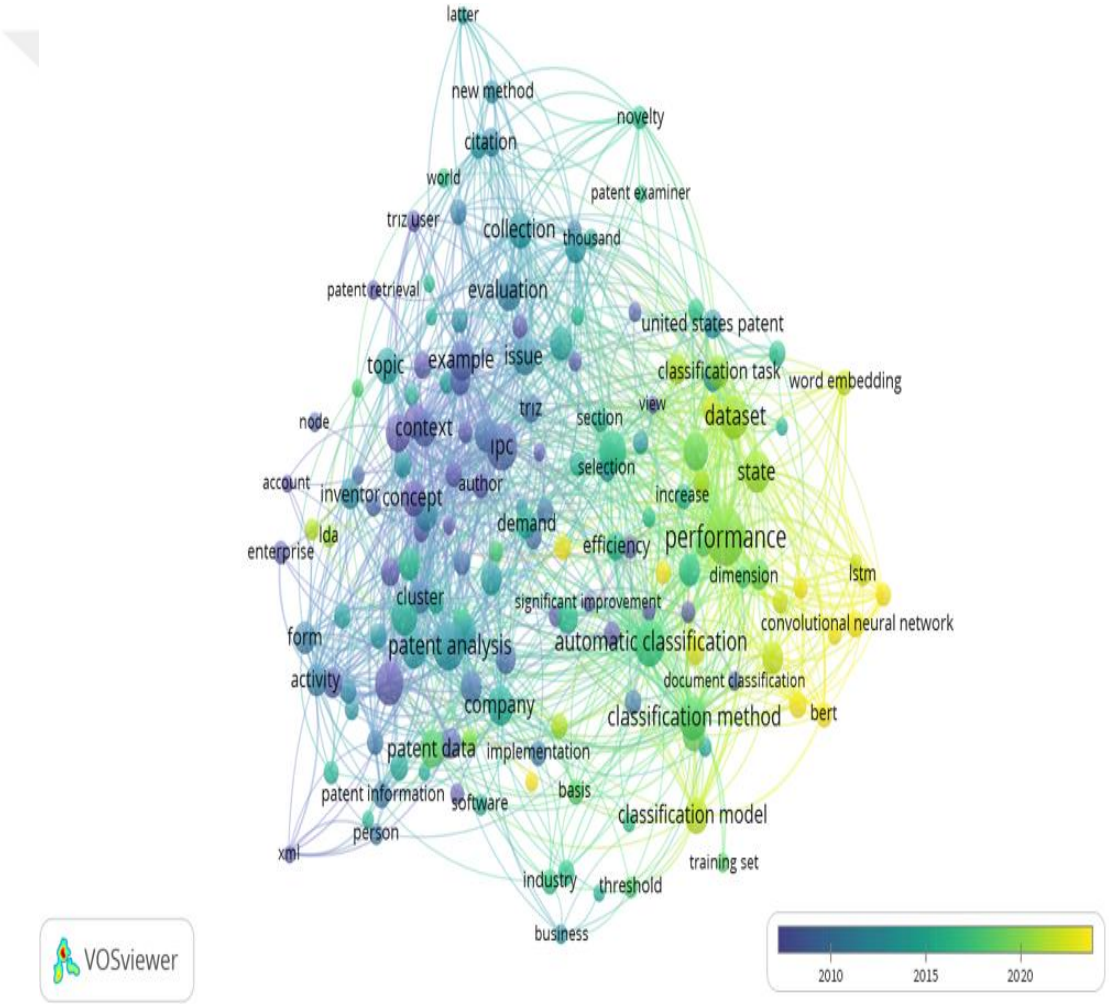


Figure 4.14. The overlay visualization of author keywords

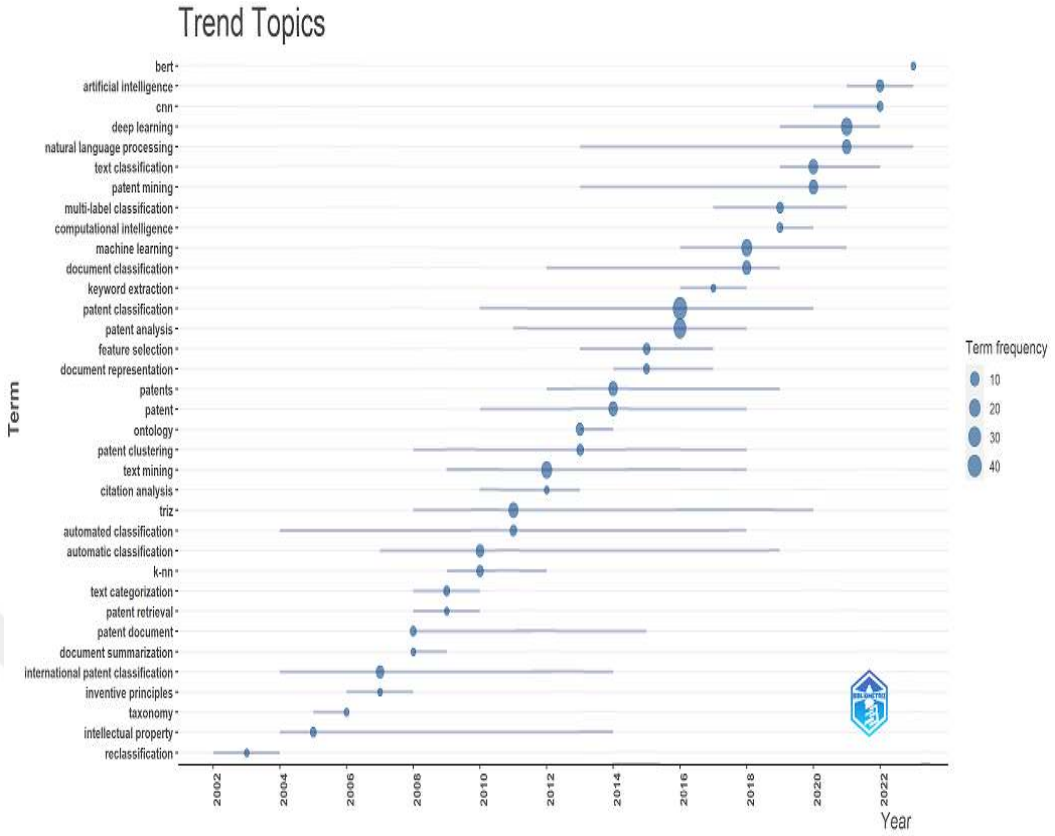


Figure 4.15. Trend topics of APC-related studies

4.4 Current APC-Related Patents Status

The Derwent Innovations Index®, accessible through the Web of Science platform, was utilized to gather patent information for the acquisition of a patent dataset pertaining to advancements in APC technologies.

4.4.1 Patents Subject Area

The distribution of patents across different subjects is illustrated in Figure 4.16, based on the findings of the analysis. The Engineering category encompasses all 71 listed patents. 69 of them fall into Computer Science. This is consistent with the findings of the APC paper research. There are nine patents within the Instruments/Instrumentation category. This category of patents is centered on enhancing categorization performance. Seven patents are in Telecommunications. Advancements in patent classification devices are encompassed in the patents related to this subject. Additionally, it has one patent registered in the expert patent classification system for the Automation Control Systems it has developed.

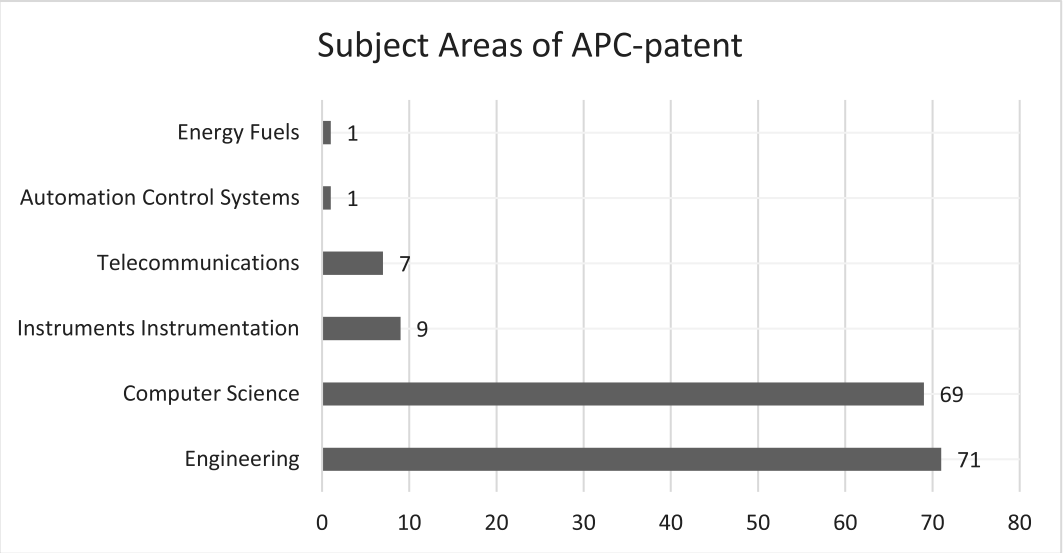


Figure 4.16 Patent subject area

4.4.2 Patent Assignees/Country/IPC Codes

Table 4.6 displays information regarding the assignee, country, and IPC code for patents published in the APC field. Among the top 10 patent holders with the most patents, technology companies are leading the list (academia was prominent in the analysis of APC publications). The list includes three examples from academic sources. The assignees represent four distinct countries: China(5), South Korea (3), Japan(5), and the United States(1). When comparing the list of academics with the highest number of publications, it is evident that the motivation to obtain a patent differs from the motivation to publish. There is no single author or assignee who simultaneously possesses a patent in the field of APC and publishes academic literature in the same field. It appears that China, South Korea, and Japan exhibit a greater propensity than other nations to engage in both academic publishing and patent acquisition within the field of APC. Taiwan is notable for its presence in academic publications, yet it is not included in the list of patents. Another significant issue is the absence of European universities or technology institutions in the top 10 rankings. The primary cause of this issue is believed to be the concentration of major company headquarters in the United States and Far Eastern nations like Japan, China, and South Korea[143].

Table 4.6 APC-Related Patents: Descriptive Results

Assignee Names	Record Count	Country	IPC Codes	Patent Count
Jiangsu Runtong Data Services Co Ltd	3	China	G06f-017/30	28
Fujitsu Ltd	2	Japan	G06f-016/35	20
Hitachi Ltd	2	Japan	G06q-050/18	16
Korea Inst Patent Information	2	South Korea	G06f-017/27	10
Korea Inst Sci Technology Information	2	South Korea	G06k-009/62	10
Lg Electronics Inc	2	South Korea	G06n-003/08	9
Nec Corp	2	Japan	G06f-016/33	8
Toshiba Kk	2	Japan	G06f-040/30	7
Toshiba Solutions Corp	2	Japan	G06n-003/04	6
Ubiq Inc	2	USA	G06f-017/00	4

The IPC codes in Table 4.6 have been analyzed, revealing that all of them fall within the G section and G06 class. There exist four distinct subclasses. The categories include G06f, G06k, G06q, and G06n. These subclasses play a crucial role in categorizing patents related to the APC field, and researchers will utilize them when cataloging patents in this domain.

4.4.3 Keywords Analysis of Patents on APC Study

In this section, the words found in patents related to APC are clustered based on co-occurrence networks. The analysis involves identifying the subject that these clusters are attempting to convey.

The most commonly used keywords in the APC, as indicated in Figure 4.17, include "document classification devices, information retrieval, and computer system." Through the analysis of these words, it is possible to identify eight major clusters for proper categorization.

The terms "classification device, classification process, computer program product, and computer system" clearly pertain to the development of computer programming that can be utilized in classification tasks within Cluster 1 (red color). New gadgets produced in this field should have the capability to be utilized in the categorization of

large datasets and in patent classification process. The terms "improper investigation, information leakage, efficient classification, clustering result identified in Cluster 2 (green color) denote the mechanisms designed to mitigate inaccurate classifications resulting from insufficient information. In Cluster 3 (blue color), keywords "automatic classification," "computer device," "patent text," "search," and "operation" are centered on the "automatic classification of computer devices." Within Cluster 4 (indicated by the yellow color), the keywords "information retrieval method," "identification item," and "word vector," among others, are associated with the subject of generating information retrieval methods. Meanwhile, Cluster 5 (purple color) includes keywords such as "computer-readable storage medium, correlation, comprehensiveness, etc.," which are primarily focused on the advancement of information storage technologies. Cluster 6 (turquoise color) is centered on the creation of electronic devices through the utilization of deep features for automated classification, which is associated with concepts such as "output, electronic device, advance, automated classification, etc." In Cluster 7 (orange color), terms like "machine learning," "classification information," and "processor" are linked to techniques developed through the use of machine learning. Cluster 8, represented by the brown color, includes keywords such as "problem," "attribute," "category number," and others, centering on the concept of "automatic code generation attribute."

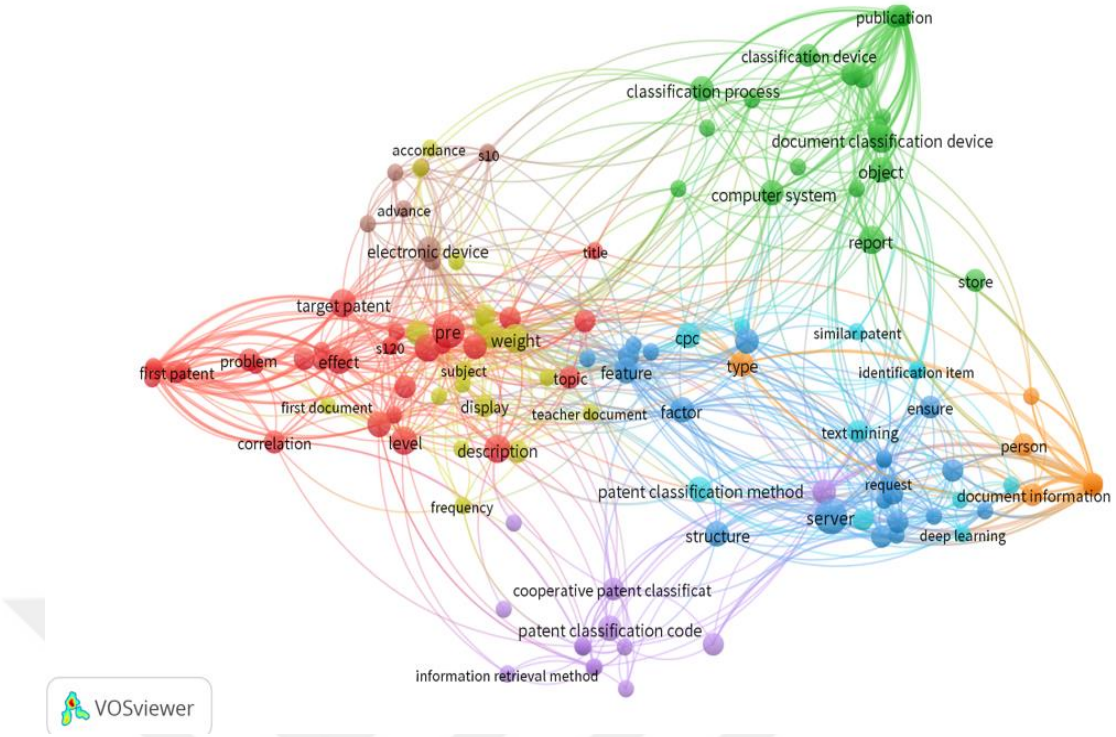


Figure 4.17. Representation of APC-related patents in the network

4.5 Limitations of the Bibliometric Analysis

The primary constraint of this study is the inability to conduct nonprobability sampling. A specific number of research articles were retrieved from the database by employing specific keywords. Nevertheless, generalizing the results is not feasible. The research results may not include all relevant records that should be included in the list. A comparable constraint may also apply to patent searches. The patent search conducted in the Scopus database resulted in approximately 17,000 pieces of patent data. This value corresponds to 78 in the Web of Science (WOS). WOS output was favoured in order to attain more targeted outcomes. In this instance, similar to paper analysis, generalizing the results is not feasible. Nevertheless, this represents the primary issue in bibliometric studies.

4.6 Conclusion

This study systematically and extensively analyzes 235 academic papers on the subject of APC from the Scopus database, along with 78 patents from the Derwent Innovations Index® database within Web of Science. This study utilized bibliometric analysis tools such as Bibliometrics and VOSviewer to analyze data spanning from 1982 to 2023.

This study investigated all scientific articles and patents published in the field of APC, in line with the specified search criteria. To our knowledge, there is no existing publication in the literature that has undertaken a thorough bibliometric, citation, and keyword analysis of patents and articles within this particular field. This study seeks to rectify this inadequacy. Numerous research areas seek to be comprehensively elucidated, including the most cited studies in the field, the most productive authors, the most cited countries, and the most preferred journals. This study encompasses more than a mere bibliometric analysis. The keyword analysis delves into the 42-year evolution of the APC problem, providing a detailed examination and revealing predictions regarding areas for future study. Several findings have been obtained regarding the publications on APC, which can be summarized as follows:

- The total number of publications (including papers and patents) on APC published in the past decade constitutes half of all publications within the corresponding research field. This phenomenon can be attributed to advancements in computer technology and the direct correlation of the subject of APC with computer science.
- National Tsing Hua University is renowned for its high volume of publications. Seven institutes in China are ranked in the top 20 based on the number of publications in the APC. The number of American authors is almost half that of Chinese authors; however, the situation is reversed when comparing the citation numbers of the two countries. Therefore, it is essential for Chinese scientists to prioritize the quality of their articles. The absence of an impact factor does not necessarily imply a lack of journal preference. The World Patent Information journal provides an intriguing illustration of this phenomenon. This journal achieved its impact factor in 2023, establishing itself as the primary publisher of most publications in this field.
- When analyzing the top 10 most cited papers, it is evident that all publications, with the exception of one, are the result of collaborative efforts involving more than one author. Furthermore, it has been disclosed that the publications originating from the APC domain not only garner local citations, but also attract a greater number of global citations compared to local citations. This

situation demonstrates the significance of the APC subject in the context of global literature.

- The patent assignee list data show that corporations are more dedicated in acquiring patents than institutions. Companies aiming to develop new products try to gather an entire set of patents in the technology class of their choice. APC acts as a workspace where companies can find solutions to these demands. As a result, it is highly reasonable for companies to develop technologies and get patents in the area of APC.
- Based on patent assignee list, it appears that academia is less inclined than companies to acquire APC patents. The absence of universities incentivizing academics to participate in industry and providing sufficient funding may be a contributing factor to this phenomenon. Universities are expected to take the lead in translating their fundamental research into tangible economic outcomes. Broadening criteria, such as patent licensing rights and commercialization, among universities and faculty members, is likely to result in an increase in the number of university patents. Universities ought to promote and support their faculty members in this particular field by providing research funding. There is a need for greater prevalence of university-industry collaborations. The issue of APC is a practical concern that necessitates the integration of theoretical knowledge from academia with industry experience for its resolution.
- An analysis of keywords in academic papers and patents indicates that these documents represent similar scientific outputs. Both documents contain clusters that indicate a tendency for the development of new procedures. Both document categories involve collections of solutions aimed at addressing APC issues using innovative methods such as deep learning. Both analyses include clusters that focus on developing software or tools to enhance the APC system's capability to produce higher-quality results. This is a notable example of the positive impact of a newly proposed method in one scientific article on another.
- Keyword analysis has once again proven to be an excellent approach for identifying hotspots and investigating trends on the topic of interest. The initial research in the field of APC focused on the classification of patents using

various parameters and testing various approaches (such as Triz, SVM, ontology, LDA, and so on), with efforts made to improve the classification results. With the rise of deep learning, research has shifted its focus to various deep learning techniques such as CNN, RNN, LSTM, BERT, and others, leading to widespread adoption in this field. Based on the publishing pattern, it appears that future research will continue to prioritize deep learning. It is possible that heuristic methods could be applied in deep learning. Researchers that publish in this discipline should keep this trend in mind..

This study can be important in the preliminary phase. This guide offers new and young researchers comprehensive access to current data in the field, providing a detailed overview. In future research, patent analysis can be enhanced by combining patents from multiple databases. Patent classification systems, with their hierarchical structure, still lack adequate classifications at the group/subgroup level. Further research is required in these areas. Patent texts use specialized jargon, so unique word embedding graphs are necessary for analysis. These graphs should be regularly updated. Otherwise, the classification accuracy will be low. It is essential to quickly integrate computer technology advancements into the APC research field. There is still a significant amount of time wasted processing big data. The adoption of cloud technology needs to accelerate.

CHAPTER 5
PATENT CLASSIFICATION WITH HIERARCHICAL ATTENTION NETWORK

5.1 Introduction

The fact that the action takes too much time in manual execution of the patent classification process causes the already long process to be further prolonged. Misclassification not only takes time but also creates difficulties in providing up-to-date knowledge of the technique with which the invention is concerned. Automatic patent classification methods have been developed to avoid such difficulties. Patent text can be problematic even with automatic classification. Because the sentences are long and have a complex structure that includes many technical and legal terms. The Hierarchical Attention Network method is used in this chapter of thesis. This method makes classification by including both word and sentence analysis.

