Department of Electrical and Information Engineering (DEI)

Interuniversity Ph.D. Program in Industry 4.0

SSD: FIS-07

Final Dissertation

EXPLORING TECHNOLOGICAL INNOVATION USING INTELLECTUAL PROPERTY ANALYTICS: A FOCUS ON HEALTHCARE 4.0

---

by

DEMARINIS LOIOTILE ANNAMARIA:

Supervisors:

Prof. Roberto Bellotti

Prof. Nicola Amoroso

Coordinator of Ph.D. Program:

Prof. CATERINA CIMINELLI

Course n°36, 01/11/2020-31/10/2023

Politecnico di Bari

LIBERATORIA PER L'ARCHIVIAZIONE DELLA TESI DI DOTTORATO

Al Magnifico Rettore
del Politecnico di Bari

La sottoscritta DEMARINIS LOIOTILE ANNAMARIA, nata il 15/12/1981 a BARI (BA), residente in CORSO MAZZINI N.51 a RUTIGLIANO (BA), C.a.p. 70018, e-mail personale ANNAMARIADEMARINIS@GMAIL.COM, iscritto al 3° anno di Corso di Dottorato di Ricerca in INDUSTRIA 4,0 ciclo XXXVI

ed essendo stato ammesso a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

EXPLORING TECHNOLOGICAL INNOVATION USING INTELLECTUAL PROPERTY ANALYTICS: A FOCUS ON HEALTHCARE 4.0

DICHIARA

1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
2) di essere iscritto al Corso di Dottorato di ricerca INDUSTRIA 4,0 ciclo XXXVI, corso attivato ai sensi del "*Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari*", emanato con D.R. n.286 del 01.07.2013;
3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archivierà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito http://www.creativecommons.it/Licenze), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviate/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dal sottoscritto tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

BARI, 22/12/2023 Firma Annamaria Demarinis Loiotile

Il/La sottoscritto, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

BARI, 22/12/2023 Firma Annamaria Demarinis Loiotile

Department of Electrical and Information Engineering (DEI)
Interuniversity Ph.D. Program in Industry 4.0
SSD: FIS-07

Final Dissertation

EXPLORING TECHNOLOGICAL INNOVATION USING INTELLECTUAL PROPERTY ANALYTICS: A FOCUS ON HEALTHCARE 4.0

---

by

DEMARINIS LOIOTILE ANNAMARIA

Referees:

Prof. Davide Iannuzzi

Prof. Armida Sodo

Supervisors:

Prof. Roberto Bellotti

Prof. Nicola Amoroso

Coordinator of Ph.D Program:

Prof. CATERINA CIMINELLI

Course n°36, 01/11/2020-31/10/2023

Artificial intelligence is revolutionizing the way healthcare is administered. It has the capacity to enhance healthcare outcomes, elevate the patient experience, and improve access to healthcare services. AI can boost the productivity and efficiency of healthcare delivery, enabling healthcare systems to offer higher quality care to a larger population. AI can expedite care delivery, especially by reducing the time it takes for diagnosis, and assist healthcare systems in taking a more proactive approach to managing population health, ensuring that resources are allocated where they can make the most significant impact.

Thanks to knowledge transfer and collaboration among the key players in the innovation process, research results can reach individual citizens, society, and communities to enhance people's lives.

INDICE

Executive summary and reading guide

A great mine of innovation is represented by the excellence of the scientific know-how of the universities and research centers. But very often the research results remain unvalued and unexploited, in the so-called “Valley of death”, which represents “the gap between where publicly available research funding stops and where private investment or commercial funding starts” (Hockaday, 2020). The Valley of Death is a metaphor for the absence of resources and skills that hinders new ideas as they move from the laboratory of research centers to the marketplace.

Recently, many universities and research centers are trying to enhance the value of their research results to a greater extent, increase the valorization of innovative results and the market uptake of new solutions, through different mechanisms, conventional and unconventional, of knowledge transfer (KT), in the framework of the broader concept of universities third mission (TM).

This research work responds to the need to understand which are the characteristics of the universities that are most successful in the technology transfer action and which are the key factors of success, subsequently aiming at the proposition and experimentation of new models and good practices useful for overcoming the "Valley of Death".

A first goal was therefore to identify the best universities and research centers worldwide that stand out for their performance in the third mission. The difficulties in this regard immediately became evident. In fact, in the complex and multifaceted scenario of the third mission, while performance indicators relative to research (in terms of quality of publications, number of citations, etc.) and teaching (in terms of student-to-staff ratio, student evaluation, etc.) are widely known and used, less is known about how KT or, more generally, TM can be characterized and evaluated.

Widely used tools for evaluating and comparing universities performance are the so called “global rankings”. Unfortunately, most of the best-known global rankings lack instruments to evaluate KT activities and fail to properly capture and evaluate the peculiarities of KT and TM. In fact, defining the activities and quantifying the TM requires the design of a complex model of analysis that is able to determine the map of indicators related to the diversified dimensions of the third mission.

Thus, in this scenario, the first research question (RQ1) faced was: according to global university rankings, how do the world's top universities perform from a knowledge transfer point of view? In other words, are the rankings currently most used able to characterize universities performance from a third mission point of view? In an attempt to answer this research question, the best-known global university rankings (The Academic Ranking of World Universities (ARWU), The QS World University Rankings® (QSWUR), The Times Higher Education World University Rankings (THEWUR)) were analyzed in order to identify the world’s top universities. A first consideration that emerges is that they lack of specific KT indicators.

On the other hand, the European Commission, starting from the same consideration, has invested efforts in defining a specific set of indicators capable of capturing and evaluating

the complexity of the universities third mission: U-Multirank (UMR) ranking, proposed by the European Commission is based on a different approach compared with the existing global university rankings. It includes a set of 9 indicators focused on KT. This initiative has already produced a very rich data set over the years.
Thus, an initial analysis carried out during the research period was aimed at verifying whether the top universities identified through the global rankings are also the best from the third mission point of view. Each ranking has its own specificities and thus we decided to evaluate a set consisting of the top 100 universities included in each of the selected rankings: ARWU, QSWUR and THEWUR. Furthermore, the coherence between rankings and their level of agreement, was evaluated by using Spearman's correlation among the top 100 positions in each ranking. The results obtained show that all rankings are strongly correlated and thus they exhibit an underlying coherence.
The second step was to search and select a set of specialized KT indicators, starting from those proposed by UMR2020, for evaluating the world top universities in the global rankings from the KT point of view. Among the 9 indicators proposed by UMR2020 indicators only 5 were selected for being used, due to the fact that they had the minimum number of null value (Co-publications with industrial partners, Patents awarded, Patents awarded, Industry co-patents, Publications cited in patents). After a data normalization and cleaning process of the available data set, we introduced a composite indicator called Global Performance Indicator KT (GPI KT), obtained as a combination of the 5 indicators selected, was also defined and used for evaluating the performance in knowledge transfer for each university included in the top 100. Then a comparison was made, and the results obtained were not as expected: the best universities in global rankings appear to drop many places when evaluated from the point of view of the third mission. Thus, global rankings are unable to properly evaluate the performance in the third mission.
A further step was to investigate the top universities in order to identify groups of similar universities, in terms of KT indicators, through a data-driven approach based on clustering algorithms. The goal was to understand from the natural aggregation in groups, the presence of common characteristics capable of explaining the different levels of performance in knowledge transfer.
The results obtained show an interesting composition of the clusters. In some cases, it appears to take into account the geographical dimension, i.e. the industrial, social and economic environment itself can affect the third mission activities, thus suggesting that there are contextual factors that the purely quantitative analysis used by global university rankings fail to grasp or bring out.
Another key point in this first research line was to understand what are the KT indicators that best discriminate the performance of the world's universities. The analysis carried out identified 3 indicators that seems to be able to discriminate more strongly the performance of the world's universities from KT point of view: co-publications with industrial partners, patents granted, and publications cited in patents.

Of these three indicators, those referring to publications (co-publications with industrial partners and publications cited in patents) are difficult to retrieve and, moreover, there is an ongoing debate in the community about their real usefulness. The number of granted patents, on the other hand, is more available data and there are numerous patent databases that can be freely consulted. Since between 70% and 90% of the information about technologies is not published anywhere except in patent documents, patents are among the best sources of information.

For these motivations, since the number of patents appears to be an important indicator, with data currently readily available, the research activity focused mainly on patent analysis and patent matchmaking platforms.

To this end, the main methodologies of Intellectual Property Analytics were adopted, as a multidisciplinary approach used to gain valuable insight about intellectual property data.

In order to address the lack of valorization of research results and help universities and research centers promote their research results, specific initiatives such as online patent platforms have emerged over time as convenient channels for patent transfer, with the joint effort of both academic, industrial and political partners. Existing initiatives/platform used for matchmaking between supply and demand for innovation are sometimes ineffective and not easily available, mainly for the following reasons:

- they are paid services, not open access - often open innovation platforms;
- they report the patent document as such, without a usable "translation" for all that facilitates matching;
- the classification of the patents in a given technological area or sector is a challenging task: the content producers choose a category based on those already proposed by platforms, but they often do not know how to choose best, and the available categories are often ineffective for patent classification.

Often the classification of patents and, therefore, the search and consultation method used, are based on taxonomies and keywords self-defined by users, experts or database managers and are not very effective. This often leads to multidisciplinary categories containing a significant number of poorly characterized and classified patents, which can be called “monster classes”. Monster categories are thus ineffective, not discriminating, and difficult to explore. A typical phenomenon of “monster class” often occurs in the case of healthcare related innovations, which, due to multidisciplinary and particularly innovative nature, tends to generate classes with a large number of patents. Also for this reason during the research period, the healthcare sector was investigated with particular attention, given its strategic importance and the poorly accurate exploration techniques.

In consideration of the above issues, further research questions that this thesis aimed to answer were:

- are the classification taxonomies used in the patent platforms effective in classifying the whole landscape of academic patents (RQ2)?

- is it possible to support the user in correctly classifying a patent entered into the platforms in order to improve the matchmaking between demand and supply of innovation (RQ3)?
- is it possible to draw up an attempted vocabulary of technological fields from the keywords that emerged from an applied AI-based approach (RQ4)?

In order to answer to these 3 research questions, the following approaches were defined and experimented:

1. Natural Language Processing and clustering techniques were used to improve the taxonomy-based classification - answer to RQ2;
2. Regression was used to build a multi-label classification system to support users - answer to RQ3;
3. An AI-based approach on the most frequent words was defined in order to improve the keywords-based classifications - answer to RQ4.

The proposed approaches were experimented on the Italy's largest and most relevant patent platform, **Knowledge Share (KS)**, a public web platform with completely free access.
The KS database includes patents registered from 12/28/1999 to 8/16/2021, consisting of 1694 patents, uploaded to the platform by 89 Italian Research Centers, both public and private (Universities, Research Centers, Scientific Institute for Research, Hospitalization and Healthcare, etc), covering 90% of institutions nationwide. This platform can be easily queried by users aiming at obtaining an overview on the state-of-the-art about particular technologies in Italy. On one hand, this service lowers firms and investors' entry barriers for innovations in fundamental and applied science, letting them overcome R&D&I challenges more easily. On the other hand, this platform helps scientists and startups/spin-offs in achieving visibility, expressing their innovative potential and gaining interests from private and public investors.
In KS the patents, accurately "translated" in a simple and self-explanatory language, are categorized in ten technological domains:

- Aerospace and aviation;
- Agrifood;
- Architecture and design;
- Chemistry, Physics, New materials and Workflows (Basic Science);
- Energy and Renewables (Green Energy);
- Environment and Constructions (Environment);
- Health and Biomedical (Biomed);
- Informatics, Electronics and Communication System (Electronics);
- Manufacturing and Packaging (Packaging);
- Transports.

In this research activity, Natural Language Processing (NLP) techniques have been applied on the corpus of 1694 patents. In order to answer RQ2, once the text has been cleaned up and processed, the TF_IDF matrix was constructed. Since the matrix appears sparse, in order to reduce the sparsity and make the clustering process less prone to the curse of dimensionality,

we applied the Singular Value Decomposition (SVD) for dimension's reduction. Finally, we used the K-means clustering algorithm and we performed a grid-search exploration of the parameters' space and used the Silhouette for optimization.
The clustering analysis reveals the presence of 8 homogeneous clusters instead of the 10 proposed by the KS platform:

1. Technologies 4.0 (mechanics and robotics)
2. Material science
3. Cancer treatment
4. Optics - Image processing
5. Sensor technology - ICT
6. New molecules - new compounds - pharmacology
7. Energy/green Technologies
8. Biomedical

This suggest the presence of possible inhomogeneities within the traditional KS classifications, probably due to the emergence of novel technologies or cross-domain areas, e.g., Healthcare 4.0 This implies that the taxonomy used by KS could be improved significantly in order to better collect/expose the patents content.
In order to answer RQ3 and RQ4, we proposed an artificial intelligence-based system that recognizes technological areas, thus classifying patents as correctly as possible, and recommends the right direction to the user, eliminating the "subjectivity" of choice. We use a combined approach of NLP and machine learning (ML) in order to support the patent platforms by addressing two main aspects: it is of paramount importance to create an automated recommender system that can identify the most suitable and correct technological area(s) a patent must be assigned, this is both useful for users looking for specific technologies or patent owners who want to reach the largest fraction of potentially interested users; on the other hand, the methodology allows for the identification of keywords characterizing a patent in an objective and human-independent way. This aspect is particularly useful to create an initial vocabulary of words extracted from patents, thus eventually leading to redefine the available categories and supporting the portal management and, again, the matchmaking among users and patent-owners.
Through the combined framework NLP-ML, an explainable patent classification system was proposed for multi-label classification of patents, improve the user friendliness of a platform and enhance the selection of suitable patents by a company or, in general, booster the matchmaking of innovation leading to social and economic impact.
It is interesting to note that patents related to the health sector, originally located in only one technological area in KS, "Health and Biomedical" category, that account for a large portion (about 30 percent of the total), after the processing proposed were rearranged into multiple and more specialized categories. In fact, in addition to the three clusters directly related to health, namely No. 3, No. 6 and No. 8, Cluster No. 1 - Technologies 4.0 - contains several patents connected to new technologies applied to healthcare.

For this reason, they were further investigated because of their proximity to **Healthcare 4.0** applications. Healthcare 4.0 (HC4.0) is a recently emerged term derived from Industry 4.0, used to describe the progressive emergence of typical Industry 4.0 technologies, such as Internet of Things (IoT), Industrial IoT (IIoT), cognitive computing, artificial intelligence, cloud computing, fog computing, and edge computing, applied to healthcare domain. In the context of this new revolution, Cyber-Physical Systems (CPS) are shaping digital health systems involving products, technologies, services, and businesses. HC4.0 must enable stepwise virtualization to support the near real-time personalization of healthcare for patients, workers, and formal and informal janitors.
The methods of patent analysis illustrated have been applied to this especially impactful field, HC4.0, that is therefore a critical sector, in continuous turmoil and, above all, sees the application of transversal and multidisciplinary technologies and innovations within it. In addition to more performing taxonomies for classifying patents, it would therefore be useful to have effective keywords, to define a vocabulary useful for better characterizing patents in this broad sector. Thus, the last research question (RQ5) that was faced during this research work was - is it possible to define a first draft of HC4.0 vocabulary for characterizing the innovations and innovative technologies in healthcare 4.0?
In order to answer this question, the analysis and study of the most frequent words contained in the 4 healthcare related clusters (Technologies 4.0, Cancer treatment, New molecules - new compounds and Biomedical), was carried out and a first attempt of Healthcare 4.0 vocabulary was drafted.
The document is organized as follows:

- The "Introduction" chapter introduces the reader to the topics of knowledge transfer and the third university mission, contextualizing them in the international literature of the sector and clarifying which are the objectives it pursues and which are the difficulties encountered.
- The chapter "Evaluate the third mission and technology transfer", contextualizing the discussion in the bibliographic scenario of the sector, addresses the issue of the use of Global Rankings and the difficulty of evaluating universities and research centers from the point of view of the third mission. Introduces and motivates RQ1.
- The chapter "From university rankings to Intellectual Property Analytics", starting from highlighting why patents are in fact one of the few truly exploitable indicators to evaluate the third mission, introduces the IPA and the methods used for the experimental investigations during the research. It defines and motivates RQ2, RQ3 and RQ4, in the light of the problems of the sector, also providing a rationale for how the various IPA methods have been used to answer the research questions.
- The chapter "The experimental field" describes the experimental context in which the various proposed approaches were tested, i.e. the Knowledge Share platform with its dataset and the focus on Healthcare 4.0.

The remaining chapters, for each of the research questions defined, after having enucleated the underlying issues in detail, present the proposed approaches/methodologies in order to

provide answers and the main results obtained. Each chapter closes with a discussion of the results.
Finally, the last chapter presents an overall discussion of the result obtained and attempts to draw the conclusions of the entire research process by providing an overview of the results obtained.

CHAPTER 1. Introduction

1.1. The new role of Universities in the society

The university's contribution to solving the grand societal challenges foregrounds the issues of environmental sustainability, social justice, and inclusion, which the United Nations 2030 Agenda places in a systemic framework, pushing universities in the direction of commitment to economic, social, and environmental sustainability (Goddard et al., 2016), going so far as to state that "none of the SDGs can be achieved without the contribution of universities through research, teaching and Third Mission (TM)" (UN, 2019). Universities in fulfilling their core mission, which is to produce knowledge and skills, serve as catalysts for political, economic, social, technological, and environmental change; their leadership becomes crucial particularly for complex issues with broad goals that cannot be confined to specific sectors but cross national boundaries, such as environmental issues, climate change, migration, illiteracy, extreme poverty, and human rights. Through research and teaching, universities play a key role in producing new knowledge, innovation, and developing generations of new leaders and skilled professionals who can drive this innovation towards social development. Through engagement in communities, universities work with a wide variety of stakeholders including governments, the private sector, and civil society to contribute toward local, national, and global impact (UN, 2019) (Blasi, 2023).

With the emergence of the knowledge economy (Dasgupta and David, 1987), global competitiveness and economic development are played out primarily on the side of innovation. Universities assume a key role in the cycles of production and circulation of knowledge and innovation, and are called upon to participate directly in the economic development of the country and to make the industrial and service system more competitive. University is more of a protagonist on the economic scene and closer to enterprises, and the processes of knowledge production, ways of transferring scientific research results and sharing the knowledge produced are being rethought through new organizational and management models (Blasi, 2023).

Universities (and more in general research institutions) carry out a more fluid and dynamic model, based on the development of multi- and trans-disciplinary research, where the boundaries between the research world and the industrial world are less defined; the involvement of different actors grows as the dimension of innovation and economic exploitation of research results; and the distinction between basic research, applied research and development thins (Limoges et al., 1994). The boundaries between science and technology and between public and private are becoming blurred, and there are no longer clear dividing lines between academic and industrial research. Thus, universities and research institutions play a crucial role in innovation processes and move within an ecosystem of diverse actors and organizations that work in synergy, cross institutional and disciplinary boundaries and pool heterogeneous objects and expertise (Gherardini, 2015; Marra, 2022).

Within a framework of increasingly permeable institutional boundaries, society interacts with universities, expressing a broader range of demands for innovation, providing a more diverse spectrum of scientific expertise, and thickening networks of inter-institutional collaboration. The role of academics changes: research-takers are called upon to build bridges between science and technology, assume entrepreneurial postures, and commercialize the technologies that emerge from scientific research (Clark, 1998; Shane, 2004; Etzkowitz, 2003). The idea of the entrepreneurial university is gaining ground, characterized by a considerable degree of independence from the state and industry, but at the same time linked to these two institutional spheres by intense interaction. Universities are seen as part of an organic system, and the state is joined by other 'stakeholders', requiring the development of specific professional figures or engagement in applied research projects, in a joint effort of innovation and economic progress (Blasi, 2023).

The interactions, the relationships and the impact of a university in a given territory within this system are classically measured by the Triple Helix Model (THM), that represents the interactions between institutional actors, university and business, a "helix" that has been established since the second half of the 1990s (Etzkowitz and Leydesdorff, 1997) and embraced by the European Union in the Lisbon Strategy.

Over time, the three helices are enriched with new ones: first, the civil society sub-system (fourth helix), represented by the production of culture and the media system, is added to the triple helix (Carayannis et al., 2012). Subsequently, due to the centrality that the evolution of the natural environment in society assumes (mainly related to climate transformations and their consequences), the socio-ecological transition aspects of society and economy (fifth helix) are integrated, with universities at the forefront as producers of knowledge and in-innovation (Carayannis et al., 2012; Blasi, 2023). See figure 1.

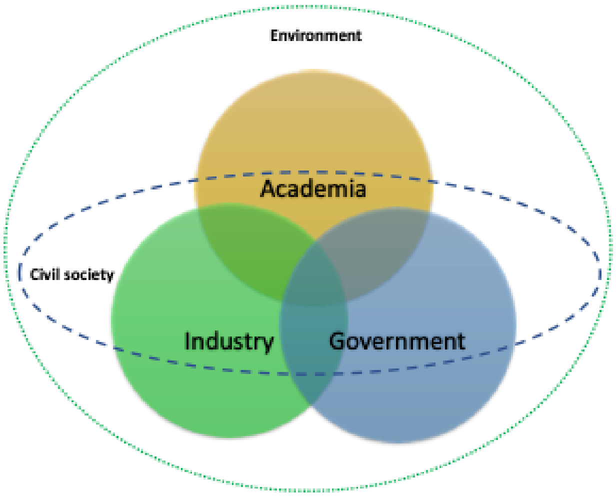


Figure 1: quintuple helix model

In recent years, universities have been required to shift from predominantly teaching and research activities also to TM activities, understood as "contributions to society" (Abreu et al., 2016; Urdari et al., 2017; Backs et al., 2019; Zawdie, 2010; Compagnucci and Spigarelli, 2020). Through the third mission, universities and research centres act as motors that make a contribution to the social, economic and cultural development of the regions in which they are active, bringing knowledge and technology to industry and society (De Jong et al., 2014; Secundo et al., 2017; Agasisti et al., 2019; Compagnucci and Spigarelli, 2020; Research & innovation valorisation channels and tools, 2020; World Economic Forum, 2019; Trippl et al., 2015; Cesaroni and Piccaluga, 2016; Di Berardino and Corsi, 2018). Figure 2 summarizes the three fondamental missions of Universities:

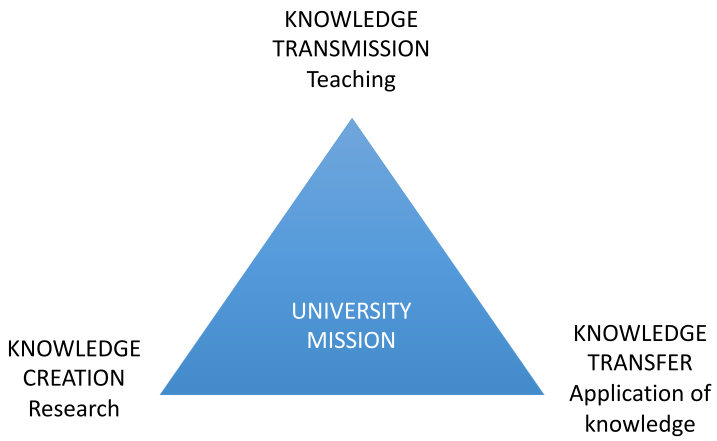


Figure 2: the three fundamental missions of Universities (Scanlan, 2018)

1.2. From Knowledge Transfer to Third Mission

Among the three university missions, the research work is focused on Knowledge transfer (KT) - or often also referred to as technology transfer (TT) (the distinction is better explained in the following paragraphs) - that is a well-recognised activity in which research organizations are expected to engage and it is seen as an essential source of innovation and a mechanism for the dissemination of research results (Campbell et al., 2020). The public and private research universities are recognized as the seed capital for creating know-how and technologies that foster economic and social development (DeVol et al., 2017). KT plays a vital role in translating academic research into practical applications that drive economic growth, innovation, and societal development. KT aims to maximize the two-way flow of technology, intellectual property, and ideas. KT thus empowers companies, the government sector, and other nonacademic organizations to promote innovation with consequent economic and social benefits (Campbell et al., 2020). This is an essential function for universities, enabling them to bridge the gap between academia and industry. By transferring knowledge and innovations to the commercial sector, universities can drive economic growth, foster industry collaboration, disseminate knowledge, generate funding, and make a positive impact on society (Godonoga and Sporn, 2023). KT serves as a catalyst for universities to realize the full potential of their research and innovation. It strengthens their economic contributions, enhances their reputation, facilitates entrepreneurship, protects intellectual property, and enables collaboration with industry, government, and international partners.

In the past two decades there has been an evolution that has seen the concept of knowledge transfer change from the more traditional concept of simple commercialization and monetization to a more comprehensive approach that supports both co-creation and dissemination of research results with and to nonacademic third parties (Campbell et al., 2020).

For that reason, over time different terms have been used to talk about knowledge transfer and third mission:

- Technology Transfer
- Third Mission
- Knowledge Transfer
- Knowledge Exchange
- Engagement.

The Figure 3 illustrates the relationship and evolution of these concepts, from which it can be seen that the term Third Mission is the one that caps it all and encompasses everything (Hockaday, 2020):

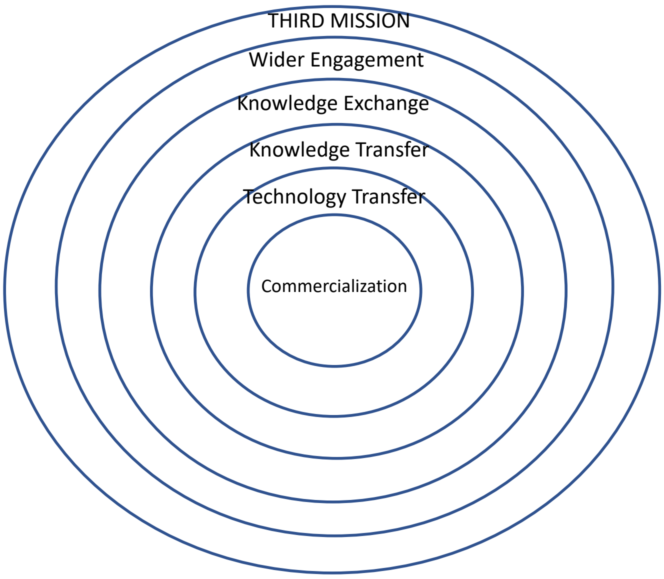


Figure 3: the evolution and extension of university technology transfer

The term TM is rather "nebulous" (Gregersen et al., 2009) and ambiguous (Laredo, 2007; Pinheiro et al., 2015), a complex and rapidly evolving phenomenon (Giuri et al., 2019). On the one hand, the concept is related to the themes of the "entrepreneurial university" (Trencher et al., 2014; Markuerkiaga et al., 2016); on the other hand, it refers to a wide range of activities carried out by universities in transferring knowledge to society at large and to external organizations (Marzocchi et al., 2023), as well as promoting entrepreneurial skills, innovation, social welfare and human capital formation. In addition, another key piece concerns the various forms of “science communication and social engagement" (Compagnucci and Spigarelli, 2020; Rothaermel et al., 2007; Di Berardino and Corsi, 2018). In other words, of course, the third mission is a complex and evolving phenomenon that, in the last years, has been articulated in policies resulting from the dialogue between universities, industry, government and society (Vorley and Nelles, 2009; Predazzi, 2012; Giuri et al., 2019).

Therefore, through the third mission, universities move away from the traditional 'ivory tower' position in which teaching and research have always been treated as aims in themselves (Nakwa and Zawdie, 2016; Mahrl and Pausits, 2011; Etzkowitz et al., 2000). Universities, abandoning their ivory towers, move closer to society to meet social needs and industrial goals (Kapetaniou and Lee, 2017; Florida and Cohen, 1999; Etzkowitz, 1998; Molas-Gallart et al., 2002).

The Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) defines the third mission in this way: “Third mission is intended as the degree of openness of the HE institutions towards the socio-economic context through the valorization and transfer of knowledge. TM is a process of knowledge exchange, not only related to technology and encompassing social and cultural benefits”.

The connection with the territory, in fact, can mainly take on two faces:

- one more directly linked to economic reasons and thus to relations with industry and the commercialization of intellectual property, as well as the promotion of entrepreneurship;
- another more closely linked to purposes of a social nature and the dissemination in the socioeconomic context of knowledge and qualified skills (ANVUR).

The areas of the third mission according to ANVUR are (Figure 4):

A: Valorization of research

(a) Intellectual property management

(b) Academic entrepreneurship (spin-off companies).

(c) Third-party activities

(d) Collaboration with territorial intermediaries.

B. Production of public goods of a social, educational and cultural nature

(a) Production and management of cultural goods

(b) Clinical trials, research infrastructure and medical training

(c) Continuing education

(d) Public engagement

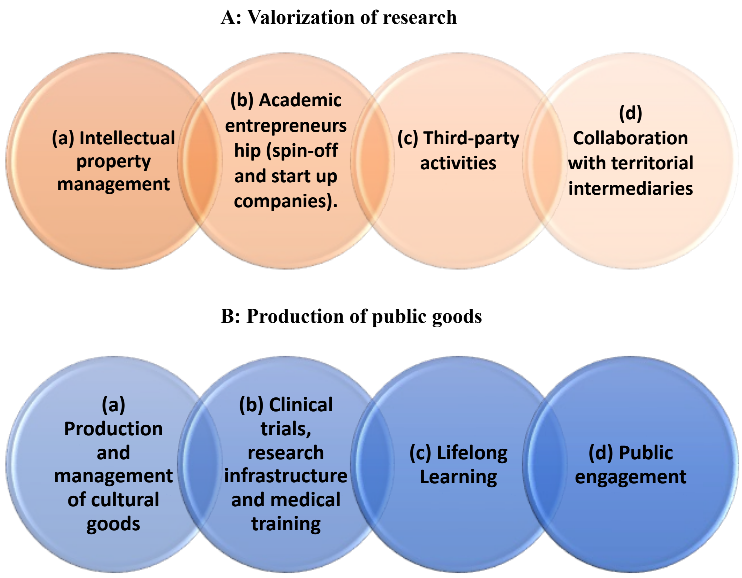


Figure 4: the areas of the third mission according to ANVUR

For the purposes of the research questions outlined for this thesis, the focus is on the activities defined in Group A: valorization of research.