The hierarchical attention network model was developed based on the recognition that documents possess a hierarchical structure. Words combine to form sentences. As the sentences combine, paragraphs form. Finally, the paragraphs come together to form the document. Words or sentences alone can have multiple interpretations. The same word or sentence may have a completely different meaning when used in a different context. The hierarchical attention network algorithm aims to interpret the varying sensitivity based on context by incorporating word attention and sentence attention mechanisms. [76].

In this study, Wipo-alpha patent data [160] is analyzed using the HAN method. Analysis results are compared with CNN and RNN methods. Academic studies conducted with the Hierarchical Attention Network method are described under the title of literature review. Therefore, in this chapter, only the explanation of the method and the details of the analysis will be explained. Readers who are curious about the

literature on this method. They can read the relevant section under the literature review heading.

5.2 Materials and Methods

In this study, CNN and RNN methods are used to make comparisons. For both methods, pre-processing is implemented in the same way as HAN. Word embedding application provides vectorization of words. Then in the CNN application; 1 dimensional convolution layers are applied on top of the embedding layer. These may have different sizes and filters. Max-pooling process is applied to the output from the convolution layers. Then the output of the layers feeds softmax. Thus, the classification results are determined by a probability distribution. In the RNN application, the bidirectional LSTM method is applied, in this method the data, propagates backward as well as forward in time. It works with hidden layer so it can direct data in two ways.

5.2.1 Hierarchical Attention Network (HAN)

The Hierarchical Attention Network (HAN) method emerges from the assumption that not all of the sentences in a document can reflect the subject in the document equally, and in the same way, not all words in a sentence can explain the meaning of that sentence equally [161], [162]. The HAN method has four main parts. In the "Word Encoder" phase, a word vector is created based on the interactions of words with each other. Words are encoded using the embedding matrix. Not all words in a sentence are used. Stop words are pre-extracted. Plain text is split into tokens and these are sent to the embedding layers. Word weights are formed as the model calculates word importance. The most important words that contribute the most to the meaning of the sentence are brought together and these are used in constructing the sentence vector. Since this is a hierarchical analysis, the outputs of vectors with low hierarchies will be the input of the next hierarchy. In “Word Attention” phase, an attention mechanism is established in order to understand the meaning that words add to that sentence [76]. The attention mechanism not only improves classification performance but also highlights which word has more importance. In the HAN method, "hidden state" is the combination of backward and forward hidden state which gives an abstract of the information that is presented in the sentence that made up of both directions of the

words. GRU with dynamic bidirectional attention mechanism are used in the HAN algorithm. GRU operates in a selectively permeable structure. With the two-gate access system, HAN keeps control of the information. It protects the information that needs to be stored and leaves other information. Reset gate controls how much previous information will affect the new state. If it has no effect, the value will be zero and the reset gate erases all data from the memory. The latest status is updated with the update gate[76], [161], [162]. Hidden layers (h_{it}) are calculated by the one-layer perceptron method. This allows the HAN method to learn the starting weight (W_w) and decide the bias value (b_w), and with tanh, the data are kept between -1 and 1. At the word level, the attention mechanism is calculated in Eq. 5.1-5.3 as follows:

$$u_{it} = tanh\,(W_w h_{it} + b_w) \quad 5.1$$

$$\alpha_{it} = \frac{\exp\,(u_{it}^T u_w)}{\sum_t u_{it}^T u_w} \quad 5.2$$

$$s_i = \sum_t \alpha_{it} h_{it} \quad 5.3$$

where "u_{it}" is importance of words, "u_w" the word-level meaning vector, "α_{it}" the normalized weight and vector s_i is the sentence vector[76]

The “Sentence Encoder” part has similar steps with the word encoder phase. The hidden layer for sentences provides a summary indicating the center sentence i and neighbouring sentences around it. Sentences are coded using the bidirectional GRU. Sentences are made up sequence of words so, they are vectorized as same logic as word vectors. In the “Sentence Attention” part, attention mechanism is used to conclude that the sentences are the ones representing the document. Document vectors are reached by using sentence vectors [76]. Figure 5.1 shows the general structure of Hierarchical attention network developed by [76].

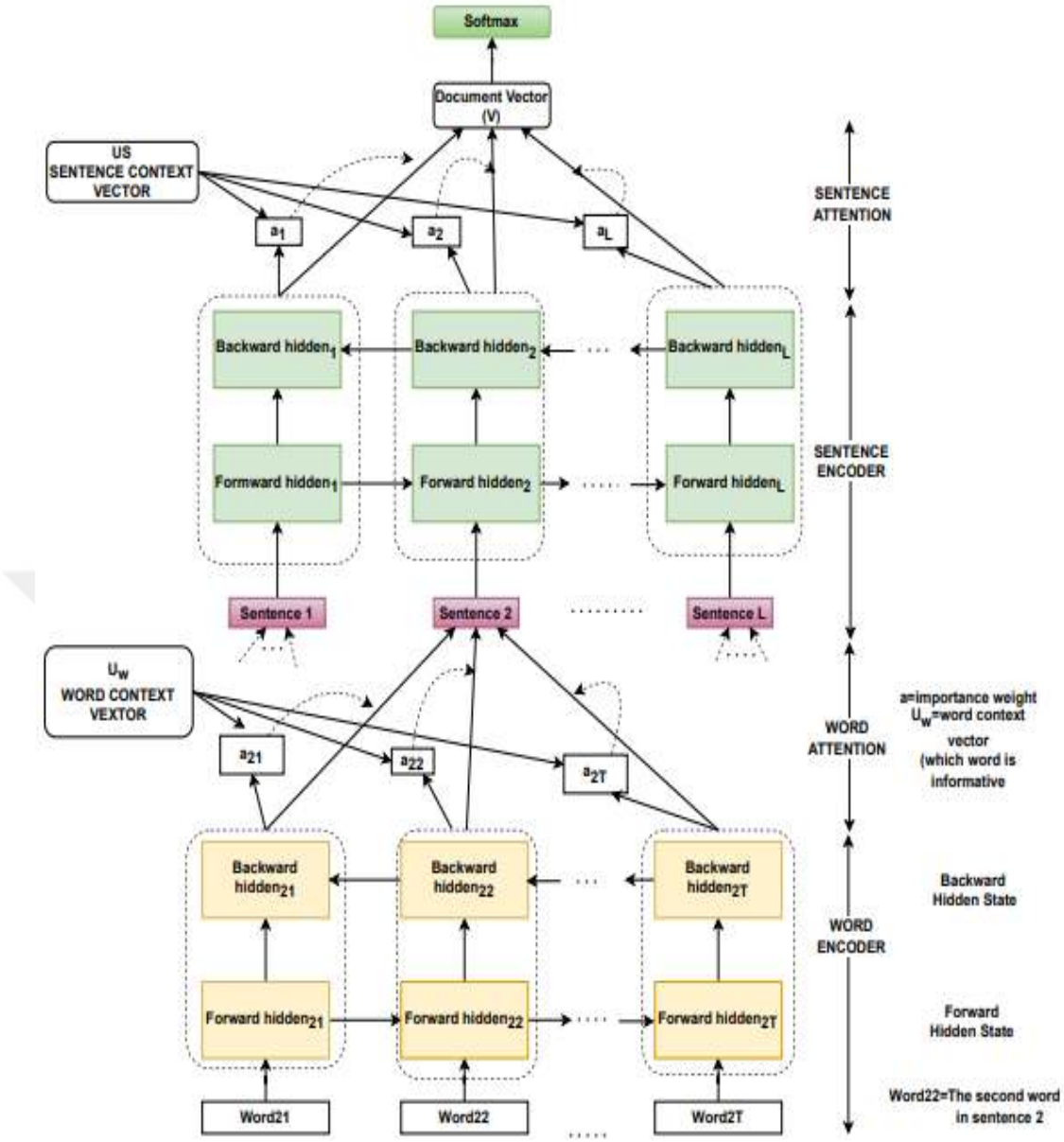


Figure 5.1 HAN structure (Source [76])

The following parts describe the comparison methods used for analysis in this part of the thesis. Moreover, the optimizers used in this analysis is also explained.

5.2.2 Convolutional Neural Network

The CNN method can be applied in hierarchical document classification method. Although it is originally introduced for computer vision analysis, CNN is also very common for using text classification analysis in deep learning studies [163], [164]. CNN method, which is one of the artificial neural network models, has feed-forward structure with multilayer concept. It contains different layer types such as

convolutional layer and pooling layer. In the first step of CNN method, convolutional layers are used as n-gram detectors [165]. Convolutional layers have set of filters. Convolutional layer is used to extract the features of the input data. It has set of learnable filters. In the convolutional layer, the dot product operation between this kernels and input data is made to find the features of data [163], [166]. In order to make a decision, in second step max pooling is used. In the pooling, the different dimensions of the vectors are combined and transformed into an only one-dimensional vector. In literature there are various pooling methods. One of the most applied one is the max-pooling method. In this pooling method, the algorithm computes the max value of the vector to make a decision. Transferring feature maps carried by the pool layer to the next layer uses the flat layer so that the layers are combined and converted into a single layer [167]. In the last step network classifies the text according to information from the outputs of previous steps of the algorithm [164].

5.2.3 Recurrent Neural Networks

A recurrent neural network (RNN) is a kind of feed forward neural network. It has multi hidden layer structure. The RNN method has memory to process sequences of input data. When output calculated, it is sent back into the recurrent network in order to make a decision. The decision is chosen between the current output result and previous output. The weight assignment is different in RNN algorithm. The predecessor data has more weight than posterior [164]. Because of this feature of RNN, it is a suitable method for text or sequential data classification. The Long Short-term Memory (LSTM) is a special kind of RNN algorithm. LSTM is a method that solves the vanished gradient problem of RNN. This problem occurs when more layers added to activation functions, the loss function's gradient value converges to zero which makes the training of that network hard. LSTM applies multiple gates in order to coordinate the entrance of information pass into each node. The LSTM algorithm equations (Eq.5.4-5.9) are as follows [164]:

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i), \quad 5.4$$

$$\mathcal{C}_t = tanh(W_c[x_t, h_{t-1}] + b_c), \quad 5.5$$

$$f_t = \sigma\big(W_f[x_t, h_{t-1}] + b_f\big), \quad 5.6$$

$$C_t = i_t * \check{C}_t + f_t C_{t-1}, \quad 5.7$$

$$o_t = \sigma(W_o[x_t, h_{t-1}] + b_o), \quad 5.8$$

$$h_t = o_t tanh(C_t), \quad 5.9$$

where W symbolize the weight matrix and b stands for bias vector, *t* is time and *xt* is input vector. *i*, is input gate, *c* is cell memory gate, *f* refers to forget gate and *o* is output gates. The Equation (4) defines input gate, Equation (5) stand for the candid memory cell value, Equation (6) represent forget-gate activation, Equation (7) computes the new memory cell value, and Equations (8) and (9) represent the last output gate value.

In this study BILSTM, method is applied. BILSTM means bidirectional LSTM, in this method the data, propagates backward as well as forward in time. It works with two hidden layers so it can direct data in two ways [168].

The following part, describes pre-processing, word embedding and optimizers methods used which are used inside of the algorithms.

5.2.4 Pre-processing

Data pre-processing is the process of preparing data before subjecting a data set to analysis. It is a process of transforming data suitable for the use of the model by removing unwanted variability in the data [169]. Data is pre-processed before the word embedding process is performed [170]. Wipo-alpha patent dataset was prepared for analysis. The applied pre-processing steps can be listed as follows: First, xml parsing and tag removal are applied to the dataset. Missing data is cleared. Input data is selected. All letters are converted to lowercase letters. Lemmatization and tokenization process is applied respectively. Finally, tokens are ready to be converted into word embedding matrix. The pre-processing step of the data before embedding is illustrated in Fig. 5.2.

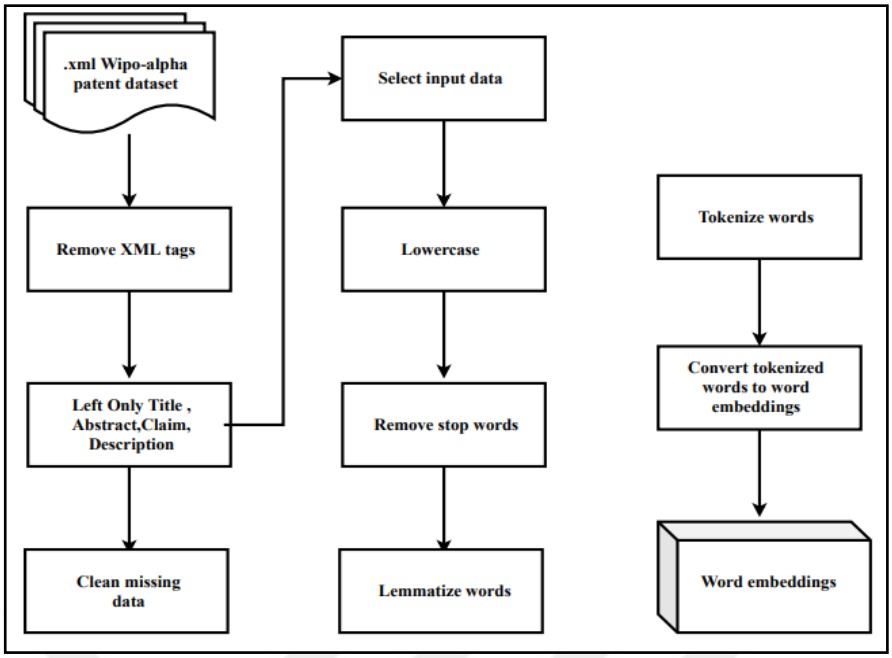


Figure 5.2 Preprocessing steps for Wipo-alpha patent dataset

5.2.5 Word embeddings

Word embedding is the operation of converting words and expressions into vectors made up of real numbers [171]. It is the placement of words from a multidimensional space created for each word to a lower dimensional and continuous vector space. One of its most important advantages is that it can reflect the semantic knowledge of words and expressions. Words with similar meanings are also positioned close to each other in distance. Words with different meanings are positioned further in distance. There is different word embedding methods in the literature [172]. For example, GloVe (Global Vectors), Word2Vec, FastText and ULMFit (Universal Language Model Fine-Tuning) are some of them. They have several different dimensions within themselves. The Glove word embedding model was created by researchers at Stanford University. It is derived from the words Global Vector. It is a model that reflects the number of words in a document and the probability of them occurring simultaneously. When two embedded words are multiplied scalarly, they are likely to occur at the same time throughout the entire document. In word embedding, a single word is defined in an n-dimensional vector space. GloVe with 100-dimension word embedding is used in this study [173].

5.2.6 RMSPROP (Root Mean Square Propagation) Optimizer

The RMSProp is a kind of stochastic gradient optimization algorithm. For each parameter, it adopts its learning rate. This adaptation is calculated as dividing the learning rate into the average of recent grants weights [174].

5.2.7 ADAM (Adaptive Moment Estimation) Optimizer

It is one of the methods, which is used for optimization. It is an extension of stochastic gradient descent algorithm [174]. Adam stores exponentially decaying average of past gradients and it keep in memory only two moments. It overcome the problem of non-stationary objective function [167]

5.3 Experimental Analysis

In this chapter of the thesis, the study investigates the impact of different hyperparameters on the classification accuracy of the applied model through experimental analysis. This study focuses on the single-label multi-class classification problem. Patent documents contain multiple labels and classes. However, this study considered the "main code" assigned to patent documents. The details of the experimental analysis are presented in Table 5.1, which consists of 10 steps.

Table 5.1 The experimental analysis definition

#	**Experimental Analysis Details**
1	Two different dataset is used in the analysis. First one is Wipo-alpha with whole sections. Second one is Wipo-alpha D section, with a total of 1,710 documents.
2	Wipo-alpha whole dataset is used to compare the HAN method with other deep learning methods.
3	Experimental analysis is performed with Wipo-alpha D-section data which is smaller than Wipo-alpha whole dataset.
4	The different texts in the patent document are evaluated individually and in different combinations as input.
5	In order to understand the difference between the gates methods GRU and LSTM are tried with three different values (50, 100, and 150).
6	In order to understand their classification performances two optimization methods namely Adam and Rmsprop are compared with each other.
7	Two different dropout values are tried in the classification.
8	In the attention layer, initial values are determined using a uniform distribution.
9	For each classification trial, experimental analysis has run for 50 epochs. But experimental analysis for epochs also is made for values 20, 50 and 100.
10	Different batch sizes (2, 10, 15, 20, 25, and 64) are tried in model HAN.

The study is performed with Google Collaboratory -on the Tesla V100-SXM2-16GB and python programming language is utilized.

5.3.1 Evaluation Metrics

The performance of the models in this thesis study is evaluated based on the formulas in the Eq 5.10-5.13. Accuracy, F1 score, sensitivity, and precision values are used in the evaluation process to interpret the results. The accuracy value is calculated by dividing the number of correctly predicted patents by the total number of predictions. It provides a comprehensive result from the confusion matrix, which contains all the outcomes. The precision value indicates whether the predicted patent classes are actually correct for those that were correctly predicted. The precision value indicates the accuracy of predicting the correct class. The F1 value is calculated by taking the harmonic mean of sensitivity and precision values. The evaluation criteria include precision, sensitivity, F1 score, and accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad 5.10$$

$$Precision = \frac{TP}{TP+FP} \quad 5.11$$

$$Recall = \frac{TP}{TP+FN} \quad 5.12$$

$$F1 = 2 \times \frac{precision \times recall}{precision+recall} \quad 5.13$$

True Positive (TP) and True Negative (TN) are the areas that the model predicts correctly, False Positive (FP) and False Negative (FN) are the areas that the model predicts incorrectly.

5.4 Results and Discussion

A series of patent classification experiments on the patent documents at the section and subclass level (8 section and 451 subclasses) of the Wipo-alpha dataset are carried out. These experiments can be listed as follows: a) the whole data set is classified at section level and the effect of HAN method on the results is examined, comparison analysis is made with CNN and RNN methods, b) the subset of the Wipo-alpha data set (Wipo-alpha D-section) is classified at the subclass level and the effect of the HAN method on the results is examined, c) which gate method is more effective in HAN classification LSTM and GRU comparison is made, d) the effect of different dropout values on the classification are tested, e) the change in classification performance

according to different epoch values is searched, f) classification accuracy variation based on different batch size values is investigated, g) condition evaluation on the classification accuracy value according to different optimizer values is made, h) the effect of using patent texts separately as an input value or combination of them on the classification performance are examined.

5.4.1 The Whole Data Set Classification at Section Level

Patent data has four different text parts. These are title, abstract, claim and description. In the HAN method these parts are used as an input data and words are extracted from them separately. To make a comprehensive analysis of the classification performance of HAN, the comparison is performed with two state-of-the-art methods, CNN and RNN. The result of this calculation is given in Table 5.2

Table 5.2 The comparison results of experimental analysis for various patent text part

Method	**Abstract**		**Claim**		**Description**	
	Acc	Time (s)	Acc	Time (s)	Acc	Time (s)
CNN	0.67	1381	0.68	227	0.61	383
RNN	0.65	1420	0.68	219	0.63	356
HAN	**0.70**	962	**0.74**	1963	**0.78**	12186
Input=Wipo alpha-Section Analysis-75250 data- Epoch=50, Batch Size=50, Dropout=0.2, Optimizer=Adam, Acc=Accuracy						

HAN method provides superiority over CNN and RNN methods in abstract, claim and description classification. It gave the best results in terms of accuracy in classification of abstracts with a score of 0.70, classification of claims with a score of 0.74 and classification of description with a score of 0.78. Title classification is not considered alone. This is because the HAN method works with sentences and the algorithm will not work since the title will not be perceived as sentences. However, when combining texts in the other comparison Table 10, it is thought that the title could be perceived as another sentence and is included in other analysis. When the methods are compared in terms of time expenditure, HAN method has achieved results by spending less time in abstract texts compared to other methods. However, this situation does not give the same result in texts with long sentences such as claim or description texts. The most problematic situation encountered in the implementation of this algorithm is that the time increases too much as the text size increases. On the contrary, CNN and RNN

methods have made classification in a much shorter time than HAN in claim and description analysis. They could not analyze the abstract texts at the same speed.

5.4.2 The Whole Data Set Classification at Subclass Level

The comparison results of classification at subclass level are given in Table 5.3.

Table 5.3 The comparison results of experimental analysis for different methods at subclass level

Method	Accuracy	Time(s)
CNN	0.33	1713
RNN	0.28	430
HAN	0.40	2795
Input =Abstract, Wipo alpha 75250-Subclass level, Epoch=50, Batch size= 128, Optimizer =adam		

As a result of the analysis, the highest classification accuracy is achieved with the HAN method. However, this result is below the expected level. In terms of the time spent, 6.5 times as much time is spent in the HAN method compared to the RNN method. In the HAN method, 60% more than the time spent in the CNN method is spent.

5.4.3 The Comparison of two Gate Unit LSTM vs. GRU

LSTM and GRU are two methods seen as substitutes for each other. In terms of the time spent, it can be said that GRU generally takes a little shorter than LSTM. This may be due to the parameters it has; in analysis with GRU, while processing occurs with 2 gates, LSTM performs with 3 gates. As the gate unit value increased, the time spent for analysis both in LSTM and GRU increase. However, although the increase in the value of the gate unit appears to increase the LSTM classification performance, the same is not true for the GRU. GRU has its lowest performance when the gate unit is 100. The comparison results of GRU / LSTM are shown in Table 5.4.