The valorization of research promotes the dissemination and use of new technologies with the aim to increase the impact, economic and/or social, of the research for all the stakeholders and partners involved (Scanlan, 2018; Alexander et al., 2020).

In this framework, the research results could be valorized through a variety of complex channels (Azagra-Caro et al., 2017) or mechanisms, intentional and unintentional (Scarrà and Piccaluga, 2022), including research conversion to intellectual property (IP) and its patenting and licensing activity, creation of academic start-ups or externally formed entrepreneurial entities, collaborative research with private sector firms or contract research consulting, etc. Figure 5 describes the flow of knowledge and technologies towards the creation of impact for a multitude of actors (Campbell et al., 2020):

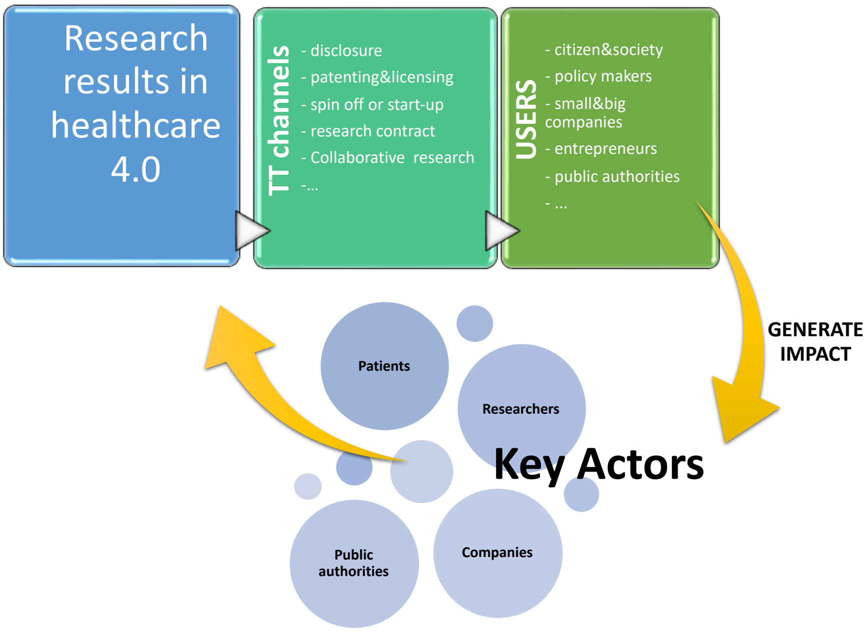


Figure 5: flow of knowledge and technologies towards the creation of impact for a multitude of actors (Campbell et al., 2020)

Thus, promoting the effective use and development of Intellectual Property and ensuring easier access to and sharing of IP-protected assets in times of crisis are key priorities of the Union's IP Action Plan (Intellectual property action plan implementation, 2022). Universities are a mine of innovation and play an important role in its development, in fact they are recognized as a critical element in the global competitiveness of enterprises (Bellantuono et al., 2022; Demarinis Loiotile et al., 2022; Hu and Zhang, 2021). Patent transfer between universities and companies has been the subject of much attention recently. Previous studies demonstrated that the patent transfer has significant impacts not only on both academia and industry (Deng and Ma, 2022), but also on national economies (Chen and Deng, 2023; Lee et al., 2017; Wang et al., 2014). To name a few studies, McDevitt et al. (2014) showed that patent transfer provides public benefits and influences economic development in addition to generating revenue, increasing funding opportunities, and promoting a culture of entrepreneurship and innovation for universities; Roessner et al. (2013) combined U.S. university licensing data from 1996 to 2010 with input-output economic models and found that the contribution of university licensing to the U.S. economy during the period was at least $162.1 billion (Deng and Ma, 2022; Chen and Deng, 2023).

The later stages of technology transfer and commercialization into marketable products and services, however, tend not to be easy (Shove, 1998).

1.3. The lack of valorization and the Death Valley

Many reports, such as the ASTP Survey Report on KT Activities FY2019, show a significant lack of valorization of the patented technologies from the Universities and Public Research Centers across Europe, since only 18% of those inventions are licensed or optioned (ASTP Report, 2019). Too many research results remain stuck in the so-called Valley of Death (Hudson and Khazragui, 2013), which represents "the gap between where publicly available research funding stops and where private investment or commercial funding starts" (Hockaday, 2020).

The Valley of Death is a metaphor for the absence of resources and skills that hinders new ideas as they move from the laboratory to the marketplace (Hensen et al., 2015) (Figure 6).

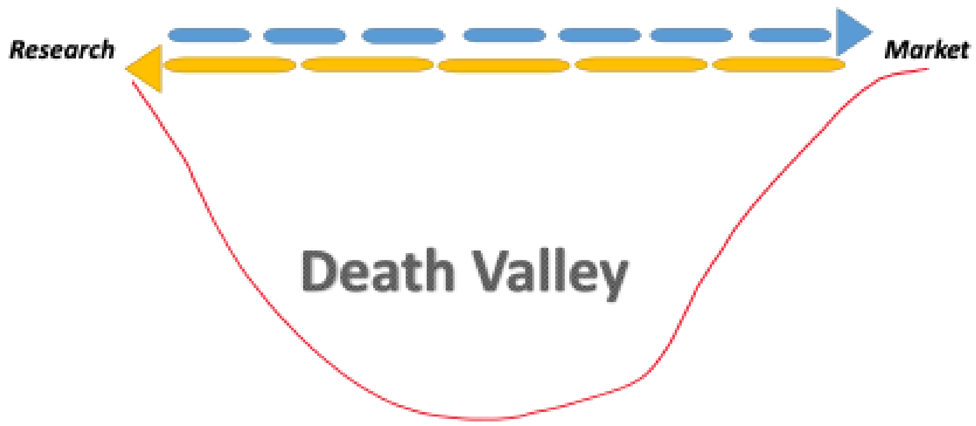


Figure 6: the Valley of Death

Typically, academic research groups succeed in developing concepts from discovery to a point with a low level of technological readiness (TRL).

Technology Readiness Levels (TRL) are used to assess the maturity of a new technology towards full economic operation. TRL starts at stage one, where the technology is in the embryonic stage, and progresses to the most mature stage, level nine, where the technology has been tested and launched. This tool provides a quick view of technology maturity and helps management make decisions about the development and transition of a technology. Investors and national funding agencies use this tool to identify the type of project that best fits their objectives.

An explanation by European Commission of the different TRL values is:

- TRL 1 – basic principles observed
- TRL 2 – technology concept formulated
- TRL 3 – experimental proof of concept
- TRL 4 – technology validated in lab
- TRL 5 – technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)

- TRL 6 – technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- TRL 7 – system prototype demonstration in operational environment
- TRL 8 – system complete and qualified
- TRL 9 – actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space).

In general, new technologies go through the various stages of the TRL scale during their life cycle. During the research and development phases, it is possible to have iterations between the different TRL levels.

In the Healthcare sector, TRL assessment is even more complex because, in addition to technology deliverables for each milestone, deliverables for clinical, market/commercial, and regulatory aspects are defined to help manage risk.

The gap lies between TRL 4 and 7. The reasons for this are diverse:

- Basic research is carried out mainly with public funding, after which investment is needed, sometimes generally high, with a relatively low success rate. Only a few technological concepts will be transformed into successful commercial products.
- The process requires an interdisciplinary approach. The right combination of skills is not always available.

The Death Valley is sometimes used as an analogy to describe this discontinuity in innovation processes (Hensen et al., 2015). The role of technology and knowledge transfer is exactly to be a catalyst to bridge this gap and help overcome the Valley of Death.

Finally, KT an, more in general, third mission could help Universities and Research Centers in the valorization process for the following reasons:

- KT plays a crucial role in driving economic growth and development. Universities are often at the forefront of cutting-edge research and innovation, and technology transfer allows them to capitalize on their intellectual assets. By commercializing their research outcomes, universities can generate revenue through licensing agreements, spin-off companies, or partnerships, thereby fostering economic development in their local communities and beyond.
- KT facilitates collaboration between academia and industry. By transferring technologies to the commercial sector, universities can establish partnerships with businesses, which can lead to joint research projects, funding opportunities, and access to industry expertise (Public-Private Partnerships). Such collaborations not only enhance the quality and relevance of research but also enable universities to address real-world challenges and create practical solutions. Public-private partnerships allow universities to tap into industry insights, gain access to funding, and create a pathway for the practical application of research outcomes.
- Universities are knowledge hubs, and KT helps disseminate their research findings and innovations to a wider audience. By commercializing technologies, universities

make them accessible to industries, entrepreneurs, and the general public. This promotes the utilization of scientific advancements, fosters innovation in various sectors, and drives social progress.

- KT provides universities with additional avenues for funding and resource generation. The revenue generated through licensing fees, royalties, and equity stakes in spin-off companies can be reinvested into further research and academic programs. This financial support enhances the capacity of universities to attract and retain talented faculty members, improve infrastructure, and expand their research capabilities.
- KT enables universities to have a tangible impact on society addressing pressing social and environmental challenges. By translating research into practical applications, universities contribute to the development of new products, processes, and services that address societal needs and challenges. This can encompass a wide range of areas, such as healthcare, energy, agriculture, environmental sustainability, and information technology, leading to improvements in people's lives and the overall well-being of communities. By transferring technologies developed in these domains, universities can contribute to solving global issues, improving quality of life, and promoting sustainable development.
- TM initiatives can enhance the reputation of universities as hubs of innovation and research excellence. By demonstrating their ability to translate theoretical knowledge into practical applications, universities gain recognition as contributors to economic and social advancement. This can attract top-tier faculty, students, and research collaborations, further bolstering the institution's prestige.
- KT often leads to the creation of new ventures and spin off/startups. Universities can support aspiring entrepreneurs by providing access to their intellectual property, research facilities, mentoring, and business development resources. By fostering an entrepreneurial culture, universities stimulate job creation, promote self-employment, and contribute to the growth of local and national economies.
- KT activities often require collaboration between multiple disciplines and departments within a university. This fosters interdisciplinary research and collaboration, breaking down silos and encouraging cross-pollination of ideas. Collaborative research networks formed through technology transfer initiatives can lead to groundbreaking discoveries and solutions to complex problems that require diverse expertise.
- KT encourages open innovation practices, where universities actively collaborate with external partners to exchange knowledge and resources. By engaging with industry, government agencies, and other stakeholders, universities can leverage external expertise and resources, accelerating the development and commercialization of technologies. Open innovation enables a broader range of perspectives and inputs, leading to more robust and impactful outcomes.

- KT is crucial for maintaining and enhancing the global competitiveness of universities. In an increasingly interconnected and knowledge-driven world, universities need to translate their research into practical applications that can be globally relevant. By transferring technologies, universities can contribute to economic competitiveness, drive innovation, and position themselves as key players in the global innovation landscape.

In conclusion, knowledge/technology transfer is a multifaceted and dynamic process that holds immense importance for universities. It supports alumni engagement, fosters public-private partnerships, facilitates knowledge commercialization, strengthens institutional capacity, and enables global collaboration. By actively embracing KT, universities can enhance their impact, relevance, and sustainability while contributing to economic growth, societal progress, and the advancement of knowledge.

CHAPTER 2. Evaluate the third mission and knowledge/technology transfer

2.1. The role of university rankings and the limits in the evaluation of KT

A first goal was to identify the best universities and research centers worldwide that stand out for their performance in the third mission. The difficulties in this regard immediately became evident. In fact, in the complex and multifaceted scenario of the third mission, while performance indicators relative to research (in terms of quality of publications, number of citations, etc.) and teaching (in terms of student-to-staff ratio, student evaluation, etc.) are widely known and used, less is known about how KT or, more generally, TM can be characterized and evaluated.

A widely used tool for evaluating and comparing universities performance are the so called "global rankings".

The rise in relevance of rankings concerning academic institutions is a rather recent phenomenon that emerged since the late 1980s (Sauder and Espeland, 2009) mainly due to the demand for information on academic quality by potential students, triggered by the worldwide expansion of access to higher education (Dill and Soo, 2005; Bellantuono et al., 2022). However, rankings have become an increasingly internalized tool for comparison and success quantification far beyond the matter of student's choice, influencing researchers, employers and, most relevantly, academic evaluators and companies (Oțoiu and Țițan, 2021; Johnes, 2018; Frondizi et al., 2019; Hazelkorn et al., 2014; Bellantuono et al., 2022).

Nowadays, rankings permeate multiple sectors and address multiple dimensions of individual and organizational behavior. There is a wide range of public measures such as different types of ratings, benchmarks, and rankings, especially for universities (Marhl and Pausits, 2011; Ringel et al., 2021; Dill and Soo, 2005; Bougnol and Dulá, 2015).

Rankings have become a tool for promoting the growth of universities in the international context, increasing their competitiveness, enhancing the attractiveness of the educational and research system (Forti and Meoli, 2020) namely, a true marketing, benchmarking and branding tool (Olcay and Bulu, 2017).

Efficient higher education institutions can actually prompt a spillover process in which regions collect knowledge and human resources that can contribute to foster their economic progress (Rodrigues, 2011; Saxenian, 1996; Mas Verdú et al., 2020; Agasisti et al., 2019). Therefore, it is evident that the evaluation of academic activity should be configured as a multi-purpose assessment (Oancea, 2019), which takes into account not only the scientific impact, but also the benefits brought to the territory, measured by an empirical quantification of the third mission outcomes (Benneworth and Hospers, 2007).

Although the scientific literature has already highlighted the presence of some critical aspects related to university rankings, such as the inhibition of regional contributions from universities (Salomaa et al., 2021), in recent years university rankings have been able to create impact (Hazelkorn et al., 2014; Rauhvargers, 2013) and influence higher education, policy and public opinion (Loukkola, 2016; Pusser and Marginson, 2013).

In fact, rankings, whatever they are meant to measure, are not neutral tools, and their use has a series of relevant drawbacks. Problematic aspects mainly stem from the effects of positive feedback between the prestige of an institution, certified by its ranked score, and the possibility to receive public funding on an awarding basis or to attract investments by private companies (Oancea, 2019; Hicks, 2012; Jonkers and Zacharewicz, 2016). The strong causal relation between ranking outcomes and funding triggers off undesired phenomena. The first problem is reactivity to rankings, namely the development of adaptation strategies to gain competitive advantages with respect to the evaluation criteria, leading to academic conformism (Oancea, 2019; Espeland and Sauder, 2007; Livan, 2019; Li et al., 2019; Fire and Guestrin, 2019). Another critical aspect is represented by territorial biases, that reward universities placed in an advantageous socioeconomic context (Rodrigues, 2011; Trippl et al., 2015; Smith and Bagchi-Sen, 2012; Rodrigues et al., 2001; Charles, 2006; Gunasekara, 2006), that can be, for example, more receptive than others with respect to third mission activities. The third issue is the onset of a “Matthew effect”, that, through the feedback between ranking and funding, consolidates existing gaps in third mission (Heher, 2006), internationalization (Rauhvargers, 2013; Van Vught, 2008), research (Clauset et al., 2015; Way et al., 2019), scholarships (Pusser and Marginson, 2013), and even diffusion of scientific ideas (Morgan et al., 2018).

Lately, rankings have prompted deep changes in the higher education system, affecting resource distribution, decision making and status definition (Sauder and Espeland, 2009; Espeland and Sauder, 2007; Johnson Jr, 2006; Stake, 2006). However, they fail to capture individual specificities and tend to marginalize parts of the academic community whose distinctive traits are not suited to the general rating framework (Pusser and Marginson, 2013; Sugimoto and Larivière, 2018). Though the methodology behind them is criticized, and their overall role is questioned (Hazelkorn and Gibson, 2017), rankings nowadays represent a consolidated evaluation framework, mainly due to their simplicity and practicality, combined with the lack of suitable alternatives (Coates, 2016). A different kind of rankings, in which the effects of structural factors are mitigated, can be relevant for academic evaluators and policy makers. Such redefined rankings could actually help identify both virtuous cases of outstanding institutions emerging in a difficult context, and cases in which the performance is below expectations, which therefore require intervention. Therefore, it would be desired to define transparent, data-driven, shared and reproducible procedures to evaluate academic performance, taking into account the effect of structural features, such as the territorial embedding of universities and their educational mission.

University rankings are compiled taking into account the variety of missions higher education is called to, which are not limited to teaching and research, but also involve the third mission (Laredo, 2007) or knowledge transfer (Bekkers and Freitas, 2008; Abreu et al., 2016).

In this complex scenario, about the issue of TM characterization and evaluation (Scanlan, 2018; O'Reilly et al., 2019), most of the best-known global university rankings completely lack instruments to evaluate KT activities (Olcay and Bulu, 2017; Landinez et al., 2019) and probably fail to properly capture and evaluate the peculiarities of KT and TM.
Defining the activities and quantifying the TM requires the design of a complex model of analysis that is able to determine the map of indicators related to the diversified dimensions of the third mission.
Several studies have tried to identify TM indicators; only a few are mentioned below.
For example, a research project called E3M "European Indicators and Ranking Methodology for University Third Mission" aimed to create a ranking methodology to measure universities' third mission activities, constructing 54 indicators, 20 of which were identified for the dimension of technology transfer and innovation (Carrión et al., 2012).
Marhl and Pausits (2011) point out in their study that TM activities pertain to more than one dimension, so it is not easy to obtain independent dimensions. However, they emphasize that these activities are important as components of institutional performance in rankings and therefore have devised a set of indicators to measure third mission activities using the Delphi method.
Lee et al. (2020) evaluated the engagement of universities in the third mission through their strategic plans.
Finne et al. (2011) attempted to design a composite indicator for knowledge transfer, considering three main sets of transfer mechanisms: through people (specially trained people), through cooperation (institutional cooperation in R&I), through university-university cooperation (Dip, 2021).
As can be seen, there is a lot of literature on KT indicators that could be used, but the correct formulation of the key indicators to measure the performance of universities in knowledge transfer activities is weakly developed in the literature (Rossi and Rosli, 2015; Dip, 2021).

2.2. From global to specialized rankings

Numerous global university rankings have been proposed so far, they are based on different parameters and indicators (Frondizi et al., 2019; Aguillo et al., 2010; Moed, 2017) and try to provide an all-round evaluation. In this research work, three of the best-known world university rankings were investigated in order to identify the best universities and be able to understand if and how they perform from the the third mission and knowledge transfer point of view:

- The Academic Ranking of World Universities (ARWU) was first published in June 2003 by the Center for World-Class Universities (CWCU), Graduate School of Education (formerly the Institute of Higher Education) of Shanghai Jiao Tong University, China, and updated on an annual basis. Since 2009 the Academic Ranking of World Universities (ARWU) has been published and copyrighted by Shanghai Ranking Consultancy, a fully independent organization on higher education intelligence that is not legally subordinated to any universities or government agencies. ARWU uses six objective indicators to rank world universities, including the number of alumni and staff winning Nobel Prizes and Fields Medals, the number of highly cited researchers selected by Clarivate Analytics, the number of articles published in journals like Nature and Science, the number of articles indexed in Science Citation Index - Expanded and Social Sciences Citation Index, and the per capita performance of a university. More than 1800 universities are ranked by ARWU every year and the best 1000 are published.

- The QS World University Rankings® (QSWUR) lists and ranks over 1000 universities from around the world, covering 80 different locations; it continues to rely on a remarkably consistent methodological framework, compiled using six simple metrics that effectively capture university performance. Universities are evaluated according to the following six metrics: Academic Reputation, Employer Reputation, Faculty/Student Ratio, Citations per faculty, International Faculty Ratio, International Student Ratio.
- The Times Higher Education World University Rankings (THEWUR) includes almost 1400 universities across 92 countries, standing as the largest and most diverse university rankings ever to date. It is based on 13 carefully balanced and comprehensive performance indicators and is trusted by students, academics, university leaders, industry and governments. Its performance indicators are grouped into five main areas: Teaching (the learning environment); Research (volume, income and reputation); Citations (research influence); International outlook (staff, students and research) and Industry Income (knowledge transfer).

However, as discussed in the previous paragraph, very often the global university rankings have been criticized (Olcay and Bulu, 2017; Bougnol and Dulá, 2015; Johnes, 2018; Moed, 2017) because, for example, using a single set of indicators, they compare different types of institutions (Hazelkorn and Gibson, 2017), evolving from a “semi-academic exercise” to an international business tool (Frondizi et al., 2019) and an important “instrument for the exercise of power” (Pusser and Marginson, 2013). Unfortunately, most of the best-known global rankings, probably lack instruments to evaluate KT activities and fail to properly capture and evaluate the peculiarities of KT and TM.

Thus, in this scenario, the first research question (RQ1) faced was: according to global university rankings, how do the world's top universities perform from a knowledge transfer point of view? In other words, are the rankings currently most used able to characterize universities performance from third mission point of view?

At the European level, given the importance of the issue related to the wider circulation of knowledge, increased access to knowledge and talent, and the rise of new products, services and markets, there is a growing interlocution regarding the need to identify effective and

objective indicators to measure knowledge transfer and the consequent need for harmonized indicators at the European level.
However, the KT indicators used in global university rankings are not recognized as the best indicators at the European level where the European Commission, in an attempt to provide answers, has promoted two separate initiatives.
The European Commission (European Commission Press release) has invested a significant effort in producing the U-Multirank (UMR) ranking (Dip, 2021). It is based on a different approach compared with the existing global university rankings and furthermore it includes a set of indicators focused on KT. This initiative has already produced a very rich data set over the years.
In a further effort the European Commission with the Joint Research Centre (JRC) published two reports about this issue: in 2020 "Towards a European-wide set of harmonised indicators" (Campbell et. al., 2020) and in 2022 "Knowledge Transfer Metrics: Phase II. Exploration of composite indicators for knowledge transfer" (Campbell et. al., 2022). This initiative, to date still in the experimental stage, has not yet produced meaningful datasets useful for conducting comparative analyses.
For these reasons in the present research work, in an attempt to answer the previous research question, we focused on UMR which uses a set of specialized KT indicators for evaluating the world's top universities from the KT point of view and which, thanks to the rich dataset, enables comparisons to be made with the most popular global rankings.
The following paragraphs provide a brief description of the initiative promoted by the JRC and U-Multirank.

2.2.1 JRC initiative

The European Commission Joint Research Centre (JRC) published two reports about this issue: in 2020 "Towards a European-wide set of harmonised indicators" (Campbell et. al., 2020) and in 2022 "Knowledge Transfer Metrics: Phase II. Exploration of composite indicators for knowledge transfer" (Campbell et. al., 2022). In these reports, the activity indicators used are:

- Number of invention disclosures (IDF),
- Number of licenses,
- Revenue from licensing,
- Number of spin-off companies,
- Revenue from equity sale in spin-off companies,
- Number of research collaboration agreements with non-academic entities,
- Revenue from research collaboration with non-academic entities.

It is quite well recognized that purely quantitative indicators, such as financial data or IP assets, are not sufficient to describe the complexity of knowledge transfer and their long-

term impact. See what is also reported by the reports of leading international KT associations, such as, AUTM ("Better World") or Knowledge Transfer Ireland.
Despite this, having a basic data set, together with cohesive basic definitions, gives a pathway for comparative and longitudinal analyses, provided the observer is adequately aware of the complexity of the field to appreciate an informed analysis (Campbell et. al., 2020).
In fact, the indicators used represent a subset of the "Quadrant KT Indicators Model" (Campbell et. al.) (2020), shown in Figure 7, which represent the four quadrants affecting input and output KT Indicators:

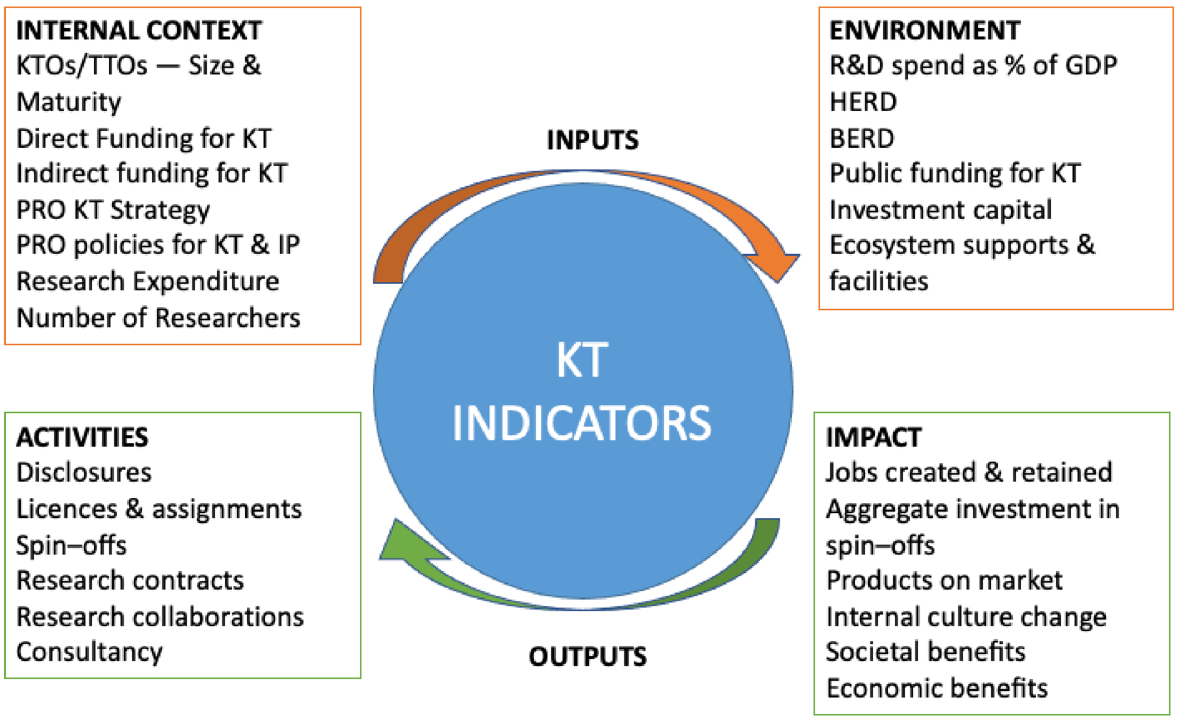


Figure 7: the four quadrants affecting input and output KT Indicators

Composite indicators are becoming more widely acknowledged as valuable tools, as they offer straightforward comparisons of countries and organizations, shedding light on intricate and often elusive matters in various domains. However, composite indicators can be misleading if not constructed and interpreted effectively. The quality of a composite indicator is primarily contingent on the excellence of the framework and data employed, as well as the methodology developed and implemented (Campbell et. al., 2022).

2.2.2 U-Multirank (UMR)

UMR was launched by the European Commission in collaboration with Bertelsmann Foundation and Banco Santander, based on the results of a feasibility study covering 150 universities which was carried out in 2010/11 (van Vught and Ziegele, 2011). It is based on a different approach compared with the existing global university rankings. It compares university performances considering different activities that they are engaged in, taking into account the diversity of the higher education sector and the complexity of evaluating educational performance (Dip, 2021; Prado, 2021; Decuypere and Landri, 2021). UMR has been developed based on a number of design principles, user-driven, multidimensionality, comparability, multilevel nature of higher education, and methodological soundness (Kaiser and Zeeman, 2017). It is considered a transparency tool for higher education stakeholders (Westerheijden and Federkeil, 2018) and takes into account five aspects and dimensions of the universities' performance: (1) teaching and learning, (2) research, (3) knowledge transfer, (4) international orientation and (5) regional engagement. The UMR web tool allows users to compare universities but also study programs. Based on empirical data, it compares institutions with similar profiles ('like-with-like') and allows users to develop their own personalized rankings by selecting indicators in terms of their own preferences. Each of the five aspects evaluated by UMR are ranked from A to E, with A and E indicating "very good" and "weak" performance, respectively.

In this research work, the choice to use of UMR was motivated by the fact that it uses a focused quality model for an in-depth evaluation of KT dimension based on the following indicators:

- Co-publications with industrial partners: the percentage of a department's research publications that list an author affiliated with an address that refers to a for-profit business enterprise or private sector R&D unit (excluding for-profit hospitals and education organizations).
- Income from private sources: the percentage of external research revenues (including not-for-profit organizations) coming from private sources, excluding tuition fees. Measured in €1.000s using Purchasing Power Parities and computed per fte (full time equivalent) academic staff.
- Patents awarded (absolute numbers): the number of patents assigned to inventors working at the university in the respective reference period.
- Patents awarded (size-normalized): the number of patents assigned to inventors working at the university over the respective reference period, computed per 1.000 students to take into consideration the size of the institution.
- Industry co-patents: the percentage of the number of patents assigned to inventors working at the university during the respective reference period, which were obtained in cooperation with at least one applicant from the industry.
- Spinoffs: the number of spinoffs (i.e. firms established on the basis of a formal KT arrangement with the university) recently created by the university (computed per 1000 fte academic staff).
- Publications cited in patents: the percentage of the university's research publications that were cited in at least one international patent (as included in the PATSTAT database).

- Income from continuous professional development: The percentage of the university's total revenues that is generated from activities delivering Continuous Professional Development courses and training.
- Graduate companies: The number of companies newly founded by graduates and computed per 1000 graduates.

CHAPTER 3. From university rankings to Intellectual Property Analytics

3.1. The importance of the intellectual property management

Despite the flourishing of indicators and the numerous rankings that exist, for at least two main reasons, the most investigated indicator in relation to KT is the number of patents granted.

Indeed, on the one hand, it is certainly if not exclusively among the most correlated and discriminating indicators with respect to KT and the performance of a research Institution in knowledge transfer.

On the other hand, the only databases available for analysis, and certainly those containing the most objective and universal data, are "again" the patents granted.

In the framework of the Fourth industrial revolution, identified with exponential evolutions, integration of technologies and holistic system impact across society, industry, and countries (Schwab, 2017), the increased data availability represents an opportunity to better support decision-making processes and introduce disruptive technologies (Baglieri & Cesaroni, 2013; Aristodemou & Tietze, 2018). The integration of Industry 4.0 technologies within society is pivotal for resolving many challenges that the world and its population are currently facing (Bartoloni et al., 2022). In this multidisciplinary and complex context, the technology transfer plays a key role for the adsorption and dissemination of technologies, resources, and knowledge to transform each invention into tangible and useful innovation.

Under the EU valorization policy, the use of knowledge and technology, the management of intellectual property, and the involvement of citizens, academia, and industry, through different channels, are highly promoted (EU valorisation policy 2020). In modern knowledge economies, intellectual property (IP) assets are both engines of development and drivers of social transition. Industries that make intensive use of intellectual property rights (IPRs), such as patents, trademarks, industrial designs and copyrights, generate 45 percent of annual GDP (€6.6 trillion) in the EU and account for 63 million jobs (29 percent of all jobs) (EU valorisation policy 2020).

As reported by the European Commission, the volume of annual investment in "intellectual property assets" has increased by 87 percent in the EU over the past two decades, in contrast to the volume of tangible (non-residential) investment. Thus, industries that make intensive use of intellectual property play an essential role in the EU economy and provide good and sustainable jobs for society (Press release EU, 2020). Seemingly, according to a study conducted by the Ponemon Institute LLC looking at the S&P500 (Standard and Poor's 500) companies, the relative importance of the intangible assets over the total patrimonial value of those organizations has increased dramatically in the last 40 years, passing from representing about 20% (122B$ intangibles vs. 594B$ tangibles) in 1975 to a ratio of more than 5 to 1 in 2018 (21T$ intangibles vs. 4T$ tangibles) (Ponemon, 2019).

With rapid changes in technology and industry value chains, it is vital for companies to be able to identify promising emerging technologies that can better respond to rapid external changes and be used to launch new businesses or improve current ones.
One of the most widely used approaches to identifying promising emerging technologies is patent analysis (Choi et al., 2021).
Patent documents are considered as a valuable database for understanding technology trends and design innovation strategies. They contain information about almost all relevant technological fields and record the direction of technological development and R&D activities (Wang and Lin, 2023). IP related documents represent valuable and recognized sources of technological and legal knowledge (Aristodemou et al., 2017). As a matter of fact, different documents, such as patents, trademarks and other IP registered in national and international IP systems, contain important research results which are of great value for industry, legal researchers, and policy advocates in science and technology R&D (Trappey et al., 2017).
Intangible assets such as R&D, inventions, artistic and cultural creations, brands, software, know-how, business processes and data "are the cornerstones of today's knowledge economy" (Press release EU, 2020).
Since between 70% and 90% of the information about technologies is not published anywhere except in patent documents (Asche, 2017; Giordano et al., 2021), patents are among the best sources of information (Puccetti et al., 2023). Organizations analyze patents for, but not limited to:

- technological mapping and forecasting (Daim et al., 2006),
- predicting core and emerging technologies (Huang et al., 2020; Altuntas et al., 2015; Kim and Bae, 2017; Kyebambe et al., 2017),
- diffusion of technologies (Daim et al., 2006),
- convergence of technologies (Karvonen and K¨assi, 2013),
- identification of technological vacuums and hotspots (Abbas et al., 2014)
- portfolio analysis (Ernst, 2003),
- competitive analysis (Thorleuchter et al., 2010; Aristodemou and Tietze, 2018),
- technology trend analysis (Tseng et al., 2011; Trappey et al., 2019; Choi et al., 2021),
- avoiding infringement (Yu and Zhang, 2019),
- identifying technological competitors (Abbas et al., 2014).

However, technologies patented by universities and research centers notoriously, at least at the European level, are poorly exploited; in fact, as reported in ASTP Survey Report on KT Activities FY2019, only 18% of inventions are licensed or optioned (ASTP Survey Report, 2019).
In order to address this lack of valorization and help universities and research centers promote their research results, specific initiatives such as online patent platforms have emerged over time as convenient channels for patent transfer, with the joint effort of both academic and political partners (Chen et al., 2020; Chen and Deng, 2023). The platforms

enable the integration of isolated patents and provide communication and negotiation services to enhance the patent transfer (Deng and Ma, 2022; Chen and Deng, 2023). An active and efficient patent marketplace could help reduce the transaction cost and facilitate technology transfer, alleviating also the information asymmetry problem (Du et al., 2021).
Existing initiatives/platform that are able to create matchmaking between supply and demand for innovation are sometimes ineffective, mainly for the following reasons:

- they are paid services, not open access - often open innovation platforms;
- they report the patent document as such, without a usable "translation" for all that facilitates matching;
- the classification of the patent in a given technological area is a challenging task: users choose a category based on those proposed, but users often do not know how to choose best, and it is not true that the proposed choices are necessarily the best.

Often the classification of patents and, therefore, the search and consultation method used, are based on taxonomies and keywords self-defined by users, experts or database managers and are not very effective. This often leads to multidisciplinary categories containing a significant number of poorly characterized and classified patents, which can be defined as “monster class”. Monster categories are thus ineffective, not discriminating, and difficult to explore. A typical phenomena of “monster class” often occur in the case of healthcare related innovations, which, due to multidisciplinary and particularly innovative nature, tends to generate classes with a great number of patents.
Thus, one of the most prominent step in managing patents is the classification, a particularly expensive and time consuming phase, above all due to the increase in the number of filed patents and the complexity of the contents; this task is in fact conventionally performed by domain experts (Haghighian Roudsari et al., 2022; Krestel et al., 2021). Moreover, patent classification is almost always a multi-label classification task, which makes the problem even more complicated. Therefore, finding ways to automate this costly and labor-intensive task is essential to assist domain experts in managing patent documents, facilitating reliable searching, better matching of innovation demand and supply, especially by companies (Haghighian Roudsari et al., 2022; Yun and Geum, 2020; Souza et al., 2021; Gomez and Moens, 2014). In particular, an invention may be related to different technological areas or be cross domain, so a patent may be assigned to different classification labels. Therefore, patent classification becomes a multi-label classification problem which is more challenging.

In consideration of the above issues, further research questions that this thesis aimed to answer are:

- RQ2: are the classification taxonomies used in the patent platforms effective in classifying the whole landscape of academic patents?
- RQ3: is it possible to support the user in correctly classifying a patent entered into the platforms in order to improve the matchmaking between demand and supply of innovation?

- RQ4: is it possible to draw up an attempted vocabulary of some technological fields from the keywords that emerged from the applied AI-based approaches?

3.2. Intellectual Property Analytics (IPA)

To answer these research questions, this research work proposes the use of a specific workflow in Intellectual Property Analytics (IPA).

In recent years, Intellectual Property Analytics IPA has emerged as a multidisciplinary approach used to gain valuable insight about intellectual property data (Trippe, 2015; Aristodemou & Tietze, 2018).

Intellectual property knowledge databases come in heterogeneous forms (text, data, images, colors, and smells), and it is becoming increasingly difficult to analyze, synthesize and classify their contents (Trappey et al., 2020a).

For this purpose, automated approaches, such as natural language processing for data, text and graph mining, clustering and neural networks, are increasingly used for IP knowledge processing and various tools have been developed for supporting patent analysis experts, business managers, and technology offices (Trappey et al., 2009; Lei et al., 2019; Yoon & Park, 2007; Rodriguez et al., 2016; Puccetti et al., 2021; Kang et al., 2020; Trappey et al., 2020b; Song et al., 2022).

Analyzing patent data using the automated tools to discover the patent intelligence through visualization, citation analysis, and other techniques, such as text mining is termed as "patent informatics" (Trippe, 2003). These techniques can be broadly classified into text mining techniques and visualization techniques and involve three steps: in the first stage, patent documents are retrieved from patent databases; next, the patent documents are transformed into structured data using text mining techniques. In the third step, working on the structured data, big data learning approaches, such as classification, regression and clustering, etc., are used for deriving logical conclusions, such as patent novelty detection and patent quality identification, trend analysis and technology forecasting, R&D planning management, etc (Abbas et al., 2014). Figure 8 shows the generic workflow of patent analysis.

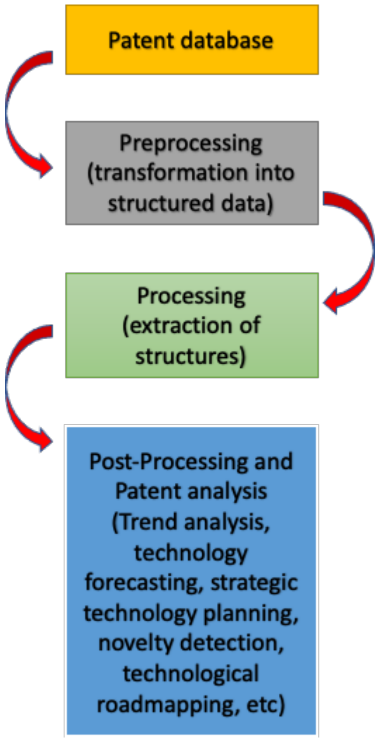


Figure 8: workflow of patent analysis

Several analytical methods have been used for the analysis of intellectual property data and, in particular, patents (Abbas et al., 2014; Trippe, 2015).

Abbas et al. (Abbas et al., 2014) provide a comprehensive literature review of patent analysis techniques, in which they distinguish between text mining and visualization approaches and applicability to structured and unstructured data (Bonino et al., 2010).

Aristodemou and Tietze (2018) summarize in their review the main methods of intellectual property analysis in the literature (Table 1).

Table 1: the main methods of intellectual property analysis in the literature (Aristodemou and Tietze, 2018)

Approach	**Method**
Artificial Neural Networks	Evolutionary sigmoidal unit, Evolutionalry product unit Extension theory Extreme learning machine (ELM) Growing cell structure, paired with Girvan-Newman

	clustering algorithm Restricted Boltzmann machines
Clustering	K-means (and derivations)
Deep Learning (DL)	Deep Belief Networks (DBN)
Ensemble	Bootstrapping Random Forest Stacking
Decision tree	Classification and Regression Tree (CART)
Dimensionality Reduction	Linear Discriminant Analysis (LDA) Multi-dimensional scaling (MDS) Principal Component Analysis (PCA) Quadratic Discriminant Analysis (QDA) Singular Value Decomposition (SVD)
Regression	Linear Logistic
Statistical and probabilistic modelling	Conditional random fields (CRF) Latent Dirichlet Allocation (LDA) Naive Bayes Hidden Markov Model (HMM)
Support Vector Networks (SVN)	Support Vector Clustering (SVC) Support Vector Machine (SVM) Semantic Support Vector Machine (SVM)
Text mining approaches	Dictionary-based approach Natural Language Processing (NLP)

	Rule-based approach Semantic based ontology

In this scenario we focus attention on the following methods for providing answer to the research questions defined:

1. Natural Language Processing and clustering techniques are used to improve the taxonomy-based classification - RQ2;
2. Regression is used to build a multi-label classification system - RQ3;
3. A complex network analysis on the most frequent words is used in order to improve the keywords-based classifications - RQ4.

3.3. NLP and ML for clustering and regression

In recent years, artificial intelligence (AI) and machine learning (ML) have seen a new wave of publicity fueled by the huge and ever-increasing amount of data and computing power, as well as the discovery of better learning algorithms. However, the idea of a computer learning some abstract concept from data and applying it to as-yet-unknown situations is not new and has been around since at least the 1950s (Rätsch, 2004).
Artificial intelligence is the ability of a device to mimic intelligent human behavior (Xu et al., 2021). AI performs tasks that in the past could only be done by the human mind such as thinking, reasoning, learning from experience, and especially making decisions.
AI can be defined as the science that develops the architecture necessary for machines to function like the human brain and related neural networks (Sheikh et al., 2023). It is a computer system that attempts to simulate biological neural networks. The ultimate goal of AI is to create computers with reasoning abilities similar (if not equal) to humans.
Machine learning is the algorithm that allows intelligent machines to improve with time, just as it does with the human brain. Without advanced learning, in fact, it would not be possible to put artificial intelligence "in motion" (Jyothi and Khare, 2023).
AI is a field focused on automating intellectual tasks normally performed by humans, and machine learning and deep learning (DL) are specific methods to achieve this goal.
Figure 9 shows the relationship among AI, ML, DL and NLP:

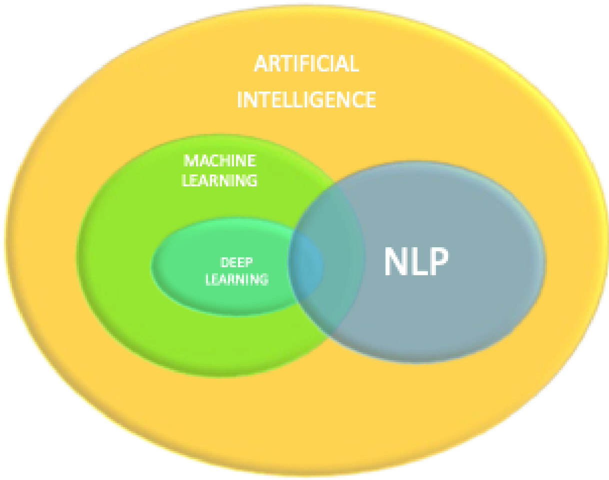


Figure 9: relationship among AI, ML, DL e NLP

3.3.1 NLP

Natural Language Processing is the field of AI that focuses on language. NLP is defined as "the ability of systems to analyze, understand, and generate human language, including speech and text" (Panesar, 2019).
NLP is the set of methods for "making human language accessible to computers". NLP is embedded in our daily lives: automatic machine translation on the web and social media, text classification in our email inboxes; dialog system and so on (Eisenstein, 2019).
NLP is an interdisciplinary field of research, begun in the 1950s, that embraces computer science, artificial intelligence, and linguistics (Nadkarni, 2011), that deals with the interaction between computers and human language, specifically on how to program computers to process and analyze large amounts of natural language data. The goal is to make technology capable of "understanding" the content of documents and their contextual nuances, so that it can then accurately extract information and ideas contained in documents, as well as classify and categorize them. Its aim is to develop algorithms capable of analyzing, representing, and then "understanding" natural language, written or spoken, in a manner similar to or even better performing than humans. Such "understanding" is determined by understanding, and then being able to use, language at various granularities, from words, in relation to their meaning and appropriateness of use with respect to a context, to grammar and the rules of structuring both sentences from words and paragraphs and pages from sentences.
Language processing challenges often involve speech recognition, natural language understanding, and natural language generation.
In more detail, firstly, NLP provides solutions for analyzing the syntactic structure of the text, associating individual words with their respective morphological categories (e.g., noun,

verb, adjective), identifying entities and classifying them into predefined categories (e.g., person, date, place), extracting syntactic dependencies (e.g., subjects and complements) and semantic relations (e.g., hyperonymy, meronymy). Secondly, it provides insight into the semantics of the text, identifying the meaning of words, also related to context and usage patterns (e.g., irony, sarcasm, sentiment, mood), classifying it into predefined categories (e.g., sports, geography, medicine) or summarizing its content.

This process is made particularly difficult and complex because of the inherent ambiguity characteristics of human language. For this reason, the processing is broken down into different stages, however similar to those that can be encountered in the processing of a programming language (see Figure 10) (Indurkhya and Damerau, 2010):

- lexical analysis: decomposition of a linguistic expression into tokens (in this case words);
- grammatical analysis: association of parts of speech with each word in the text;
- syntactic analysis: arrangement of tokens into a syntactic structure;
- semantic analysis: assigning meaning (semantics) to the syntactic structure and, consequently, to the linguistic expression.

In semantic analysis, the automatic procedure that assigns a meaning from among several possible meanings to the linguistic expression is called disambiguation (Indurkhya and Damerau, 2010).

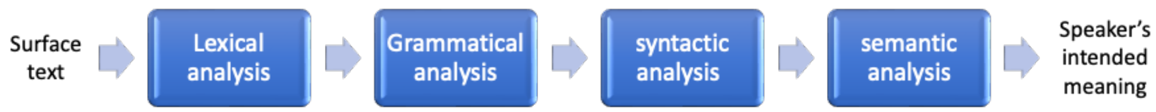


Figure 10: phases of processing of a programming language

NLP is useful in the following tasks:

- The retrieval of structured and unstructured data in a dataset;
- Social media monitoring;
- Interpretation of natural language from humans as in virtual assistants or speech recognition;
- Ability to analyze and interpret a text to get a sense of feeling and mood;
- Image to text recognition;
- Topic modelling;
- Understanding sentiment from social media;
- Etc. (Panesar, 2019).

Despite convincing results in different applications, for example with search engines and in knowledge mining, the need to further improve the automatic comprehension capabilities of

natural language content, reaching human-like levels, is, still, an open challenge that the research world is working on.
NLP has a toolkit of text processing procedures including a range of data mining methods that can be used for model development (Panesar, 2019). The main steps of NLP are performed in the following order:

- Tokenization.
- Stop-words removal.
- Stemming.

In the *Tokenization* phase, texts were subdivided into single words (also called “tokens”). A tokenizer breaks unstructured data and natural language text into chunks of information that can be viewed as discrete elements. The occurrences of tokens in a document can be used directly as a vector representing that document. This step is then able to transform an unstructured string (text document) into a numeric data structure that can be used directly by a computer to trigger useful actions and responses or instead suitable for machine learning. Tokenization is used to separate sentences, words, characters, or subwords. When the text is splitted into sentences, we talk about sentence tokenization.
Regarding tokenization of words, it is also important to consider:

- Bigrams: Tokens consist of two consecutive words, known as bigrams.
- Trigrams: Tokens consist of three consecutive words, known as trigrams.
- Ngrams: Tokens consist of an 'N' number of consecutive words, known as ngrams.

In the second step the noise from the data was removed. In the *Stop-words removal* phase all the useless tokens (such as articles, prepositions, conjunctions, punctuation, numbers etc.) were removed. Thus, some ubiquitous words, which seem of little value for analysis purposes but increase the dimensionality of the feature set, are excluded completely from the vocabulary as part of the stopword removal process. Stopwords refer to the most common words in a language (such as "the," "a," "in") that aid in sentence formation, but these words provide less or no meaning in language processing.
The removal of these words occurs essentially for two reasons:

- Irrelevance: It allows only content-bearing words to be analyzed. Stopwords, also called empty words because they generally do not have much meaning, introduce noise into the analysis/modeling process.
- Dimensionality: Removing stopwords allows tokens in documents to be significantly reduced, thus decreasing the size of features.

NLTK library consists of a list of words that are considered stopwords for the English language. Some examples are: [i, me, my, myself, we, our, ours, ourselves, you, you’re, you’ve, you’ll, you’d, your, yours, yourself, yourselves, he, most, other, some, such, no, nor, not, only, own, same, so, then, too, very, s, t, can, will, just, don, don’t, should, should’ve, now, d, ll, m, o, re, ve, y, ain, aren’t, could, couldn’t, didn’t, didn’t]. But not only the list provided by the library was used as stopwords, as they were chosen wisely based on our patent text.

In the *Stemming* phase, the remaining words were "stemmed" so that only the root-words were kept; stemming is a normalization technique in which words are stemmed or reduced to their basic root/form to remove redundancy. A computer program that reduces words can be called a stemmer. A stemmer reduces words such as 'programmer', 'programming', 'program' to 'program'; 'fished' and 'fishing' were transformed into their common root-word "fish" (Panesar, 2019).

In particular, the pre-processing phase in patent analysis is useful to extract the information from structured and unstructured data contained into the patent documents. In other words, NLP is used to transform technological information into simple linguistic structures by extracting grammatical structures from text data and creating structural relationships between components (Abbas et al., 2014; Masiakowski and Wang, 2013). The text mining techniques help in this task because it is a knowledge-based process using analytical tools in order to derive meaningful information from natural language text (Abbas et al., 2014). The most widely used text mining techniques in the literature for patent analysis are mainly based on NLP approaches, property function-based approaches, rule-based approaches, neural network-based approaches and semantic-based approaches (Park et al., 2013; Abbas et al., 2014).

3.3.2 ML

Machine learning algorithms are designed to learn patterns and relationships from data, and then use that knowledge to make informed decisions or predictions.
The fundamental concept behind machine learning is to build mathematical models that can automatically learn and improve from experience. These models are trained on labeled data, which consists of input examples along with their corresponding correct output or target values. During the training process, the machine learning algorithm analyzes the data, identifies patterns, and adjusts its internal parameters to optimize its performance in making predictions or decisions (Smola, 2008; Alpaydin, 2020).
In ML, there are four used learning methods, each useful for solving different tasks: supervised, unsupervised, semisupervised, and reinforcement learning - see Figure 11 (Choi et al., 2020; James et al., 2013; Hastie et al., 2009; Badillo et al., 2020).

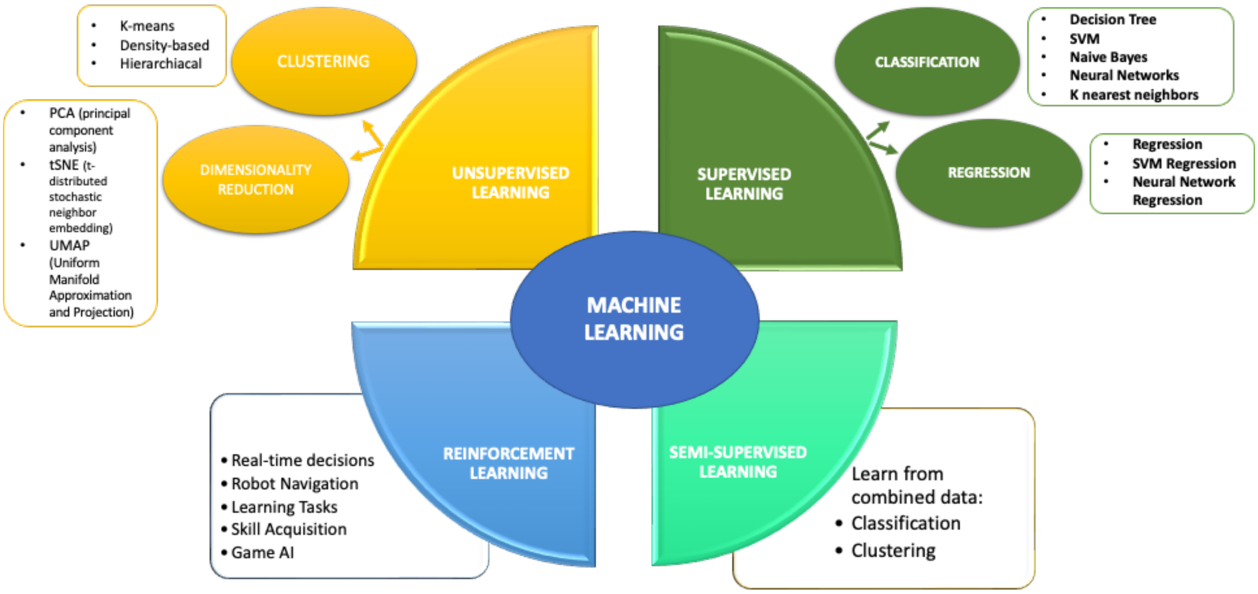


Figure 11: four used learning methods in ML

Clustering can be considered the most important unsupervised learning problem.
Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition (Madhulatha, 2012).
Clustering consists of a set of methods for grouping objects into homogeneous classes. A cluster is a set of objects that have similarities with each other but, conversely, have dissimilarities with objects in other clusters. The input of a clustering algorithm is a sample of items, while the output is given by a number of clusters into which the items in the sample are divided according to a measure of similarity. Clustering algorithms also provide as output a description of the characteristics of each cluster, which is crucial for then making strategic decisions about actions to be taken toward these clusters (targeted marketing, ad-hoc promotions, creation of new products/services) (Badillo et al., 2020; Alpaydin, 2020).
Cluster analysis is used for numerous applications:

- Market research
- Pattern recognition
- Clustering of customers based on buying behavior (market segmentation)
- Product positioning
- Social network analysis, for recognition of user communities
- Identification of outliers.

Data clustering algorithms can be of two types: hierarchical or partitional. Hierarchical algorithms find successive clusters using previously established clusters, while partitional algorithms determine all clusters at one moment in time. Hierarchical algorithms can be agglomerative (bottom-up) or divisive (top-down). Agglomerative algorithms start with each element as a separate cluster and join them into successively larger clusters. Divisive algorithms start with the entire set and proceed to divide it into smaller clusters ((Madhulatha, 2012; Milligan and Cooper, 1987).

The regression is one of the most important supervised machine learning techniques.
The supervised regression methodologies lead to numerical representation of output variables in order to predict a number. Regression is mainly used for market forecasting, growth prediction, and life expectancy calculation (Chang, 2020).
The linear regression delineates the strength of the relationship between two continuous variables. The method for fitting a regression line in linear regression is the method of least squares with a correlation coefficient r. This regression is named “simple” when there is a single input variable and “multiple” when there are multiple input variables.
The logistic regression is the adaptation of the aforementioned linear regression to a binary classification (via a logistic function to yeld maximum likelihood) (Chang, 2020).

CHAPTER 4. The experimental fields

4.1. Knowledge Share

In the framework of the online patent platforms born in order to enhance the transfer and exploitation of intellectual property, in the Italian landscape, in order to overcome the difficulties that many Italian universities face in effectively promoting their research results and their IP assets, the Knowledge-Share platform was developed as a joint project involving Politecnico di Torino, the Italian Patent and Trademark Office at Ministry of Enterprises and Made in Italy and Netval (the Italian Network for the Valorization of Public Research). KS is a platform designed for Italian Universities, Research Centers (PROs), Scientific Institute for Research, Hospitalization and Healthcare (IRCCS) to showcase their patented technologies and spin-off projects seeking commercialization opportunities, and for businesses to find solutions and expertise to overcome R&D&I challenges (Technology Transfer System Handbook, 2019). It is specifically aimed to "translate" the contents of academic patented inventions into a self-speaking language which anybody can understand (the so called "patent marketing annex"), thus obtaining three important results: i) to generate a real social and economic impact at national level, in accordance with the objectives of the Third Mission; ii) to provide a tangible support to (not only) Italian businesses to accelerate their innovation processes; iii) to drive economic return for Universities and PROs to be re-invested in new technology transfer activities within the public research system.
Particularly the platform's key objectives are to:

- become the touchpoint between corporations, SMEs and public research;
- create a national standard to foster the exploitation of intellectual property;
- create an innovation network for technological excellence at an international level;
- provide industry scouting teams with an easy and effective way to tap into the Italian research landscape;
- provide a service for technology transfer offices (market intelligence);
- promote and foster events and initiatives related to innovation and exploitation of research;
- generate spin-offs and innovative technology projects.

Existing initiatives/platforms that are able to create matchmaking between supply and demand are classically open innovation platforms or marketplace of patents/technologies.
Classically existing traditional open innovation platforms:

- gather challenges from companies;
- find and allow "solvers" to propose their own solutions;
- solvers may be companies, start-ups and physical people (including researchers).

Their main pain results in poor scalability.
The patent or technology marketplaces offer paid or free solutions: paid services provide various quality and value-added solutions, such as "edited" content for more effective and

"user friendly" representation, while free offerings are characterized by low value-added offerings, often simply aggregation of patent content from public databases with little/no rework and quality control on content. Their main pain is "value for money" (high cost or low interest offerings).
Among the existing initiatives, KS shows some critical success factors:

- the "guarantee" of updating and feeding content;
- the content quality assurance process (total quality approach);
- the simplicity of language and description of patented technology;
- the gratuitousness of all services for both technology providers and technology seekers;
- the presence of an established community in Italy;
- the organization of events with synergistic function to the promotion actions;
- institutional collaboration and synergy with other initiatives launched by Ministries and other stakeholders at national and international level.

Knowledge Share was recognized as best practice by the European Union - “Promoting IP valorization through the IP platform – Knowledge Share run by the national network NETVAL” and was chosen as one of 30 case studies presented in 'How did COVID-19 shape co-creation', the report published by the OECD on the crucial role of co-creation during the COVID-19 pandemic emergency. Industry, research, government and civil society worked together to reactively and prolifically initiate support for innovation and technology transfer (De Silva et al., 2022).
The KS database we use as an experimental field in this research work includes patents registered from 12/28/1999 to 8/16/2021, consisting of 1694 patents. These documents are uploaded to the platform by 89 Italian Research Centers, both public and private (Universities, Research Centers, Scientific Institute for Research, Hospitalization and Healthcare, etc). The yearly number of patents is highly variable, depending on the filing date and the consequent policy about the protection of intellectual property at national and international level. This platform can be easily queried by users aiming at obtaining an overview on the state-of-the-art about particular technologies and ground-breaking startups in Italy. On one hand, this service lowers firms and investors’ entry barriers for innovations in fundamental and applied science, letting them overcome R&D&I challenges more easily (Technology Transfer System Handbook, 2019). On the other hand, this platform helps scientists and startups in achieving visibility, expressing their innovative potential and gaining interests from private and public investors.
One or more labels can be assigned to each patent; these labels represent the following ten technological domains:

- Aerospace and aviation;
- Agrifood;
- Architecture and design;
- Chemistry, Physics, New materials and Workflows (Basic Science);

- Energy and Renewables (Green Energy);
- Environment and Constructions (Environment);
- Health and Biomedical (Biomed);
- Informatics, Electronics and Communication System (Electronics);
- Manufacturing and Packaging (Packaging);
- Transports.

The website allows the search for patents according to several criteria: the name, the organization it comes from (i.e. the patent owner), the area of application, the events at which they were shown and free full text research.

The Figure 12 shows the platform homepage and the possibility to search patents starting from a keyword, the name of the Institution, the technological area.