Table 5.4 The comparison of two gate unit LSTM vs. GRU

		Gate Unit Value 50		**Gate Unit Value 100**		**Gate Unit Value 150**	
Dropout	Gate Unit	Acc	Time(s)	Acc	Time(s)	Acc	Time(s)
0.2	LSTM	0.57	61	0.60	73	0.61	91
	GRU	0.58	61	0.55	61	0.58	90
0.5	LSTM	0.59	61	0.59	73	0.61	92
	GRU	0.60	61	0.56	62	0.58	90
Input = Wipo-D Section Claim, (Subclass level), Batch Size=64, epoch=50, optimizer=adam, Acc=Accuracy							

5.4.4 The Change in Dropout and Optimizer Values

Dropout prevents data from being over fitted by keeping the data out of training. This function eliminates 20 percent (if dropout=0.20) of the input and the remaining part of the sequence became a new input data of following layer. The result of different dropout values effect on the classification is shown in Table 5.5. With two different optimizers, the change of dropout value is also controlled in Table 5.5. Adam optimizer generally gives better results than Rmsprop optimizer. It is observed that, for the Rmsprop optimizer, the higher the dropout value, the lower the accuracy value. However, a similar effect is not observed in the Adam. No major change is observed in terms of duration except for a few seconds.

Table 5.5 The effect of different dropout values on the classification

Optimizer	**Dropout**	**Accuracy**	**Time (s)**
Adam	0.01	0.67	150
	0.1	0.67	150
	0.25	0.68	151
	0.2	0.67	151
	0.5	0.68	152
Rmsprop	0.01	0.66	152
	0.1	0.66	157
	0.25	0.67	154
	0.2	0.63	154
	0.5	0.64	153
Input= Wipo-D Section –Description (Subclass level), Batch size=64, epoch=50			

5.4.5 The Change in Epoch Values on the Classification Performance

In machine learning applications, the number of times the model will run is an important issue. If it is run less often than it should, adequate learning may not be achieved and the accuracy rate may decrease. On the contrary, if over fitting emerge;

memorization may occur due to too much learning. Trials in Table 5.6 are made to see how many times the HAN algorithm needs to be run for this.

Table 5.6 The effect of change in epoch values on the classification performance

	Epoch=20		**Epoch 50**		**Epoch 100**	
Batch Size	**Acc**	**Time (s)**	**Acc**	**Time(s)**	**Acc**	**Time (s)**
10	0.57	43	0.67 %	358	**0.73**	610
15	0.61	30	0.67 %	327	0.65	467
20	0.58	27	0.64 %	320	0.65	400
25	0.61	28	0.66 %	306	0.68	383
Input= Wipo-alpha-D-section data, full-text value (Subclass level), Dropout=0.5						

In this analysis, "full-Text" is created by combining all textual parts (title, abstract, claim, description) in the patent text. Analysis is performed using different batch sizes. Epoch values have been changed to 20, 50 and 100. While the epoch value is 20, the accuracy value reaches a maximum of 61%. In this case, the training accuracy remains at 90%. When the Epoch value is 50, the accuracy value increases and the accuracy of the training reach 99%. When the Epoch value is 100, the training accuracy value reaches 1.0. As the Epoch value increases, the time spent increases as expected.

5.4.6 Classification Accuracy Variation Based on Different Batch Size

Batch size is one of the special characters of the algorithm. It divides the training data set into batches. Batch size is limiting the number of samples to the predetermined size to be shown to the network before the weight calculation. Before conducting an experimental analysis for batch size, there was an expectation that as batch size increased, the model would have learning difficulties and would achieve lower accuracy. It is thought that batch sizes with high value would have low accuracy. However, as a result of the analysis, it is concluded that this situation will change depending on the input text. For example, there is no significant change in the accuracy of abstract and description texts. However, in claim texts, when the batch size value is 10, the accuracy value, which increased to 67%, decreased to 60% when the batch size value is 64. It is thought that one of the reasons for this situation may be due to the complex structure of the HAN algorithm and its attention mechanism. The result of batch size effect on classification performance is in Table 5.7.

Table 5.7 The effect of change in batch sizes on the classification

Batch Size	Abstract		Claim		Description	
	Acc	**Time (s)**	**Acc**	**Time (s)**	**Acc**	**Time (s)**
2	0.53	921	0.67	921	0.64	566
10	0.52	331	0.60	142	0.67	329
15	0.54	295	0.58	109	0.67	289
20	0.53	264	0.59	97	0.67	263
25	0.53	247	0.59	100	0.67	247
64	0.53	195	0.60	73	0.67	89
Input=Wipo-alpha-D-section (Subclass level), Optimizer= Rmsprop, Epoch=50, Dropout=0.2						

5.4.7 The Classification Performance of Different Patent Text Parts

To achieve the highest classification accuracy in the patent document, should we analyze all the textual contents of the patent separately or make use of the various combinations created from these texts? The answer to this question is sought with the HAN method. In this analysis, title, abstract, claim, description and different combinations of these are used as input respectively. Analysis results using two different dropout values and four different lot size values are listed in Table 5.7. When the results in Table 5.8 are evaluated, the accuracy value of 70% and above is always in "Ab_Cl_Desc", "Cl_Desc" and "full-TEXT" entries with high text sizes. It is thought that this result is achieved due to the structural features and attention mechanisms of the HAN method. Longer sentence structures results in higher accuracy. Thus, the importance of text combinations becomes evident. Using only a part of the patent text in the analysis to be made may prevent the analysis result from reaching the expected classification accuracy.

Table 5.8 The effect of using different parts of patent text on the classification

Drop out	Batch Size	Abstract		Claim		Description		Ti_Ab		Ab-Cl		Ab_Cl_Desc		Cl_Desc		full-TEXT	
		Acc. %	Time (s)	Acc. %	Time (s)	Acc. %	Time (s)	Acc. %	Time (s)	Acc. %	Time (s)	Acc. %	Time (s)	Acc. %	Time (s)	Acc. %	Time (s)
0.2	10	0.68	288	0.61	129	0.66	361	0.62	351	0.66	144	0.70	454	0.70	453	0.66	602
	15	0.63	223	0.63	106	0.63	327	0.62	321	0.66	118	0.67	333	0.70	393	0.67	478
	20	0.59	46	0.62	96	0.67	322	0.65	314	0.66	116	0.68	387	0.69	386	0.72	584
	25	0.62	178	0.64	95	0.65	308	0.60	325	0.64	115	0.71	312	0.66	367	0.68	379
0.5	10	0.64	578	0.61	129	0.64	362	0.65	354	0.64	144	0.66	454	0.71	445	**0.73**	610
	15	0.66	468	0.64	105	0.65	330	0.64	321	0.63	118	0.66	342	0.67	394	0.65	467
	20	0.62	196	0.63	96	0.65	320	0.63	314	0.62	111	0.70	332	0.67	384	0.65	400
	25	0.53	179	0.62	96	0.63	307	0.64	304	0.67	115	0.70	313	0.66	327	0.68	383

Optimizer: Adam, Epoch=50, Wipo alpha-D-section data HAN-(subclass) Ti_Ab: Title and Abstract, Ab_Cl: Abstract and Claim, Ab_Cl_Desc: Abstract +Claim+Description, Full-Text: Title+ Abstract +Claim+Description

5.5 Conclusion

The problem of automatic classification of patents is an area where solutions are frequently searched and different suggestions are presented in the literature. With the increasing number of patent applications coming to patent offices around the world, this problem gradually turning into an issue whose solution is sought. In this study, the performance of the hierarchical attention network model on the APC problem is examined. Due to the nature of patent texts, they are texts that contain both heavy technical terms and legal terms, and contain a lot of new words in terms of the specific language they are used in. The HAN method is an algorithm suitable for the complex structure of the patent text with its attention structure that emphasizes the understanding of both words and sentences. With the word embedding application on which the method is based, the distance of the word vectors to each other will catch the word similarity better. Thus, it will be easier to classify the documents. The HAN method gives much better results than other methods used in terms of accuracy. However, it is observed that the working speed is much slower than them. Parameter tests have been carried out to increase the classification accuracy with experimental analysis. In addition, the HAN method is tested for different input combinations. It has been observed that the HAN method is suitable for the patent text in terms of its operation. The attention mechanism in its content brings together important words and important phrases respectively, making it easy to find the appropriate class. The study will be tested in the future with a different embedding algorithm that will adapt more to the patent text .

CHAPTER 6
COMPARING TRANSFORMER ENCODER MODEL WITH ENSEMBLING A STATIC AND DYNAMIC VERSIONS OF DEEP LEARNING MODELS FOR AUTOMATED PATENT CLASSIFICATIONS

6.1 Introduction

Keeping up to date with technological changes is the most important task of a company, technical personnel, researcher, academician, or people interested in these issues. Those who cannot meet this requirement have difficulty adapting to rapid change [175]. Because in this technology market, where a new product is launched and an invention is exhibited every day, they will not be able to gain a share of the market and will be gradually erased. Predicting the future success of technology is extremely complex due to the limited information squeeze and market instability [176]. Patents are at the forefront of the tools that direct and follow technological developments[177]. Because patents are an effective source of information for technological development in a specific field [178], [179], [180].

In order to access rich information of patents, it is necessary to bring them together. It is also necessary to use the appropriate technological classes to collect all the patents published in the technology of interest in a pool. It is possible that an incomplete and/or incorrect patent will fall into the created patent pool in research made with words only. Because in the writing of patents, heavy language is used in which there are many different words and legal terms are used a lot[181], [182]. At this stage, automatic classification systems come into play. With the help of these systems, the patents desired to be classified in the field of interest can be assigned to an appropriate class. Patent specialists are among those who have the most problems in this field. Assigning a patent application to the class should be their primary action. Thanks to these text classification systems, they will not only save time and effort but also prevent possible classification errors [183].

Patent texts are technical documents in semi-structured form. In addition to containing many structural data from the information of the inventor to all the information in the patent's tag, it also includes unstructured text data such as title, abstract, claims, and description. Different techniques have been used in the automatic classification of patents from the past to the present. Among these, many techniques can be listed, from classical machine learning methods to deep learning methods and various transformer-based classifiers.

In this chapter of thesis, patent classification is carried out using the title, abstract, and claim sections of the patent and its different combinations. In the classification process, 1) deep learning methods CNN and RNN, as well as their static and dynamic versions, the classification is performed and 2) the patents are classified at the subclass level using the transformer encoder model. Wipo-alpha is used as the data set.

The contributions of this study are (1) the evaluation of single and collective performances of different deep learning models to be used in the ensemble process and selection of the appropriate model, (2) the developing of a deep learning model to effectively predict multiclass patent text classification in terms of different performance measures, (3) ensuring the performance of the transformer encoder model in different parameters to be measured.

The literature review section in chapter 3 provides a summary of the published literature on patent classification using the transformer encoder model and its variations. Readers seeking information about this subject would be better off referring to the relevant section in chapter 3. It will not be mentioned again in this section to avoid repetition.
In the following section, the Transformer encoder model that is used in this study will be introduced.

6.2 Transformer Model

The transformer is the architecture used to convert one string to another with the help of "Encoder" and "Decoder". This architecture differs from existing recurring networks in that it does not use networks such as LSTM, RNN, and GRU (Gated Recurrent Unit). There is an encoder and decoder architecture in the transformer mechanism. Both the encoder and decoder consist of modules that can be

interconnected multiple times. The modules mainly consist of “Multi-Head Attention” and FFN layers. The Transformer relies on a simple but powerful mechanism called "attention" that allows AI models to selectively focus on certain parts of their input, thus learning more effectively. The function of the attention mechanism is to determine the relationship of a word with other words ("word to word"). The reason why it is called multi-head attention is that attention mechanisms are operated in parallel with each other during the determination of the relationship between one word and another. In the original paper, 8 attention mechanisms were used. Thus, the training time is shortened and each attention mechanism discovers different information about the word [184], [185].

Consisting of 6 identical layers, the encoder has 2 sublayers. The first of these is the multi-head self-attention mechanism and the second is the position-wise fully connected feed-forward network. There are residual connections between these two sublayers. The layer normalization process is applied to the output of these sublayers. It produces d-dimensional output, including embedding layers, to facilitate all residual connections. Residual connections are responsible for transmitting unprocessed inputs to the layer normalization function. Thus, the possibility of losing key information such as positional encoding on the way is minimized. Although the structure of the encoders’ 6 layers is the same, the information they carry is different. For example, the embedding layer is only visible in the first phase and embedding is not reapplied in other phases, thus ensuring that the encoded input is stable. There is a similar situation in multi-head attention. If the same process is applied to all layers, each layer acquires new information using the information from the previous layer and transmits this information. In multi-head attention, the input is a vector containing embedding and positional encoding data. Each attention sub-layer and each FFN sub-layer are followed by the post-layer normalization layer. This normalization layer has the add function and the layer normalization process. The add function retrieves the data from the residual connection. Then the normalization process works. The encoder and decoder are all connected with FFN. Each position is handled separately. It has 2 layers and a Relu activation function. Position-wise-FFN has 2 kernel sizes with 1 convolution. The FFN output is processed by going to the Position normalization layer

and from there the output is sent to the next layer of the encoder and the attention layer of the decoder [184]

The decoder consists of 6 layers, just like the encoder. Each sublayer and function of the decoder is similar to the encoder [184]. In this, too, there are two sub-layers after each coding layer. In the decoder, unlike the encoder, there is a third sublayer that performs multi-head attention on the output of the encoder stack. The decoder layers' names are Multi headed attention layer (masked), a multi-headed layer, and a fully connected-feed-forward network layer (position-wised). Self-attention sublayers are placed to prevent locations from linking to subsequent locations. In the output of the multi-head masked attention layer, the following words are masked and they no longer continue the process. This masking process ensures that the estimation of output embeds depends only on outputs in subordinate locations[186].

6.3 Methodology

6.3.1 Transformer Encoder Model Description

This study uses the encoder part of the transformer model [186]. The transformer encoder model is preferred because of its self-attention feature. Thus, the importance of all the words in the patents will be reached. The decoder section is not preferred because it contains a masked self-attention layer. The reason for this is the desire to reach and use all tokens. In order to understand which attention score is significant among the word pairs in the patent documents, different patterns are obtained with multi-headed self-attention layers.

The transformer encoder consists of a combination of several different layers. The layers in this architecture are as follows; self-attention, dropout, layer normalization, and two fully connected layers. First, the words are vectorized with word embedding, and then the places of the words in the sentence are determined with positional embedding. Thus, contextual word embeddings are created. With layer normalization and dropout, the input of the transformer model is prepared. Then the Transformer model works. In the multi-head attention layer, multiple models work simultaneously to reach the matrices of the words with different patterns of attention scores. From here, 2 fully-connected layers are connected to access different attention score

combinations. Then, the pooler, dropout, and dense layers are connected and finally, the patent classification is made. Figure 6.1 shows the transformer model architecture.

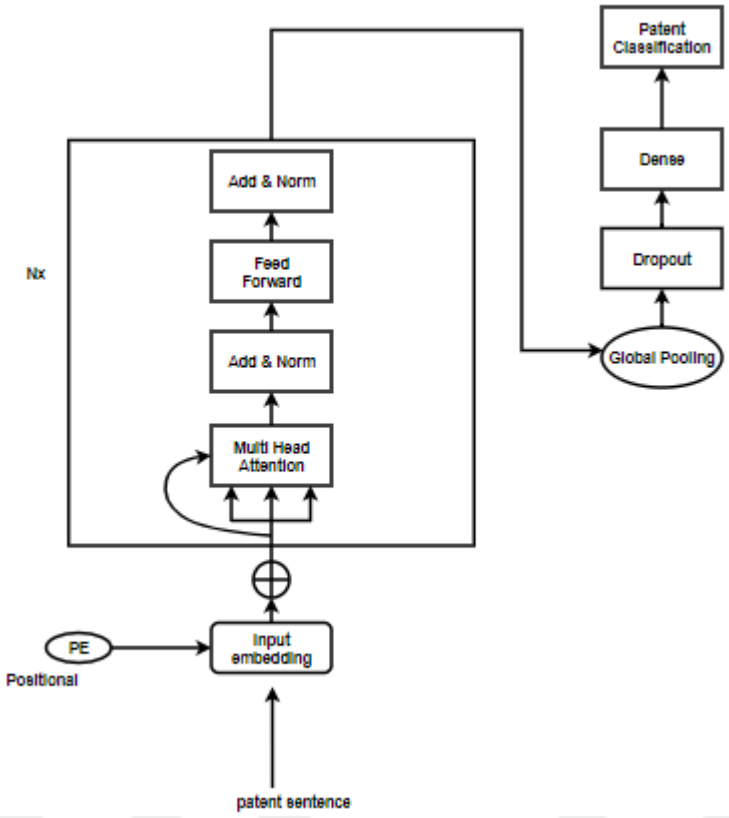


Figure 6.1 Transformer model architecture

6.3.2 Ensemble Deep learning model for patent classification

Supervised learning is a type of machine learning. In this type of learning, algorithms learn the labels to make predictions from the training data previously given to them. Next, this algorithm encounters unlabeled test data it has never seen before. Predicts what the label of this test data should be. Thus, the system has learned and has reached the capacity to interpret the new data. The classification process has been performing methods with a linear infrastructure based on machine learning, such as the classical methods Naive Bayes and SVM (support vector machine) methods, for many years. Later, when deep learning methods began to be studied in the literature, the problem that classical methods could not work with high-dimensional data was resolved. Many deep learning methods have overfitting problems. In this study, the ensemble learning method will also be used to prevent encountering overfitting. The Ensemble method is the process of combining multiple base classifiers under a common meta-classifier.

It has different applications. In this study, a method called stacking, which is based on the combination of individual outputs produced by the base classifiers by the meta-classifier and reaching the final classification output, is used. CNN and LSTM, which are deep learning methods, are the two methods used in this study. Static and dynamic versions of both methods were used in this study. Static and dynamic here are just a distinction about how word embedding can be trained. If it cannot be trained, it is static, otherwise it is dynamic.

6.3.2.1 CNN Architecture of This Study

In the first stage of deep learning, stop word removal and lemmatization processes were applied. At this stage, the spacy library was used. The lemmatization process is the conversion of words into their basic root. The initial layer of the model was created with the 100-dimensional Glove embedding pre-trained on all classifiers. Glove embedding is the sum of Wikipedia 2014 and Gigaword 5 archives with a size of approximately 6 billion words [173]. Both static and dynamic models began the classification process with embedding. However, later on dynamic models were allowed to make changes during backpropagation, while static models were not. There are several hyperparameters in the CNN classification architecture. These are values such as kernel values, activation functions, dropout values, and pooling values. In this study, 2, 3 and 5 values were used as kernel size. As the convolution layer value, the results for 1-dimensional 64, 128, and 256 values were investigated. Xavier uniform initializer for calculating kernel initial weights. A batch normalization process was applied to make the model a regular form. A corrected linear unit (ReLU) was used for the activation process. Global pooling was used in the pooling step. Mean pooling was excluded as it resulted in overfitting in this classification study. The algorithm was run for 2 different values, 0.2 and 0.5, within the dropout value. Finally, experiments were made with the values of 100, 128, 256, and 512 for the dense layer. Relu was used in the activation process. Softmax was used in class probability calculation.

6.3.2.2 LSTM Architecture of This Study

While creating the LSTM architecture, two 500 and 200-dimensional layers were used as the LSTM layer. The first of these will be created for all units, while the second will be used only for the last unit. The values of 0.2 and 0.5 were used as the dropout value. A hyperbolic tangent was preferred for the activation function. While Xavier uniform

is used to initialize the kernel weights, vertical initializers are used for the initial value of the duplicate kernel weights.

6.3.2.3 Meta-classifier

A meta-classifier was used to reach a common output result by combining the outputs of the CNN and RNN models used in this study and their static and dynamic versions under a single model. In the meta-classifier, a single 100-dimensional output layer and Relu are preferred as the activation function. Outcome probabilities are found with softmax. The deep learning application made in this study is shown in Figure 2.