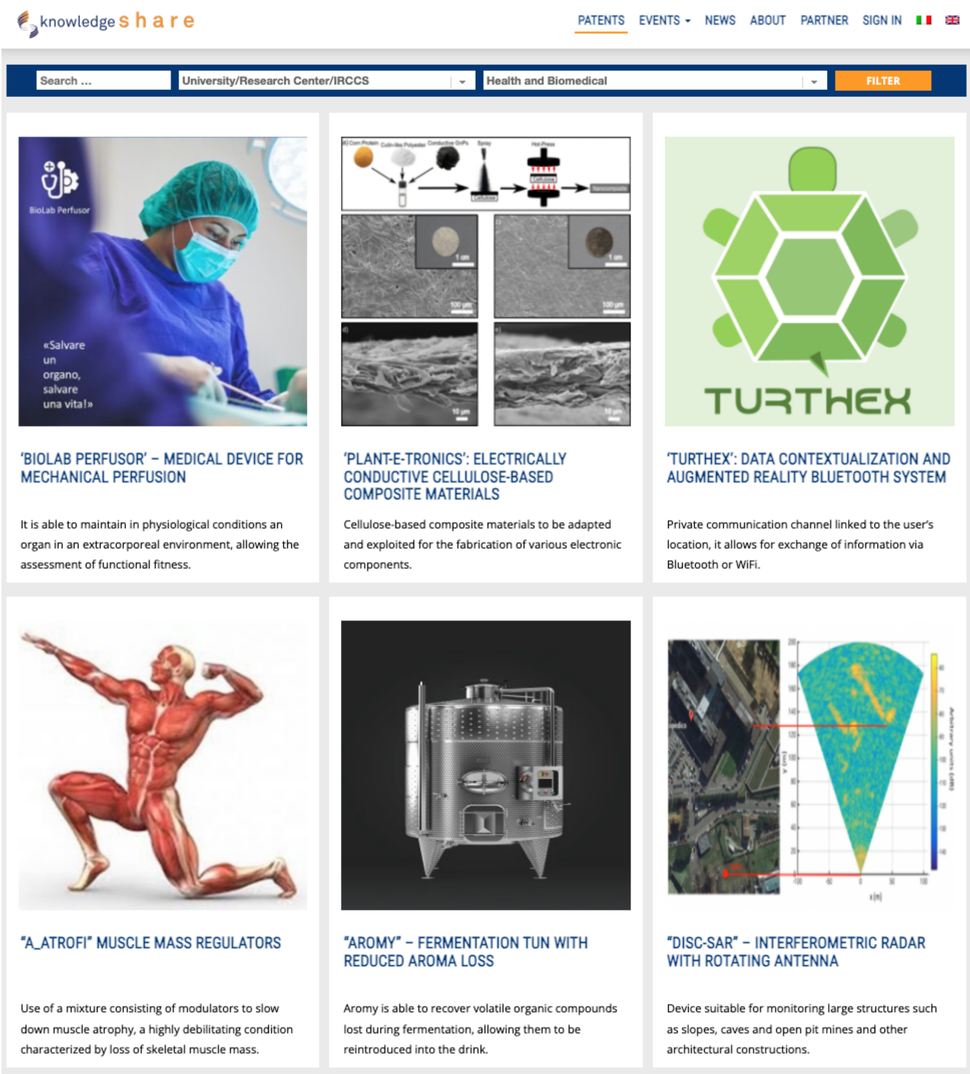
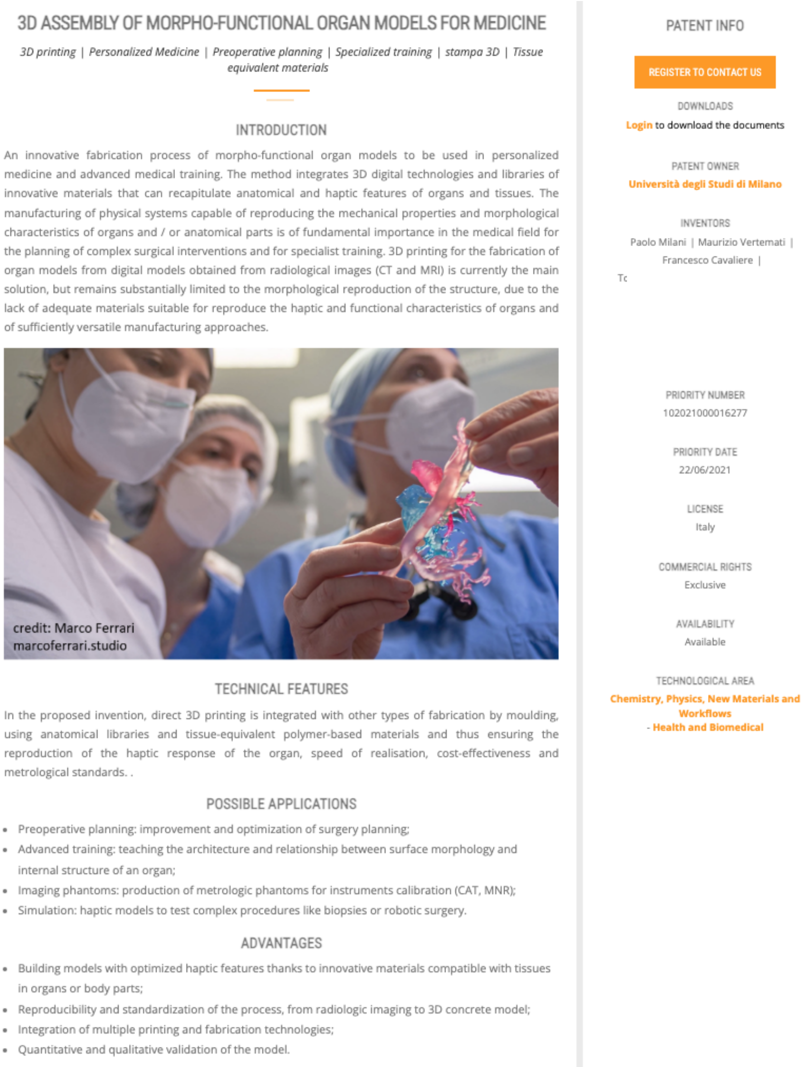


Figure 12: Knowledgeshare home page

For each patent, described in an informative and non-technical language, the technical features, the applications, uses and characteristics, and the benefits deriving from the adoption of the technology are illustrated. Furthermore, information about the inventors, the priority number, the priority date, the license, the commercial rights can be found, and it is possible to download a "marketing annex", i.e., a sheet that contains the basic information on the patent, conceived to be a functional and brief communication tool to share and circulate outside the platform (Figure 13).



3D ASSEMBLY OF MORPHO-FUNCTIONAL ORGAN MODELS FOR MEDICINE

3D printing | Personalized Medicine | Preoperative planning | Specialized training | stampa 3D | Tissue equivalent materials

INTRODUCTION

An innovative fabrication process of morpho-functional organ models to be used in personalized medicine and advanced medical training. The method integrates 3D digital technologies and libraries of innovative materials that can recapitulate anatomical and haptic features of organs and tissues. The manufacturing of physical systems capable of reproducing the mechanical properties and morphological characteristics of organs and / or anatomical parts is of fundamental importance in the medical field for the planning of complex surgical interventions and for specialist training. 3D printing for the fabrication of organ models from digital models obtained from radiological images (CT and MRI) is currently the main solution, but remains substantially limited to the morphological reproduction of the structure, due to the lack of adequate materials suitable for reproduce the haptic and functional characteristics of organs and of sufficiently versatile manufacturing approaches.

credit: Marco Ferrari
marcoferrari.studio

TECHNICAL FEATURES

In the proposed invention, direct 3D printing is integrated with other types of fabrication by moulding, using anatomical libraries and tissue-equivalent polymer-based materials and thus ensuring the reproduction of the haptic response of the organ, speed of realisation, cost-effectiveness and metrological standards. .

POSSIBLE APPLICATIONS

- Preoperative planning: improvement and optimization of surgery planning;
- Advanced training: teaching the architecture and relationship between surface morphology and internal structure of an organ;
- Imaging phantoms: production of metrologic phantoms for instruments calibration (CAT, MNR);
- Simulation: haptic models to test complex procedures like biopsies or robotic surgery.

ADVANTAGES

- Building models with optimized haptic features thanks to innovative materials compatible with tissues in organs or body parts;
- Reproducibility and standardization of the process, from radiologic imaging to 3D concrete model;
- Integration of multiple printing and fabrication technologies;
- Quantitative and qualitative validation of the model.

PATENT INFO

REGISTER TO CONTACT US

DOWNLOADS
Login to download the documents

PATENT OWNER
Università degli Studi di Milano

INVENTORS
Paolo Milani | Maurizio Vertemati | Francesco Cavaliere |
Tc

PRIORITY NUMBER
102021000016277

PRIORITY DATE
22/06/2021

LICENSE
Italy

COMMERCIAL RIGHTS
Exclusive

AVAILABILITY
Available

TECHNOLOGICAL AREA
Chemistry, Physics, New Materials and Workflows
- Health and Biomedical

Figure 13: Marketing annex of each patent uploaded on Knowledge-Share platform

The process of uploading a new patent to the platform starts from the individual university, which invites inventors to fill out, with the support of technology transfer offices, the marketing annex in two languages (Italian and English), with the following logic:

- Describe what is the technical problem underlying the invention. Generally, describe the technology, its main functionalities, and what is the purpose for which it was designed and patented.
- Highlight what the main limitations of current technologies are: give immediate evidence to the reader of issues they are sensitive to and index the content to be easily found in search engines.
- Emphasize how the technology solves these limitations, creating value for the user: make the proposed solution clear and find a match with the search of the average user who, tends to be looking for a solution to a problem.
- Make pointed references to Possible Applications and Benefits: often a user's search may start with a specific application specific or a topic. These fields must represent the most suitable match with respect to the user's search habits.
- Choose two very explanatory and "talking" images of the patented technology that can easily capture the user's attention.

Currently, there are more than 1520 registered users on the KS platform, including companies, investors, banks, stakeholders and so on, which make KS the largest patent platform in Italy, the most accessed platform for patent investigation and technological transfer in Italy. In recent years, more than 250 contacts have been initiated between universities and companies and some of them have already led to signed contracts and multiple forms of collaborations (co-development agreement, new research agreement leveraging on the same know-how of a particular invention, license or option agreements, etc.).

KS data is not equally distributed among labels, a well-known phenomenon in patent analysis studies. In particular, the most populated area is “Health and Biomedical” which accounts for over 30% of the entire database, while the least one is “Aerospace and aviation” with less than 3%, see Figure 14.

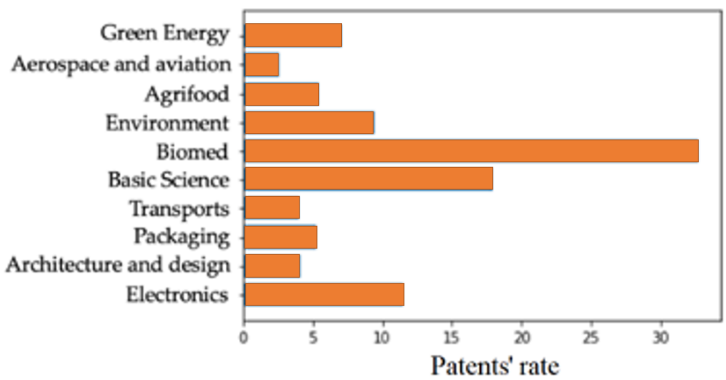


Figure 14: KS percentage distribution of patents across the different technological domains

KS was used as the experimental field during the research work. Its dataset was firstly pre-processed in order to be used for the experimentation. Natural Language Processing (NLP) techniques have been applied on the "marketing annex" of the 1694 patents, and in particular the sections "introduction", "technical features" and "application" were processed. These fields include all the relevant information about the patent without substantial redundancies.

4.2. Healthcare 4.0

According to the World Health Organization (WHO) in its Digital Health Strategy (2020-2025), digital health is described as the application of digital technologies and data to enhance health outcomes, enhance the effectiveness of health systems, and empower individuals in making informed decisions regarding their health and overall well-being. Additionally, it underscores fundamental principles including transparency, accessibility, scalability, reproducibility, interoperability, privacy, security, and confidentiality.

"Artificial intelligence has the potential to transform how care is delivered. It can support improvements in care outcomes, patient experience and access to healthcare services.
It can increase productivity and the efficiency of care delivery and allow healthcare systems to provide more and better care to more people.
AI can help improve the experience of healthcare practitioners, enabling them to spend more time in direct patient care and reducing burnout. Finally, it can support the faster delivery of care, mainly by accelerating diagnosis time, and help healthcare systems manage population health more proactively, allocating resources to where they can have the largest impact" (McKinsey & Company Report, 2020).

AI is already making an impact in health care in six major areas, illustrated in Figure 15 (McKinsey & Company Report, 2020):

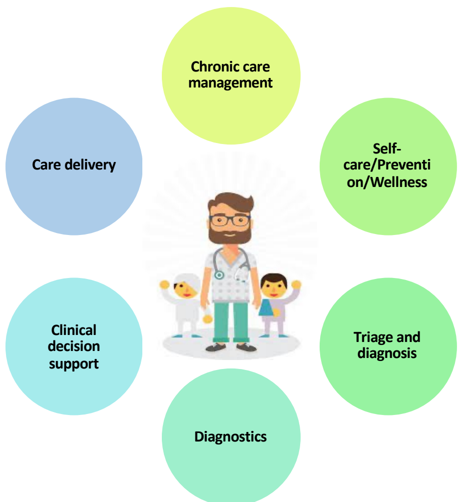


Figure 15: impact areas of AI

The success of this paradigm is reshaping modern healthcare, with promising technological, economic, and social prospects: IoT is arguably the main enabler for distributed healthcare applications, thus giving a significant contribute to the overall decrease of healthcare costs while increasing the health outcomes, although behavioral changes of the stakeholders in the system are needed (Couturier et al., 2012).

Healthcare 4.0 (HC4.0) is a recently emerged term derived from Industry 4.0 (Sannino et al., 2018; Thuemmler and Bai, 2017), used to describe the progressive emergence of typical Industry 4.0 technologies, such as Internet of Things (IoT), Industrial IoT (IIoT), cognitive computing, artificial intelligence, cloud computing, fog computing, and edge computing, applied to healthcare domain (Ayer et al., 2019). In the context of this new revolution, Cyber-Physical Systems (CPS) are shaping digital health systems involving products, technologies, services, and businesses (Yang et al., 2020; Wan et al., 2018). HC4.0 must enable stepwise virtualization to support the near real-time personalization of healthcare for patients, workers, and formal and informal janitors. This personalization of health care requires the substantial use of CPS, cloud computing, extended specialized IoT, Internet of Everything (IoE), which includes devices, services, people, and 5G communication networks (Monteiro et al., 2018).

Virtualization makes it possible to inspect small spatiotemporal windows of the real world in real time and, as a result, enables the theragnostics (Needham and Glasby, 2015; Jeelani et al., 2014) in personalized and precise medicine (Monteiro et al., 2018).

Healthcare has emerged as one of the most interesting areas for the application of IoT (Islam et al., 2015; Botta et al., 2016). The IoT is probably the main enabler for distributed healthcare applications (Couturier et al., 2012) and is therefore helping to reshape modern healthcare, with hopeful technological, economic and social perspectives; it is thus contributing significantly to the overall reduction of healthcare costs and the increase of health outcomes, if combined, however, with necessary behavioural changes of system actors (Couturier et al., 2012; Osmani et al., 2008; Aceto et al., 2020).

Market trends and scientific literature testify to the role of healthcare as a driver of the main I4.0 pillars. The IoT is being exploited for remote monitoring in all its facets, thus enabling the implementation of healthcare in various contexts, ranging from long-term elderly care and home surveillance to acute healthcare rehabilitation systems. The consequent production of large volumes of data, and thus high-speed acquisition, discovery and analysis, requires next-generation big-data technologies and architectures to extract value from them (need to move to cloud architectures) (Costa, 2014; Aceto et al., 2020).

Globally, health care systems are facing multiple challenges: increasing burden of disease, multimorbidity and disability due to aging and epidemiological transition, increasing demand for health care services, higher societal expectations and rising health care expenditures (Atun, 2015), as well as inefficiency and low productivity (Panch et al., 2018; Kocher and Sahni, 2011).

A radical transformation of health systems is essential to surmount these challenges and reach universal health coverage (UHC) by 2030. Machine learning, the most tangible manifestation of artificial intelligence and the newest growth area of digital technology, promises to achieve more with less. It could be the enabler of this transformation (Jones et al., 2012; Panch et al., 2018).

In recent years, health care costs have been rising constantly around the world. According to the World Health Organization (WHO) (2019), spending on health is growing faster than the rest of the global economy, accounting for 10 percent of the world's gross domestic product. In countries with developing economies, this upward trend is even more critical, with health spending growing at an average of 6 percent per year, compared to 4 percent in countries with developed economies (Tortorella et al., 2022).

The increase in health spending in developing economies can be linked to several reasons and represents a combination of unstoppable forces (McKinsey & Company Report, 2020):

- the increasing longevity of the population: 65 percent of the world's population aged 60 or older currently resides in developing countries; this percentage is expected to rise to 79 percent by 2050 (United Nations) (Tortorella et al., 2022);
- increase in chronic diseases, such as diabetes, heart disease, and neurological disorders, which drive up health care costs (Peltzer et al., 2014) and may be partly associated with lifestyle (Rtveladze et al., 2013) and living conditions (Arora et al., 2019);
- increased utilization of medical services and related prices: increased health care spending is often accompanied by lower levels of efficiency and productivity of health care systems (Visconti et al., 2017; Tortorella et al., 2022).

For these reasons, healthcare organizations are recapping new solutions and management approaches for the improvement of operational effectiveness and the reduction of costs (Tortorella et al., 2017; Tortorella et al., 2022).

Health systems also need a larger workforce, but although the global economy may create 40 million new health care jobs by 2030, a shortage of 9.9 million doctors, nurses and midwives is projected globally over the same period. So, there is a need not only to attract, train and retain more health care professionals, but also ensure that their time is used where it has added value, namely to care for patients (McKinsey & Company Report, 2020).

HC4.0 represents a continuous but disruptive process of transforming the entire healthcare value chain, ranging from drug and medical equipment manufacturing, hospital care, out-of-hospital care, healthcare logistics and healthy living environment, to financial and social systems.

As technology has advanced, accelerated, and converged, massive new sources of data have emerged, ranging from wearable devices to personal genomics, to information contained in our electronic health records, including a wide range of "real world" data that increasingly comes from beyond the traditional four hospital walls. The way and place we obtain, analyze, and use these data, along with the growing capabilities of artificial intelligence and machine learning, have the potential to drastically shift the practice of medicine from a fundamentally reactive system of "patient care" based on intermittent data historically collected only in the clinical setting, to a continuous, proactive, personalized, information-rich, and increasingly crowd-sourced and truly "health care"- centered system (Chang, 2020).

The final aim is the so called “P4 Medicine” (Sobradillo et al., 2011), i.e. predictive, preventive, personalized and participatory, by means of the radical change in medicine enabled by these I4.0 new technologies (Aceto et al., 2020).
The final goal is: Healthcare-as-a-Service mentality, understood in terms of offering healthcare services to patients, and testing, diagnostic and communication services to healthcare professionals. Healthcare professionals are able to provide front-office and remote consulting services to patients, thus creating a high impact in terms of time, transport and comfort for patients and being able to cover a much larger population at a fraction of the cost compared to in-person activities, resulting in an improved quality of life for patients (and some categories of professionals) and a competitive advantage for private-sector providers (Aceto et al., 2020).
Parallel to the exponential growth of enabling technologies, patent production has grown, reflecting a convergence of the three main technologies: digital communication, medical technology, and computer technology (EPO, 2022). The rise of HC4.0 goes hand in hand with advances in AI, which have reshaped the healthcare sector. Between 2017 and 2021, there was more private investment in AI in the medical and healthcare sectors than any other globally. Contextually, the great growth in computer processing capacity has helped reduce the cost and time of AI training. This has facilitated the deployment of machine learning and image data processing in many sectors, helping to gather new insights from big data. This has resulted in a sharp increase in patent activity related to the use of AI, especially in diagnostic technologies, but also in areas such as digital surgery and novel therapies (EPO, 2022).
In the past ten years, the Food and Drug Administration (FDA) has evaluated and approved an increasing array of medical devices that incorporate machine learning techniques. These devices have been legally marketed through various pathways, covering a wide range of medical specialties. The FDA anticipates that this trend will persist and expand further in the future.

Italy lands in 11th place in the Patent Index 2022-the annual global ranking compiled by Epo-with +2.5 percent of new patents, just a whisker behind Sweden, tenth with 2.6 percent, and the UK, ninth with 2.9 percent. In first place in the ranking of international patents is confirmed by the United States with 25 percent, out of a total of 193,460 applications filed. Italy, on the other hand, is fifth among the 27 countries of the European Union (EPO, 2023).
For the purposes of this thesis, it is important to focus on patent production by Italian universities and research centers, and the PATIRIS platform helps to have a permanent observatory of patenting by universities and public research institutes in Italy (PATIRIS).
It is also interesting to go and look at the main technological areas of Italy's patent assets, classified according to the IPC - International Patent Classification - which is based on an international agreement between 52 countries and 4 international organizations. The hierarchical system of patent document classification is operationally managed by WIPO (World Intellectual Property Organization) and is based on a taxonomy made of technology classes and subclasses.

The chart, in Figure 16, takes into account all patent families of Italian institutions by constructing a ranking of the most highly patented technology areas (according to IPC) (PATIRIS). The first class in the ranking is A61 - medical or veterinary science; hygiene – which included medical technologies.

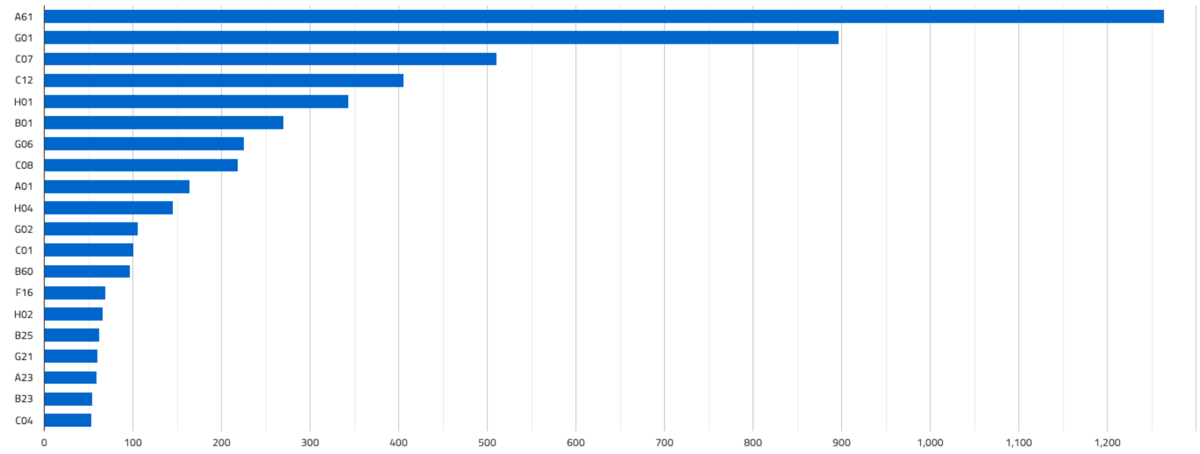


Figure 16: Ranking of the top 20 technology areas covered by the patent families of Italian research institutions.

The next chart takes into account all patent families of Italian institutions by constructing a ranking of the technology areas that have been most heavily patented over the past ten years.

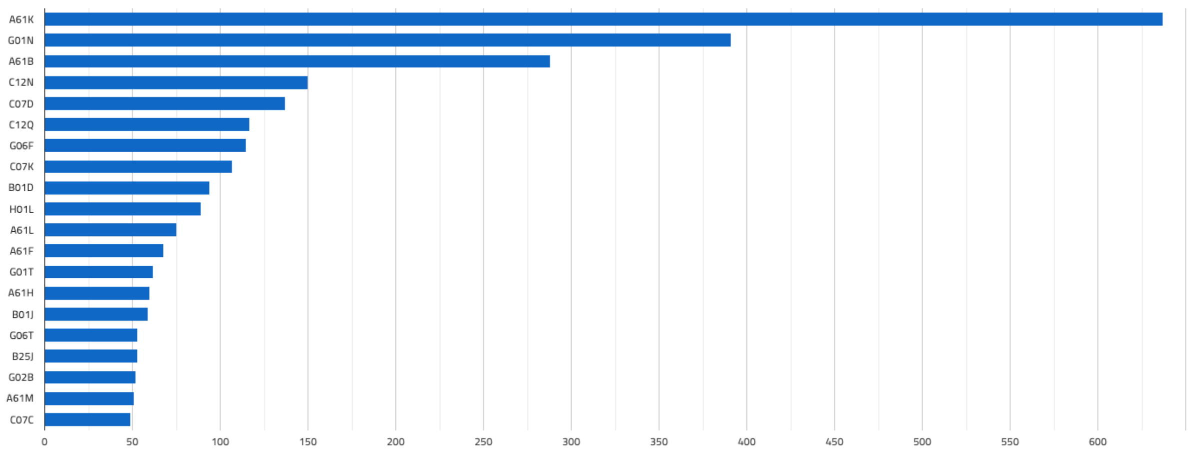


Figure 17: Ranking of the top 20 technology areas covered by the patent families of Italian research institutions for the last ten years.

It can be seen that 6 out of 20 classes are subclasses of A61:

- the first A61K is: PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES (devices or methods specially adapted for bringing pharmaceutical products into particular physical or administering forms; chemical aspects of, or use of materials for deodorisation of air, for disinfection or sterilisation, or for bandages, dressings, absorbent pads or surgical articles; soap compositions);

- the second one is A61B: DIAGNOSIS; SURGERY; IDENTIFICATION – it includes medicinal methods (non-surgical), instruments for psycho-physical tests, medical instruments, surgical instruments, devices or methods, other instruments, implements or accessories for surgery or diagnosis;
- the third one (A61L) is methods or apparatus for sterilising materials or objects in general; disinfection, sterilisation, or deodorisation of air; chemical aspects of bandages, dressings, absorbent pads, or surgical articles; materials for bandages, dressings, absorbent pads, or surgical articles;
- the fourth is A61F: filters implantable into blood vessels; prostheses; devices providing patency to, or preventing collapsing of, tubular structures of the body, e.g. stents; orthopaedic, nursing or contraceptive devices; fomentation; treatment or protection of eyes or ears; bandages, dressings or absorbent pads; first-aid kits;
- the fifth (A61H) is physical therapy apparatus, e.g. devices for locating or stimulating reflex points in the body; artificial respiration; massage; bathing devices for special therapeutic or hygienic purposes or specific parts of the body;
- the last one, introduced last years ago, is A61M: devices for introducing media into, or onto, the body; devices for transducing body media or for taking media from the body; devices for producing or ending sleep or stupor; introducing media into or onto the bodies of animals; means for inserting tampons ; devices for administering food or medicines orally ; containers for collecting, storing or administering blood or medical fluids.

This shows that the patent production of Italian research institutions is also focused on the HC4.0 sector.

In this overall scenario, the effort of this research work focuses on the impactful field of healthcare. In synthesis the reasons of this choice are essentially the following:

- the valorization of research results is even more important in this field where the advancement of research and technology can truly create an impact on society by improving people's lives, redesigning care services, implementing the capacity for emergency management, rapid information analysis, widespread archiving and future projection of diseases, increasing the monitoring of medication intake, personalization of medicine, home rehabilitation, etc;
- HC4.0 is a multidisciplinary and complex context where the knowledge/technology transfer and the creation of networks are especially important in order to develop applied research and make it quickly "usable" innovation according to continuous open innovation schemes. In particular, the involvement of all actors in the innovation chain, which in the health sector becomes even more important, represents the emblem of the quadruple

helix model where, with stakeholders from the public and private sectors and academia, a strong emphasis is placed on citizens and their needs, especially in the development of health, social and other related services. This model brings greater social benefits and empowers citizens who are not only passive consumers of content / services but take on the role of creators of innovation;

- the healthcare sector has been the first to face the impact of Industry 4.0 revolution, where the Internet of Things, Cloud and Fog Computing, and Big Data technologies are revolutionizing eHealth and its whole ecosystem, also considering that Healthcare proved to be among the most attractive areas for IoT application, effectively moving eHealth towards HC4.0;
- due to its multidisciplinary and particularly innovative nature, healthcare related patents represent generally a "monster class", multidisciplinary categories containing a significant number of poorly characterized and classified patents. Monster categories are thus ineffective, not discriminating, and difficult to explore.

CHAPTER 5. How the world top universities, evaluated according to global university rankings, perform from KT point of view

The research question addressed in this chapter is RQ1: according to global university rankings, how do the world's top universities perform from a knowledge transfer point of view? In other words, are the rankings currently most used able to characterize universities performance from third mission point of view?

Accordingly, we:

- identify and analyze three of the best-known global university rankings in order to identify the world top universities and, at the same time, evaluate the coherence between rankings;
- search and select a set of specialized KT indicators for evaluating the world top universities from the KT point of view;
- verify if the world top universities, according to the global rankings, continue to best perform from KT point of view.

And after having answered to the RQ, the final goal is to identify the best practices in knowledge transfer that can be adopted by universities that want to improve their performances.

The contents of this chapter were published on the articles:

- Demarinis Loiotile, A., De Nicolò, F., Agrimi, A., Bellantuono, L., La Rocca, M., Monaco, A., ... & Bellotti, R. (2022). Best Practices in Knowledge Transfer: Insights from Top Universities. Sustainability, 14(22), 15427.
- Bellantuono, L., Monaco, A., Amoroso, N., Aquaro, V., Bardoscia, M., Demarinis Loiotile, A., ... & Bellotti, R. (2022). Territorial bias in university rankings: a complex network approach. Scientific reports, 12(1), 4995.

5.1 Proposed methodology

In order to answer to RQ the methodology described in this section (Figure 18) was defined and followed.

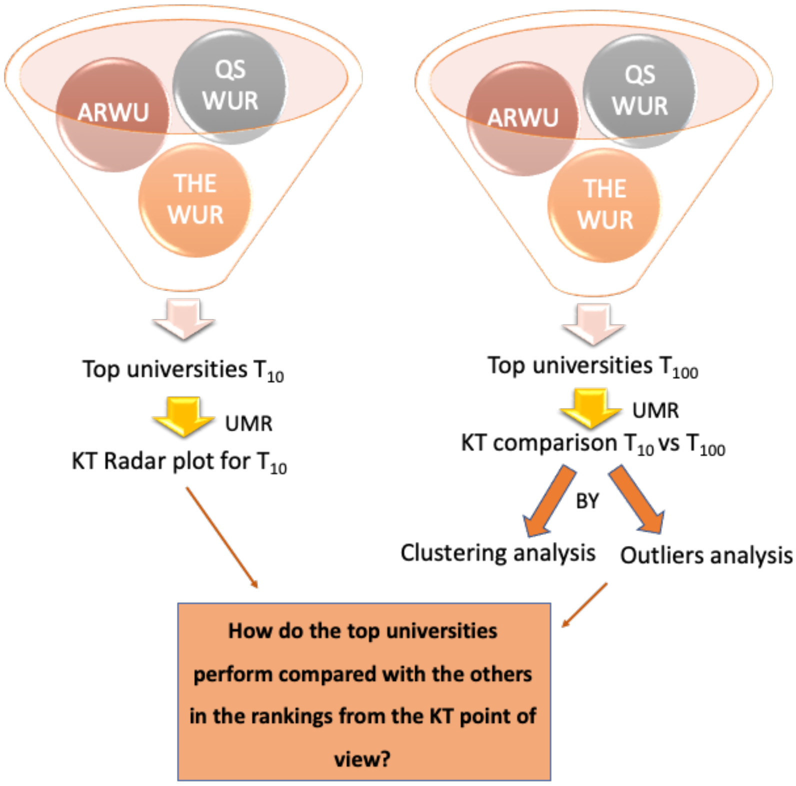


Figure 18: flowchart of methodology

The first step was to identify and analyze three of the best-known global university rankings in order to extract the world top universities and, at the same time, evaluate the coherence between rankings.
The ranking selected, already described in the previous section, are: the Academic Ranking of World Universities (ARWU), the QS World University Rankings® (QSWUR) and the Times Higher Education World University Rankings (THEWUR).
Each ranking has its own specificities; thus, to define "top universities", we considered the union Tk of all three rankings in 2020, so that:

$$T_k = T_k^{ARWU} \cup T_k^{QSWUR} \cup T_k^{THEWUR} \tag{1}$$

where T_k^{ARWU}, T_k^{QSWUR} and T_k^{THEWUR} are the top k universities in the rankings ARWU, QSWUR and THEWUR.

The sets T_{10} and T_{100} were then determined and furthermore the coherence between rankings was tested by using Spearman's correlation among the top 100 positions in QRWU, QSWUR and THEWUR.

The second step was to search and select a set of specialized KT indicators for evaluating the world top universities from the KT point of view. For this aim the UMR2020 previously described was used. Among the 9 UMR2020 indicators the following 5 were selected for being used for the T_{10} and T_{100} elaboration, this is due to the fact that they had the minimum number (less than 8%) of null value:

1) Co-publications with industrial partners
2) Patents awarded (absolute numbers)
3) Patents awarded (size-normalized)
4) Industry co-patents
5) Publications cited in patents

The five indicators were quantified for all universities included in T_{10} and T_{100} and a composite indicator called Global Performance Indicator KT (GPI KT), obtained as the average of the 5 previous indicators, was also defined and used for determining the global performance in knowledge transfer for each university. Then a comparison was made between the T_{10} obtained from the global universities ranking and the top performer universities in KT included in T_{100}, also by using radar plots for graphically expressing the macroscopic differences.

The goal of this analysis is to verify if the world top universities, according to the global rankings, continue to best perform from KT point of view.

The third step was to investigate the universities included in T_{100} in order to identify groups of similar universities, in terms of KT indicators, through a data-driven approach based on the hierarchical clustering. The goal is to understand from the natural aggregation in groups the presence of common characteristics capable of explaining the different levels of performance in knowledge transfer. Hierarchical clustering algorithms allow to group similar items in an unsupervised way (Roux, 2015). Compared with optimization-based clustering methods, such as K-means (Ahmed et al., 2020), this particular class of algorithms follows an alternative approach that entails the advantages of being deterministic and not requiring to fix the number of clusters a priori. An agglomerative hierarchy linkage algorithm was used, which starts by considering each point, corresponding to a data vector, as a cluster, and proceeds by iteratively merging the closest pairs of clusters, until ending up with one cluster that includes all data points. Vicinity of two points and is quantified by their Euclidean distance, while the distance between clusters and is evaluated as namely the minimum distance between points in the two clusters. The algorithm used allows to obtain dendrograms, which can help with the interpretation of the results.

As last step, we tried to understand how the universities included in T_{10} perform if compared to the others in T_{100} for each of the 5 UMR2020 indicators. This is in order to understand from the KT perspective what are the strengths of those universities that the global rankings identify as the best performers. For all the universities in T_{100} and for each of the 5 UMR 2020 indicators, the median absolute deviations (MAD) criterion (Robust Statistics, 2009) was used in order to compare each university in T_{10} with the distribution of the remaining universities in T_{100}. In details, we computed the scaled MAD factor for each KT indicator as:

$$\mathrm{MAD} = c \cdot \mathrm{median}\left(|T100 - \mathrm{median}(T100)|\right) \quad (2)$$

where c = -1/(·erfcinv(3/2)); then we determined whether each university in T_{10} represented an outlier for the distribution of items in T_{100}, considering three scaled median absolute deviations (MAD) away from the median as the threshold for outlier detection (Leys et al., 2013; Simmons et al., 2011).

5.2 Methodology application and results

In this section the step by step application of the methodology presented in section 5.1 is described together with the obtained results.

5.2.1 First step

The first step of the methodology is to identify the set T_{10} and T_{100}. For k=10 the set T_{10} results to be the union of the top ten universities included in each of the selected rankings: QRWU, QSWUR and THEWUR (Table 2). Similarly, the T_{100} was determined, according to eq. (1).

Table 2: The University Ranking: World Top Ten Universities

	ARWU		QSWUR		THEWUR	
Rank	University	Country	University	Country	University	Country
1	Harvard University	US	Massachusetts Institute of Technology (MIT)	US	University of Oxford	UK
2	Stanford University	US	Stanford University	US	California Institute of Technology (CALTECH)	US
3	University of Cambridge	UK	Harvard University	US	University of Cambridge	UK
4	Massachusetts Institute of Technology (MIT)	US	University of Oxford	UK	Stanford University	US

5	University of California, Berkeley	US	California Institute of Technology (CALTECH)	US	Massachusetts Institute of Technology (MIT)	US
6	Princeton University	US	ETH Zurich – Swiss Federal Institute of Technology	CH	Princeton University	US
7	Columbia University	US	University of Cambridge	UK	Harvard University	US
8	California Institute of Technology (CALTECH)	US	University College London (UCL)	UK	Yale University	US
9	University of Oxford	UK	Imperial College London	UK	University of Chicago	US
10	University of Chicago	US	University of Chicago	US	Imperial College London	UK

Thus T_{10} = {California Institute of Technology (CALTECH), Columbia University, ETH Zurich – Swiss Federal Institute of Technology, Harvard University, Imperial College London, Massachusetts Institute of Technology (MIT), Princeton University, Stanford University, University College London (UCL), University of California, Berkeley, University of Cambridge, University of Chicago, University of Oxford, Yale University} included 14 universities.

The consistency in rankings, indicating how closely they agree with each other, was assessed by calculating Spearman's correlation among the top 100 positions in each ranking. Spearman's rank correlation coefficient, also known as Spearman's ρ, is a nonparametric measure of rank correlation (Van de Wiel and Di Bucchianico, 2001). It helps us determine how well we can describe the relationship between two variables using a monotonically increasing or decreasing function. The resulting Spearman's correlation coefficients are as follows: 0.89 for the correlation between ARWU and QSWUR, 0.91 for ARWU and THEWUR, and 0.95 for QSWUR and THEWUR. These high correlation values indicate that all the rankings are strongly related to each other, demonstrating a fundamental consistency among them.

5.2.2 Second Step

In the second step we identified a set of specialized KT indicators for evaluating the world top universities from the KT point of view. Starting from UMR2020, 5 indicators were selected:

1) Co-publications with industrial partners
2) Patents awarded (absolute numbers)
3) Patents awarded (size-normalized)
4) Industry co-patents
5) Publications cited in patents.

The indicators are primarily represented as percentages, with the exceptions of two metrics: 2) Patents awarded (measured in absolute numbers) and 3) Patents awarded (size-normalized). These two metrics were converted into percentage scales to ensure comparability with the others and enable mathematical operations. Additionally, a composite index was introduced, calculated as the arithmetic mean of the five indicators from the UMR2020 dataset. This composite index, known as the KT GPI (Knowledge Transfer Global Performance Index), can be regarded as an overarching measure of universities' performance in knowledge transfer activities.

Table 3 shows, in descending order, how the universities in T_{10} perform.

Table 3: Summary of UMR2020 KT and GPI KT indicators for T_{10} Universities

	Co-publications with industrial partners	**Patents awarded (absolute number)**	**Patents awarded (size-normalized)**	**Industry co-patents**	**Publications cited in patents**	**GPI KT**
University of California, Berkeley (UCB)	7.10%	98.94%	2.00%	10.02%	2.30%	24.07%
Harvard University	8.00%	66.12%	1.45%	8.26%	3.30%	17.43%
Massachusetts Institute of Technology (MIT)	10.50%	57.72%	4.04%	5.86%	4.90%	16.60%
Standford University	9.40%	40.63%	1.68%	11.13%	2.80%	13.13%

California Institute of Technology (CALTECH)	7.60%	29.42%	10.88%	7.33%	2.10%	11.47%
ETH Zurich – Swiss Federal Institute of Technology	8.70%	6.94%	0.29%	38.91%	2.00%	11.37%
University of Cambridge	8.20%	4.86%	0.22%	25.59%	1.70%	8.11%
University of Chicago	6.80%	4.89%	0.24%	20.76%	2.00%	6.94%
Columbia University	7.80%	19.97%	0.54%	4.23%	1.70%	6.85%
University College London (UCL)	7.40%	8.30%	0.18%	14.88%	1.50%	6.45%
Yale University	6.70%	8.20%	0.51%	11.97%	2.10%	5.90%
University of Oxford	7.10%	6.38%	0.22%	12.75%	1.80%	5.65%
Imperial College London	10.20%	4.17%	0.20%	10.95%	1.90%	5.48%
Princeton University	7.30%	5.88%	0.61%	11.11%	1.20%	5.22%

The top-performing university is the University of California, Berkeley (UCB). Specifically, when looking at the first indicator, which measures co-publications with industrial partners, the Massachusetts Institute of Technology (MIT) and Imperial College London excel. When considering the indicator "Patents awarded as absolute numbers", the University of California, Berkeley is the leader, followed by Harvard University. On the other hand, if we examine the indicator "Patents awarded (size-normalized)", the top spot goes to the California Institute of Technology (CALTECH), followed by MIT.
For "Industry co-patents" the ETH Zurich - Swiss Federal Institute of Technology ranks first, followed by the University of Cambridge. Lastly, in the indicator "Publications cited in patents" MIT and Harvard University emerge as the most influential.

It is also intriguing to assess how the 14 universities in the T_{10} group compare to other universities. Using equation (1), it is determined that the cardinality of T_{100} is 151. After eliminating entries with missing or null data, the cardinality of T_{100} becomes 123. For each of the universities in T_{100}, the five UMR2020 indicators from Table 3 were computed, along with the composite indicator GPI KT. Subsequently, the list was sorted in descending order. At this point, it's noteworthy to examine how the universities in T_{10} are ranked in comparison to those in T_{100} and whether they maintain their high positions. The results are surprising. Only two of the universities from T_{10}, namely Berkeley and Harvard, are placed in the top 14 positions, while the others seem to experience a significant drop in the rankings (see Table 4). The Massachusetts Institute of Technology (MIT) appears in the 15th position. Table 5 provides detailed information on the five UMR2020 indicators and the GPI KT for the first 14 universities included in T_{100}.

Table 4. Performance of universities in T_{10} according to GPI KT

	Position in T_{10}	**Position in T_{100} according to GPI KT**
University of California, Berkeley (UCB)	1	6
Harvard University	2	14
Massachusetts Institute of Technology (MIT)	3	15
Stanford University	4	26
California Institute of Technology (CALTECH)	5	30
ETH Zurich – Swiss Federal Institute of Technology	6	31
University of Cambridge	7	45
University of Chicago	8	62
Columbia University	9	65
University College London (UCL)	10	71
Yale University	11	79

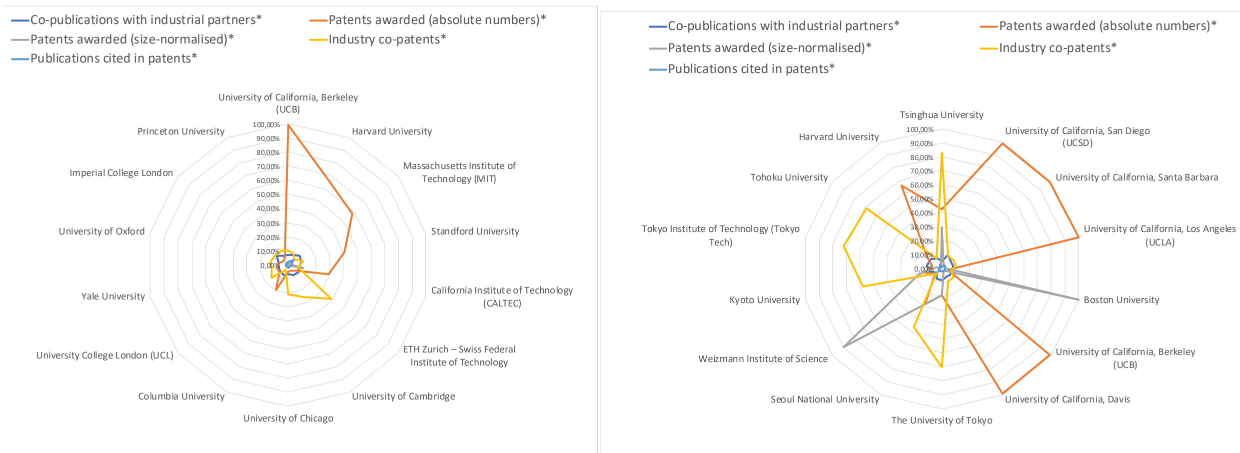
University of Oxford	12	84
Imperial College London	13	88
Princeton University	14	93

Table 5: Best performer universities in T_{100} according to GPI KT

	Co-publications with industrial partners	Patents awarded (absolute number)	Patents awarded (size-normalized)	Industry co-patents	Publications cited in patents	GPI KT
Tsinghua University	5.60%	42.18%	29.09%	82.28%	1.20%	32.07%
University of California, San Diego (UCSD)	10.60%	98.98%	2.34%	10.02%	2.70%	24.93%
University of California, Santa Barbara	8.20%	98.94%	3.30%	10.02%	2.30%	24.55%
University of California, Los Angeles (UCLA)	8.00%	100.00%	1.85%	10.15%	2.50%	24.50%
Boston University	7.90%	4.46%	100.00%	6.85%	1.80%	24.20%
University of California, Berkeley (UCB)	7.10%	98.94%	2.00%	10.02%	2.30%	24.07%
University of California, Davis	6.70%	98.94%	2.19%	10.02%	1.70%	23.91%
The University of Tokyo	8.30%	19.11%	19.36%	70.63%	1.90%	23.86%

Seoul National University	8.10%	28.13%	28.27%	46.52%	1.80%	22.56%
Weizmann Institute of Science	4.90%	6.74%	89.68%	5.12%	3.50%	21.99%
Kyoto University	8.80%	12.93%	17.01%	57.71%	1.80%	19.65%
Tokyo Institute of Technology (Tokyo Tech)	11.10%	7.24%	3.62%	72.17%	1.70%	19.17%
Tohoku University	10.20%	13.98%	0.69%	68.89%	1.60%	19.07%
Harvard University	8.00%	66.12%	1.45%	8.26%	3.30%	17.43%

We can see that the Californian universities and, to a lesser extent the Japanese and some Asian ones, stand out in the KT even though they are not among the best if we consider the traditional global university rankings.



a) b)

Figure 19: Radar plot for five UMR knowledge transfer indicators for a) the 14 universities in T_{10} b) the top performer in T_{100} according to GPI KT

The radar plot depicted in Figure 19 illustrates a notable distinction in Knowledge Transfer (KT) performance between the universities in the T_{10} group and the top universities in T_{100}. The leading universities in KT exhibit a distinct technological orientation and capability.

5.2.3 Third step

In the third phase, an examination is carried out on the universities encompassed within T_{100}. The aim is to uncover clusters of universities that exhibit similar patterns in their KT indicators through the application of the hierarchical clustering algorithm. This procedure results in the creation of four clusters, which are detailed in Table 6.

Table 6: The clusters obtained on T_{100}

N. cluster	Universities contained into the cluster
1	University of California, Santa Barbara University of California, Los Angeles (UCLA) University of California, Berkeley (UCB) University of California, Davis University of California, San Diego (UCSD) Harvard University KAIST - Korea Advanced Institute of Science &Technology
2	KU Leuven University of Toronto Boston University Weizmann Institute of Science The University of Queensland Université Grenoble Alpes The University of New South Wales (UNSW Sydney) The University of Melbourne University of Copenhagen Aarhus University

	The Hong Kong University of Science and Technology Sungkyunkwan University (SKKU) The University of Tokyo Korea University Kyoto University Seoul National University Tsinghua University
3	Stanford University California Institute of Technology (CALTECH)
4	Universities in the rest of the world

In Figure 20 the dendrograms resulted by the analysis are reported.

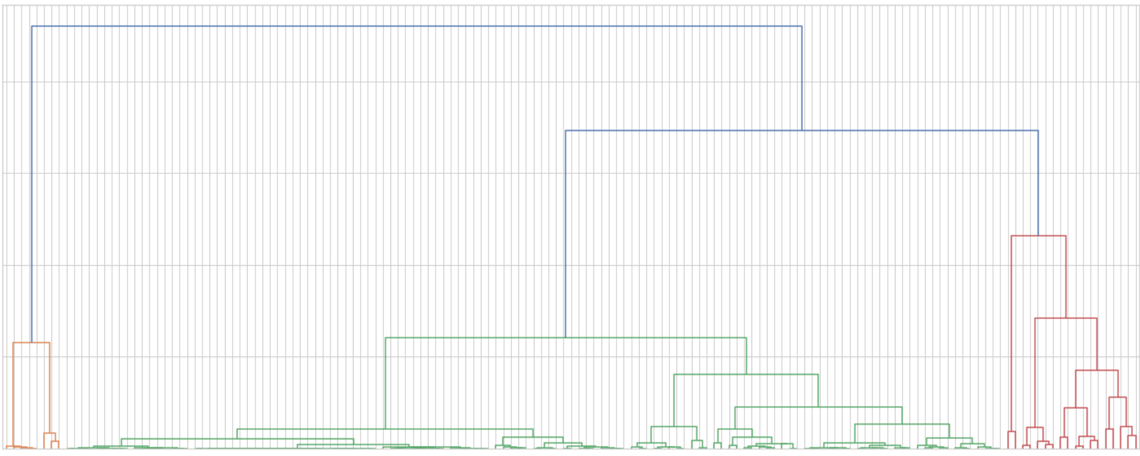


Figure 20: The dendrograms resulted by the hierarchical clustering on T_{100}

Among the top 14 universities found within the three identified clusters (excluding the fourth cluster, representing the rest of the world), only 4 out of 14, which amounts to 29%, fall into this category. Specifically, MIT and ETH are notably absent, as are institutions like Imperial College London, Cambridge University, Princeton University, University of Chicago, Columbia University, Yale University, University of Oxford, and University College London. Of particular interest is the formation of a distinct cluster (cluster n.3) comprising Stanford University and CALTECH. Furthermore, the emergence of another cluster (cluster n.1), consisting primarily of California-based universities, encompassing 71% of the cluster, along with Harvard University and the Korean KAIST, is a noteworthy observation and aligns in part with the findings from section about second step. These clustering results are consistent with the rankings presented in Table 5, demonstrating alignment with the GPI KT. In fact, the highest-performing institutions are concentrated within clusters n.1 and n.2. The presence of California universities in both analyses suggests that the surrounding industrial, social, and economic environment can indeed influence the third mission activities of these institutions.

5.2.4 Fourth step

The objective of the fourth step, guided by the median absolute deviations (MAD) criterion as outlined in formula (2), is to assess the performance of universities within the T_{10} group in comparison to those in the broader T_{100} category across all five UMR2020 indicators. Figures 21, 22, 23, 24 and 25 help show the difference between the MAD value of each KT indicator for the 14 top universities in T_{10} and the outlier threshold: a positive difference is indicated with a red bar and denotes that the corresponding university is classified as an outlier with respect to the distribution of the specific KT indicator of the T_{100} universities. In Figure 21, with respect to the "Co-publication with industrial partners" indicator, 7 (bar in red) over 14 universities in T_{10}, result to perform much better than the remaining ones. Figure 22, that refer to the indicator named "Patents awarded (absolute numbers)", point out that only 3 universities result to outperform with respect to the remaining ones. Finally, figure 25 illustrates the result for the "Publications cited in patents" indicator and points out that only 4 universities in T10 outperform. The figure 23 and 24 show that there are no outperformers universities.

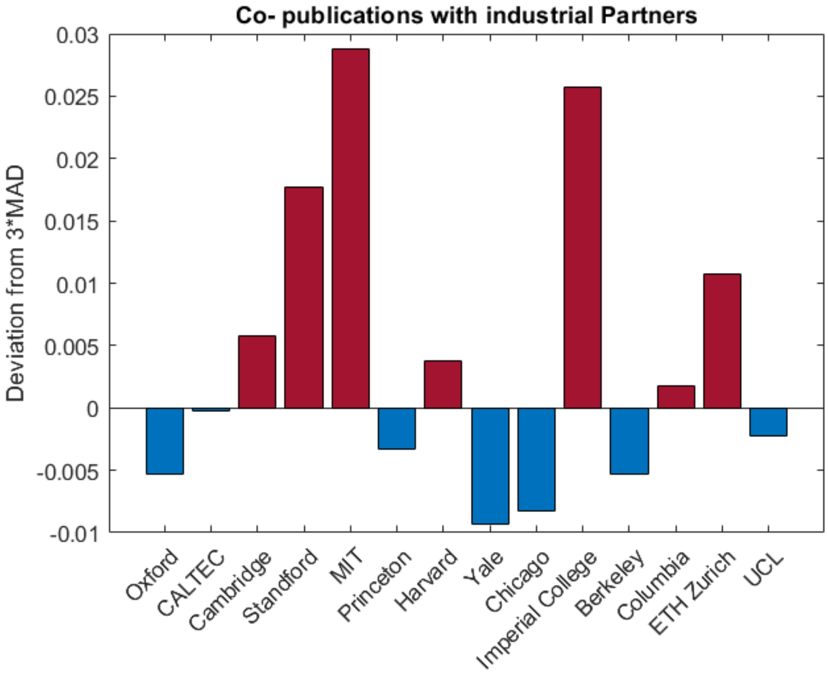


Figure 21: Difference between the value of the indicator "Co-publication with industrial partners" for the 14 top universities in T_{10} and the outlier threshold

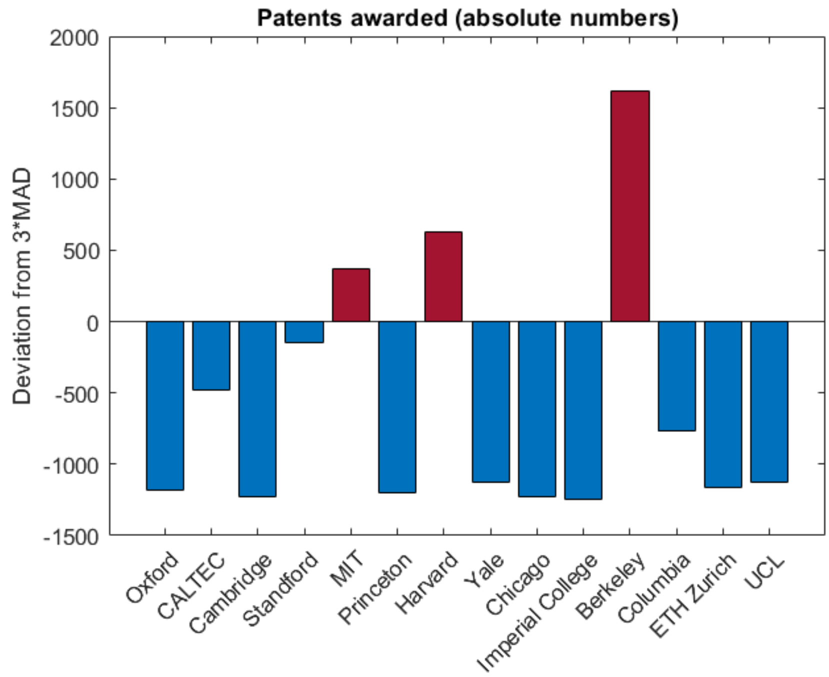


Figure 22: Difference between the value of the indicator "Patents awarded (absolute numbers)" for the 14 top universities in T_{10} and the outlier threshold

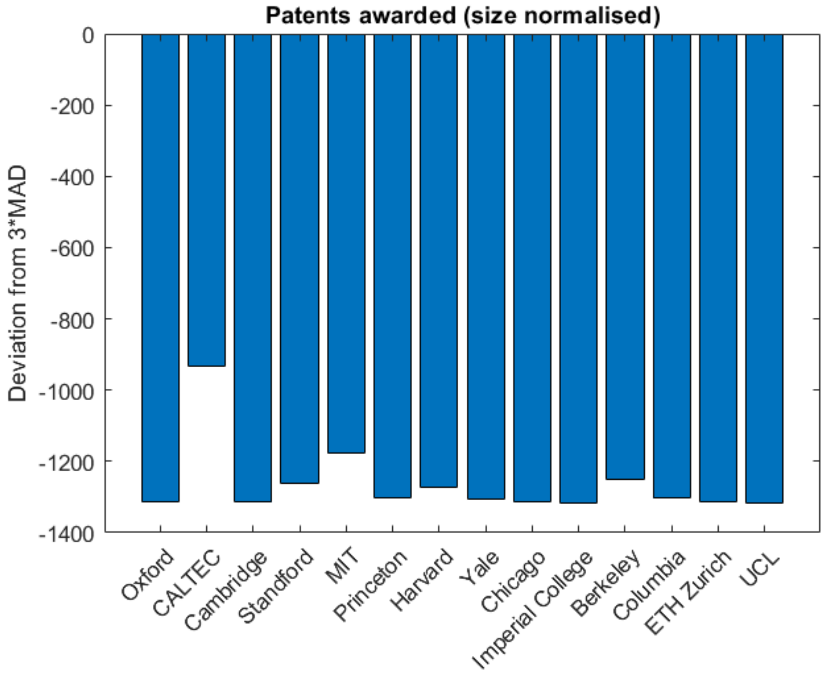


Figure 23: Difference between the value of the indicator "Patents awarded (size normalized)" for the 14 top universities in T_{10} and the outlier threshold

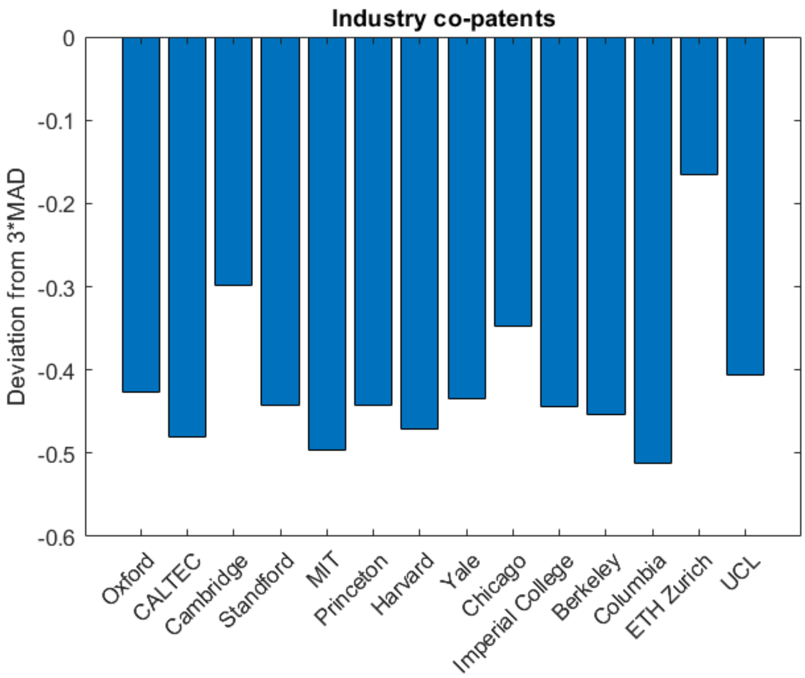


Figure 24: Difference between the value of the indicator "Industry co-patents" for the 14 top universities in T_{10} and the outlier threshold

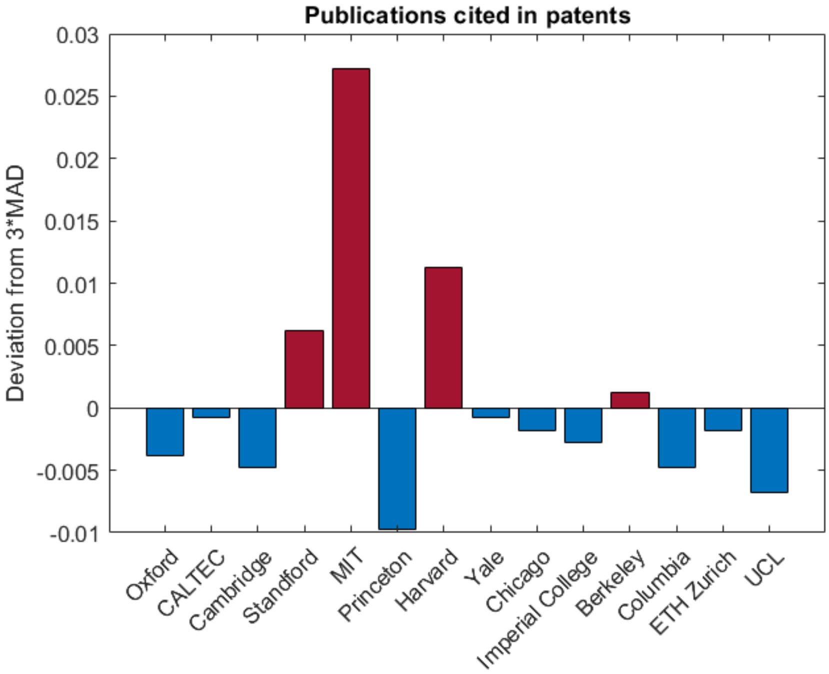


Figure 25: Difference between the value of the indicator "Publications cited in patents" for the 14 top universities in T_{10} and the outlier threshold

5.3 Discussion

The findings obtained are far from straightforward and, in certain instances, quite unexpected. The clustering analysis indicates that the top 14 universities globally, as determined by global university rankings, do not form an isolated or distinctive group when considering knowledge transfer and, more broadly, third mission activities. With the exception of Stanford University and CALTECH, which constitute a distinct cluster, and Harvard and Stanford presence within the California cluster, the remaining top universities do not exhibit a significant deviation from the others when examining the available KT indicators.