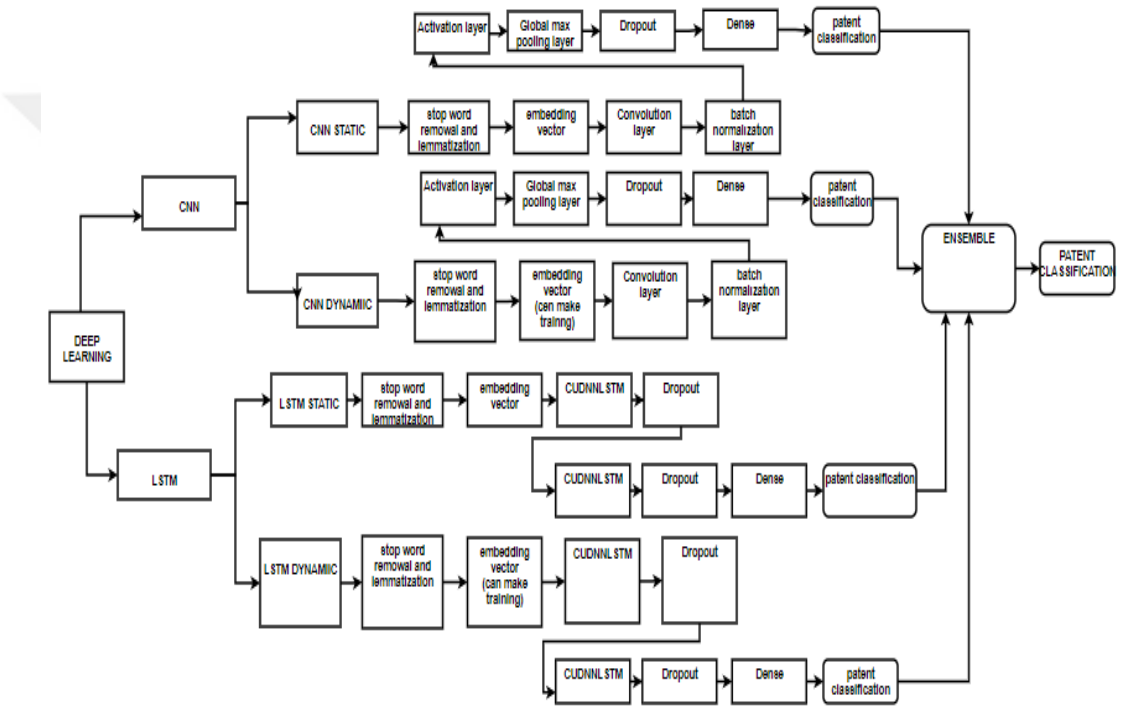


Figure 6.2 Deep Learning architecture

6.3.3 Experimental Settings

This study consists of two different sets of models. A number of different systems were developed for both models and these systems were tested with different combinations of parameters. The first system includes deep learning methods such as CNN, LSTM and their static and dynamic model versions. The second system is based on the encoder phase of the transformer model. The first system was tested with 35 different parameter combinations. In the second system, 40 different analyses were made. In the analysis made using deep learning methods, the abstract, title, and claim sections of the patent text are included. The description section is excluded from the

analysis. The reason for this is the long analysis time. In the analyses made with the Transformer encoder, all text parts of the patent document were combined under the "text" heading, and analyses were made. In addition, analyses also were carried out with the abstract and claim sections.

Of the decoder and encoder sections in the standard transformer model, only the encoder is used. The encoder part consists of self-attention and feed-forward network. However, the decoder part has a masked layer of self-attention that prevents the attention layer from looking at the entire input queue. If the decoder part is used, it will not be possible to participate in every token. That is, the information will be lost. However, it is necessary to reach each token to classify the given array. Because it is enough to have text features to be used in the classification process. To process the text data, the part of each patent that will be used in that experiment is extracted, its paragraphs are divided into markers, and embeddings of the markers are created. Embeddings of tokens are generated using Word2Vec. What is sought in this architecture is to train the query and key matrices of the attention layers to reach the importance of each word in the patent texts. Therefore, an attention score is calculated for each word in a patent document to determine its usefulness for each class. Many different attention patterns are obtained with multi-headed self-attention layers that allow the model to decide which attention score is important among the pairs of patent terms in the patent document. Next, the class estimate of the patents is found after the flattening and the last fully connected layer have passed. Class probabilities are calculated using the categorical cross-entropy classifier. The transformer encoder model has been tried to be optimized with different hyperparameters. Adjusted the head number of multi-headed self-attention to two. Analyses were performed for 64, 128, and 256 values for sequence length. For the Embedding dimension value, the algorithm was run for 32, 64,128, and 256 values. Apart from the trials in which all text parts of the patent (title, abstract, claim, description) were used in the analysis, separate trials were conducted within the abstract and claim sections of the patent. Mostly all texts were included in the analyses. The list of parameters used in the analyses performed in this study is shown in Table 6.1.

Table 6.1 Hyperparameter setting of methods

Parameter	CNN Static	CNN Dynamic	LSTM Static	LSTM Dynamic	Transformer Model
Activation	ReLu	ReLu	ReLu	ReLu	ReLu
Feed forward	-	-	-	-	32, 64 and 128
Batch Size	128	128	128	128	128
Epoch Number	15	15	15	15	15
Optimizer	Adam	Adam	Adam	Adam	Adam and rmsprop
Kernel Size	3, 4, 5	3, 4, 5	3, 4, 5	3, 4, 5	-
Dropout	0.02, 0.2, 0.3 and 0.5	0.02, 0.2, 0.3 and 0.5	0.02, 0.2, 0.3 and 0.5	0.02, 0.2, 0.3 and 0.5	0.2
Max-length	50 and 100	50 and 100	50 and 100	50 and 100	50, 100, 500, 1000, 1500 and 2000
Patent Part	abstract, claim, abstract + title, abstract+ claim	abstract, claim, abstract + title, abstract+ claim	abstract, claim, abstract + title, abstract+ claim	abstract, claim, abstract + title, abstract+ claim	abstract, claim, text (title, abstract + claim +description)

In the deep learning analysis, after the words passed through the pre-processing stage, experiments were made with two different values in deep learning models for maximum sequence length. These are values of 50 and 100.

All experiments in this article were run with Google Colab Pro+. All experiments are programmed in Python. The pandas and Nltk toolsets were used for text pre-processing, and the TensorFlow-based Keras framework was used to generate the main experiment, deep learning, and transformer encoder algorithms.

6.4 Results and Discussion

6.4.1 Deep learning results

Deep learning analysis was performed by combining different parts of the patent or these parts as inputs one by one the results of the classification analysis are shown in Table 6.2. The table, which includes 35 different trial results, provides a summary of all analyzes performed. (1) In the table where accuracy values are presented, the CNN-dynamic model obtained the highest value with the combination of abstract+claim. Other parameter values that ensure the success of this combination are filter with a value of 256, dropout with a probability value of 0.2, and max length with a length value of 100. (2) This study was based on the assumption that the results of the

ensemble model constructed by combining the results of CNN and LSTM models and their static and dynamic versions would be better than the results of CNN and LSTM. The results support this situation. (3) Dynamic CNN not only gives the best results but also competes with the ensemble model in other trials. (4) In trials where the claim is a stand-alone input, the highest accuracy value is 66%, which can be obtained with dynamic CNN. (5) In cases where the abstract is alone, it was obtained from the ensemble model with 65%. (6) The highest accuracy value that the Abstract and title combination could achieve was obtained from the ensemble model at 68%. (7) When the results are evaluated in terms of parameter values, the accuracy value increases as the filter value increases. When examined in terms of dropout, the accuracy value increases as the dropout value decreases. (8) In this study, there are two different values taken as the max length value. In the results obtained according to these values, while the max length value is low, the accuracy values are better.

Table 6.2 Deep learning analysis accuracy results

Patent part	Filter	Dropout	Max-length	CNN-static	CNN-Dynamic	LSTM - Static	LSTM-Dynamic	Ensemble
Abstract	128	0.2	50	0.6225	0.6459	0.6125	0.6203	0.6520
Abstract	128	0.2	100	0.6181	0.6353	0.5928	0.6160	0.6441
Abstract	128	0.5	100	0.5877	0.6194	0.5851	0.6071	0.6254
Abstract+Claim	128	0.5	50	0.6257	0.6521	0.6295	0.6381	0.6505
Abstract+Claim	128	0.5	100	0.6257	0.6521	0.6295	0.6381	0.6505
Abstract+Claim	256	0.2	100	0.6760	**0.6927**	0.6501	0.6401	**0.6858**
Abstract+Claim	256	0.2	50	0.6458	0.6687	0.6385	0.6353	0.6717
Claim	64	0.2	50	0.6452	0.6648	0.5734	0.6266	0.6537
Claim	128	0.2	50	0.6475	0.6670	0.6199	0.6369	0.6592
Claim	128	0.5	50	0.6289	0.6418	0.6421	0.6195	0.6541
Claim	128	0.2	100	0.6361	0.6621	0.5350	0.5945	0.6488
Claim	128	0.5	100	0.6129	0.6283	0.5543	0.6068	0.6399
Abstract+Title	64	0.02	100	0.6215	0.6198	**0.6714**	0.6670	0.6641
Abstract+Title	64	0.2	100	0.6347	0.6231	0.6700	0.6677	0.6613
Abstract+Title	64	0.3	100	0.6311	0.6484	0.6662	0.6549	0.6594
Abstract+Title	64	0.5	100	0.5834	0.6189	0.6352	0.6356	0.6297
Abstract+Title	64	0.2	50	0.6292	0.6476	0.6563	0.6516	0.6743
Abstract+Title	64	0.02	50	0.5967	0.5997	0.6515	0.6494	0.6488
Abstract+Title	64	0.3	50	0.6105	0.6361	0.6542	0.6373	0.6406
Abstract+Title	128	0.02	50	0.5972	0.6202	0.6528	0.6451	0.6655
Abstract+Title	128	0.02	100	0.6278	0.6570	0.6669	0.6656	**0.6827**
Abstract+Title	128	0.2	50	0.6271	0.6532	0.6523	0.6476	0.6672
Abstract+Title	128	0.2	100	0.6576	0.6624	0.6687	0.6682	**0.6861**
Abstract+Title	128	0.3	100	0.6358	0.6773	0.6671	0.6657	0.6793
Abstract+Title	128	0.3	50	0.6145	0.6291	0.6403	0.6392	0.6724
Abstract+Title	128	0.5	100	0.6368	0.6445	0.6658	0.6549	0.6729
Abstract+Title	128	0.5	50	0.6296	0.6516	0.6692	0.6665	0.6647

Table 6.2 Deep learning analysis accuracy results *(continued)*

Patent part	Filter	Dropout	Max-length	CNN-static	CNN-Dynamic	LSTM - Static	LSTM-Dynamic	Ensemble
Abstract+Title	256	0.2	50	0.6480	0.6512	0.6498	0.6497	0.6720
Abstract+Title	256	0.3	50	0.6428	0.6643	0.6520	0.6437	0.6714
Abstract+Title	256	0.2	100	**0.6716**	0.6774	0.6690	0.6647	**0.6828**
Abstract+Title	256	0.02	50	0.5958	0.6276	0.6549	0.6381	0.6626
Abstract+Title	256	0.3	100	0.6606	0.6780	0.6652	0.6674	**0.6825**
Abstract+Title	256	0.02	100	0.6266	0.6611	0.6640	**0.6699**	0.6774
Abstract+Title	256	0.5	50	0.6383	0.6532	0.6327	0.6336	0.6529
Abstract+Title	256	0.5	100	0.6580	0.6639	0.6648	0.6658	0.6724

6.4.2 Transformer Encoder Results

The analyses for the problem of classifying patents at the subclass level using the Transformer encoder model are listed in Table 6.3. In the analyses in which the abstract, claim, and text (title + abstract + claim + description) sections of the patents are used as input data, the model was run on different hyperparameters. According to the classification results obtained as a result of the analysis: (1) the highest accuracy value (0.776) and the highest F1 value (0.798) were achieved by using the "text" section as input. (2) Too low or too high maxlen value decreases accuracy. However, the maxlen value alone is not enough to affect the accuracy value. The embedding dimension value should also be low. Otherwise, applications with the same maxlen value can reach a lower accuracy value with a higher embedding dimension value. (3) The low feed-forward dimension value increases the accuracy value. In trials with the same maxlen and embedding dimension value, the one with the lower feed-forward dimension value has a higher accuracy value. (4) When a high embedding dimension value is given, the accuracy of the model is very low. The transformer model shows itself as a model in which the parameters alone do not affect the classification performance much, but the parameters together increase the performance of the model.

Table 6.3 Transformer encoder results

Patent part	maxlen	Embed-dim	Feed forward dim	Accuracy	F1
Abstract	50	32	64	0.7477	0.644
Claim	100	32	64	0.7574	0.683
Claim	256	32	64	0.7356	0.6704
Claim	500	64	128	0.7306	0.67
Claim	100	32	64	0.7574	0.683
Claim	256	32	64	0.7356	0.6704
Text	50	64	32	0.7676	0.716
Text	50	64	64	0.7663	0.697
Text	50	64	128	0.7627	0.6789
Text	50	32	64	0.7534	0.66
Text	75	64	128	**0.7713**	0.722
Text	75	128	32	0.7676	0.7488
Text	100	32	64	**0.7764**	0.710
Text	100	128	64	0.7616	0.73
Text	100	128	128	0.7557	0.723
Text	500	32	64	0.742	**0.798**
Text	500	64	32	0.7321	**0.782**
Text	500	128	32	0.666	0.47
Text	500	128	64	0.6416	0.461
Text	1500	64	128	0.6742	0.634
Text	2000	64	32	0.7101	0.603
Text	2000	64	128	0.6701	0.6650

6.4.3 Overall Results

When the transformer encoder result and embedding deep learning results are compared, when the patent sections are analyzed alone, the transformer model reaches higher accuracy than the embedding model. In the model of the Transformer patent texts, the binary combinations could not be analyzed because they were not run in Google colab pro+ (crash issue). Therefore, information about binary results cannot be given. In the deep learning model, 70% accuracy could not be reached in any trial. In the Transformer model, high values such as 1500 and 2000 were used for the maxlen value, but these values did not increase the accuracy value. Therefore, the highest maxlen value was accepted as 256 during the deep learning study. Considering the F1 (0.798) and accuracy (0.7764) values of the analysis performed with the WIPO-alpha data set, it reaches higher values when compared to the studies performed at the same classification level with the same data set in the literature which has accuracy result at subclass level 76.7 [187].

6.5 Conclusion

In this study, the transformer encoder model and ensembled deep learning model are tested on patent classification problems. As deep learning models, CNN and LSTM models and the static and dynamic versions of these models, which differ according to the trainability of the word embedding value they use, were preferred as deep learning models. The results of the deep learning models were combined with the ensemble model and the classification results were calculated with the meta-classifier. It was analyzed that ensembled deep learning methods could outperform commonly used deep learning models in patent text classification problems. The effect of different parameters on the performance of the model was observed with the hyperparameter setting. The classification performance of the transformer encoder model was observed when different parameters were used. The results of the analyses for both the deep learning and transformer model show that the transformer encoder model outperforms the deep learning models for patent text classification. In our future studies, it is planned to run attention-specific deep learning models together with transformer models and to conduct a study on combining the performances of these two models, not comparing them.

CHAPTER 7
PATENT CLASSIFICATION WITH PRE-TRAINED BERT MODEL

7.1 Introduction

The automated patent classification system facilitates the categorization of patent texts based on their respective technical fields, thereby simplifying access to technical information. Automated patent classification systems are crucial as they streamline the process of identifying, categorizing, and analyzing intellectual property. With the exponential increase in patent applications, manual classification is no longer feasible. There is a critical need for an efficient and reliable automatic patent classification system. Machine learning algorithms enable these systems to efficiently analyze and categorize large volumes of patent documents, offering valuable insights into emerging technologies, industries, and business trends. According to the general consensus in the scientific community, patent texts are the primary source of technological information, with 70% to 90% of technical information being published [188]. Therefore, compared to scientific articles, it is much easier to establish technological networks and examine the evolution of technology by using patent metadata to provide detailed information. Patent texts can offer valuable information to inventors, investors, and policymakers, enabling them to make informed decisions about intellectual property assets and market opportunities [189], [190], [191]. Overall, APC systems are essential to the intellectual property ecosystem and contribute to progress and innovation in a knowledge-based economy.

Creating a system by classifying technology and assigning patents to the classes to which they belong using this structure is the most basic feature of automatic patent classification logic. The suitability of patents to the aforementioned technology class they may belong to and the similarity of patents written on the same subject should be determined. One of the most important problems that researchers may experience in this process is that the compatibility of a predetermined technology class with a new invention may be limited. Sometimes, a technological invention that has never been

encountered before and will break new ground in its field enters the process of becoming a patent [193]. Because of this innovation, patent researchers may have trouble ensuring the compatibility of new technologies with limited predetermined classes. In addition, due to technological changes, new inventions may not belong to the technology classes they belonged to in the past and may show similarities with patents in more distant classes. Current classification schemes struggle to catch up with technology at the class and subclass levels [193], [194]. The scientific classification tools used by automatic patent classification systems should be compatible with the current literature. In this way, a contribution can be made to solving problems caused by incorrect and/or incomplete classification of the classification systems.

Patent experts responsible for assigning patents to the appropriate class do not make decisions based solely on the title or abstract sections of the patent. In most cases, they also have to review the claims and descriptions. Understanding and interpreting patent texts manually, which contain various legal and technological jargon and are constantly infused with new terminology to keep from infringing on existing patents, is a exhausting task for patent experts. To fully understand modern technologies, professionals need to dedicate part of their time to keep up with developments. This traditional manual approach needs a considerable amount of time and effort, resulting in a waste of resources. Experts, patent researchers, inventors, and scientists interested in the issue might utilize automatic patent classification techniques to expedite their research and reallocate their attention on other areas [155], [195].

Comparing studies in the field of automatic patent classification is challenging for several reasons. The first difference is the dimensional variations in the data sets. In some studies, the volume of data can reach millions [45], [58]. Another issue is the variability of problem definitions. A patent document contains multiple labels. While some studies focus on multi-label classification, others prefer binary or multi-class classification [45], [196], [197]. Differences are also evident in the selection of text sections in patent documents [68]. It may also vary based on the subcategory level to which the classification belongs. While some studies choose to classify only patent sections, others may prefer classification at the section, subclass, or subgroup level within patent classes [62].

In this study, the aim is to classify patent documents using only textual data. The analyses utilized a pre-trained BERT model. The WIPO-alpha dataset, is utilized for the analysis. The literature pertaining to Bert is discussed in the literature review section in Chapter 3 of this thesis.

7.2 Material and Method

7.2.1 BERT (Bidirectional Encoder Representations from Transformers)

BERT is a language representation model proposed by Google researchers in 2018. Bert analyzes entire layers of text that were previously unlabeled. It works bidirectionally, utilizing both the right and left contexts. The pre-trained Bert model, with only one output layer, can offer solutions to current problems. To implement the BERT model, there is no need to modify the algorithm's architecture for each model [34]. The BERT model, which has been tested in 11 different areas and topics related to natural language processing problems, has proven itself by significantly improving prediction scores in many of these areas[83], [198].

BERT is trained using large amounts of pre-labeled raw data for predicting masked words and finding the next sentence. It can then fine-tune the labeled dataset. The process of developing a pre-trained BERT model is challenging due to the substantial amount of data required. The developers of the BERT model used 12 layers and 110 million parameters for training. Large amounts of data are required to develop complex structures [200]. Bert's model architecture is adapted from the transformer structure published by Vaswani et al. [91], with some variations. This structure consists of two sides: the encoder on the left and the decoder on the right. The encoder generates representations of the input sequence, while the decoder maps these representations to output sequences. The encoder and decoder consist of multiple layers, with each layer containing sublayers. The encoder has 2 sublayers in its layers, while the decoder has 3 sublayers.

There are three different types of attention in the structure. These include Multi-head Attention, Masked-Multihead Attention (only in the decoder), and feedforward networks. In the multi-head attention structure, the keys, values, and queries from self-attention, as well as the output of the previous layer, are sent linearly to perform

attention processing in parallel. Masked multi-head attention is only the first sublayer in the decoder. Its task is to prevent making predictions about a specific location that is not yet known. The Feed Forward Network is responsible for applying two linear transformations separately and equally at each position. Before the attention processing and information input layers, spatial coding provides relative information about the position of a token in an array. The transformer utilizes a token sequence derived from this information. After the attention and transformation processes are performed on the layers, a Linear transformation is applied to perform another linear transformation, and a Softmax transformation is applied to convert the output into probabilities.

Bert innovates by using two-way education. There are two pre-training tasks: masked language modeling (MLM) and next-sentence prediction (NSP). The BERT algorithm uses the MLM technique during the training phase. Some words are masked in this technique. These words are suggested based on previously unmasked words. The main purpose of this process is to analyze the words in the sentence and make predictions. MLM randomly conceals 15% of the words in each string and sends the remainder to the system. However, of this 15% of tokens, only 80% were replaced by the [MASK] token, 10% were replaced by another random token, and 10% remained unchanged. Next Sentence Prediction (NSP) also trains the model to understand the relationship between pairs of sentences, with each fragment consisting of several sentences. The structural and semantic relationships of a sentence with the preceding and subsequent sentences are revealed [201]. In the NSP transaction, Bert splits the inputs into two. During the training phase, it is taught that the second sentence follows and depends on the first sentence. However, in the testing phase, the second sentence is chosen randomly and asked to determine if it is dependent on the first sentence [202]. BERT uses the WordPiece embeddings developed by Wu et al. [203].

BERT first separates the sentences by using the token ([SEP]) to distinguish between the sentences. Secondly, it uses the token [CLS] to define the expression to which each token belongs [83]. In Figure 7.1, the term "CLS" represents the token utilized for the classification process. Additionally, the SEP token is used to separate two simple, sequential sentences. Bert uses three different types of input embeddings. Word token

embedding, position embedding, and segment embedding are utilized to determine the sentence to which the token belongs. For example, the word “cat” is represented in the token embedding by the vector for the token "E_{cat}," and its position in the sentence is indicated by E_3. To associate it with the correct sentence, it is represented in the E_A vector in the segment embedding. Bert masks the word to a certain extent; here, the word “wonderful” is replaced by the word “mask”. Since Bert's output will be the result of the guessing process to find this word, the output vector is Q_5 because the word "wonderful" is in the fourth position. Then, it calculates the probability values of the words using feed-forward neural networks and the softmax activation function. Here, Bert can suggest words such as "amazing," "beautiful," and "cute" in various possibilities.