The analysis based on the median absolute deviations (MAD) highlights that three indicators seem to be particularly effective in assessing universities' performance in KT activities: Co-publications with industrial partners, Patents awarded (measured in absolute numbers), and Publications cited in patents. This discovery is particularly valuable since these aspects have received comparatively less attention in previous research (Rossi and Rosli, 2015). The significance of these findings should be understood within the context of similar observations made in the academic literature from a knowledge transfer perspective.

Regarding "Co-publications with industrial partners," it is worth noting that in recent years, collaborations between academia and industry have seen significant growth. These collaborations, which play a pivotal role in research, development, and innovation within the framework of the third mission (Piirainen et al., 2016), have taken on diverse forms and generated various outcomes that can sometimes be challenging to quantify (Kohus et al., 2020).

Several indicators can be used to measure the outputs of the university-industry collaboration, such as patents/intellectual property rights, publications, and learning metrics (Perkmann et al., 2011). Among these, the analysis of the co-publications with industrial partners is used as an explicit proxy for evaluating university-industry collaboration (Kohus et al., 2020; Tijssen, 2011; Tijssen et al., 2009; Giunta et al., 2016; Levy et al., 2009) and for studying the university's entrepreneurial orientation (Tijssen, 2006). Collaborative publications between universities and industry, where researchers from both academia and business enterprises co-author papers, can play a crucial role in addressing the challenge of limited publicly accessible information. This challenge often relates to factors like the number of research contracts, the extent and nature of joint industry projects, and the count of licenses issued (Yegros-Yegros et al., 2016). Furthermore, some studies have demonstrated that these university-industry collaborative publications exert a significantly positive influence on universities' outcomes in terms of technology commercialization. This impact is observable in various aspects, including increased patenting activity, the establishment of spin-off companies, and enhanced technology licensing activities (Wong and Singh, 2013).

Historically, there was a prevailing belief that research publications involving corporate collaboration held less significance compared to those involving academic partners. However, a global examination utilizing a field-weighted citation impact metric has convincingly debunked this notion. In fact, it has been established that publications resulting from collaborations with industry partners tend to have a higher citation impact, effectively dispelling the "urban myth" that such collaborations receive less respect ("The urban myth of less respect for collaborating with industry is busted") (University-industry collaboration, 2021).

However, co-publications with industrial partners, as well as other quantitative measures, are still a long way from being regarded as perfect measures of university-firm collaboration (Yegros-Yegros et al., 2016). Relying solely on the sheer number of publications is an insufficient and unreliable method for gauging the effectiveness of university-industry collaboration, as emphasized by Seppo and Lilles in 2012. It's important to acknowledge that not all research collaborations result in co-publications (Katz and Martin, 1997). Moreover, while co-publications with industrial partners do provide a valuable new source of data for assessing the interaction between industry and academia, it is crucial to apply this data judiciously. Tijssen et al. (2009) recommend using it primarily within non-evaluative, multidimensional benchmarking frameworks for the purpose of domestic and international comparisons of research universities. This approach is more appropriate than attempting to construct traditional university league tables based solely on these metrics.

About “Patent awarded”, it has always been used as an indicator of invention (Finne et al., 2009), a valuable estimator of technology development (Huang et al., 2003), an indicator of KT outputs or a signal of capability of exploitation and commercialization of research results

(Finne et al., 2011). Eurostat reports: “A count of patents is one measure of a country’s inventive activity and also shows its capacity to exploit knowledge and translate it into potential economic gains” (Archive: Patent statistics). Although patents have grown over time and are a tool to create economic profit, there is still a disproportionately small number of real cases of technology transfers (Choi et al., 2015). The universities that stand out for this indicator are generally small and private, like Harvard and MIT, or very specialized and home of numerous Nobel Prize winners as Berkeley.
Lastly, regarding "Publications cited in patents," it's worth noting that several studies have highlighted their pivotal role in establishing a bridge between scientific knowledge and technological applications (Hammarfelt, 2021). In essence, they serve as a representation of the knowledge exchange between structured scientific information (scientific papers) and structured technological knowledge (as reflected in patents), as elucidated by Yamashita in 2018. As a result, the indicator "Publications cited in patents" is gaining increasing prominence as a statistical parameter (as seen in the OECD Science, Technology, and Industry Scoreboard 2015). It can be interpreted as a valuable measure of Knowledge Transfer (KT) efficiency, essentially reflecting the pipeline from research to practical application and market exploitation.
Excluding Patent awarded (size-normalized) that do not allow the distinction between the 14 top universities and the others in the rankings, the other indicator, Industry co-patents, probably suggests that patenting in partnership between universities and businesses is not the most useful and efficient form of collaboration. The indicator "Industry co-patents" tends to spark debate and controversy. There are several reasons for which patenting is not considered an optimal KT indicator: the poor understanding of the needs of the market by academics, their need for publication (publish or perish), very cumbersome academic procedures for patenting, unrealistic royalties, the company’s necessity to take high risk and large investments to bring the technology to the market (Hamano, 2018). Li et al. stated that the co-ownership has a negative impact on patent commercialization: industry-academia patents are less probable to be commercialized (Li et al., 2021). Cerulli et al. focalized their attention on the impact of academic patents on firm’s performance and they stated that there is a positive impact on market power but a lower profitability (Cerulli et al., 2021). Certainly, academic scientific research is beneficial to industry because allows it to enlarge their capability to explore and develop new solutions and technological fields (Peeters et al., 2020), however there is still dramatically limited empirical evidence on the impact of academic patents on business performance (Cerulli et al., 2021).

In attempting to address the research question of "how do the world's top universities, as assessed by global university rankings, fare in terms of knowledge transfer," through an examination of the five knowledge transfer indicators sourced from the UMR2020 dataset and synthesizing the outcomes derived from GPI KT calculation, clustering analysis, and MAD-based analysis, it is feasible to assert that:

- only 4 universities in T_{10} are clearly present in the identified clusters: Stanford University, CALTECH, University of Berkeley (UCB) and Harvard University;
- only 2 university over 14 (MIT, Harvard) result to be stand out with respect to the remaining ones if all the three indicators, “co-publications with industrial partners”, “patent awarded (as absolute number)” and “publications cited in patents”, are jointly used;
- by combining the two previous results, it is possible to state that only 5 over 14 universities in T_{10} (Berkeley, Stanford, MIT, Harvard, CALTECH) exhibit a high-level performance (they are included in the first 30 position over the 123) in KT if compared with the remaining universities in T_{100};
- a comprehensive analysis of the findings reveals a lack of a coherent and unequivocal interpretation. The third mission and the intricate process of knowledge transfer occurring within a region can manifest in diverse ways that defy straightforward quantification using simplistic indicators. The multifaceted nature and intricacy of knowledge transfer are such that they cannot be distilled into a limited set of significant elements, and it's conceivable that the quantitative data gathered may require supplementation with qualitative and contextual information.
- it is evident that the most widely recognized global university rankings may fall short in adequately capturing the essence of third mission activities. While they excel in assessing other university missions, notably teaching and research, their effectiveness appears to wane when applied to the realm of third mission endeavors.

To better grasp the key factors contributing to their success and to gain insights into best practices for enhancing knowledge transfer performance, we'll provide a brief overview of the five universities mentioned, specifically focusing on their technology transfer service organization, business relationships, entrepreneurship programs, and broader third mission activities. In a broader context, these universities demonstrate a remarkable ability to bridge the gap between research outcomes and the market. They effectively motivate businesses and stakeholders to engage in collaborative efforts with them. Essentially, as depicted in Figure 26, they have established organizational structures, methodologies, and approaches that enable them to navigate the challenging terrain of the Valley of Death (Hudson and Khazragui, 2013; Hockaday, 2020). These universities serve as examples to emulate and offer an inspiring repository of best practices that others can adopt.

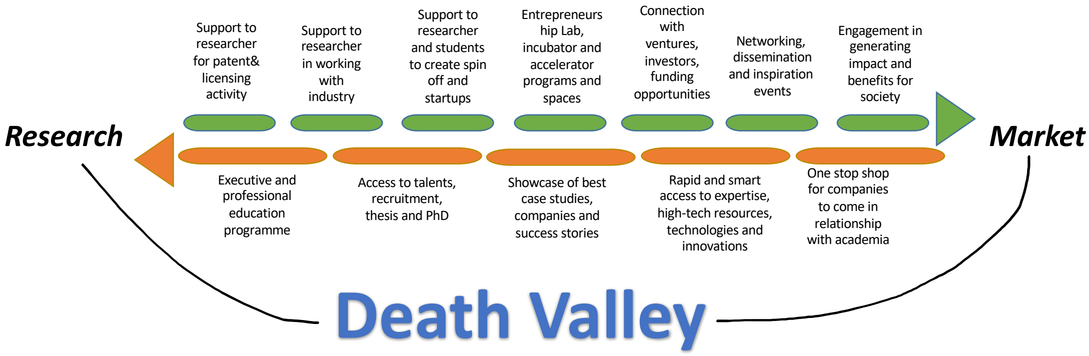


Figure 26: The steps on a bridge in order to cross the Valley of Death: the research towards the market (green steps) and the market towards the research (orange steps)

Harvard University. At Harvard University, the Office of Technology Development (OTD) connects innovators with industry partners, giving support to researchers/innovators in the advancement of their research through corporate partnerships, collaborations, and accelerator programs; the Office helps with the protection of IP to create a clear path forward for commercial development, with business development strategies for licensing or for the creation of new companies. The Office also provides support for industry partners by offering a single point of entry for engaging with Harvard researchers, accelerators, technology licensing, and new ventures. In order to bridge development gaps, Harvard University offers Accelerator programs that combine funding strategies, technical support, and business expertise to help promising innovations make the leap from the lab to the commercial sphere. Regarding entrepreneurship, OTD's Entrepreneurs in Residence (EIRs) engage directly with Harvard research groups to help advance technologies toward the launch of a startup.
Finally, the section "Impact" reports the number of new startup companies with their impact on society in terms of education, health care, food and agriculture, energy, sustainability, high-tech goods, and much more.

Massachusetts Institute of Technology. The Massachusetts Institute of Technology (MIT), one of the most vibrant hubs of innovation and entrepreneurship on Earth, dedicates a large part of its web site to the topic "Innovation". MIT's TLO (Technology Licensing Office) is engaged in the cultivation of an inclusive environment of scientific and entrepreneurial excellence and bridges connections from MIT's research community to industry and startups by strategically evaluating, protecting, and licensing technology.
A separate website is dedicated to MIT's Industrial Liaison Program (ILP) that is "industry's most comprehensive portal to MIT, enabling companies world-wide to harness MIT resources to address current challenges and to anticipate future needs". Nowadays more than 800 of the world's leading companies collaborate with MIT researchers and together bring knowledge to bear on the world's great challenges.

MIT Corporate Relations, the organizational parent of the ILP at MIT, is dedicated to finding connections to MIT faculty, departments, labs, and centers.
Great emphasis is placed on entrepreneurship with Martin Trust Center for MIT Entrepreneurship that seeks to advance knowledge and educate students in innovation-driven entrepreneurship by providing proven frameworks, courses, programs, facilities, and mentorship, and with the Program “Entrepreneur in Residence (EIR)”, a centerpiece of the Trust Center, where accomplished business leaders advise students on the challenges and benefits of startup life. MIT Startup Exchange is a program of MIT Corporate Relations that actively promotes collaborations and partnerships between MIT-connected startups and industry, principally ILP members.

Stanford University. At Stanford University, the Office of Technology Licensing (OTL) receives invention disclosures from Stanford faculty, staff and students and evaluates them for their commercial possibilities, and when possible, license them to industry. The office supports researchers providing numerous guides, for instance the “Inventor’s Guide” or the “Researcher’s Guide to Working with Industry”.
Great relevance is given to the concept of IMPACT; in fact, every year, OTL drafts an annual report where the number of issued patents, executed technology licenses, formed startups, and the amount of license income generated are reported.
The office “University Corporate and Foundation Relations" is a central university office that helps to foster relationships between Stanford University, companies and private professional foundations. For corporations, there are engagement opportunities to collaborate with Stanford University, to connect to and recruit students, and to get executive education.
In the framework of Professional Education, Stanford University has the “Innovation and Entrepreneurship (SI&E) Certificate Program” to learn innovation and entrepreneurship as practiced at Stanford and in the Silicon Valley, and the “Stanford Idea-to-Market (I2M) course” to learn tools, techniques and real-world expertise to make a business idea a reality.

University of California, Berkeley. The Office of Intellectual Property and Industry Research Alliances (IPIRA) provides a "one-stop shop" for industry research partners to interact with the campus. IPIRA's mission is to establish and maintain multifaceted relationships with private companies, and thereby enhance the research enterprise of Berkeley campus. IPIRA has promulgated technology transfer that generated billions of dollars in revenue and has created IP policies to promote social impact. Noticeably, the “Socially Responsible Licensing Program” serves as the gold standard for universities in the public health space.
About entrepreneurship, UC Berkeley helps students, faculty, researchers, and other innovators access a deep, interconnected ecosystem of resources for educating entrepreneurs, commercializing research, and advancing startups. The office supports entrepreneurs looking for funding, legal services, start-up guidelines, connection with ventures and so on. The

section “Berkeley Startups” provides a partial list of companies born out of licensing UC Berkeley IP rights.

CALTECH. In the California Institute of Technology, the Office of Technology Transfer and Corporate Partnerships (OTTCP) has the mission to drive the transfer of scientific and engineering knowledge created by the researchers to “maximize societal impact by developing partnerships with industry through the creation of new ventures, collaborations with corporations, and transfer of IP while nurturing an entrepreneurial environment” (https://innovation.caltech.edu/). The Istitute’s homepage showcases the following sections: “Corporate Partnerships” for productive collaborations and long-term partnerships with industry partners, in order to accelerate progress towards shared goals; “New Venture Creation & Entrepreneurship” illustrating the support to the formation of startup companies based on Caltech and JPL technologies by maintaining close relationships in the entrepreneurial community; “Patents & Licensing” whose goal is to make the technology transfer process and working with industry as easy as possible.
In a section dedicated to start ups, the program ”Entrepreneurs in Residence” is illustrated, i.e., a path to Launching a Venture, the Caltech Funding Sources, the Entrepreneurship Resources and so on.
Section “Impact”, with pay-off “Pushing interdisciplinary boundaries in the service of discovery”, described the inventions made by Caltech researchers since its founding. In the same page the “Impact Report” and the “CALTECH Impact” report the outsized impact on science, technology, and society: its numbers convey the extent of the impact generated by research innovations, commercialization activities, and overall output of invention disclosures, patents, licenses, and startup companies. OTTCP teams connect companies, industry leaders, and other financial partners to Caltech and to the Jet Propulsion Laboratory's research communities and help identify the types of strategic partnerships and opportunities that can advance business and investment goals.

5.4 Conclusion

This research work proposes a four steps methodology for answering the following research question: “how do the world top universities, evaluated according to global university rankings, perform from a knowledge transfer point of view?” and, as direct effect, to point out the success factors and best practices that can be adopted for improving performances in knowledge transfer.
Starting from the top universities in the most important global universities ranking (ARWU, QSWUR and THEWUR), this study delves into their performance from the perspective of third mission activities, knowledge transfer, business engagement, and entrepreneurship. To accomplish this, a set of specialized Knowledge Transfer (KT) indicators sourced from U-multirank 2020 was carefully chosen and employed to assess the achievements of these top-tier universities. The outcomes of this comparative analysis reveal that the prevailing global

university rankings often fall short in adequately assessing knowledge transfer and third mission endeavors. In other words, the universities that consistently secure top positions in global rankings do not invariably exhibit equally outstanding performance from a third mission standpoint. Intriguingly, when evaluated against the specific U-multirank 2020 knowledge transfer indicators, only three of the highest-ranked universities globally emerge as leaders in this domain. Among the top 30 universities excelling in Knowledge Transfer, a mere five belong to the elite T_{10} group of universities.
The research work, in an attempt to discern potential attributes that distinguish top universities concerning KT, investigates the sample through a hierarchical clustering algorithm. The results obtained do not show a particular relationship between the clusters obtained and the top universities of global university rankings. On the other hand, the composition of the clusters is interesting. In some cases, it appears to be based on basis of geographical position (such as the presence of numerous Californian universities in the same cluster), thus suggesting that there are contextual factors that the purely quantitative analysis used by global university rankings fail to grasp or bring out. Finally, the analysis based on the MAD indicator helps to identify three indicators that best help assess the performance of universities in terms of KT activities: Co-publications with industrial partners, Patents awarded (absolute numbers), Publications cited in patents. The work also tries to explain, through a targeted bibliographic analysis, why these indicators are interesting and useful for interpreting the performances in the KT.
This research endeavors to contribute a modest yet meaningful advancement in the academic exploration of university evaluations in terms of technology transfer and the third mission. This evaluation remains a complex task, given the limited tools and indicators currently available, as indicated by previous authors (Olcay and Bulu, 2017; Rossi and Rosli, 2015). In fact, KT indicators are poorly considered in the most popular global university rankings (the QS World University Rankings, the Academic Ranking of World Universities, and The Times Higher Education) except for U-multirank (Dip, 2021).
Several attempts have been made to develop patterns of KT indicators at the European level (Campbell et al., 2020; Hockaday, 2020), as illustrated in Chapter 2. Often the difficulty lies in being able to measure only what is measurable, i.e. Numerous activities still elude measurement and quantification, especially those related to Knowledge Transfer (KT) that unfold through unintentional mechanisms (Azagra-Caro et al., 2017). For example, learning, contacts, friendship, networks, impact, reputation and publicity are "non-monetary currencies" (Hockaday, 2020). In this regard, there is also the scientific debate about the measurement of the impact and benefits of the KT activities on society (e.g. the UK REF Impact Case Studies).
On the other hand, the challenge of evaluating universities from the third mission point of view is becoming increasingly relevant, since the topic of the impact of the third mission activities and research results on society and the territory is becoming more and more central, even in international policies. In general, governments and politicians are interested in evaluating outcomes of public investment in research and maximise the impact of R&I

investment (Council Recommendation (EU), 2022), university actors want to demonstrate their contribution to society and different types of stakeholders (Campbell et al., 2020), and industry uses university evaluation in order to assess who to partner with.

In the following chapter, the attention and the analysis are focused on one of the KT indicators emerged: the number of granted patents, because of the availability of data and the richness of information contained. For these motivations, the research activity focused mainly on patent analysis and patent matchmaking platforms.

CHAPTER 6. Classification taxonomies used in patent platforms within the academic patent landscape

In order to answer RQ2 - are the classification taxonomies used in the patent platforms effective in classifying the whole landscape of academic patents? - we performed the following steps:

- pre-processing phase through NLP realized on the patents data contained in the Knowledge Share platform;
- construction of the Matrix TF_IDF;
- application of Singular Value Decomposition (SVD) for dimension's reduction;
- K-means clustering algorithm;
- grid-search exploration of the parameters' space;
- Silhouette algorithm for optimization.

The ultimate goal is to understand whether the classification proposed by KS in 10 technological areas is sufficient to capture the whole landscape of Italian academic patents or the content of the patents suggests a different classification.

The contents of this chapter were published in:
Demarinis Loiotile, A., De Nicolò, F., Monaco, A., Tangaro, S., Loccisano, S., Conti, G., ... & Bellotti, R. (2023). Innovations and Emerging Technologies: A Study of the Italian Intellectual Property Knowledge Database. DOI: 10.5220/0011627000003393 In Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 2, pages 75-86 ISBN: 978-989-758-623-1; ISSN: 2184-433X Copyright c 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0).

6.1 Proposed methodology

The proposed approach for patent analysis and its application to KS is illustrated in Figure 27:

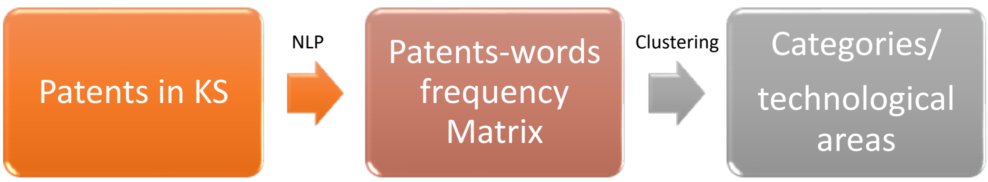


Figure 27: proposed workflow for KS patent analysis

As already explained in section 3.3.1, we performed some of the main steps of NLP on the "marketing annex" of the 1694 patents, and in particular on the sections "introduction", "technical features" and "application", using python library NLTK, in the following order:

- Tokenization.
- Stop-words removal.
- Stemming.

NLTK is a leading platform for creating Python programs for working with human language data. It provides a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, semantic reasoning and more, as well as easy-to-use interfaces to over 50 corpora and lexical resources.

Once all of the above steps were completed and text had been cleaned up and processed, a final step was to combine it all into a simple, generalized function to be executed on the text. Thus, it was possible to reconstruct the treated text. The final result was a list of 1.694 texts "cleaned up and processed" of the Italian patents contained in the KS platform.

6.2 Methodology application and results

After the pre-processing phase through NLP performed on the patents data contained in the Knowledge Share platform, the Matrix TF_IDF was constructed in order to obtain a Term Frequency-Inverse Document Frequency matrix constructed in this way (Figure 28):

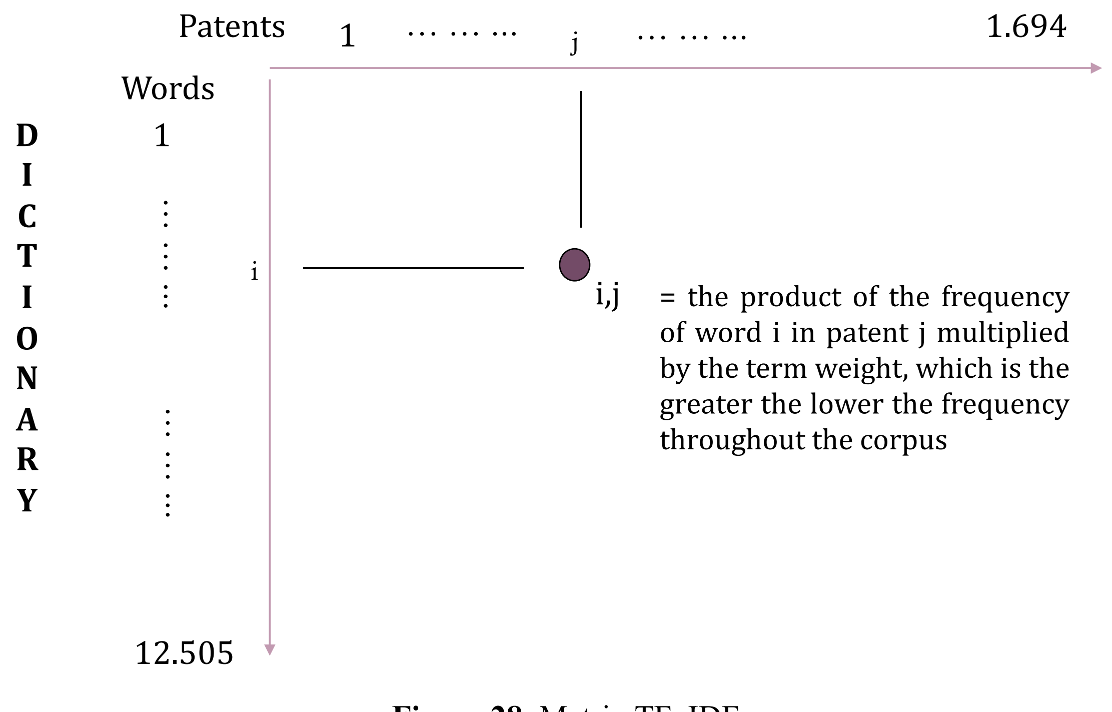


Figure 28: Matrix TF_IDF

The generic element of this matrix is a product of two terms:

$$(TF - IDF)_{ij} = (TF)_{ij} \times (IDF)_{ij};$$

where

$$(TF)_{ij} = \frac{f_{ij}}{\sum_k f_{ij}} \quad ;$$

$$(IDF)_{ij} = log \frac{N}{\left|\{d \in D \middle| w_j \in d\}\right|} \quad ;$$

where:

- f_{ij} = frequency of word j in patent i;
- N = total number of patents in the corpus;
- D = set of patents, so that |D| = N;
- w_j = j – th word;
- d = document in D.

We obtained a matrix of 1.694 rows (patents) and 12.505 columns (words).
The elements of this matrix indicate the frequency of occurrence of each term in each document. The rationale behind this approach can be summarized as follows: the first factor rewards words that appear frequently within a patent, signifying that words cited more often carry greater importance. Conversely, the second factor penalizes words that have high frequency across the entire set of patents because a word used in all patents would contribute little to discrimination. This emphasizes the significance of terms that occur infrequently.
Generally, in this type of matrix there are two problems: too high a number of features because the size of the rows (terms) is greater than that of the columns (patents); sparseness problem due to the fact that many of the elements of the array have zero value (Jun et al., 2014). Therefore, although the matrix is suitable for statistics and machine learning, it is difficult to analyze it because it has a very sparse data structure.
Accordingly, in order to reduce the sparsity and make the clustering process less prone to the curse of dimensionality, we applied the Singular Value Decomposition (SVD) (Abdi, 2007) for dimension's reduction. Finally, we used the K-means clustering algorithm (Trappey et al., 2017; Bock, 2007; Trappey et al., 2013) and we performed a grid-search exploration of the parameters' space and used the Silhouette (Jun et al., 2014) for optimization.
K-means is one of the most popular and best performing clustering algorithms. Despite this, it is a very simple algorithm to implement and use. K-means is based on so-called centroids. The centroid is a point belonging to the feature space that averages the distances between all the data belonging to the cluster associated with it. It therefore represents a sort of center of

gravity of the cluster and in general, due to its characteristics, it is not one of the points of the dataset. The K-means clustering algorithm is very simple:

- Not knowing the classes present in the input dataset, the first thing to do is decide the number of classes (or rather clusters, in this case) into which you want to divide the dataset itself. This number is called K, hence the name of the K-means method (the term means implies the use of centroids, i.e. midpoints).
- K centroids belonging to the feature space are randomly selected. The only condition is that they are not coincident, indeed usually we make sure that they are far enough from each other. Otherwise, the algorithm may have problems converging.
- The distance of each point of the dataset with respect to each centroid is calculated.
- Each point in the dataset is associated with the cluster connected to the closest centroid.
- The position of each centroid is recalculated by averaging the positions of all the points of the associated cluster (only of these points!).
- Iterates from step 3 until there are no more cluster-changing inputs.

Since both SVD and K-means depend on user-defined parameters, optimal values for these parameters must be chosen. In particular, since SVD shrinks the dimension of matrices by using linear combinations of columns, it must be decided the optimal number of these linear combinations; for what concerns k-means, the most important parameter to choose is the number of clusters to retrieve. Accordingly, we performed a grid-search exploration of the parameters' space and used the Silhouette (Jun et al., 2014) for optimization.

Silhouette comprises two components: (1) the average distance of each point within a cluster to other points within the same cluster, and (2) the average distance of each point within a cluster to all the points in different clusters. However, given the various distance functions available for computing Silhouette, we opted to assess three different distance measures (Euclidean, Manhattan, and Cosine) and selected the function that maximizes Silhouette for a given set of SVD and K-means parameters. Therefore, the grid-search explored three quantities:

(1) the number of linear combinations in SVD - word combinations interval between 10 to 1694 (step of 10);
(2) the number of clusters in K-MEANS - the number of clusters from 2 to 10;
(3) the distance function in Silhouette - the most common distance metric among the Euclidean distance, the Manhattan distance and cosine distance.

The combination having the maximum value of Silhouette was the following:

- SVD - number of linear combinations 10,
- K-means - number of clusters 8,
- Silhouette - cosine distance.

Thus, the results obtained are based on this configuration of parameters.

Figure 29 shows that, according to Silhouette, 8 clusters should be considered.

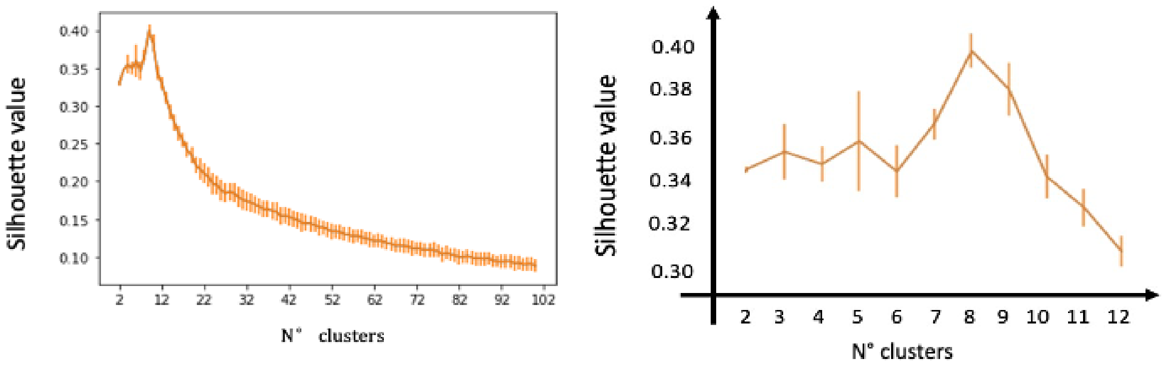


Figure 29: The variation in Silhouette value as a function of the number of clusters (the second one shows a magnification)

To fully answer the research question, we examined each cluster trying to evaluate whether there was a consistent overlap with the KS classification. Besides, we examined the keywords (in terms of frequency) for each cluster, and we assigned to each cluster a coherent although subjective classification. Table 7 shows the clustering results and Figure 30 illustrates the wordcloud (Heimerl et al., 2014) for the 8 clusters A wordcloud is a visual representation of text data, where words from a body of text are displayed in varying sizes and colors. The size of each word in the cloud is typically determined by its frequency or importance in the original text. Words that appear more frequently in the source text are typically displayed in larger and bolder fonts in the wordcloud.

Table 7: Clustering results

	CLUSTER	**N° patents**
1	Technologies 4.0 (mechanics and robotics)	271
2	Material science	341
3	Cancer treatment	108
4	Optics - Image processing	179
5	Sensor technology - ICT	288
6	New molecules - new compounds - pharmacology	242
7	Energy/green Technologies	114
8	Biomedical	133

Therefore, the available 1694 patents were finally grouped into 8 clusters.

Figure 30: Wordcloud for the 8 clusters

6.3 Discussion

The first cluster, characterized by the frequent occurrence of words such as "system," "robot", "device", "material", "biomedical", "sensor" and so on, appears to be associated with Industry 4.0 technologies. This encompasses various applications of mechanics and robotics in diverse sectors, including healthcare, transportation, design, manufacturing, and more. A closer examination of the patents within this cluster, coupled with the analysis of prevalent keywords, unveils a range of innovative developments spanning multiple research domains:

- In the automotive and transportation realm, patents related to rail safety devices, underwater drones, advancements in cycling and motorcycling technologies are notable.
- Within the aerospace sector, patents encompass anti-icing systems for aircraft, rotating aerodynamic components, drones, and intelligent structures for integrity testing, among others.
- In the field of architecture and design, innovations extend to energy-efficient construction techniques, seismic isolation methods, complex structural monitoring, and applications within the fashion industry.

- Healthcare which is the main area of application of the innovations in this cluster. The patents present are truly varied, ranging from underwater guidance for the blind, innovative orthotics, endomedicular prostheses, sensorized heart valve prostheses to devices for radiotherapy, artificial bladders, and innovative brassieres and sheaths. But the bulk of innovations revolve around robotics leading the way: robotic exoskeletons, robotic platforms for laparoscopy, wearable robots, biomimetic robots, robotic surgical simulator, robotic limbs etc.

The second cluster was primarily centered on the field of material sciences, with a particular emphasis on applications in agrifood, chemistry and physics, manufacturing, and environmental sectors. An in-depth examination of the patents within this cluster, along with an analysis of the prevailing keywords, uncovers a diverse range of innovative breakthroughs spanning various research domains. For example, we find patents that create new technology for the production of composite ceramic powders, biocompatible sandwich panel, invention employing supercritical carbon dioxide to pasteurize foods or the realization of a natural product from Rosa canina seeds obtained by CO_2 extraction in supercritical phase. Other innovations concern the synthesis of a pesticide nano-formulation from environmentally friendly materials or high mechanical performance materials from stone processing waste, micro-algal photo-bioreactor, inhibitor preparation of unpleasant odors from household waste, an economical and effective method for treating wastewater in liquid form or a multifunctional hybrid material based on natural clays for environmental recovery and bioremediation. Many patents deal with the food sector: intelligent, active and biocompatible label; process for manufacturing additives for use in making antibacterial nanocomposites; innovative coating composed of a glassy matrix and metal nanoparticles with antiviral, antibacterial and antifungal properties; ready-made base for chocolate confectionery products; device for removing proteins, metals and other instability agents from wine and vegetable beverages; production of biopolymers, exploiting agro-industrial wastes, etc.

Cluster number four was centered around optics and image processing, as it featured innovations related to: packaging of optical signals, automatic machine for real-time detection of contaminants, optoelectronic apparatus for measuring position and orientation of rigid bodies, x-ray concentrator, energy conversion device, solid-state photodetector, integrated optical device, automatic immersion ultrasonic system, new sensor for accurate pH measurements, multi-modal optical fiber communication system, fully automatic optical microscope for fast reading of samples consisting of a transparent dielectric with metal nanoparticles. In this cluster are also present some patents with medical application: textile electrode device for acquiring electrophysiological signals from the skin; video-assisted dentistry using intraoral video cameras; acquisition of surface electromyographic signals and ultrasound images from the same portion of muscle; measuring device for assessing the volume of a breast; treatment of tumors with ion beams (hadrontherapy).

Upon analyzing the most frequently occurring words within cluster number five, it becomes evident that this cluster is primarily centered on sensor technology and Information and Communication Technology (ICT) in a broader sense. In fact, the main innovations are: soft-computing techniques for aerospace, anthropogenic noise control device; low-cost portable apparatus for characterizing sensor devices integrated with RFID transponders; RFID system developed for precise localization and tracking of objects equipped with low-cost tags; device prepared for analysis, simulation and prediction of slope instability phenomena; ground-based synthetic aperture radar capable of acquiring both three-dimensional and two-dimensional images; virtual sensor organized with a trained neural network. There are also patents with healthcare application (such as wearable haptic system to guide the cadence of steps in a person through vibrotactile stimuli), agricultural application (technology for disinfection of agricultural soil through the use of microwaves or radio frequency), architectural application (innovative IoT system to manage building cooling through advanced machine learning techniques), automotive application (logic control system for automotive; platform, and related method, for the identification, classification and subsequent removal of manufacturing defects present on vehicle components; innovative system that projects vehicle information and augmented reality elements to enhance the driving experience) and economic ones (dynamic and responsive computer model capable of representing economic interactions among financial institutions).

Cluster n.7 is the one that best matches Knowledgeshare's classification; in fact, it is characterized by patents in energy and green technologies. Following are some examples: energy harvester device built into the stem of the paddles; energy converter, which uses gyroscopic effects to generate electricity from sea waves; energy conversion device, which allows electricity to be generated from wave motion; micro-wind generation system; enthalpy heat exchanger; solar-derived thermal energy storage and/or exchange device; motorized system with parallel kinematics that enables automatic cleaning of the surface of photovoltaic panels; spectrally selective solar absorber coatings with enhanced photo-thermal performance and stability; device that produces water from the air; innovative portable device capable of ionizing water taken from environments outside the home; new method for the simultaneous treatment of polluted water and power generation; system to carry out air humidification and heat recovery in air conditioning systems; electric generator that statically converts heat into electricity without the use of moving parts or matter flows. Other innovations go towards the automotive application: new cooling solution applicable to electric machines; integrated system capable of transforming an internal combustion vehicle into an electric vehicle; hybrid-electric light aircraft and so on.

Three clusters (clusters n. 3, 6, and 8) were entirely focused on the healthcare and biomedical sector, encompassing a broad range of related areas.

The cluster n.3 was more related to cancer treatment with very relevant innovation in this field, such as: photodynamic therapy as a promising noninvasive treatment for cancer and nonmalignant tumors; method for early diagnosis and/or monitoring of Mucor infection;

tumor suppressor of malignant mesothelioma of the pleura; innovative theranostic system involving a multifunctional nanoconstruct and ultrasonic activation set-up capable of treating cancer cells; use of particular strawberry extracts for the prevention, treatment, and/or control of the progression of uterine fibroids; circulating bio-marker for diagnosis and prognosis of tumor progression; multi-modular and innovative system capable of isolating stem cells from small amounts of adipose tissue; molecular markers predictive of response to immunotherapy; novel method for cryopreservation of dental pulp to isolate mesenchymal stem cells; efficient targeted delivery system of molecules with therapeutic action (e.g., cytotoxic agents) based on adipose stromal stem cells; method for identifying a biomarker of stemness in hepatocellular carcinoma cells. Some patents were focalized on regenerative medicine: non-erodible, sterilizable, biocompatible hydrogel scaffold for 3D cell cultures or recombinant protein scaffold for preparing cell culture plates for use in developing biomaterials for neuro-regenerative medicine.

The cluster n.6 was more connected with the formulation of new compounds and the discovery of new molecules and therapies. Some examples: medicine designed to counteract the progression of acute and chronic neurodegenerative diseases; RNA interference-mediated therapy for neurodegenerative diseases; synthesis of a leptin antagonist tetrapeptide; synthetic melanocortins with antimicrobial effects for the treatment of topical infectious diseases; pharmaceutical compound for the treatment of wounds for topical use; pharmaceutical composition, which contains bactericidal/permeability-increasing protein and hyaluronic acid with the purpose of treating different types of arthropathies; new peptide and its use for the treatment of Alzheimer's disease; nanostructures capable of delivering oxygen into hypoxic tissues, which are associated with various metabolic, ischemic, and infectious diseases; multifunctional biomaterial consisting of a hydrogel (hydrogel) that is administered through an injection directly into the tissue to be treated; use of irisin, a hormone secreted mainly from skeletal muscle and in smaller amounts from adipose tissue, as a drug/strategy for the preservation of the function and survival of pancreatic islets of Langerhans and, in particular, the beta-cells that are deputed to insulin production. Other innovations deal with: antiviral compounds that find application in the prevention and treatment of infections caused by coronaviruses; use of benzofurans as synthetic "natural-like" herbicides, characterized by high phytotoxic and herbicidal activity; micro and nanocapsules tannins useful for the preparation of controlled-release pharmaceutical, nutraceutical or cosmetic compositions; new yeast strain that can be used to combat fungal infections in fungi of agronomic and commercial interest; lentil extract with cholesterol-lowering and prebiotic activity, particularly useful in therapeutic applications and used as a nutraceutical; mixture of active ingredients from pomegranate seeds useful in the treatment and/or prevention of obesity and associated diseases, such as particularly insulin resistance and type 2 diabetes and hepatic steatosis.

The last cluster (n.8) was named “biomedical” as it included mainly methods and techniques for disease diagnosis and monitoring. Some of these are very interesting: new diagnostic marker for Paget's bone disease and associated bone tumors; new diagnostic test to identify the two most common mutations in chronic myeloid tumors; new test for early detection of colorectal cancer which assesses decreased expression of a protein; noninvasive method suitable for pancreatic cancer diagnosis at an early stage; fecal sample testing system able to diagnose major chronic inflammatory bowel diseases; innovative system for early diagnosis of acute renal failure; method of in vitro diagnosis of head and/or neck cancer in tissue and/or biological fluid samples; new reporter system that enables early detection of the occurrence of muscle atrophy; diagnostic for rapid and early differential diagnosis of ulcerative rectocolitis; next-generation sequencing techniques to detect specific molecular signatures of urinary miRNA, which can be used to distinguish bladder cancer cases from healthy controls; use of low field nuclear magnetic resonance for monitoring patients with cystic fibrosis; innovative method that allows separation of nucleated fetal cells from maternal peripheral blood at all gestational ages, etc.

Next, we assessed the extent of overlap between the new classification and the one employed by the KS platform. This evaluation holds strategic importance when aiming to identify emerging technologies or cross-domain areas. To address this query, we constructed a "contingency table" for comparing KS classifications (represented in rows) with those derived from our analyses (presented in columns), as illustrated in Table 8.

Table 8: Comparison between KS technological areas and clusters emerged by the proposed approach

Knowledgeshare areas vs Clusters	Technologies 4.0	Material science	Cancer treatment	Optics - Image processing	Sensor technology ICT	New molecules, compounds, pharmacology	Energy/green Technologies	Biomedical	TOTAL
Aerospace, aviation et al	18	9		15	19	2	3		66
Agrifood et al	5	72		10	18	15	5	8	133
Environment and Constructions et al	54	65	1	17	40	1	26	1	205
Architecture and design et al	22	4		1	14		6	1	48
Chemistry, Physics, New materials and Workflows et al	18	169	6	76	8	34	11	9	331
Energy and Renewables et al	3	7		5	12		60		87
Informatics, Electronics and Communication System et al	33	1	1	22	140	2	1	5	205
Manifacturing and Packaging et al	17	6		3	2		1		29
Health and Biomedical	94	14	100	29	26	188	3	112	566
Transports et al	9			2	12		1		24
TOTAL	273	347	108	180	291	242	117	136	1694

Our research outcomes underscore a discrepancy between the technological domains defined in KS, which are established based on inventor self-assignments, and the clusters identified through the proposed approach. This discovery suggests that the patent analysis workflow we have outlined could potentially yield an alternative and likely more cohesive classification system, thereby enhancing the alignment between supply and demand. We are currently in discussions with our Netval partner (here is an NDA signed with Netval for the use of KS data) to explore the feasibility of adjusting the fixed patent categories/technological areas within KS and/or developing a recommender system to assist inventors in accurately categorizing their patents.

6.4 Conclusion

In this study, a distinct workflow for patent analysis was introduced, subsequently applied and tested on the Knowledge Share KS platform, where patents span across 10 technological areas. Leveraging NLP techniques and clustering analyses, we reorganized these 1694 patents into eight clusters, namely: Technologies 4.0, Material Science, Cancer Treatment, Optics - Image Processing, Sensor Technology – ICT, New Molecules - New Compounds – Pharmacology, Energy/Green Technologies, and Biomedical. Our findings have brought to light a discrepancy between the technological categories defined within KS and the clusters generated through our proposed workflow. This mismatch likely stems from inventors' self-assignments when uploading patents onto the platform. An automated classification approach could offer more coherence and potentially enhance performance in terms of aligning supply and demand. The potential advantages are manifold, ranging from enabling companies and investors to tap into innovations generated by research institutions through an improved matchmaking system to the prospect of introducing new technological domains aligned with the current innovation landscape. This could facilitate the development of cross-disciplinary or emerging technologies.

This study shows that Italian patents represent an extraordinary source of innovation that unfortunately is not yet fully "exploited" as all these inventions do not reach the market and the population. A great effort is still needed for a more intelligent management of intellectual assets that are the only factor that can foster innovation, creativity, knowledge sharing and improve the chances that knowledge reaches the market and brings faster benefits to society. This is especially true and strategic for the Healthcare sector that represents a main challenge in the Industry 4.0 for the final improvement of quality life and wellbeing of community and territory.

Having analyzed the problem of classification in the patent platform, the focus in the next chapter is on user support through an automated recommendation system to improve matchmaking between supply and demand and on the main keywords characterizing the technological fields.

CHAPTER 7. An applied AI-based approach for an Italian patent multi-label classification system and keywords identification to support the innovation demand and supply matching

In order to answer RQ3 - is it possible to support the user in correctly classifying a patent entered into the platforms in order to improve the matchmaking between demand and supply of innovation? - and RQ4 - is it possible to draw up an attempted vocabulary of technological fields from the keywords that emerged from an applied AI-based approach? - we performed the following steps:

- a combined approach of Natural Language Processing (NLP) and machine learning (ML) to create an automated recommender system that can identify the most suitable and correct technological area(s) a patent must be assigned;
- the same methodology allows for the identification of keywords characterizing a patent in an objective and human-independent way in order to create an initial vocabulary of words extracted from patents.

The contents of this chapter are contained in the article:
Nicola Amoroso[#], Annamaria Demarinis Loiotile[#], Francesco De Nicolò, Giuseppe Conti, Shiva Loccisano, Sabina Tangaro, Alfonso Monaco and Roberto Bellotti (2023) "An Italian patent multi-label classification system to support the innovation demand and supply matching". Scientometrics (under review) (# These authors contributed equally to this work)

The solution proposed for these RQs is an artificial intelligence-based system that accurately recognizes technological areas, thus classifying patents as correctly as possible, and recommends the right direction to the user, eliminating the "subjectivity" of choice. We use a combined approach of NLP and machine learning ML in order to propose an explainable patent classification system for multi-label classification of patents, supporting so the patent platforms by addressing two main aspects: on the one hand, it is of paramount importance to create an automated recommender system that can identify the most suitable and correct technological area(s) a patent must be assigned; this is both useful for users looking for specific technologies or patent owners who want to reach the largest fraction of potentially interested users. On the other hand, the methodology allows for the identification of keywords characterizing a patent in an objective and human-independent way; this aspect is particularly useful to create an initial vocabulary of words extracted from patents, thus eventually leading to redefine the available categories and supporting the portal management and, again, the matchmaking among users and patent-owners. This approach improves the user-friendliness of the platform and facilitates the selection of suitable patents by a company or, in general, enhances the matchmaking of innovation leading to social and economic impact.

The proposed approach has been implemented and tested on KS where the labels assigned to patents do not follow the International Patent Classification (IPC) and Cooperative Patent Classification (CPC) hierarchies but represent wide technological domains established by KS experts. Thus, an important question about the significance of this categorization naturally arises. Finally, the KS database is particularly appealing for classification purposes as, on the one hand it reduces the number of available categories and mitigates the curse of dimensionality problem; on the other hand, it does not require any arbitrarily set cut-off about the classification hierarchy (e.g. class, sub-class, group, sub-group), allowing to deal with clear and well established categories.

7.1 Proposed methodology

Our primary objectives revolve around two key aspects: (1) the classification of patents within the KS database as they are fed into Artificial Intelligence models, and (2) the identification of the most crucial keywords that define the labels, corresponding to technological areas. In pursuit of these goals, we have devised and put into action a structured approach consisting of three fundamental steps: (i) textual analysis, (ii) multi-class classification, and (iii) the identification of pivotal features that form the foundation for patent categorization. A visual representation of this approach is provided in Figure 31.

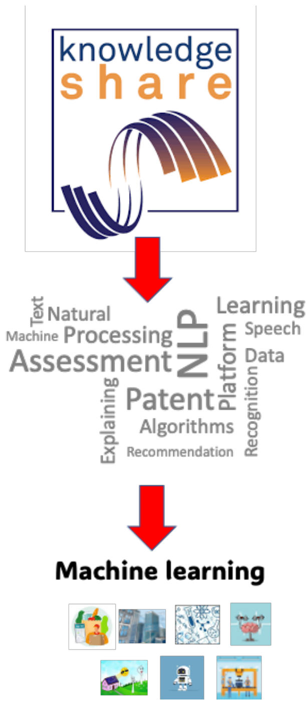


Figure 31: Overall flowchart of the proposed methodology

7.2 Methodology application and results

In the pre-processing phase, some technological domains of KS were excluded in the analysis as the number of patents was negligible with respect to other areas: "Aerospace and aviation", "Architecture and design" and "Transports". Moreover, patents associated with more than two labels, accounting for less than 5% percent of available data, were discarded. The final data consisted of 1.527 patents and 7 possible labels. As a final remark, KS patents (written in Italian) were translated into English using the deep_translator package in Python as NLP tools for English texts are more developed and consolidated than those for other languages (Chowdhary, 2020).

In the first phase, NLP tools are used to transform patents' textual data into matrix form, which is a necessary step in order to use them as input in an Artificial Intelligence pipeline. In the multiclass classification phase, we employ Machine Learning algorithms to allocate the appropriate labels (i.e., classes) to patents. Specifically, to execute these initial two phases and evaluate the performance of the classification algorithms, we adopt a zeroth-order approach, which involves disregarding the temporal information associated with patents. Instead, we implement a 5-fold cross-validation framework, repeating it 100 times (Schaffer, 1993). A visual representation of this framework is presented in Figure 32 below.

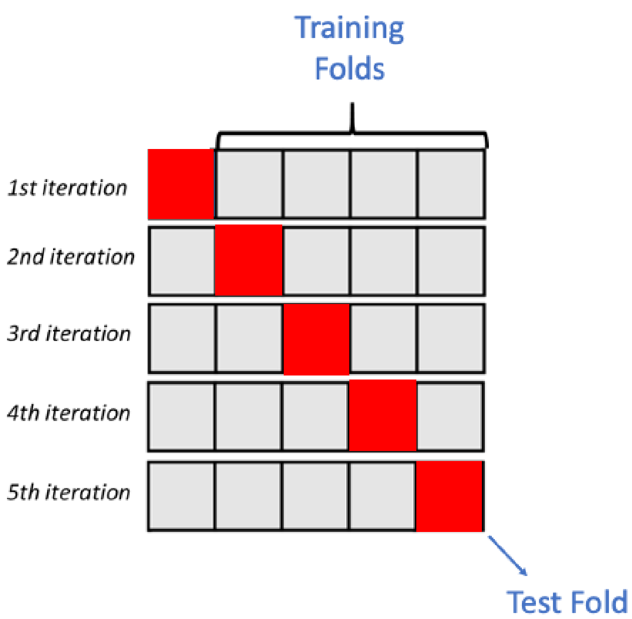


Figure 32: 5-fold cross validation framework. Schematic overview of a single repetition of a 5-fold cross-validation framework. At every iteration, four folds (the *training folds*) are used to train a model and the remaining one (the *test fold*) is used to validate it. The overall performance of the model is found averaging its performances on all the iterations. In order to have more robust results, this framework was repeated 100 times.

This approach was adopted to highlight the Italian innovation scene, as it is represented by the KS database, more than assessing the temporal evolution. This approach is justified by an ex-post validation: bad classification performances of all the algorithms would indicate the presence of a significant number of disruptive patented innovations during the considered time period, so that training algorithms on a set of patents (included in the *training folds*) would not be helpful in correctly classify the remaining patents (comprised in the *test fold*). Accordingly, we measure the performance of these algorithms using ranking-based evaluation indexes and find the most effective one in patent classification in the KS database (Wu and Zhou, 2017).

Finally, in the third phase, we accomplish a feature importance task on the top-performer algorithm: the aim is to identify the words that primarily influence the classification of patents into the corresponding classes. In the following sections, these steps are explained in detail.

7.2.1 From patents to feature representation

In every cross-validation iteration, the set of patents contained in the training folds are transformed, using the appropriate NLP tools, into a suitable representation which can feed machine learning algorithms. First, patents' words are lower-cased. In fact, it is worth noting that learning algorithms are sensitive to upper/lower-case differences, thus, lower-casing every single word is necessary (Ebert et al., 2016; Camacho-Collados and Pilehvar, 2018); the result of this first pre-processing phase is threefold: uniforming data, obtaining a substantial dimensionality reduction and yielding a correct and meaningful tokenization. Then, we removed stop-words, as they are useful in setting up phrases but meaningless for semantic purposes (Gerlach et al., 2019; Sarica and Luo, 2021). Finally, we performed lemmatization. Even in this case the objective is to facilitate the subsequent learning phase. By grouping words that share a common root, we can mitigate the sparsity of the data, resulting in a more concise representation that is less susceptible to the challenges posed by high dimensionality (Jivani, 2011; Hassani and Lee, 2016). This process ultimately yields a set of words that constitute the *vocabulary* associated with the patents within the training folds.

All these pre-processing phases are propaedeutic for setting up the TF-IDF matrix (Baeza-Yates and Ribeiro-Neto, 1999; Aizawa, 2003), as already explained. The result is a matrix with a number of columns equal to the patents in the training folds and a number of rows equivalent to the length of the respective vocabulary. In instances where a patent is assigned two categories, we replicate the corresponding column, associating one category with the original column and the other with its duplicate. As for patents in the test fold, a corresponding TF-IDF matrix is constructed using the vocabulary derived from the training folds. This choice is made to prevent data leakage between the training and testing phases, thereby ensuring that the generalization capabilities of the Machine Learning algorithms are assessed without over-optimistic biases.

7.2.2 Patent categorization as a multi-class classification problem

The TF − IDF representation can be suitably exploited by any machine learning algorithm, with patents representing the available samples, words' occurrences being the discriminating features and the target variable the labels associated with each patent. Since, according to the previous section, at most two out of seven labels are associated with every patent, we are faced with a *non-standard* multi-class classification problem. In fact, in a *standard* multi-class classification problem, only one label (out of more than two labels) is associated with every sample. Nonetheless, since multiclass classification algorithms return a ranking of all the possible labels for every sample, we can safely use classical Machine Learning methods but we need performance measures different from those usually used in standard multiclass classification. Accordingly, we have to refer to measures usually adopted in Information Retrieval tasks to evaluate Recommendation Systems (Li et al., 2018; Jung et al., 2023). In particular, we first consider two common metrics in these fields, that are defined for each sample:

- *Precision-top-k* (P@k): represents the rate of correct labels for the i-th sample in the top-k labels as predicted by the classification model. In formula:

$$(P@k)_i = \frac{\text{number of correct labels in top k for the i−th sample}}{k}$$

- *Recall-top-k* (R@k): is the ratio between the number of corrected labels for the considered sample in the top-k labels as predicted by the classification model and the number of true labels for the sample. In mathematical terms:

$$(R@k)_i = \frac{\text{number of correct labels in top k for the i−th sample}}{\text{number of true labels}}$$

Both these local metrics have values ranging from 0 to 1.

It is feasible to derive two global metrics from the preceding local ones by computing the average of the individual local values obtained for each sample:

$$\overline{P@k} = \frac{1}{N}\sum_{i=1}^{N}(P@k)_i$$

$$\overline{R@k} = \frac{1}{N}\sum_{i=1}^{N}(R@k)_i$$

where *N* is the number of samples under consideration.

It seems wise to underline that the information carried by the average Precision-at-k and the average Recall-at-k can be condensed in just one metric: the averaged F1-top-k ($\overline{F1@k}$). This metric is defined as the harmonic mean $\overline{P@k}$ and $\overline{R@k}$:

$$\overline{F1@k} = \frac{2 \times \overline{P@k} \times \overline{R@k}}{\overline{P@k} + \overline{R@k}}$$

F1@k ranges from 0 to 1 and is equal to 0 when at least one of the two metrics is 0, while it is equal to 1 when both of them are equal to 1. Since a precision value could be very high at the cost of a very low recall and viceversa, it seems wise to use F1@k as an unbiased performance measure (Hu et al., 2018; Zuva and Zuva, 2012). In particular, following the scientific literature on automatic patent classification (Lee and Hsiang, 2020) we consider F1@k with k=1 as our main performance metric. Accordingly, we can use this score to benchmark our work with the best F1 scores in other articles.

We consider and evaluate five different models in order to evaluate the robustness of the defined pipeline: a Random Classifier (RC), serving as a performance baseline, Logistic Regression (LR) (Hosmer et al., 2013), Random Forest (RF) (Shi and Horvath, 2006), Support Vector Machine with linear kernel (SVM) (Rameshbhai and Paulose, 2019) and XGBoost (XGB) (Chen and Guestrin, 2016). We consider these Machine Learning models since they are all able to return a measure of features' importance, although through different mechanisms (Bentéjac et al., 2021; Hastie et al., 2009). In fact, LR associates a coefficient to every input feature, while tree-based models define the importance of a feature according to its ability in increasing the purity of a node in the tree, as measured by the Gini index. Moreover, SVM with linear kernel, being a geometrical method that aims at finding a hyperplane in the feature space that separates classes, gives a weight to every dimension (i.e. to every feature). It seems wise to underline that, even though all the models can be used to accomplish a multi-class classification, not all of them natively support it (Hilbe, 2009; Chauhan et al., 2019). Actually, LR and SVM are able to carry out a binary classification task, not a multi-class one. Nonetheless, this problem can be circumvented by adopting a "One-versus-Rest" (O-v-R) or a "One-versus-all" (O-v-A) framework. In the first case, for each class, a binary model is trained in order to recognize the considered class against all the others. Accordingly, we end up with a model for each class. In the second case, models are trained to distinguish every class from each other, obtaining a much greater number of models with respect to the O-v-R case. In this work we adopt a O-v-R approach because we can obtain which features (i.e. words) characterize every category with respect to all the others.

7.2.3 Mitigating class imbalance: SMOTE

In the previous sections we describe how to transform textual data into a matrix form and how to set up a multi-class classification problem, together with presenting the appropriate performance metrics. However, it should be noted that there is an imbalance among classes that can severely impact algorithms' performance (Japkowicz and Stephen, 2002; Guo et al., 2008). In order to mitigate this effect, we leverage the mathematical form of textual data and

use *SMOTE* (Synthetic Minority Oversampling TEchnique) in order to obtain a perfect balance among classes (Chawla et al., 2002; Blagus and Lusa, 2013). In particular, we apply this technique in the training folds of every iteration in our cross-validation framework. Then, for every iteration, we train our models on these class-balanced training folds and validate them on the test fold. We then replicate this workflow for every repetition of the cross-validation framework.
The results obtained applying our pipeline to the KS database are shown here. In particular, we will show (P@k)¯, (R@k)¯and (F1@k)¯for k=1 for completeness, but we will use (F1@1) to determine the best model. Once the best classification model is identified, we deepen its performance analysis highlighting the most important words: those that mainly influence it in classifying patents. Moreover, we show how reliable it is in classifying patents in each of the seven categories of the KnowledgeShare database and which are the most error-prone.

7.2.4 Top-k performance measures

In table 9 we report, for each considered classification algorithm, the mean and standard deviation values (in brackets) of $\overline{P@1}$, $\overline{R@1}$ and $\overline{F1@1}$ as determined by our 5-fold cross-validation approach repeated 100 times.

Table 9: Mean $\overline{P@2}$, $\overline{R@2}$ and $\overline{F1@2}$ values for the classification algorithms. The corresponding standard deviations are reported in brackets. $\overline{F1@2}$ values are in bold.