Figure 7.2 depicts the transformer layer, which constitutes the second stage of the BERT architecture. At this stage, the output data is generated by processing the data from the input embedding layer. This process is repeated twelve times. Thanks to its multi-headed attention mechanism, the Transformer reads the incoming data bidirectionally and performs multiple calculations simultaneously. The model is trained in this way. Here, the attention mechanism prevents data loss by transferring each processed datum to the hidden layers and the encoder [91], [204].

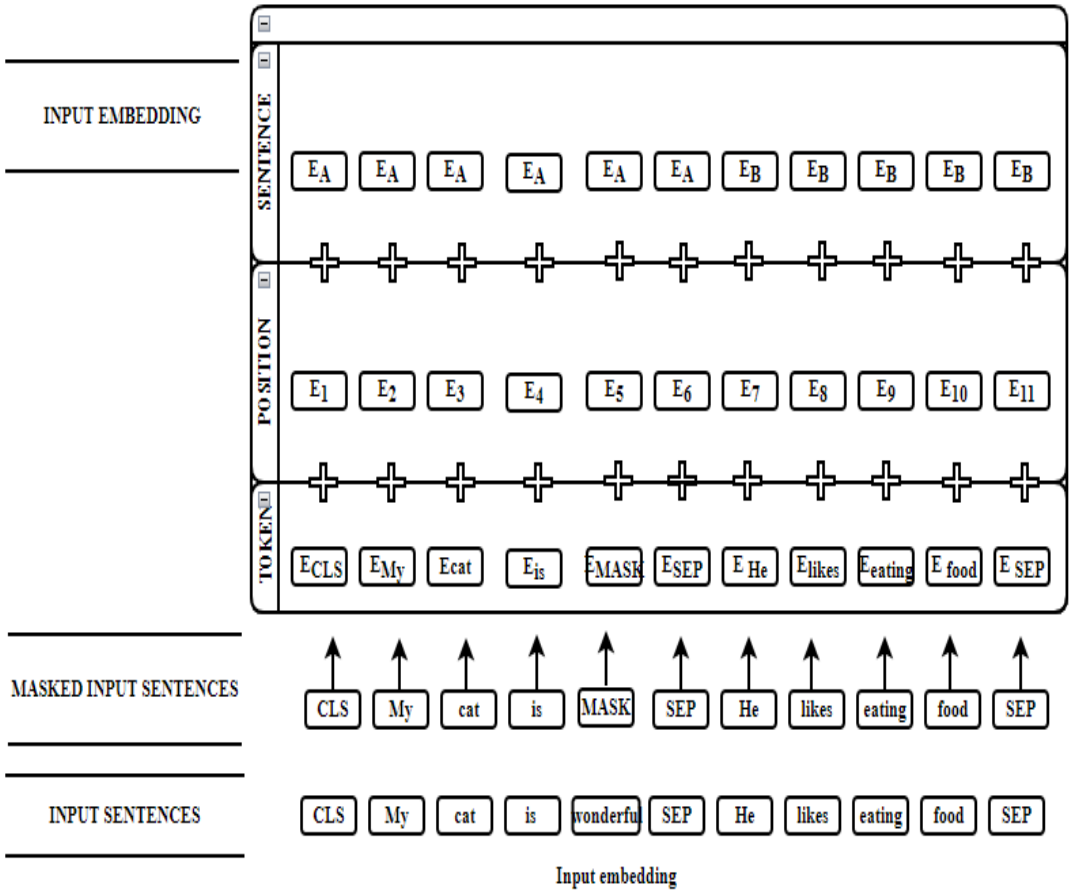


Figure 7.1 Input representation of Bert

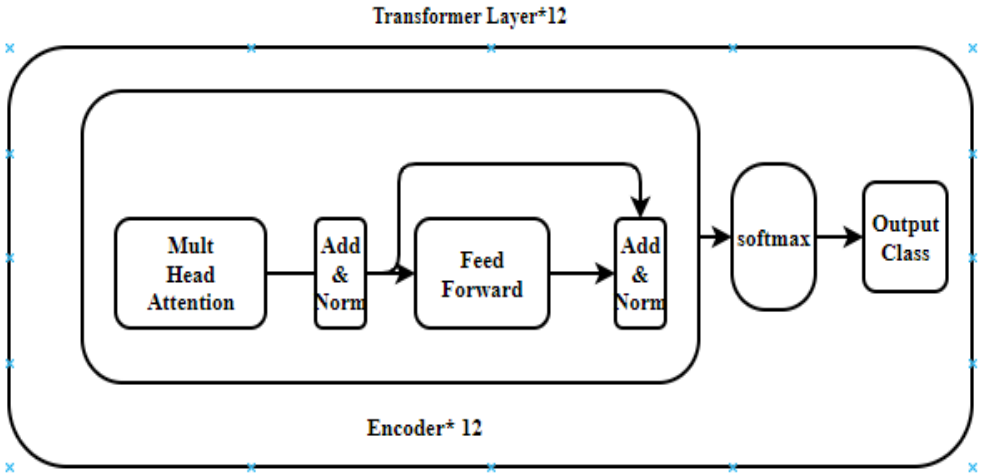


Figure 7.2 Transformer phase of Bert

Bert was trained by using a combination of two different datasets: BookCorpus (800 million words) and English Wikipedia (2,500 million words), totaling 16 GB. BERT's performance has inspired many researchers to work in this field, leading to the creation of models with similar architecture. Thus, a new family of BERT models emerged,

including CamemBERT, RoBERTa, RuBERT, DistilBERT, and AlBERT. Members of the Bert family have differences among themselves and with BERT. For example, RoBERTa developers have argued that increasing the model's training time, dataset size, and the number of batches has a positive impact on performance. They also suggested that the goal of next sentence prediction should be removed and dynamic masking should be used [108]. RoBERTa is trained by using a combination of datasets that are ten times larger than those used for BERT [199]. DistilBERT is a version of the Bert model that is 40% smaller and 60% faster [205]. Bert's skill is provided 97% [107]. AlBERT uses the same datasets as BERT and DistilBERT. While RoBERTa primarily aims to improve performance, DistillBERT focuses more on speed. In contrast, AlBERT aims to achieve both simultaneously. It predicts the order of sentences instead of predicting the next sentence [206]. CamemBERT has the same structure as RoBERTa and was developed for the French [207]. BERT utilizes 12 layers, 768 hidden layers, and 12 self-attention heads. While BERT-base has 110 million parameters [208], BERT-large has 340 million parameters (24 layers, 1024 hidden layers 16 self-attention heads). Although the layer structures of RoBERTa-base and RoBERTa-large are the same as BERT, the number of parameters differs. RoBERTa-base has 125 million parameters, while RoBERTa-large has 355 million parameters [108]. It is based on DistilBERT architecture and consists of 66 million parameters, six layers, 768 hidden layers and 12 self-attention structures. In comparison, the base model of AlBERT had 12 million parameters, while the larger model had 18 million parameters. The layer structure is similar to that of RoBERTa [206].

In this study, the performance of BERT for multi-class patent classification is examined. An attempt is made to improve the prediction accuracy by using different hyperparameters. Additionally, various textual contents from the patent document and their combinations are used as input data. The studies observed the performance of the Bert algorithm on texts of varying lengths.

7.2.2 Multi-class patent classification

This study proposes the use of a pre-trained language model to display patent texts and to fine-tune the downstream task of the multiclass patent classification problem. It is required to fine-tune the parameters of the pre-trained language model.

7.2.2.1 Pre-processing

A simple preprocessing step is required to prepare the input text of the patent data to be used in the analysis to match the input format of the pre-trained language model. Standard text-cleaning procedures were not used in this study. Some of these procedures are steps, such as lowering all text, removing punctuation, removing HTML links/email addresses, and removing numbers. However, these steps have not been implemented. Because this is an unnecessary step for Bert. The general text preprocessing steps are as follows. First, it was necessary to perform data tokenization. This stage separates the entire word/word group existing in the raw text into subwords and punctuation marks. This step provides a smaller set of inputs that can be defined and is appropriate for the dictionary infrastructure of the pre-trained model. The second step is to add the classification tokens [CLS] and the separator tokens [SEP]. The tokens were converted into numeric values using the indices in the pre-trained table. A sample of the tokens is shown in Figure 7.3.

```
Tokenize the first sentence:
['[CLS]', 'an', 'apparatus', 'for', 'pre', '##fa', '##bri', '##cating', 'furniture', 'having', 'shelves', '(', '2', ')', 'or',
'guide', 'rails', '(', '60', ')', 'for', 'drawers', 'is', 'disclosed', '.', 'the', 'apparatus', 'comprises', ',', 'two', 'side', '
plates', '(', '10', ')', 'having', 'one', 'or', 'two', 'more', 'flutes', '(', '14', ')', 'formed', 'vertically', 'in', 'the', 'insi
de', 'surface', '(', '12', ')', 'thereof', ',', 'guide', 'bars', '(', '20', ')', 'being', 'inserted', 'in', 'the', 'flutes', '(', '14',
')', 'and', 'having', 'a', 'vertical', 'guide', 'groove', '(', '22', ')', 'with', 'a', 'slit', '(', '28', ')', ',', 'and', 'supporting
', 'members', '(', '40', ')', 'being', 'installed', 'in', 'the', 'guide', 'bars', '(', '20', ')', 'across', 'the', 'slit', '(', '28', '
)', 'and', 'supporting', 'the', 'shelves', '(', '2', ')', 'or', 'said', 'guide', 'rails', '(', '60', ')', 'for', 'drawers', 'there', '
##on', '.', 'the', 'supporting', 'members', '(', '40', ')', 'can', 'be', 'moved', 'up', 'and', 'down', 'along', 'the', 'gui
de', 'bar', '(', '20', ')', 'and', 'be', 'fixed', 'to', 'a', 'desirable', 'position', 'thereof', '.', '[SEP]']
```

Figure 7.3 Tokenize first sentence

One of the steps before starting the algorithm is to fix the input sequence, i.e. to limit the maximum number of tokens. This fixation method creates a standard length. Thus, it provides easier batching of inputs and more efficient GPU usage. The self-attention mechanism, which shows the tokens that the model should and should not use, is the last step before the data enters the model. Tokens that should be used in the analysis are denoted by 1, whereas tokens that should not be included in the analysis are 0. Tokens with zero values are called padded tokens. The basic BERT model consists of 12 encoders. Each encoder is a feed-forward network structure, in each encoder self-attention mechanism works and the data is transmitted to the next one [110].

Bert uses three embedding steps. The first is the token embedding. In a predetermined vector space, each word is represented by an n-dimensional vector. This vector also contains the semantic meanings of the words. The second embedding step is segment embedding. This segment also checks the previous and next sentences around it to find the target sentence. In other words, it is an adaptation of the skip-gram models created by embedding words in a sentence. The third type is position embedding. Embedding memorizes the position of a word in a sentence. Thus, the relative location information is stored. This situation is similar to the application logic of LSTM.

7.3 Results and Discussion

This section presents the experimental results. A series of patent classification experiments were conducted at the subclass level using WIPO alpha patent documents. These experiments can be summarized as follows: examining the impact of different learning rates on classification performance, exploring the influence of various batch sizes, and assessing the effect of using patent texts as separate input values on classification performance. The hyperparameter values utilized in the study are summarized in Table 3. Some hyperparameters used in all experiments conducted in this study are not included in the tables. For example, the Adam optimizer is used as the optimization function in all experiments. The maximum string length is 128. The dropout rate used to prevent overfitting is set at 0.1.

Table 7.1 Different hyper parameter values used in the study

Hyperparameter	Values
Learning Rate	1e-5, 2e-5, 3e-5, 4e-5, 5e-5
Input Type	Title, Abstract, Claim, Description, Full text and various combinations
Batch Size	16, 32, 64, 128

This study is carried out with Google Collaboratory on NVIDIA Tesla V100 GPU with 51 GB RAM and the data is analyzed using the Python programming language.

7.3.1 The Changes in Classification Performance According to Different Batch Sizes

The batch size is considered one of the important hyperparameters in deep learning systems. In this study, the impact of changes in batch size on classification performance is tested. These values are shown in Table 7.2. The batch size value is kept within the recommended range for BERT. However, upon evaluating this study,

it is observed that the classification performance decreases as the batch size increases when the classifications are made using the values of 16, 32, 64, and 128. Increasing the batch size accelerates the training speed of the model. However, in this case, the updates are made less frequently. The classification performance cannot be affected solely by the batch size value. Factors such as the input type, the established model, and the optimization algorithm also impact classification performance.

Table 7.2 Classification performance according to different batch size values

Learning Rate	Batch Size	Precision	Recall	F1	Accuracy
lr=1e-05	16	0,55	0,55	0,54	0,55
	32	0,52	0,53	0,51	0,53
	64	0,53	0,53	0,51	0,53
	128	0,49	0,51	0,48	0,51
Input=Claim, Epoch=5					

7.3.2 The Change in Classification Performance According to the Different Learning Rates

One of the hyperparameters that has the most significant impact on classification performance is the change in the learning rate. The learning rate is a parameter that determines the extent to which the model weights are adjusted during training on classification problems. When examining the data in Table 7.3, it is evident that the decrease in the learning rate generally has an inverse relationship with the increase in classification accuracy. In this case, as the learning rate decreases, the accuracy values generally increase. However, it is also important to adjust the learning rate based on the problem. For example, using a very small learning rate, such as 2e-06, significantly reduces the classification performance of the model. A standard value is not valid for this parameter.

When comparing the iteration values in this study (as shown in Table 7.3) for the analysis of patent texts, it is not evident that increasing the iteration value had a positive impact on the accuracy. As the iteration value increases from 1 to 2, there is an improvement in classification performance. However, this rate of increase remains constant as the iteration values increase. Additionally, if the impact of the number of iterations on prediction accuracy is significantly higher in studies using BERT, a much higher accuracy value should be found in analyses with 10 iterations. Therefore, increasing the iteration value after 5 iterations does not improve the accuracy of the predictions. This test confirms similar explanations found in the literature [204], [209]. Figure 7.4 illustrates the variation in learning rates across different cycles.

Table 7.3 The classification performance according to different learning range

Iteration	Learning Rate	Precision	Recall	F1	Accuracy
Iteration=5	2e-06	0.26	0.34	0.26	0.34
	1e-05	0.49	0.51	0.48	0.51
	2e-05	0.54	0.54	0.53	0.54
	3e-05	0.55	0.55	0.54	0.55
	4e-05	0.55	0.54	0.53	0.54
	5e-05	0.55	0.54	0.53	0.54
Iteration=10	2e-06	0.2	0.29	0.21	0.21
	1e-05	0.55	0.54	0.54	0.54
	2e-05	0.55	0.54	0.54	0.54
	3e-05	0.55	0.54	0.54	0.54
	4e-05	0.55	0.53	0.53	0.53
	5e-05	0.54	0.53	0.52	0.53
Input=Abstract. Batch Size= 64					

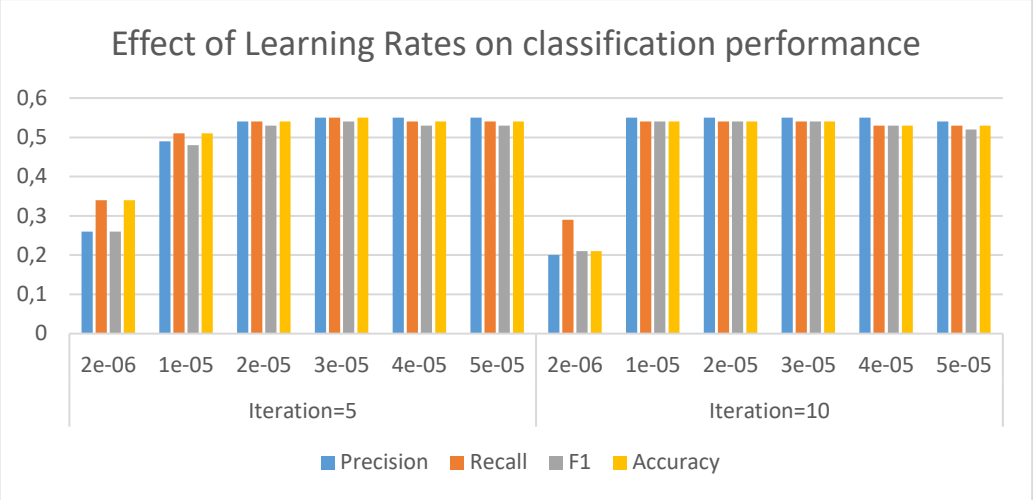


Figure 7.4 The effect of different learning rates on classification performance

7.3.3 The Change in Classification Performance According to Different Patent Text Parts

The results of the studies conducted with different sections of the patent document and various combinations of these sections are shown in Table 7.4. When the sections are examined individually, the section with the longest text content achieved a higher accuracy in the classification performance. The reason for this may be that the number of words and sentences in the short sections, such as the title and abstract, is lower than in the relatively longer sections, such as the claim and description. The Bert algorithm performs classification based on the sequential sentences and the position of the words in these sentences, due to its structure. Therefore, the lower prediction accuracy in the abstract section, which contains a limited number of sentences and is shorter than the claims, aligns with the performance of the BERT algorithm.

Table 7.4 Classification performance for different patent text parts

Input	Precision	Recall	F1	Accuracy
Title	0.47	0.48	0.46	0.49
Abstract	0.54	0.54	0.53	0.54
Claim	0.55	0.54	0.53	0.54
Description	0.57	0.57	0.56	0.57
Full text	0.58	0.58	0.56	**0.58**
Title + Abstract	0.55	0.56	0.54	0.56
Abstract + Claim	0.55	0.55	0.54	0.55
Title + Claim	0.57	0.56	0.55	0.56
Title + Abstract + Claim	0.55	0.55	0.54	0.55
Learning Rate=2e-05, Batch size=64, epoch=5				

7.4 Conclusion

Patent classification is a system used to categorize patent applications based on their technical fields. Thanks to this system, patent applications can be more easily searched and evaluated. Nowadays, classification systems are constantly trying to keep pace with the record-keeping of new inventions and updating those involved in this field. Research in the field of deep learning is also advancing in finding solutions to patent classification problems. The BERT model, which entered the deep learning world in 2018, continues to be involved in research in the field of patent classification.

In this section of the thesis, a patent classification study is conducted using the BERT-based model. In this study, it is tried to measure the classification performance using various hyperparameter variables and different sections of patent texts. The highest accuracy value (58%) is achieved in the model where the "specification" is used as input, the learning rate is set to "2e-05", and the batch size is "64". Although this accuracy value may be considered low for many other patent classification problems, it is noteworthy that for the Wipo-alpha data set, it is competitive with similar studies in the literature and contributes to the existing body of literature in this regard. In this study, we emphasize the importance of hyperparameter analysis. Presenting the parameters used in detail not only adds to the existing literature but also provides guidance for researchers conducting similar studies. This type of hyperparameter analysis provides the information that researchers need for preliminary analysis in any research article. The variation in the number of examples across different classes in this dataset may impact the performance of the classification model, leading to better or worse recognition of certain classes. This may lead to a decrease in classification accuracy.

This study contributes to the literature by conducting experiments with four different batch sizes and six different learning rates. It also includes an extensive hyperparameter analysis, incorporating four different text sections of the patent document and their various combinations as input data for the model. In this study, the importance of hyperparameter analysis is emphasized. Presenting the parameters used in detail not only contributes to the literature but also guides researchers conducting similar studies. Because this type of hyperparameter analysis provides the information

the researcher needs for the analysis in the preliminary research stage of any article The fact that some classes have more or fewer samples than others in this data set may have caused the classification model to recognize some classes better or worse. This may cause the classification accuracy to decrease. While similar results are obtained in the hyperparameter analysis with the studies in the literature in terms of learning rate and number of cycles (it is observed that the number of 4-5 iterations is sufficient and increasing the number of iterations did not make a significant contribution to the classification performance). Different results are obtained in the literature regarding the batch size. It is thought that this situation will be important for future studies by repeating the experiments by keeping the batch size values lower. Although there are many studies in the literature that conclude that the patent abstract or patent claim section is sufficient in terms of selection of patent texts, according to the results obtained from this study; It is not enough to use only the abstract or only the claim section. Using the full textual content of the patent document allowed achieving a higher classification accuracy. However, this causes a high computation time. A technical problem is also encountered in this study. The study is carried out in the Google Colab environment. Restrictions such as limited RAM usage and GPU value disrupted the operation on some models.

It is desirable to try the classification process with BERT-large and other BERT family members in the future. Although it is not necessary to use word embedding in studies on BERT, in some studies this situation can be ignored and word embedding can be used. For the future, it is desirable to classify with a specific word embedding algorithm derived from patent texts and containing many patent-specific words, from legal terms to technological terms

CHAPTER 8
CONCLUSION

8.1 Introduction

This chapter consists of a content in which the doctoral thesis is summarized, its objective, its contributions to the literature are listed, its limitations are stated, and potential future studies are described.

In this thesis, Chapter 1 introduces the scope of the thesis and provides an overview of the topic of APC. Chapter 2 presents diverse data on global patenting. By examining this data, the statistical evidence supporting the rationale for conducting this thesis is revealed. Chapter 3 includes a review of academic studies in the field of APC. In the other chapters (Chapters 4 - 7), the studies carried out within the scope of this thesis in the area of APC, the solution approaches offered by these studies, and recommendations for future research are elaborated. This thesis includes a total of four research studies, one of which is bibliometric. Early versions of these works have been presented at various conferences to discuss their technical validity (Please see references). The extent to which it fulfills the stated objectives is discussed in detail in this section. In this context, this chapter provides an overview of the main issues discussed in each analysis carried out in this thesis.

Section 8.2 of this chapter discusses separately the research objectives and results of four different studies covering various approaches to the APC problem. The next section discusses the contributions of this research to knowledge. After stating the limitations in the implementation of this thesis, studies recommended for future implementation are presented. Finally, Section 8.3 concludes the thesis with a brief summary of the findings.

8.2 Objective of the Thesis

The main objective of the thesis is to present approaches that can provide systematic, effective and efficient solutions to automatic patent classification problems. To this

end, four different studies were conducted to provide solutions to automatic patent classification problems. (1) Proposing a systematic approach that brings together all bibliometric elements regarding academic studies and patents published in the field of APC and summarizes these studies in order to show trends and changes in the field. (2) Using the hierarchical attention network method to adapt it to the problem of classification of patents and also to show the effect of the change in hyperparameters on the classification performance of the algorithm. (3) To address the patent classification problem, we aim to develop a solution using the Transformer Encoder approach, which closely resembles human thinking logic. Comparing the results of deep learning techniques and their combined outputs. (4) Perform patent classification using the BERT method, an advanced version of the Transformer Encoder that includes its own specialized tokens.

8.3 Contributions of Thesis

8.3.1 Bibliometric Analysis Study

The bibliometric analysis study is expected to contribute the following to the literature.

- While explaining the background of this thesis in Chapter 2, it is evident that the number of patent applications worldwide is increasing rapidly. Concurrently with this increase, there has been a rise in the number of studies conducted in the field of APC. Therefore, this study reveals that the issue of APC is an active area of research.
- By outlining the direction in which the trends in the field of APC have shifted and the research methods used, the aspects where research in this field is lacking have been addressed, shedding light on the future of research in this field through word analysis.
- The detailed analysis of the prominent countries and authors in this field provided the researchers with a preliminary idea for potential study collaborations.
- Patent researchers were provided with initial information about the prominent affiliations, particularly universities, and inventors in this field.
- A comprehensive and detailed bibliometric analysis study was presented to researchers interested in conducting bibliometric analysis.

- Researchers were able to determine in which journals their publications would have the best chance of being published, along with the technical specifications of these journals.
- There are approaches that can facilitate the decision-making process for academics or researchers who are either continuing or have just started their studies in this field. These approaches can help them determine the direction in which their efforts will be directed

8.3.2 Hierarchical Attention Network Study

The HAN analysis study is expected to contribute the following to the literature.

- Experimental analyses were conducted systematically, aiming to identify the best parameter combination through various hyperparameter analyses.
- The analysis was repeated using different sections of the patent text, revealing the impact of text content on classification performance.
- The results of the classification process, which combines both word and sentence analysis, are presented. The results were compared with CNN and RNN, which are the most widely used deep learning methods in the literature.
- The Han method has not yet been applied to the Wipo-Alpha dataset and this study makes a significant contribution to the existing literature in this area.

8.3.3 Transformer Encoder Study

The transformer encoder analysis study is expected to contribute the following to the literature.

- To compare the results of the transformer encoder model, not only traditional deep learning methods were used, but also an ensemble deep learning approach was proposed and the comparison was based on this.
- To assess the performance of the ensemble deep learning approach and the transformer encoder method on the individual CNN and LSTM models, a number of experiments (35 experiments for deep learning architectures and 40 experiments for transformer encoder model architectures) are being carried out.

- The difference between the static and dynamic versions of the deep learning methods proposed in this study is related to the trainability of the word embedding factor.
- The Transformer encoder model yielded better results than the ensembled models. When comparing ensemble models internally, it is not possible to make a generalization for static and dynamic versions. While the dynamic CNN is better than the static CNN, the opposite is true for the LSTM.
- When only text sections of the patent were used as input in the analysis, the resulting accuracy values were consistently lower in all experiments compared to the versions where various combinations were used as input. This situation shows that the view in the literature of "performing the analysis using only one patent section" is not correct.

8.3.4 Bert study

- This study contributes to the literature by conducting experiments with different hyperparameters. Presenting the parameters used in detail both contributes to the literature and guides researchers conducting similar studies.
- It also involves incorporating four different text sections of the patent document and various combinations of them as input data for the model.
- Although there are many studies in the literature that conclude that the patent abstract or the patent claim section is sufficient in terms of selection of patent texts, the results obtained from this study are parallel to the results in Chapter 6. It is not enough to use only the summary or only the claim section. Using the full text content of the patent document allowed achieving higher classification accuracy. However, this causes high calculation time

8.4 Limitations of the Thesis

There are some limitations in this study.

- Limited hardware (better efficiency can be achieved by running methods on better hardware). In some studies, the study could not be completed due to insufficient hardware and returned an error.
- The one of the primary constraints of this study is the inability to carry out nonprobability sampling. A specific number of research articles were retrieved

from the database using particular keywords. Nevertheless, it is not feasible to generalize the results. The research results may not encompass all pertinent records that should be listed. A similar limitation may also apply to patent searches. The patent search conducted in the Scopus database yielded approximately 17,000 pieces of patent data. This value corresponds to 78 in the Web of Science (WOS). We favored using WOS output to achieve more targeted outcomes. In this case, like paper analysis, it is not feasible to generalize the results. However, this is the main issue in bibliometric studies.

- Although analyses are conducted in various experiments in all studies, it may be possible to find an analysis of a parameter value that was used in the literature but not evaluated in this study.

8.5 Recommendations for Future Research

- It is envisaged that creating a reliable patent word embedding vector and using it in deep learning studies will influence studies in this field in the future.
- Studies can be conducted that consider the comparison of different embedding vectors with each other and the impact of this situation on classification accuracy.
- The classification performance achieved by combining different models into an ensemble to create a specific model and comparing these methods can be examined.
- In this study, analyzes were carried out by considering only the textual content of the patent document. In the future, meta-data of patent texts may also be used as input.
- It is believed that focusing on case analysis in patent text classification is a task that will reduce the workload of patent experts. In these classification studies, it is thought that studies will be carried out on the detection of adversarial attacks with combinations of image and text classification.

Researchers can benefit from this thesis as it will enable them to direct their research path towards this prominent field.

REFERENCES

[1] P. Kurz, “Brunelleschi and Galilei: Super-Early Patents in Florence and Venice,” *2023 8th IEEE History of Electrotechnology Conference (HISTELCON)*, pp. 143–146, Sep. 2023, doi: 10.1109/HISTELCON56357.2023.10365949.

[2] WIPO, “Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore: Sixth Session,” in *Patents As A Source Of Technological Information In The Technology Transfer Process*, Switzerland: WIPO, 2004. Accessed: Dec. 31, 2023. [Online]. Available: https://www.wipo.int/meetings/en/details.jsp?meeting_id=5412&la=EN

[3] “International Classifications - IT Support Area.” Accessed: Dec. 31, 2023. [Online]. Available: https://www.wipo.int/classifications/ipc/en/ITsupport/Version20240101/transformations/stats.html

[4] C. Righi and T. Simcoe, “Patent examiner specialization,” *Res Policy*, vol. 48, no. 1, pp. 137–148, Feb. 2019, doi: 10.1016/J.RESPOL.2018.08.003.

[5] L. Leydesdorff, D. F. Kogler, and B. Yan, “Mapping patent classifications: portfolio and statistical analysis, and the comparison of strengths and weaknesses,” *Scientometrics*, vol. 112, no. 3, pp. 1573–1591, Sep. 2017, doi: 10.1007/S11192-017-2449-0/TABLES/7.

[6] “Cooperative Patent Classification (CPC) | Epo.org.” Accessed: Dec. 31, 2023. [Online]. Available: https://www.epo.org/en/searching-for-patents/helpful-resources/first-time-here/classification/cpc

[7] WIPO, “International Classifications - IT Support Area-2014.” Accessed: Dec. 31, 2023. [Online]. Available: https://www.wipo.int/classifications/ipc/en/ITsupport/Version20140101/transformations/stats.html

[8] WIPO, “International Classifications - IT Support Area-2024.” Accessed: Dec. 31, 2023. [Online]. Available: https://www.wipo.int/classifications/ipc/en/ITsupport/Version20240101/transformations/stats.html

[9] J. H. Kim and K. S. Choi, “Patent document categorization based on semantic structural information,” *Inf Process Manag*, vol. 43, no. 5, pp. 1200–1215, Sep. 2007, doi: 10.1016/J.IPM.2007.02.002.

[10] Y. H. Tseng, C. J. Lin, and Y. I. Lin, “Text mining techniques for patent analysis,” *Inf Process Manag*, vol. 43, no. 5, pp. 1216–1247, Sep. 2007, doi: 10.1016/j.ipm.2006.11.011.

[11] M. Iwayama, A. Fujii, N. Kando, and Y. Marukawa, “An empirical study on retrieval models for different document genres,” *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 251–258, Jul. 2003, doi: 10.1145/860435.860482.

[12] M. Iwayama, A. Fujii, N. Kando, and Y. Marukawa, “Evaluating patent retrieval in the third NTCIR workshop,” *Inf Process Manag*, vol. 42, no. 1, pp. 207–221, Jan. 2006, doi: 10.1016/J.IPM.2004.08.012.

[13] D. Eisinger, G. Tsatsaronis, M. Bundschus, U. Wieneke, and M. Schroeder, “Automated Patent Categorization and Guided Patent Search using IPC as Inspired by MeSH and PubMed,” *J Biomed Semantics*, vol. 4 Suppl 1, no. Suppl 1, Apr. 2013, doi: 10.1186/2041-1480-4-S1-S3.

[14] World Intellectual Property Organization, “World Intellectual Property Indicators 2023,” 2023. doi: 10.34667/TIND.48541.

[15] C. J. Fall, A. Törcsvári, K. Benzineb, and G. Karetka, “Automated categorization in the international patent classification,” *ACM SIGIR Forum*, vol. 37, no. 1, pp. 10–25, Apr. 2003, doi: 10.1145/945546.945547.

[16] R. Chikkamath, V. R. Parmar, Y. Otiefy, and M. Endres, “Patent Classification Using BERT-for-Patents on USPTO,” *ACM International Conference Proceeding Series*, pp. 20–28, Dec. 2022, doi: 10.1145/3578741.3578746.

[17] S. Don and D. Min, “Feature Selection for Automatic Categorization of Patent Documents,” *Indian J Sci Technol*, vol. 9, no. 37, Oct. 2016, doi: 10.17485/IJST/2016/V9I37/98974.

[18] S. Chakrabarti, B. Dom, and P. Indyk, “Enhanced hypertext categorization using hyperlinks,” *ACM SIGMOD Record*, vol. 27, no. 2, pp. 1–12, 1998, doi: 10.1145/276305.276332.

[19] A. J. C. Trappey, C. V. Trappey, T. A. Chiang, and Y. H. Huang, “Ontology-based neural network for patent knowledge management in design collaboration,” *Int J Prod Res*, vol. 51, no. 7, pp. 1992–2005, Apr. 2013, doi: 10.1080/00207543.2012.701775.

[20] A. J. C. Trappey, C. V. Trappey, and C. Y. Wu, “Automatic patent document summarization for collaborative knowledge systems and services,” *J Syst Sci Syst Eng*, vol. 18, no. 1, pp. 71–94, Mar. 2009, doi: 10.1007/S11518-009-5100-7.

[21] H. T. Loh, C. He, and L. Shen, “Automatic classification of patent documents for TRIZ users,” *World Patent Information*, vol. 28, no. 1, 2006, doi: 10.1016/j.wpi.2005.07.007.

[22] H. Cong and L. H. Tong, “Grouping of TRIZ Inventive Principles to facilitate automatic patent classification,” *Expert Syst Appl*, vol. 34, no. 1, 2008, doi: 10.1016/j.eswa.2006.10.015.

[23] S. Ardiansyah, M. A. Majid, and J. M. Zain, “Knowledge of extraction from trained neural network by using decision tree,” *Proceeding - 2016 2nd*

International Conference on Science in Information Technology, ICSITech 2016: Information Science for Green Society and Environment, pp. 220–225, Feb. 2017, doi: 10.1109/ICSITECH.2016.7852637.

[24] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevill-Manning, "KEA: Practical Automatic Keyphrase Extraction," *Proceedings of the ACM International Conference on Digital Libraries*, vol. 1999-January, pp. 254–255, Feb. 1999, doi: 10.48550/ARXIV.CS/9902007.

[25] J. Sangeetha and S. Jothilakshmi, "A novel spoken keyword spotting system using support vector machine," *Eng Appl Artif Intell*, vol. 36, pp. 287–293, Nov. 2014, doi: 10.1016/J.ENGAPPAI.2014.07.014.

[26] Y. Chen, J. Yin, W. Zhu, and S. Qiu, "Novel word features for keyword extraction," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9098, pp. 148–160, 2015, doi: 10.1007/978-3-319-21042-1_12/COVER.

[27] B. Armouty and S. Tedmori, "Automated keyword extraction using support vector machine from Arabic news documents," *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings*, pp. 342–346, May 2019, doi: 10.1109/JEEIT.2019.8717420.

[28] Z. Cailan and L. Shasha, "Research of Information Extraction Algorithm based on Hidden Markov Model," *2nd International Conference on Information Science and Engineering, ICISE2010 - Proceedings*, 2010, doi: 10.1109/ICISE.2010.5690348.

[29] H. G. Woo, J. Yeom, and C. Lee, "Screening early stage ideas in technology development processes: a text mining and k-nearest neighbours approach using patent information," *Technol Anal Strateg Manag*, vol. 31, no. 5, pp. 532–545, May 2019, doi: 10.1080/09537325.2018.1523386.

[30] Y. R. Li, L. H. Wang, and C. F. Hong, “Extracting the significant-rare keywords for patent analysis,” *Expert Syst Appl*, vol. 36, no. 3, pp. 5200–5204, Apr. 2009, doi: 10.1016/J.ESWA.2008.06.131.

[31] R. Mihalcea and P. Tarau, “TextRank: Bringing Order into Text.” pp. 404–411, 2004. Accessed: Dec. 25, 2023. [Online]. Available: https://aclanthology.org/W04-3252

[32] “Rapid automatic keyword extraction for information retrieval and analysis,” Sep. 2009.

[33] M. G. Thushara, M. S. Krishnapriya, and S. S. Nair, “Domain classification of research papers using hybrid keyphrase extraction method,” *Advances in Intelligent Systems and Computing*, vol. 708, pp. 387–398, 2018, doi: 10.1007/978-981-10-8636-6_40/COVER.

[34] J. Hu, S. Li, Y. Yao, L. Yu, G. Yang, and J. Hu, “Patent Keyword Extraction Algorithm Based on Distributed Representation for Patent Classification,” *Entropy (Basel)*, vol. 20, no. 2, Feb. 2018, doi: 10.3390/E20020104.

[35] J. Hu, S. Li, Y. Yao, L. Yu, G. Yang, and J. Hu, “Patent keyword extraction algorithm based on distributed representation for patent classification,” *Entropy*, vol. 20, no. 2, Feb. 2018, doi: 10.3390/E20020104.

[36] M. G. Thushara, T. Mownika, and R. Mangamuru, “A comparative study on different keyword extraction algorithms,” *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, pp. 969–973, Mar. 2019, doi: 10.1109/ICCMC.2019.8819630.

[37] C. H. Wu, Y. Ken, and T. Huang, “Patent classification system using a new hybrid genetic algorithm support vector machine,” *Applied Soft Computing Journal*, vol. 10, no. 4, pp. 1164–1177, 2010, doi: 10.1016/J.ASOC.2009.11.033.

[38] C. Xue, Q. Y. Qiu, P. E. Feng, and Z. N. Yao, “An automatic classification method for patents,” *Proceedings - 2010 7th International Conference on Fuzzy*

Systems and Knowledge Discovery, FSKD 2010, vol. 4, pp. 1497–1501, 2010, doi: 10.1109/FSKD.2010.5569326.

[39] B. Nugroho, M. Aritsugi, Y. Otachi, and Y. Manabe, "Combined graph kernels for automatic patent classification: A hybrid approach," *World Patent Information*, vol. 57, pp. 18–24, Jun. 2019, doi: 10.1016/J.WPI.2019.03.002.

[40] S. Venugopalan and V. Rai, "Topic based classification and pattern identification in patents," *Technol Forecast Soc Change*, vol. 94, pp. 236–250, 2015, doi: 10.1016/J.TECHFORE.2014.10.006.

[41] J. Yun and Y. Geum, "Automated classification of patents: A topic modeling approach," *Comput Ind Eng*, vol. 147, Sep. 2020, doi: 10.1016/J.CIE.2020.106636.

[42] K. V. Neethu, R. Binu, N. Vahab, S. Mathew, and L. Khamari, "An NLP based information extraction system for patents," *ACM International Conference Proceeding Series*, vol. 25-26-August-2016, Aug. 2016, doi: 10.1145/2980258.2982100.

[43] D. De Clercq, N. F. Diop, D. Jain, B. Tan, and Z. Wen, "Multi-label classification and interactive NLP-based visualization of electric vehicle patent data," *World Patent Information*, vol. 58, p. 101903, Sep. 2019, doi: 10.1016/J.WPI.2019.101903.

[44] M. F. Grawe, C. A. Martins, and A. G. Bonfante, "Automated patent classification using word embedding," *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, vol. 2017-December, pp. 408–411, 2017, doi: 10.1109/ICMLA.2017.0-127.

[45] S. Li, J. Hu, Y. Cui, and J. Hu, "DeepPatent: patent classification with convolutional neural networks and word embedding," *Scientometrics*, vol. 117, no. 2, pp. 721–744, Nov. 2018, doi: 10.1007/s11192-018-2905-5.

[46] M. Shalaby, J. Stutzki, M. Schubert, and S. Günnemann, "An LSTM approach to patent classification based on fixed hierarchy vectors," *SIAM International*

Conference on Data Mining, SDM 2018, pp. 495–503, 2018, doi: 10.1137/1.9781611975321.56.

[47] S. Choi, H. Lee, E. Park, and S. Choi, "Deep learning for patent landscaping using transformer and graph embedding," *Technol Forecast Soc Change*, vol. 175, p. 121413, Feb. 2022, doi: 10.1016/J.TECHFORE.2021.121413.

[48] T. Xinyu, Z. Ruijie, and L. Yonghe, "Multi-label Patent Classification with Pre-training Model," *Data Analysis and Knowledge Discovery*, vol. 6, no. 2–3, pp. 129–137, Mar. 2022, doi: 10.11925/INFOTECH.2096-3467.2021.0930.

[49] R. Krestel, R. Chikkamath, C. Hewel, and J. Risch, "A survey on deep learning for patent analysis," *World Patent Information*, vol. 65, p. 102035, Jun. 2021, doi: 10.1016/J.WPI.2021.102035.

[50] A. J. C. Trappey, S. C. I. Lin, and A. C. L. Wang, "Using neural network categorization method to develop an innovative knowledge management technology for patent document classification," in *Proceedings of the 9th International Conference on Computer Supported Cooperative Work in Design*, 2005. doi: 10.1109/cscwd.2005.194293.

[51] D. Tikk, G. Biró, and A. Törcsvári, "A Hierarchical Online Classifier for Patent Categorization," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-59904-373-9.ch012*, pp. 244–267, Jan. 1AD, doi: 10.4018/978-1-59904-373-9.CH012.

[52] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, "Support vector machine learning for interdependent and structured output spaces," *Twenty-first international conference on Machine learning - ICML '04*, p. 104, 2004, doi: 10.1145/1015330.1015341.

[53] S. V. N. Vishwanathan, N. N. Schraudolph, and A. J. Smola, "Step size adaptation in reproducing kernel Hilbert space," *Journal of Machine Learning Research*, vol. 7, pp. 1107–1133, Jun. 2006.

[54] L. Cai and T. Hofmann, “Hierarchical document categorization with Support Vector Machines,” in *International Conference on Information and Knowledge Management, Proceedings*, Association for Computing Machinery (ACM), 2004, pp. 78–87. doi: 10.1145/1031171.1031186.

[55] Y. L. Chen and Y. C. Chang, “A three-phase method for patent classification,” *Inf Process Manag*, vol. 48, no. 6, pp. 1017–1030, Nov. 2012, doi: 10.1016/j.ipm.2011.11.001.

[56] A. Giachanou, M. Salampasis, and G. Paltoglou, “Multilayer source selection as a tool for supporting patent search and classification,” *Inf Retr Boston*, vol. 18, no. 6, pp. 559–585, Dec. 2015, doi: 10.1007/S10791-015-9270-2.

[57] E. D’hondt, S. Verberne, C. Koster, and L. Boves, “Text representations for patent classification,” *Computational Linguistics*, vol. 39, no. 3, pp. 755–775, 2013, doi: 10.1162/COLI_A_00149.

[58] A. H. Roudsari, · Jafar Afshar, · Wookey Lee, and S. Lee, “PatentNet: multi-label classification of patent documents using deep learning based language understanding,” *Scientometrics*, vol. 127, pp. 207–231, 123AD, doi: 10.1007/s11192-021-04179-4.

[59] S. T. Aroyehun, J. Angel, N. Majumder, A. Gelbukh, and A. Hussain, “Leveraging label hierarchy using transfer and multi-task learning: A case study on patent classification,” *Neurocomputing*, vol. 464, pp. 421–431, Nov. 2021, doi: 10.1016/j.neucom.2021.07.057.

[60] J.-S. Lee and J. Hsiang, “PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model,” May 2019, Accessed: Dec. 27, 2023. [Online]. Available: https://arxiv.org/abs/1906.02124v2

[61] A. H. Roudsari, J. Afshar, C. C. Lee, and W. Lee, “Multi-label Patent Classification using Attention-Aware Deep Learning Model; Multi-label Patent Classification using Attention-Aware Deep Learning Model,” 2020, doi: 10.1109/BigComp48618.2020.000-2.

[62] J. S. Lee and J. Hsiang, “Patent classification by fine-tuning BERT language model,” *World Patent Information*, vol. 61, p. 101965, Jun. 2020, doi: 10.1016/J.WPI.2020.101965.

[63] F. Aiolli, R. Cardin, F. Sebastiani, and A. Sperduti, “Preferential text classification: Learning algorithms and evaluation measures,” *Inf Retr Boston*, vol. 12, no. 5, pp. 559–580, Oct. 2009, doi: 10.1007/s10791-008-9071-y.

[64] D. Tikk, G. Biró, and A. Törcsvári, “A hierarchical online classifier for patent categorization,” *Emerging Technologies of Text Mining: Techniques and Applications*, pp. 244–267, 2007, doi: 10.4018/978-1-59904-373-9.CH012.

[65] J. Beney, “LCI-INSA Linguistic Experiment for CLEF-IP Classification Track.,” 2010, Accessed: Jan. 19, 2023. [Online]. Available: http://ceur-ws.org/Vol-1176/CLEF2010wn-CLEF-IP-Beney2010.pdf

[66] A. J. C. Trappey, F. C. Hsu, C. V. Trappey, and C. I. Lin, “Development of a patent document classification and search platform using a back-propagation network,” *Expert Syst Appl*, vol. 31, no. 4, 2006, doi: 10.1016/j.eswa.2006.01.013.

[67] D. Hull *et al.*, “Language technologies and patent search and classification,” *World Patent Information*, vol. 23, no. 3, pp. 265–268, Sep. 2001, doi: 10.1016/S0172-2190(01)00024-2.

[68] S. Yucesoy, T. Dereli, and A. Durmusoglu, “Patent Classification via Textual Analysis Which Sections to be Included?,” in *2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018*, 2019. doi: 10.1109/IDAP.2018.8620929.

[69] J. Hu, S. Li, J. Hu, and G. Yang, “A Hierarchical Feature Extraction Model for Multi-Label Mechanical Patent Classification,” *Sustainability 2018, Vol. 10, Page 219*, vol. 10, no. 1, p. 219, Jan. 2018, doi: 10.3390/SU10010219.

[70] J. Hepburn, “Universal language model fine-tuning for patent classification,” in *Proceedings of the Australasian Language Technology Association Workshop*

2018, Sunghwan Mac Kim and Xiuzhen (Jenny) Zhang, Eds., New Zealand, 2018, pp. 93–96. Accessed: Dec. 29, 2023. [Online]. Available: https://aclanthology.org/U18-1013/

[71] S. C. Pujari, A. Friedrich, and J. Strötgen, "A Multi-task Approach to Neural Multi-label Hierarchical Patent Classification Using Transformers," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12656 LNCS, pp. 513–528, 2021, doi: 10.1007/978-3-030-72113-8_34.

[72] J. Rousu JUHOROUSU, C. Saunders, S. Szedmak, J. Shawe-Taylor, K. P. Bennett, and E. Parrado-Hernández, "Kernel-Based Learning of Hierarchical Multilabel Classification Models *," *Journal of Machine Learning Research*, vol. 7, pp. 1601–1626, 2006.

[73] D. Tikk and G. Biró, "Experiment with a hierarchical text categorization method on the WIPO-alpha patent collection," *4th International Symposium on Uncertainty Modeling and Analysis, ISUMA 2003*, pp. 104–109, 2003, doi: 10.1109/ISUMA.2003.1236148.

[74] M. Sofean, "Deep learning based pipeline with multichannel inputs for patent classification," *World Patent Information*, vol. 66, p. 102060, Sep. 2021, doi: 10.1016/J.WPI.2021.102060.

[75] Q. Li *et al.*, "A Survey on Text Classification: From Shallow to Deep Learning," Aug. 2020, doi: 10.48550/arxiv.2008.00364.

[76] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, Association for Computational Linguistics (ACL), 2016, pp. 1480–1489. doi: 10.18653/v1/n16-1174.

[77] W. Huang *et al.*, "Hierarchical multi-label text classification: An attention-based recurrent network approach," in *International Conference on Information*

and Knowledge Management, Proceedings, New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1051–1060. doi: 10.1145/3357384.3357885.

[78] J.-S. Lee and J. Hsiang, "PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model," May 2019, Accessed: Dec. 29, 2021. [Online]. Available: http://arxiv.org/abs/1906.02124

[79] Y. Dong, P. Liu, Z. Zhu, Q. Wang, and Q. Zhang, "A Fusion Model-Based Label Embedding and Self-Interaction Attention for Text Classification," *IEEE Access*, vol. 8, pp. 30548–30559, 2020, doi: 10.1109/ACCESS.2019.2954985.

[80] Y. Zeng, H. Yang, Y. Feng, Z. Wang, and D. Zhao, "A convolution BiLSTM neural network model for chinese event extraction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10102, Springer Verlag, 2016, pp. 275–287. doi: 10.1007/978-3-319-50496-4_23.

[81] J.-S. Lee and J. Hsiang, "Patent classification by fine-tuning BERT language model," *World Patent Information*, vol. 61, p. 101965, Jun. 2020, doi: 10.1016/j.wpi.2020.101965.

[82] S. Choi, H. Lee, E. L. Park, and S. Choi, "Deep Patent Landscaping Model Using Transformer and Graph Embedding," Mar. 2019, Accessed: Jun. 06, 2022. [Online]. Available: http://arxiv.org/abs/1903.05823

[83] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: https://github.com/tensorflow/tensor2tensor

[84] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?," *Lecture Notes in Computer Science*, vol. 11856 LNAI, pp. 194–206, 2019, doi: 10.1007/978-3-030-32381-3_16.

[85] A. H. Roudsari, · Jafar Afshar, · Wookey Lee, and S. Lee, "PatentNet: multi-label classification of patent documents using deep learning based language understanding," *Scientometrics*, 123AD, doi: 10.1007/s11192-021-04179-4.

[86] H. Bekamiri, D. S. Hain, and R. Jurowetzki, "PatentSBERTa: A Deep NLP based Hybrid Model for Patent Distance and Classification using Augmented SBERT," 2021, [Online]. Available: https://huggingface.co/AI-Growth-Lab/PatentSBERTa/

[87] J. S. Lee and J. Hsiang, "Patent classification by fine-tuning BERT language model," *World Patent Information*, vol. 61, 2020, doi: 10.1016/j.wpi.2020.101965.

[88] J. Wang, Z. Zhang, L. Feng, K. Y. Lin, and P. Liu, "Development of technology opportunity analysis based on technology landscape by extending technology elements with BERT and TRIZ," *Technol Forecast Soc Change*, vol. 191, p. 122481, Jun. 2023, doi: 10.1016/j.techfore.2023.122481.

[89] Z. Qiu and Z. Wang, "What is your next invention? — A framework of mining technological development rules and assisting in designing new technologies based on BERT as well as patent citations," *Comput Ind*, vol. 145, Feb. 2023, doi: 10.1016/j.compind.2022.103829.

[90] S. Choi, H. Lee, E. Park, and S. Choi, "Deep learning for patent landscaping using transformer and graph embedding," *Technol Forecast Soc Change*, vol. 175, p. 121413, Feb. 2022, doi: 10.1016/J.TECHFORE.2021.121413.

[91] A. Vaswani *et al.*, "Attention Is All You Need."

[92] J. C. Gomez, "Analysis of the effect of data properties in automated patent classification," *Scientometrics*, vol. 121, no. 3, pp. 1239–1268, Dec. 2019, doi: 10.1007/s11192-019-03246-1.

[93] L. Abdelgawad, P. Kluegl, E. Genc, S. Falkner, and F. Hutter, "Optimizing Neural Networks for Patent Classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes*

in Bioinformatics), vol. 11908 LNAI, pp. 688–703, 2020, doi: 10.1007/978-3-030-46133-1_41/COVER.

[94] L. Zhang, Z. Liu, L. Li, C. Shen, and T. Li, "PatSearch: an integrated framework for patentability retrieval," *Knowledge and Information Systems 2017 57:1*, vol. 57, no. 1, pp. 135–158, Nov. 2017, doi: 10.1007/S10115-017-1127-0.

[95] H. Smith, "Automation of patent classification," *World Patent Information*, vol. 24, no. 4. Elsevier Ltd, pp. 269–271, Dec. 01, 2002. doi: 10.1016/S0172-2190(02)00067-4.

[96] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Applied Soft Computing Journal*, vol. 86, p. 105836, Jan. 2020, doi: 10.1016/j.asoc.2019.105836.

[97] J. Yun and Y. Geum, "Automated classification of patents: A topic modeling approach," *Comput Ind Eng*, vol. 147, p. 106636, Sep. 2020, doi: 10.1016/j.cie.2020.106636.

[98] C. H. Wu, Y. Ken, and T. Huang, "Patent classification system using a new hybrid genetic algorithm support vector machine," *Appl Soft Comput*, vol. 10, no. 4, pp. 1164–1177, Sep. 2010, doi: 10.1016/J.ASOC.2009.11.033.

[99] C. H. Wu, Y. Ken, and T. Huang, "The support vector machine classification system for patient document information importance analysis," in *BioMedical Engineering and Informatics: New Development and the Future - Proceedings of the 1st International Conference on BioMedical Engineering and Informatics, BMEI 2008*, 2008, pp. 375–379. doi: 10.1109/BMEI.2008.178.

[100] Q. N. Li and T. H. Li, "Research on the application of Naive Bayes and Support Vector Machine algorithm on exercises Classification," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jan. 2020, p. 012071. doi: 10.1088/1742-6596/1437/1/012071.

[101] B. Xia, B. LI, and X. Lv, “Research on Patent Document Classification Based on Deep Learning,” 2016. doi: 10.2991/aiie-16.2016.71.

[102] Y. Liu, L. Wang, T. Shi, and J. Li, “Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM,” *Inf Syst*, vol. 103, Jan. 2022, doi: 10.1016/j.is.2021.101865.

[103] Y. Lu, X. Xiong, W. Zhang, J. Liu, and R. Zhao, “Research on classification and similarity of patent citation based on deep learning,” *Scientometrics*, vol. 123, no. 2, pp. 813–839, May 2020, doi: 10.1007/s11192-020-03385-w.

[104] S. Jiang, J. Hu, C. L. Magee, and J. Luo, “Deep Learning for Technical Document Classification,” *IEEE Trans Eng Manag*, pp. 1–17, 2022, doi: 10.1109/TEM.2022.3152216.

[105] D. Caragea *et al.*, “Identifying FinTech Innovations Using BERT; Identifying FinTech Innovations Using BERT,” *2020 IEEE International Conference on Big Data (Big Data)*, 2020, doi: 10.1109/BigData50022.2020.9378169.

[106] L. Xiaolei and N. Bin, “BERT-CNN: a Hierarchical Patent Classifier Based on a Pre-Trained Language Model.”

[107] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.” [Online]. Available: https://github.com/huggingface/transformers

[108] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” 2019. [Online]. Available: https://github.com/pytorch/fairseq

[109] S. Choi, H. Lee, E. Park, and S. Choi, “Deep learning for patent landscaping using transformer and graph embedding,” *Technol Forecast Soc Change*, vol. 175, p. 121413, Feb. 2022, doi: 10.1016/J.TECHFORE.2021.121413.

[110] A. H. Roudsari, · Jafar Afshar, · Wookey Lee, and S. Lee, “PatentNet: multi-label classification of patent documents using deep learning based language understanding,” *Scientometrics*, 123AD, doi: 10.1007/s11192-021-04179-4.

[111] D. de Clercq, N. F. Diop, D. Jain, B. Tan, and Z. Wen, “Multi-label classification and interactive NLP-based visualization of electric vehicle patent data,” *World Patent Information*, vol. 58, Sep. 2019, doi: 10.1016/j.wpi.2019.101903.

[112] S. C. Pujari, A. Friedrich, and J. Strötgen, “A Multi-task Approach to Neural Multi-label Hierarchical Patent Classification Using Transformers,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 513–528. doi: 10.1007/978-3-030-72113-8_34.

[113] M. Sofean, “Deep learning based pipeline with multichannel inputs for patent classification,” *World Patent Information*, vol. 66, p. 102060, Sep. 2021, doi: 10.1016/J.WPI.2021.102060.

[114] H. G. Woo, J. Yeom, and C. Lee, “Screening early stage ideas in technology development processes: a text mining and k-nearest neighbours approach using patent information,” *https://doi.org/10.1080/09537325.2018.1523386*, vol. 31, no. 5, pp. 532–545, May 2018, doi: 10.1080/09537325.2018.1523386.

[115] Y. Li and J. Shawe-Taylor, “Advanced learning algorithms for cross-language patent retrieval and classification,” *Inf Process Manag*, vol. 43, no. 5, pp. 1183–1199, Sep. 2007, doi: 10.1016/J.IPM.2006.11.005.

[116] Z. Ruijie, X. Ying, J. Shuaichen, and L. Yonghe, “Patent text modeling strategy and its classification based on structural features,” *World Patent Information*, vol. 67, p. 102084, Dec. 2021, doi: 10.1016/J.WPI.2021.102084.

[117] J. Kim, J. Yoon, E. Park, and S. Choi, “Patent document clustering with deep embeddings,” *Scientometrics*, vol. 123, no. 2, pp. 563–577, May 2020, doi: 10.1007/s11192-020-03396-7.

[118] J. Risch and R. Krestel, “Domain-specific word embeddings for patent classification,” *Data Technologies and Applications*, vol. 53, no. 1, 2019, doi: 10.1108/DTA-01-2019-0002.

[119] J. Risch and R. Krestel, “Learning patent speak: Investigating domain-specific word embeddings,” in *2018 13th International Conference on Digital Information Management, ICDIM 2018*, Institute of Electrical and Electronics Engineers Inc., Sep. 2018, pp. 63–68. doi: 10.1109/ICDIM.2018.8846972.

[120] A. Abbas, L. Zhang, and S. U. Khan, “A literature review on the state-of-the-art in patent analysis,” *World Patent Information*, vol. 37, pp. 3–13, Jun. 2014, doi: 10.1016/J.WPI.2013.12.006.

[121] R. Krestel, R. Chikkamath, C. Hewel, and J. Risch, “A survey on deep learning for patent analysis,” *World Patent Information*, vol. 65, p. 102035, Jun. 2021, doi: 10.1016/j.wpi.2021.102035.

[122] F. Madani and C. Weber, “The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis,” *World Patent Information*, vol. 46, pp. 32–48, Sep. 2016, doi: 10.1016/j.wpi.2016.05.008.

[123] J. C. Gomez and M. F. Moens, “A survey of automated hierarchical classification of patents,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8830, pp. 215–249, 2014, doi: 10.1007/978-3-319-12511-4_11/COVER.

[124] A. Agarwal *et al.*, “Bibliometrics: Tracking research impact by selecting the appropriate metrics,” *Asian Journal of Andrology*, vol. 18, no. 2. Medknow Publications, pp. 296–309, Mar. 01, 2016. doi: 10.4103/1008-682X.171582.

[125] O. J. de Oliveira *et al.*, “Bibliometric Method for Mapping the State-of-the-Art and Identifying Research Gaps and Trends in Literature: An Essential Instrument to Support the Development of Scientific Projects,” *Scientometrics Recent Advances*, Nov. 2019, doi: 10.5772/INTECHOPEN.85856.

[126] S. Yucesoy Kahraman, T. Dereli, and A. Durmusoglu, “Forty years of automated patent classification,” *Int J Inf Technol Decis Mak*, Jan. 2023, doi: 10.1142/S0219622023500165.

[127] G. Demir, P. Chatterjee, and D. Pamucar, “Sensitivity analysis in multi-criteria decision making: A state-of-the-art research perspective using bibliometric analysis,” *Expert Syst Appl*, vol. 237, p. 121660, Mar. 2024, doi: 10.1016/J.ESWA.2023.121660.

[128] A. Durmuşoğlu, “A pre-assessment of past research on the topic of environmental-friendly electronics,” *J Clean Prod*, vol. 129, pp. 305–314, Aug. 2016, doi: 10.1016/J.JCLEPRO.2016.04.068.

[129] M. Pradana, H. P. Elisa, and S. Syarifuddin, “Discussing ChatGPT in education: A literature review and bibliometric analysis,” *Cogent Education*, vol. 10, no. 2, Dec. 2023, doi: 10.1080/2331186X.2023.2243134.

[130] P. Yildiz Gülhan and M. Nurullah KURUTKAN, “Bibliometric Analysis of COVID-19 Publications in the Field of Chest and Infectious Diseases Göğüs ve Enfeksiyon Hastalıkları Alanındaki COVID-19 Yayınların Bibliyometrik Analizi,” *Düzce Tıp Fak Derg*, vol. 23, no. 1, pp. 30–40, 2021, doi: 10.18678/dtfd.826465.

[131] P. Kokol and H. B. Vošner, “Discrepancies among Scopus, Web of Science, and PubMed coverage of funding information in medical journal articles,” *Journal of the Medical Library Association*, vol. 106, no. 1, pp. 81–86, Jan. 2018, doi: 10.5195/jmla.2018.181.

[132] M. Tober, “PubMed, ScienceDirect, Scopus or Google Scholar - Which is the best search engine for an effective literature research in laser medicine?,” *Medical Laser Application*, vol. 26, no. 3, pp. 139–144, Aug. 2011, doi: 10.1016/j.mla.2011.05.006.

[133] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, and G. Pappas, “Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses,” *The FASEB Journal*, vol. 22, no. 2, pp. 338–342, Feb. 2008, doi: 10.1096/fj.07-9492lsf.

[134] Z. Taşkın and A. U. Aydinoglu, “Collaborative interdisciplinary astrobiology research: A bibliometric study of the nasa astrobiology institute,”

Scientometrics, vol. 103, no. 3, pp. 1003–1022, Jun. 2015, doi: 10.1007/s11192-015-1576-8.

[135] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J Informetr*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.

[136] A. ASAN and A. ASLAN, "Quartile Scores of Scientific Journals: Meaning, Importance and Usage," *Acta Medica Alanya*, vol. 4, no. 1, pp. 102–108, Mar. 2020, doi: 10.30565/MEDALANYA.653661.

[137] "Measuring a journals impact." Accessed: May 06, 2021. [Online]. Available: https://www.elsevier.com/authors/tools-and-resources/measuring-a-journals-impact

[138] W. Intellectual Property Organization, *World Intellectual Property Indicators 2020 2*.

[139] S. H. Zyoud and D. Fuchs-Hanusch, "A bibliometric-based survey on AHP and TOPSIS techniques," *Expert Systems with Applications*, vol. 78. Elsevier Ltd, pp. 158–181, Jul. 15, 2017. doi: 10.1016/j.eswa.2017.02.016.

[140] Y. H. Tseng, C. J. Lin, and Y. I. Lin, "Text mining techniques for patent analysis," *Inf Process Manag*, vol. 43, no. 5, pp. 1216–1247, Sep. 2007, doi: 10.1016/j.ipm.2006.11.011.

[141] T. Nasukawa and T. Nagano, "Text analysis and knowledge mining system," *IBM Systems Journal*, vol. 40, pp. 967–984, 2001, doi: 10.1147/sj.404.0967.

[142] Y. Zhang, A. L. Porter, Z. Hu, Y. Guo, and N. C. Newman, "'Term clumping' for technical intelligence: A case study on dye-sensitized solar cells," *Technol Forecast Soc Change*, vol. 85, pp. 26–39, 2014, doi: 10.1016/J.TECHFORE.2013.12.019.

[143] J. Joung and K. Kim, "Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data," *Technol*

Forecast Soc Change, vol. 114, pp. 281–292, Jan. 2017, doi: 10.1016/J.TECHFORE.2016.08.020.

[144] R. Bawa, "Patents and nanomedicine," *Nanomedicine*, vol. 2, no. 3, pp. 351–374, Jun. 2007, doi: 10.2217/17435889.2.3.351.

[145] T. Teichert and M. A. Mittermayer, "Text mining for technology monitoring IEEE IEMC 2002," in *IEEE International Engineering Management Conference*, 2002, pp. 596–601.

[146] C. J. Fall, A. Törcsvári, P. Fiévet, and G. Karetka, "Automated categorization of German-language patent documents," *Expert Syst Appl*, vol. 26, no. 2, pp. 269–277, Feb. 2004, doi: 10.1016/S0957-4174(03)00141-6.

[147] Y. Liang, R. Tan, C. Wang, and Z. Li, "Computer-aided classification of patents oriented to TRIZ," in *IEEM 2009 - IEEE International Conference on Industrial Engineering and Engineering Management*, 2009. doi: 10.1109/IEEM.2009.5372983.

[148] C. Xue, Q. Y. Qiu, P. E. Feng, and Z. N. Yao, "An automatic classification method for patents," in *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2010*, 2010. doi: 10.1109/FSKD.2010.5569326.

[149] C. Y. Chiu and P. T. Huang, "Application of the honeybee mating optimization algorithm to patent document classification in combination with the Support Vector Machine," *International Journal of Automation and Smart Technology*, vol. 3, no. 3, pp. 179–191, 2013, doi: 10.5875/ausmt.v3i3.183.

[150] S. H. Liu, H. L. Liao, S. M. Pi, and J. W. Hu, "Development of a patent retrieval and analysis platform - A hybrid approach," *Expert Syst Appl*, vol. 38, no. 6, pp. 7864–7868, Jun. 2011, doi: 10.1016/j.eswa.2010.12.114.

[151] M. Bermudez-Edo, M. Noguera, N. Hurtado-Torres, M. V. Hurtado, and J. L. Garrido, "Adding sense to patent ontologies: A representation of concepts and

reasoning," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2013, pp. 67–75. doi: 10.1007/978-3-319-00563-8_9.

[152] L. Huang, S. Xu, G. Hu, C. Zhang, and N. N. Xiong, "Design and analysis of a weight-lda model to extract implicit topic of database in social networks," *Journal of Internet Technology*, vol. 18, no. 6, pp. 1393–1406, 2017, doi: 10.6138/JIT.2017.18.6.20170509.

[153] F. Zhu, X. Wang, D. Zhu, and Y. Liu, "User demand-driven patent topic classification using machine learning techniques," in *Decision Making and Soft Computing - Proceedings of the 11th International FLINS Conference, FLINS 2014*, World Scientific Publishing Co. Pte Ltd, 2014, pp. 657–663. doi: 10.1142/9789814619998_0108.

[154] S. Venugopalan and V. Rai, "Topic based classification and pattern identification in patents," *Technol Forecast Soc Change*, vol. 94, 2015, doi: 10.1016/j.techfore.2014.10.006.

[155] M. F. Grawe, C. A. Martins, and A. G. Bonfante, "Automated patent classification using word embedding," in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 408–411. doi: 10.1109/ICMLA.2017.0-127.

[156] S. Li, J. Hu, Y. Cui, and J. Hu, "DeepPatent: patent classification with convolutional neural networks and word embedding," *Scientometrics*, vol. 117, pp. 721–744, 2018, doi: 10.1007/s11192-018-2905-5.

[157] H.-C. Lo and J.-M. Chu, "Pre-trained Transformer-based Classification for Automated Patentability Examination," Institute of Electrical and Electronics Engineers (IEEE), Mar. 2022, pp. 1–5. doi: 10.1109/csde53843.2021.9718474.

[158] M. Freunek and A. Bodmer, "BERT based freedom to operate patent analysis," 2021, [Online]. Available: https://worldwide.espacenet.com/classification?locale=en_EP.

[159] J. S. Lee and J. Hsiang, “Patent classification by fine-tuning BERT language model,” *World Patent Information*, vol. 61, p. 101965, Jun. 2020, doi: 10.1016/j.wpi.2020.101965.

[160] C. J. Fall, A. Törcsvári, K. Benzineb, and G. Karetka, “Automated categorization in the international patent classification,” *ACM SIGIR Forum*, vol. 37, no. 1, pp. 10–25, Apr. 2003, doi: 10.1145/945546.945547.

[161] S. Gao *et al.*, “Hierarchical attention networks for information extraction from cancer pathology reports,” *Journal of the American Medical Informatics Association*, vol. 25, no. 3, pp. 321–330, Mar. 2018, doi: 10.1093/jamia/ocx131.

[162] C. Qin and C. Zhang, “Which structure of academic articles do referees pay more attention to?: perspective of peer review and full-text of academic articles,” *Aslib Journal of Information Management*, 2022, doi: 10.1108/AJIM-05-2022-0244.

[163] A. Jacovi, O. S. Shalom, and Y. Goldberg, “Understanding Convolutional Neural Networks for Text Classification,” *EMNLP 2018 - 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Proceedings of the 1st Workshop*, pp. 56–65, Sep. 2018, doi: 10.18653/v1/w18-5408.

[164] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. E. Barnes, and D. E. Brown, “Text Classification Algorithms: A Survey,” *Information (Switzerland)*, vol. 10, no. 4, Apr. 2019, doi: 10.3390/info10040150.

[165] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, “Stacked convolutional auto-encoders for hierarchical feature extraction,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6791 LNCS, no. PART 1, pp. 52–59, 2011, doi: 10.1007/978-3-642-21735-7_7/COVER.

[166] Y. Goldberg, “A Primer on Neural Network Models for Natural Language Processing,” *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, Oct. 2015, doi: 10.1613/jair.4992.

[167] K. Kowsari, D. E. Brown, M. Heidarysafa, K. J. Meimandi, M. S. Gerber, and L. E. Barnes, "HDLTex: Hierarchical Deep Learning for Text Classification," *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, vol. 2017-December, pp. 364–371, Sep. 2017, doi: 10.1109/ICMLA.2017.0-134.

[168] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent Advances in Recurrent Neural Networks," Dec. 2017, Accessed: Jan. 03, 2024. [Online]. Available: https://arxiv.org/abs/1801.01078v3

[169] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC Trends in Analytical Chemistry*, vol. 132, p. 116045, Nov. 2020, doi: 10.1016/J.TRAC.2020.116045.

[170] D. Tiwari and B. Nagpal, "Enhanced long short-term memory (Elstm) model for sentiment analysis," *International Arab Journal of Information Technology*, vol. 18, no. 6, pp. 846–855, Nov. 2021, doi: 10.34028/iajit/18/6/12.

[171] Z. Touati-Hamad, M. R. Laouar, I. Bendib, and S. Hakak, "Arabic Quran Verses Authentication Using Deep Learning and Word Embeddings," *International Arab Journal of Information Technology*, vol. 19, no. 4, pp. 681–688, Jul. 2022, doi: 10.34028/iajit/19/4/13.

[172] J. Wu, C. Huang, and Y. Chen, "Patent Text Classification Study Based on Bi-LSTM-A Model," in *2020 5th International Conference on Control, Robotics and Cybernetics, CRC 2020*, 2020. doi: 10.1109/CRC51253.2020.9253461.

[173] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," in *In Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. Accessed: Jun. 08, 2022. [Online]. Available: https://nlp.stanford.edu/projects/glove/

[174] S. Gao *et al.*, "Hierarchical attention networks for information extraction from cancer pathology reports," *J Am Med Inform Assoc*, vol. 25, no. 3, pp. 321–330, Mar. 2018, doi: 10.1093/JAMIA/OCX131.

[175] W. F. Cascio and R. Montealegre, “How Technology Is Changing Work and Organizations,” *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 3, no. 1, pp. 349–375, Mar. 2016, doi: 10.1146/annurev-orgpsych-041015-062352.

[176] S. Altuntas, T. Dereli, and A. Kusiak, “Forecasting technology success based on patent data,” *Technol Forecast Soc Change*, vol. 96, pp. 202–214, Jul. 2015, doi: 10.1016/j.techfore.2015.03.011.

[177] M. Tsouchnika, A. Smolyak, P. Argyrakis, and S. Havlin, “Patent collaborations: From segregation to globalization,” *J Informetr*, vol. 16, no. 1, p. 101238, Feb. 2022, doi: 10.1016/j.joi.2021.101238.

[178] T. Daim *et al.*, “Patent analysis of wind energy technology using the patent alert system,” *World Patent Information*, vol. 34, no. 1, pp. 37–47, Mar. 2012, doi: 10.1016/j.wpi.2011.11.001.

[179] W. qiang Li, Y. Li, J. Chen, and C. yi Hou, “Product functional information based automatic patent classification: Method and experimental studies,” *Inf Syst*, vol. 67, 2017, doi: 10.1016/j.is.2017.03.007.

[180] Y. Zhang *et al.*, “A hybrid similarity measure method for patent portfolio analysis,” *J Informetr*, vol. 10, no. 4, pp. 1108–1130, Nov. 2016, doi: 10.1016/j.joi.2016.09.006.

[181] C. Riedl *et al.*, “Detecting figures and part labels in patents: competition-based development of graphics recognition algorithms,” *International Journal on Document Analysis and Recognition*, vol. 19, no. 2, pp. 155–172, Jun. 2016, doi: 10.1007/s10032-016-0260-8.

[182] N. Bhatti and A. Hanbury, “Image search in patents: A review,” *International Journal on Document Analysis and Recognition*, vol. 16, no. 4. pp. 309–329, Dec. 2013. doi: 10.1007/s10032-012-0197-5.

[183] C. F. Tsai and C. W. Chang, “SVOIS: Support vector oriented instance selection for text classification,” *Inf Syst*, vol. 38, no. 8, pp. 1070–1083, Nov. 2013, doi: 10.1016/j.is.2013.05.001.

[184] D. Rothman, *Transformers for Natural Language Processing: Build innovative deep neural ... - Denis Rothman - Google Kitaplar*, vol. 1. Birmingham: Packt Publishing, 2021. Accessed: Aug. 26, 2022. [Online]. Available: https://books.google.com.tr/books?hl=tr&lr=&id=Cr0YEAAAQBAJ&oi=fnd&pg=PP1&dq=Transformers+for+Natural+Language+Processing:+Build+Innovative+Deep+Neural+...+-+Denis+Rothma&ots=a9q0Vn2i-1&sig=Ald46Hb9Kkaunhwu9Qae5AIRFoQ&redir_esc=y#v=onepage&q=Transformers%20for%20Natural%20Language%20Processing%3A%20Build%20Innovative%20Deep%20Neural%20...%20-%20Denis%20Rothma&f=false

[185] M. Hamdan, H. Chaudhary, A. Bali, and M. Cheriet, “Refocus attention span networks for handwriting line recognition,” *International Journal on Document Analysis and Recognition*, 2022, doi: 10.1007/s10032-022-00422-7.

[186] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, doi: 10.48550/arxiv.1706.03762.

[187] Z. Ruijie, X. Ying, J. Shuaichen, and L. Yonghe, “Patent text modeling strategy and its classification based on structural features,” *World Patent Information*, vol. 67, p. 102084, Dec. 2021, doi: 10.1016/j.wpi.2021.102084.

[188] G. Asche, “‘80% of technical information found only in patents’ – Is there proof of this?,” *World Patent Information*, vol. 48, pp. 16–28, Mar. 2017, doi: 10.1016/J.WPI.2016.11.004.

[189] W. Seo, “A patent-based approach to identifying potential technology opportunities realizable from a firm’s internal capabilities,” *Comput Ind Eng*, vol. 171, p. 108395, Sep. 2022, doi: 10.1016/j.cie.2022.108395.

[190] M. Sofean, “Deep learning based pipeline with multichannel inputs for patent classification,” *World Patent Information*, vol. 66, p. 102060, Sep. 2021, doi: 10.1016/j.wpi.2021.102060.

[191] V. Giordano, F. Chiarello, N. Melluso, G. Fantoni, and A. Bonaccorsi, “Text and Dynamic Network Analysis for Measuring Technological Convergence: A Case Study on Defense Patent Data,” *IEEE Trans Eng Manag*, Apr. 2021, doi: 10.1109/TEM.2021.3078231.

[192] A. Durmusoglu, “Updating technology forecasting models using statistical control charts,” *Kybernetes*, vol. 47, no. 4, pp. 672–688, Mar. 2018, doi: 10.1108/K-04-2017-0144.

[193] D. S. Hain, R. Jurowetzki, T. Buchmann, and P. Wolf, “A Text-Embedding-based Approach to Measure Patent-to-Patent Technological Similarity ↤-Workflow, Code, and Applications.” [Online]. Available: https://github.com/ANONYMEOUS_FOR_REVIEW

[194] B. Yan and J. Luo, “Measuring technological distance for patent mapping,” *J Assoc Inf Sci Technol*, vol. 68, no. 2, pp. 423–437, Feb. 2017, doi: 10.1002/asi.23664.

[195] S. Altuntas, T. Dereli, and A. Kusiak, “Analysis of patent documents with weighted association rules,” *Technol Forecast Soc Change*, vol. 92, pp. 249–262, Mar. 2015, doi: 10.1016/j.techfore.2014.09.012.

[196] S. Lim and Y. Kwon, “IPC multi-label classification applying the characteristics of patent documents,” in *Lecture Notes in Electrical Engineering*, vol. 421, 2017. doi: 10.1007/978-981-10-3023-9_27.

[197] J. Hu, S. Li, J. Hu, and G. Yang, “A Hierarchical Feature Extraction Model for Multi-Label Mechanical Patent Classification,” *Sustainability*, vol. 10, no. 1, p. 219, Jan. 2018, doi: 10.3390/su10010219.

[198] A. A. Golubev and N. V. Loukachevitch, “Use of Bert Neural Network Models for Sentiment Analysis in Russian,” *Automatic Documentation and Mathematical Linguistics*, vol. 55, no. 1, pp. 17–25, Jan. 2021, doi: 10.3103/s0005105521010027.

[199] M. H. Syed and S.-T. Chung, “MenuNER: Domain-Adapted BERT Based NER Approach for a Domain with Limited Dataset and Its Application to Food Menu Domain,” 2021, doi: 10.3390/app11136007.

[200] Ş. OZAN, U. ÖZDİL, D. E. TAŞAR, B. ARSLAN, and G. POLAT, “BERT Modeli’nin Sınıflandırma Doğruluğunun Sıfır-Atış Öğrenmesi ile Artırılması,” *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, vol. 14, no. 2, pp. 99–108, Dec. 2021, doi: 10.54525/tbbmd.1004781.

[201] O. Sevli and N. Kemaloğlu, “Olağandışı Olaylar Hakkındaki Tweet’lerin Gerçek ve Gerçek Dışı Olarak Google BERT Modeli ile Sınıflandırılması,” 2021. Accessed: May 03, 2023. [Online]. Available: www.dergipark.gov.tr/veri

[202] Y. Wu *et al.*, “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.”

[203] S. Jiang, C. Benge, and W. C. King, “BERTVision -- A Parameter-Efficient Approach for Question Answering,” Feb. 2022, Accessed: May 02, 2023. [Online]. Available: http://arxiv.org/abs/2202.12210

[204] O. KANTAR and Z. H. KİLİMCİ, “Derin öğrenme temelli hibrid altın endeksi (XAU/USD) yön tahmin modeli,” *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 38, no. 2, pp. 1117–1128, May 2022, doi: 10.17341/gazimmfd.888456.

[205] Z. Lan *et al.*, “ALBERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF LANGUAGE REPRESENTATIONS.” [Online]. Available: https://github.com/google-research/ALBERT.

[206] L. Martin *et al.*, “CamemBERT: a Tasty French Language Model.” [Online]. Available: https://universaldependencies.org

[207] Z. Gao, A. Feng, X. Song, and X. Wu, “Target-dependent sentiment classification with BERT,” *IEEE Access*, vol. 7, pp. 154290–154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[208] Y. Hao *et al.*, “Visualizing and Understanding the Effectiveness of BERT,” *ArXiv*, p. arXiv:1908.05620, 2019, doi: 10.48550/ARXIV.1908.05620.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: YÜCESOY KAHRAMAN, Selen

EDUCATION

Degree	Institution	Graduation Year
MS	Gaziantep University	2013
BS	Gazi University	2010

PUBLICATIONS

International Journal Articles:

YÜCESOY S., DERELİ T., DURMUŞOĞLU A., 2023, Forty Years of Automated Patent Classification, International Journal of Information Technology & Decision Making
https://doi.org/10.1142/S0219622023500165

YÜCESOY S., DURMUŞOĞLU A, DERELİ T., 2023, Ön eğitimli Bert modeli ile patent sınıflandırılması, Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, *Accepted.*

International Conferences:

YÜCESOY S., DERELİ T., DURMUŞOĞLU A. 2018, Patent Classification via Textual Analysis Which Sections to be Included? International Conference on Artificial Intelligence and Data Processing IDAP 2018.

YÜCESOY S., DERELİ T., DURMUŞOĞLU A., 2016, A Text Mining Analysis on Environmental Friendly Electronic Products Patents. th International Conference on Advanced Technology & Sciences (ICAT'Rome), 355-359. 23.11.2016, İtalya, Roma

YÜCESOY S., DERELİ T., DURMUŞOĞLU A., 2016, A Review and Bibliometric Analysis of Automatic Patent Classification. International Conference on Advanced Technology & (ICAT'16), 01.09.2016, Türkiye, Konya

DERELİ T., **YÜCESOY S.,** DURMUŞOĞLU A., 2014, Analysis of academic papers on environmental-friendly products. CIE'44 and IMSS'14: Joint International Symposium on "The Social Impacts of Developments in Information, Manufacturing and Service Systems", TÜRKİYE, İstanbul, 14.10.2014

YÜCESOY S., GÖÇKEN M., 2012**,** Comparing University Campuses by Using Fuzzy Analytic Hierarchy Process**,** ICEOS'xx12: 13th International Conference on Econometrics, Operations Research and Statistics, Kuzey Kıbrıs Türk Cumhuriyeti, Famagusta, 25.05.2012

National Conferences:

YÜCESOY S., DURMUŞOĞLU A., DERELİ T., 2023, Giyilebilir Deprem Teknolojileri Alanında Yayınlanmış Patentlerin Derin Öğrenme Yöntemiyle İncelenmesi Üniversitelerin Rolü, Yöneylem Araştırması ve Endüstri Mühendisliği (YA/EM) 42. Ulusal Kongresi, 01.11.2023, Türkiye, Gaziantep

DERELİ T., **YÜCESOY S.,** DURMUŞOĞLU A., 2013, Cep telefonlarının teknik özelliklerinin veri zarflama analiziyle tahminlenmesi. ÜAS 2013: Üretim Araştırmaları Sempozyumu, 541-551, Sakarya.

YÜCESOY S., GÖÇKEN M., 2012, Kansei Mühendisliği Yaklaşımının Üniversitelerin Kongre Kültür Merkezi Uygulanması, UEK 18: 18. Ulusal Ergonomi Kongresi, 16.11.2012, TÜRKİYE, Gaziantep.

YÜCESOY S., GÖÇKEN M., 2012, Üniversite Kütüphanelerinin Tasarımının Kansei Mühendisliği Yaklaşımı Kullanılarak Değerlendirmesi, UEK 18: 18. Ulusal Ergonomi Kongresi, 16.11.2012, TÜRKİYE, Gaziantep.

YÜCESOY S., GÖÇKEN M., 2012, Ürün Tasarımında Kansei Mühendisliği, YAEM 12: Yöneylem Araştırması ve Endüstri Mühendisliği 32. Ulusal Kongresi (YA/EM), 20.06.2012, Türkiye, İstanbul

YÜCESOY S., GÖÇKEN M., 2012, Siparişlerin Kabulü/Reddi Kararları ve Atölye Tipi Üretim Sistemlerinin Performansına Etkisi, Y AEM 2011: Yöneylem Araştırması ve Endüstri Mühendisliği 31. Ulusal Kongresi, TÜRKİYE, Sakarya, 05.07.2011

WORKSHOPS

Public Workshop on Big Data Analytics and Security, 11 February 2016 **/** Ankara **/**Turkey

PROJECTS:

Cep Telefonlarının Teknolojik Gelişiminin Tahminlemesi

Cep telefonlarının teknolojik gelişimlerinin tahminlenmesi, 01.01.2013 - 01.01.2014, Yükseköğretim Kurumları tarafından destekli bilimsel araştırma projesi, Araştırmacı

AWARDS

2016 **Best Paper** award University of Palermo

CERTIFICATES

TS EN ISO/IEC 17025:2017 Deney ve Kalibrasyon Laboratuvarlarının Yeterliliği için Genel Gereklilikler Dokümantasyon Eğitimi

TS EN ISO/IEC 17025:2017 Deney ve Kalibrasyon Laboratuvarlarının Yeterliliği için Genel Gereklilikler Dokümantasyon Eğitimi konulu eğitim programıdır. Gaziantep Üniversitesi, *Türkiye,* Ulusal, Sertifika

TS EN ISO/IEC 17025:2017 Deney ve Kalibrasyon Laboratuvarlarının Yeterliliği İçin Genel Gereklilikler Temel Eğitimi

TS EN ISO/IEC 17025:2017 Deney ve Kalibrasyon Laboratuvarlarının Yeterliliği İçin Genel Gereklilikler Temel Eğitimi konulu eğitim programı, Gaziantep Üniversitesi, *Türkiye,* Ulusal, Sertifika

TS EN ISO/IEC 17025:2017 Deney ve Kalibrasyon Laboratuvarlarının Yetkinliği için Genel Gereklilikler İç Tetkik Eğitimi

17025:2017 Deney ve Kalibrasyon Laboratuvarlarının Yetkinliği için Genel Gereklilikler İç Tetkik Eğitimi verilmiştir. Gaziantep Üniversitesi, *Türkiye* Ulusal, Sertifika

Pazarlama Stratejileri ve Müşteri İlişkileri Yönetimi
Pazarlama Stratejileri ve Müşteri İlişkileri Yönetimi Gaziantep, *Türkiye,* Ulusal, Sertifika 2012

FOREIGN LANGUAGE

ENGLISH (ADVANCED) - YÖKDİL: 85

GERMAN(BEGINNER)