Metric	RC	LR	RF	SVM	XGB
$\overline{P@1}$	0.181 (0.022)	0.801 (0.024)	0.755 (0.026)	0.793 (0.023)	0.764 (0.023)
$\overline{R@1}$	0.139 (0.017)	0.679 (0.022)	0.645 (0.024)	0.674 (0.022)	0.648 (0.021)
$\overline{F1@1}$	0.157 (0.019)	0.735 (0.022)	0.695 (0.024)	0.728 (0.022)	0.701 (0.021)

Observing the $\overline{F1@1}$ values in Table 9, we can conclude that the Random Classifier is the least effective model, while all the other models significantly outperform it. This indicates that the latter models make effective use of word features for patent classification. Furthermore, it's worth noting that LR (Logistic Regression) and SVM (Support Vector Machine) demonstrate similar levels of performance. Accordingly, we decide to perform a Welch's t-test to establish if the mean values of $\overline{F1@1}$ of LR and SVM are equal or not in a statistically significant way. Considering a 5% significance level, we can reject the null hypothesis that the mean values of $\overline{F1@1}$ of LR and SVM are identical; but, at a 1% significance level we cannot reject this null hypothesis. In order to remove this ambiguity, following the scientific literature, we consider the performances of LR and SVM as measured by $\overline{F1@2}$. We obtain the results reported in Table 10, where, for the sake of completeness, we report the results for all the models.

Table 10: $\overline{P@2}$, $\overline{R@2}$ and $\overline{F1@2}$ values for the classification algorithms. The corresponding standard deviations are reported in brackets. values are in bold.

Metric	**RC**	**LR**	**RF**	**SVM**	**XGB**
P@2	0.184 (0.013))	0.544 (0.013)	0.503 (0.016)	0.518 (0.016)	0.508 (0.016)
R@2	0.286 (0.022)	0.867 (0.016)	0.813 (0.022)	0.830 (0.018)	0.818 (0.021)
F1@2	0.224 (0.019)	0.669 (0.013)	0.622 (0.018)	0.637 (0.016)	0.627 (0.017)

Performing a Welch's t-test at both 1% and 5% significance level, we can reject the null hypothesis and consider the mean performances of LR and SVM different in a statistically significant way. Accordingly, we can consider LR our top-performer model in classifying patents.

7.2.5 How words influence patent classification

According to the previous section, LR has proved to be the best model in classifying patents in the KS database. We can go deeper in analyzing LR performances considering how the words in the vocabulary (i.e. the features of the model) influence the classification task carried out by the LR model. To this aim, since in the O-v-R scheme a LR model is built for every category, we evaluate the corresponding coefficients, which can be seen as a measure of importance of the input features. Moreover, we also determine the words' global importance ranking concatenating the words' coefficients of every category. We end up with a global ranking of 88.144 words. For the sake of clarity, Figure 33 shows the average importance for the first 1.000 words in the global ranking. By utilizing the *elbow* method, we were able to quantitatively ascertain the number of the most significant words, resulting in a set of 260 words. For a more comprehensive list of these crucial words, along with their respective categories, please refer to Appendix A.

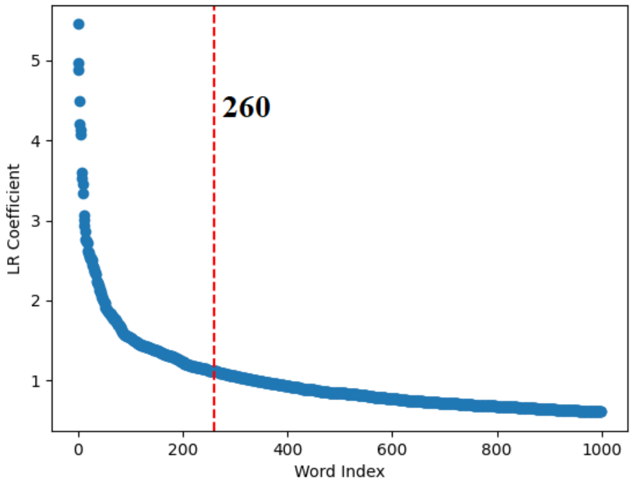


Figure 33: LR sorted importance for the first 1.000 words. Notably, the trend shows a steep decrease followed by a large plateau after 260 words.

Figure 34 illustrates the distribution of the 260 words among the seven technological areas.

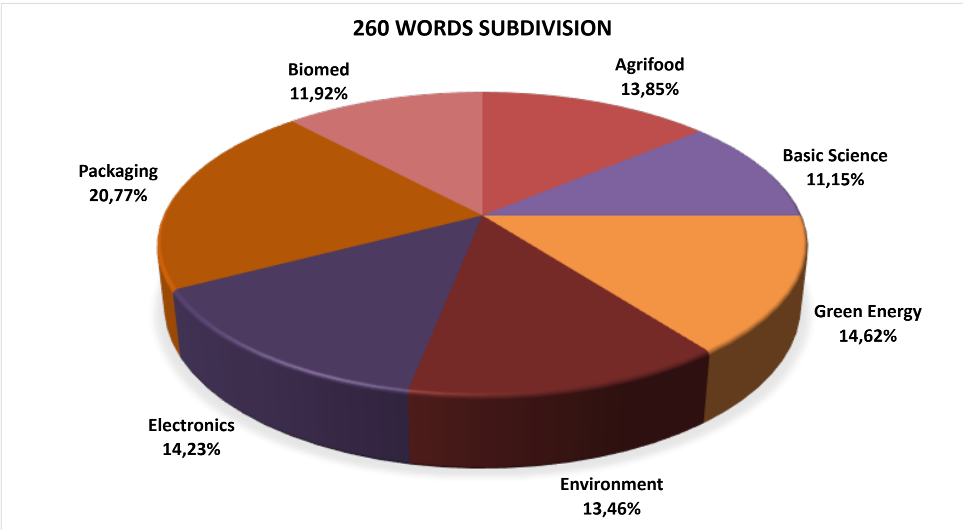


Figure 34: The partition of the 260 words in the 7 technological areas

As regards Green Energy, the top words include topics related to major renewable sources, heat, electricity, fuel and power. In the Agrifood area, the top words included specific products like wine, vegetable oil and milk. Interestingly, among the top words particular attention was given to the waste theme. In Electronics, the principal words were: user, network, circuit, software, signal, optical, device. In the Environment and Constructions area, the focus seems to be on buildings with seismic properties and energy- and water-saving solutions. The only 5 words in the area concerning the Packaging appear not so relevant: component, object, material, packaging and piece. On the contrary, the 5 words describing the Biomed sector are very meaningful because they seem to characterize the areas of greatest investment and innovation in recent years in the medical field: patient-centeredness, tissue engineering, treatment and diagnosis of cancerous and non-cancerous diseases, and stem cells and cell therapy. The last area, Basic Science, is described by only 2 words among the top 50: particle and solvent.

7.2.6 Confounding categories for the best classifier

In this section, we delve into the categories that pose the greatest challenge for the best-performing model, namely LR. To accomplish this, Table 11 displays, for each category, the average frequencies at which it is misclassified with all other categories within our cross-validation framework. Additionally, Table 11 presents the frequencies of correct classification for each category.

Table 11: Mean frequencies of confusion among categories (in percentage). The frequencies of correct labelling, in bold, are reported in the main diagonal. The corresponding standard deviations are in brackets.

Predicted Label	**Agrifood**	**Environment**	**Basic Science**	**Green Energy**	**Electronics**	**Packaging**	**Biomed**
True Label							
Agrifood	**68.1 (9.1)**	6.2 (3.3)	7.2 (2.1)	2.1 (1.0)	5.1 (2.2)	1.1 (0.4)	10.2 (4.1)
Environment	4.5 (3.2)	**42.1 (8.2)**	19.3 (6.2)	13.4 (4.4)	13.2 (4.1)	6.2 (2.2)	3.1 (1.0)
Basic Science	6.2 (2.1)	6.7 (2.2)	**51.1** (5.1)	6.2 (3.1)	7.6 (2.1)	3.1 (1.0)	20.4 (4.1)
Green Energy	3.8 (1.1)	11.2 (2.3)	11.2 (3.2)	**56.9 (8.2)**	13.2 (5.3)	2.1 (0.5)	3.4 (1.1)
Electronics	1.8 (0.2)	4.5 (0.8)	6.5 (1.9)	6.8 (1.8)	**64.3 (5.2)**	5.6 (1.2)	14.5 (2.2)
Packaging	8.7 (1.3)	8.2 (1.2)	23.8 (3.2)	5.2 (1.4)	23.9 (8.4)	**24.8 (2.1)**	10.8 (2.1))
Biomed	1.1 (0.1)	1.8 (0.2)	8.8 (1.1)	1.6 (0.1)	9.7 (1.1)	2.8 (0.7)	**77.8 (2.3)**

Observing these results, “Biomed” is the best-recognized category, meaning that it is a well-established and self-contained research area, with its own keywords and concepts, being rarely confused with other research domains. On the other hand, it can be observed that “Packaging” is the worst-recognized category being mainly confused with “Electronics” and “Basic Science”. This indicates that “Packaging” cannot be considered a standalone research area but, on the contrary, it may be seen as a cross-domain that benefits from innovations in Basic Science and Electronics. In fact, an invention may cover several technological areas or be cross-domain, and this is becoming increasingly true in emerging technologies that are becoming more and more "multi-disciplinary" and struggle to remain "confined" to a single technological area. This highlights the need to create new technological domains that are not strictly tied to a single area.

7.3 Discussion

7.3.1 Best method performance

Patent Automatic Classification (PAC) algorithms have raised an increasing interest in recent years because of the ever-growing number of patents and the consequent need of high quality analysis (Li et al., 2018). Moreover, correctly classifying patents means understanding why a patent is labelled with some categories and not with others. Then, we can query the classifier algorithm to know which words mainly influence the classification of a patent considering the model's feature importance. Since this is our work's main aim, we have to compare our best classifier's performance with the state-of-the-art found in the literature. It should be noted that there is no officially acknowledged database on which pipelines' performances should be compared. Accordingly, there is neither a common set of patents nor a shared classification system of patents (e.g. CPC, IPC). Then, the significance of the following comparisons should consider this aspect. Moreover, it should be underlined that, as far as we know, our work is among the first in using a cross-validation approach, presenting mean and standard deviation values of $\overline{P@k}$, $\overline{R@k}$ and $\overline{F1@k}$.
Beginning from articles having a dataset size similar to ours, both Fall et al. (2003) and Hepburn (2018) consider SVM together with other classification algorithms (k-NN, Naive Bayes) to classify 75.250 patents in 8 categories (IPC section level). In particular Hepburn (2018) uses a 60/40 train-test split and a particular transfer learning technique reporting $\overline{F1@1}$ = 0.784, without standard deviation, which improves the result of Fall et al. (2003), which indicates $\overline{F1@1}$ = 55%. Hu et al. (2018) use Neural Networks (both Convolutional and Recursive) to capture semantic correlations among patents. The training, validation and test datasets contain respectively 72.532, 18.133, and 2.679 patents from the CLEF-IP competition dataset with 96 labels (CLEF-IP). Their best $\overline{F1@1}$ is 63.97%. Regarding the CLEF-IP competition itself, Verberne et al. (2010) have their best F1@1 equal to 70.59% considering a series of classification experiments with the Linguistic Classification System (LCS). The training dataset has 905.458 patents and the testing has 1.000 patents. Li et al. (2018) build a Deep Learning model to classify more than 2 millions patents in the USPTO-2M dataset in 637 classes (the subclass level in IPC) reporting P@1=73,9% without disclosing F1@1. Haghighian Roudsari et al. (2022) compare different Deep Learning models on the USPTO-2M dataset, obtaining a maximum $\overline{F1@1}$ = 63.33%, with $\overline{P@1}$ = 82,72% and $\overline{F1@1}$ = 55.89%. Lee and Hsiang (2020) improve this result by considering a

novel USPTO-3M dataset, comprising more than 3 millions patents and working both at the IPC subclass level and at the CPC subclass level (656 classes). It reports a maximum $\overline{F1@1}$ = 66.71% and, correspondingly, $\overline{R@1}$ = 54.92% and $\overline{P@1}$ = 84.95%.
Considering that our best model (LR) has $\overline{F1@1}$ = 73.5% (2.2%), $\overline{P@1}$ = 80.1% (2.2%), $\overline{R@1}$ = 67.9% (2.2%), we can say that our results are in line with the state-of-the-art found in the literature, thus justifying the investigation of LR words' importance and the most confounding categories.

7.3.2 Categories' keywords and how they explain the confusion frequencies

A detailed examination of the words in each category reveals some trends in the Italian innovation scenario and can even explain the confusion frequencies reported in Section before.

As concerns the "Green Energy" category, the innovations speak about automatic systems for cleaning solar photovoltaic modules, water production from the air with solar energy, systems for concentrating the solar power, energy-efficient systems for the use of solar thermal energy, micro turbo gas systems, thermodynamic solar systems and so on. Since these innovations deal with electronic devices and the exploitation of physical processes, it is not surprising that the main areas it is confused with are "Electronics" and "Basic Science". Moreover, since green energy innovations aim at the protection of the environment, this strong relationship is reflected in the non-negligible confusion frequency of "Green Energy" with "Environment".

With reference to the "Agrifood" area, the main identified words suggest research investments in innovating the quality and production systems of wine, lab-on-chip devices for the improvement of olive oil protection, novel sustainable processes for milk production. The main innovations in this field involve biological processes and methodologies and this is reflected in the relatively high frequency of confusion of this area with Biomed.

The "Electronics" research area reveals the presence of innovations about user-centred design methods, novel architectures of artificial neural networks and, above all, deployments of synthetic biological circuits. This relatively new research branch is increasingly growing and explains the high frequency of confusion between "Electronics" and "Biomed", which mainly encompasses biological tools and methodologies.

As regards the "Environment" category, the most frequent words concern innovations in: cleantech, automatic irrigation systems, water saving, energy production from water or wave, micro-geophysics, remediation of stone buildings, dehumidification processes, building maintenance services and seismic isolation methods, and so on. It should be noted that these innovations are strictly related to the use of novel electronic devices and sensors and the pioneering exploitation of thermodynamics processes. These observations explain why

“Environment” patents are mainly labelled as “Electronics” and “Basic Science”. Moreover, as noted in the previous discussion about the “Green Energy” category, there is a strong relationship between it and “Environment” innovations, which accounts for the relatively high level of confusion between them.

In the “Packaging” area, a number of health-related innovations characterise the research effort in this field: Hybrid locomotion stair-lift wheelchair, artificial muscle, human-robot cooperation, wearable robot for lower limb, exoskeleton for upper limb rehabilitation, adaptive wearable robot, support frame for upper limb exoskeleton, biomimetic active foot and ankle prosthesis. The remaining innovations include 3D printing, multi-sensor dimensional measurements, conversion of traditional vehicles, and language learning devices. It can be seen that innovations registered in the “Packaging” area are mainly linked to the development of novel electronic devices for relieving heavy lifting and learning methods using ad-hoc Artificial Intelligence algorithms. This can be seen as the main reason behind the high confusion frequencies characterizing the “Packaging” area, already noted. In fact, it has the two highest confusion frequencies with “Electronics” and “Basic Science”. This clearly indicates that the “Packaging” category cannot be considered as a stand-alone research branch because it largely benefits from innovations both in electronics and in basic science research.

The innovations in the “Biomed” technological area range from wearable devices, method for diagnosis and/or monitoring of infection, to new molecules for cancerous and chronic diseases, interventional radiotherapy, tests for early detection of tumors, theragnostic radiopharmaceuticals, cell therapies and selection of cell populations. It should be noted that these novel techniques and devices mainly rely on Artificial Intelligence methodologies. In fact, the greatest part of patents is about methods in tissue engineering, an interdisciplinary area concerned with the development of functional 3D tissues by mixing cells, and bioactive chemicals. Accordingly, we can see that this explains why the “Biomed” category has non-negligible confusion frequencies with “Basic Science” and “Electronics”. Nonetheless, since the Biomed research is even highly characterized by biological-medical terms, the corresponding confusion frequencies are not as high as in all the other categories.

Research efforts in the “Basic Science” category are multi-faceted: they range from innovative, recyclable and low-cost solvents to field-effect transistor sensor, 3D super resolution optical microscope, and magnetic resonance imaging using metamaterials. Since many patents deal with bio-materials or describe bio-medical innovations, it is not surprising that “Basic Science” has a 20.4% mean frequency of confusion with “Biomed”.

In general, as regards the confusion frequencies, it seems evident that the technological areas so defined in KS are sometimes a bit "narrow" to categorise patents that are related to different technology domains, i.e., the 7 technology areas may not be sufficient to capture the entire landscape of academic patents. It would be interesting to use an unsupervised clustering approach, based on the textual content of patents, to develop an alternative and probably more consistent classification, improving the matching of supply and demand for innovation.

7.4 Conclusion

In conclusion, we introduce a fully automated patent multi-classification system that leverages NLP techniques and ML to highlight the keywords defining classification categories. Patent platforms like Italy's KnowledgeShare were created to bridge the gap between patent buyers and providers, addressing issues of information asymmetry and distrust in traditional patent transactions. Tools like the one we've proposed are of paramount importance in enhancing their mission and effectiveness. We've demonstrated that our framework's accuracy is on par with state-of-the-art methods. By relying on an easily interpretable model like Logistic Regression (LR), it also allows us to identify the keywords associated with each technological domain. An important feature of our approach is its ability to assign patents to one or more categories simultaneously, recognizing the potential for emerging technologies to span multiple domains.
Our method is open-access and serves both patent managers by offering reliable category recommendations for patents and their associated technologies, and end-users of technological platforms (e.g., investors, companies) by providing a quantitative evaluation of the similarity between the technologies they seek or wish to patent and the most suitable technological domains. Thus, our recommender system streamlines the matchmaking process between innovation supply and demand, a critical aspect of generating social and economic impact with new technologies.
While our model is currently tailored to the Italian patent ecosystem, it possesses broad applicability and can be extended to patents from international research centers. As a future technical enhancement, we can explore the use of Artificial Neural Network architectures, coupled with appropriate explainability tools, to further enhance classification performance and deepen our understanding of innovation. Another pivotal step forward would involve implementing the model directly within the KnowledgeShare platform itself.

In the following chapter, the analysis is concentrated on the frequent words inherent to health sector with the final aim of constructing a first draft of a HC4.0 vocabulary based on the contents of the Italian research institutions patents.

CHAPTER 8. A first draft of HC4.0 vocabulary for characterizing the innovative technologies in healthcare 4.0

In order to answer RQ5 - is it possible to define a first draft of HC4.0 vocabulary for characterizing the innovations and innovative technologies in healthcare 4.0? - we performed the following steps:

- NLP, clustering and complex network analysis on the most frequent words focusing on the clusters inherent to health sector;
- identification and characterization of the technology trends in the cross-domain field of Healthcare 4.0.

The contents of this chapter were published in the proceedings of the following conferences:

- Annamaria Demarinis Loiotile, Francesco De Nicolò, Adriana Agrimi, Giuseppe Conti, Nicola Amoroso and Roberto Bellotti (2023). "Towards a Healthcare 4.0 vocabulary: a Patent-based approach". Proceeding of the 2023 World Conference on Information Systems & Technologies (WorldCIST'23) - Lecture Notes in Networks and Systems (LNNS, volume 799) - Series ISSN 2367-3370; Softcover ISBN 978-3-031-45644-2; eBook ISBN 978-3-031-45645-9.
- Annamaria Demarinis Loiotile, Davide Veneto, Adriana Agrimi, Gianfranco Semeraro and Nicola Amoroso (2023). "An AI-based approach for the improvement of university technology transfer processes in healthcare". Proceeding of the 2023 World Conference on Information Systems & Technologies (WorldCIST'23) - Lecture Notes in Networks and Systems (LNNS, volume 799) - Series ISSN 2367-3370; Softcover ISBN 978-3-031-45644-2; eBook ISBN 978-3-031-45645-9.
- Annamaria Demarinis Loiotile, Nicola Amoroso and Roberto Bellotti (2023). "Characterization of innovative technologies in healthcare 4.0 through the analysis of Italian patents". Proceeding of the Ambient Assisted Living Forum 2023 - ForItAAL 2023 - Lecture Notes in Electrical Engineering (accepted and to be appeared).

8.1 Proposed methodology

First by using NLP and clustering techniques, and then, by means of a complex network analysis on the most frequent words, the analysis focuses on those clusters inherent to health in order to investigate the technologies used and their applications to the Healthcare 4.0 domain. The final attempt is the construction of a first draft of a HC4.0 vocabulary based on the contents of the Italian research institutions patents (see Figure 35).

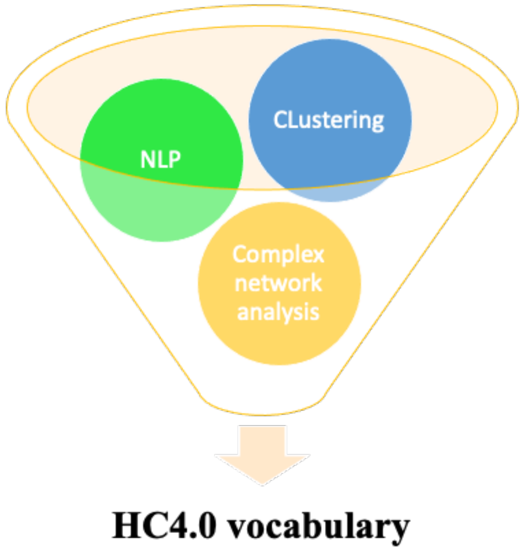


Figure 35: Overall flowchart of the proposed methodology

8.2 Methodology application and results

As already explained in Chapter n.6, each patent contained in KS was transformed by means of NLP techniques, then an unsupervised k-means cluster analysis was performed and, using a Silhouette score, the quality of clusters created using clustering algorithm was evaluated. This analysis outlined the presence of 8 clusters, that are again listed here for reading convenience:

- Technologies 4.0 (mechanics and robotics)
- Material science
- Cancer treatment
- Optics - Image processing
- Sensor technology - ICT
- New molecules - new compounds - pharmacology
- Energy/green Technologies
- Biomedical.

It is interesting to underline that patents related to the health sector account for a large portion (about 30 percent of the total) and are enclosed in 4 out of 8 clusters. Accordingly, four clusters were further investigated because of their proximity to healthcare 4.0 applications: Technologies 4.0, Cancer treatment, New molecules - new compounds and Biomedical. These clusters were then considered for subsequent complex network analyses.

For each cluster, we examined the occurrence frequency of each word to narrow the research field to the most representative words: in particular, the top 5% of words was selected within each cluster and, finally, an overall amount of 106 words was determined. At this point, to

highlight the relationships among these words we built a weighted complex network model whose nodes were the selected words, links were drawn between a pair of different words if they co-occurred in at least one patent and the number of co-occurrences represented the weight of an existing link.
This model was adopted to explore if particular words, and therefore concepts, could be related in significant patterns. The underlying assumption is that these patterns provide a first step towards the definition of a healthcare 4.0 vocabulary with the opportunity to detect which are the basic assets defining healthcare 4.0 in an unsupervised way. There is not a univocal defined ontology for healthcare 4.0, yet. Therefore, we investigate here the adoption of network metrics to address this issue.
In a complex network, by definition, the importance of nodes can be evaluated by means of centrality metrics (Boccaletti et al., 2006; Batool and Niazi, 2014); here three different options were considered: degree (d), eigenvector centrality (e) and betweenness (b) (Batool and Niazi, 2014). The degree measures the connections of a node; eigenvector centrality weighs this information according to the global degree distribution and it is a measure of the influence of a node in a network; betweenness evaluates the number of paths passing through each node, or in other words, betweenness centrality is a measure that quantifies how often a node serves as a crucial bridge along the shortest path connecting two other nodes. The reason for such choice is that, in general, three different perspectives can be used to measure centrality according to the local, global or dynamical flavour to be emphasized (Amoroso et al., 2021). Moreover, as these metrics do not take into account the links' weights, we investigated the nodal strength (s). To this aim, we examined these four metrics and compared the ranking they returned in terms of Spearman correlation, see Figure 36.

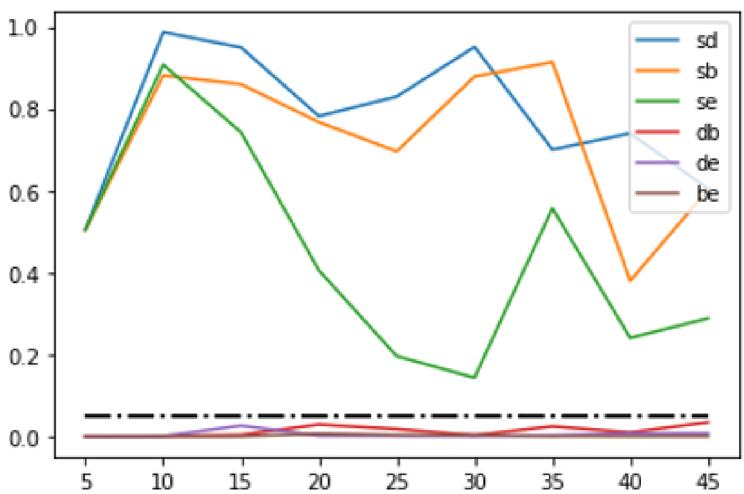


Figure 36: Spearman correlation p-values of degree(d), betweenness (b), eigenvector centrality (e) and strength(s) varying with the number of considered words. Statistical significance 0.05 is represented with a dash-dot line.

Our findings show that degree, betweenness and eigenvector centrality never show statistically significant differences, on the contrary the ranking provided by strength is significantly different independently from the number of words considered. Besides, we evaluated to which extent this result was affected by the number of considered words. We observed that, despite the varying number of words, no changes in correlation could be observed. This finding suggests that the information carried by network weights is far from being trivial and should not be neglected.

In the following Table 12 the list of the first twenty words according to the importance evaluated by degree and strength is reported.

Table 12: The top twenty words by Strength and Degree

Ranking	Degree	Strength
1	Control	Gen
2	Patient	Treat
3	Method	Cellul
4	Tissue	Effect
5	System	Pharmac
6	Gen	Articul
7	Articul	System
8	Treat	Tumor
9	Disease	Effectiv
10	Effectiv	Patient
11	Mechani	Therap
12	Therap	Molecular
13	Process	Method
14	Function	Disease
15	Technolog	Tissue
16	Model	Tumor
17	Cell	Protein
18	Compound	Compound
19	Protein	Diagnos
20	Surger/Surgical	Mechani

While degree accounts for the number of patents using a specific word, nodal strength accounts for the number of times that word has been used, therefore normalized strength can be easily interpreted as the percentage of words' co-occurrences. Top twenty words account for more than 30% of existing connections in both cases. The comparison between the two lists shows an agreement of about 60%. In the following, a possible interpretation of these lists is provided.

8.3 Discussion

An analysis of the most frequent words in these 4 clusters and the patents contained in them reveals innovative technologies in the Healthcare 4.0 field developed by Italian universities and research centers. In the illustration of each cluster, the words shown in Table 12, i.e. those constituting the vocabulary, are highlighted in bold.

The cluster n.1 "Technologies 4.0" contains new technologies including mechanics and robotics applied to different field: healthcare, transport, design, manufacturing, etc. Regarding healthcare, the technologies range from:

- innovative wheelchairs for the disabled people,
- underwater guidance for the blind,
- active hip orthosis dedicated to assisting flexion-extension movement, as well as an exoskeleton that includes said actuation **system**,
- active orthosis for shoulder and elbow assistance, with active and direct assistance on rotational degrees of freedom and with passive **mechanisms** for compensating shoulder movements that do not result in rotation,
- haptic feedback dispositive that can be integrated into a prosthesis or a defective part of the human body, includes a sensory interface for the prosthesis or stump, equipped with a contact sensor capable of generating a signal upon contact with the object,
- a wearable haptic device, in the form of a robotic thimble, that enables integrated simulation and perception of contact and temperature in virtual reality, augmented reality contexts under remote operating conditions, for example, in **surgery** and industrial maintenance tasks; this **mechanism** is integrated with a miniaturized thermal module, divided into two coplanar sectors, each equipped with an independent temperature control system,
- innovative sensorized heart valve prosthesis for continuous monitoring of valve leaflet motion,
- interventional radiot**herapy** applicator device,
- artificial pneumatic muscles, capable of producing force and bi-directional motion with variable stiffness,
- artificial brassiere for women who exhibit temporary/permanent breast asymmetry due to differences in volume, shape, or size after breast **surgery**,
- easily transportable robotic exoskeleton designed for lower limb rehabilitation in **patient**s with central nervous **system** injury outcomes,
- wearable robotic device that has the **function** of aiding or extending grasping abilities,
- wearable exoskeletons for robot-assisted neuro-motor and orthopedic rehabilitation and limb movement guidance during the performance of daily activities,

- single-body, sutureless heart valve using an innovative material and spray-robotized **technology** capable of coating a cast generated through 3D **modeli**ng and simulation with polymeric solutions,
- a device and **method** that provides foot movement mobility that is highly compatible with the physiological movements of the ankle musculoskeletal **system**,
- robotic platform for laparoscopy, an extremely useful tool for single-port **surgery**,
- implantable artificial bladder, characterized by the ability both to vary its internal volume and to offer high resistance to urine and fouling,
- cardiac simulator, which can be used to reproduce as closely as possible to reality the structural and **functional** behavior of the heart muscle and ventricles during the phases of systole and diastole,
- assistance to **patient**s with motor disabilities with a personalized modality, based on innovative machine learning techniques, with an automatic **control** scheme, based on **model** predictive **control** and thus such as to provide predictive support action.

The cluster n.3 is more related to cancer treatment with very relevant innovation in this field, such as:

- new three-dimensional co-culture **method** of podocytes and endothelial cells for the study of in vitro **model**s of kidney **disease**,
- novel **method** for cryopreservation of dental pulp to isolate mesenchymal stem cells,
- use of a recombinant **protein** scaffold to prepare **cell** culture plates for use in the development of biomaterials for neuro-re**gen**erative medicine,
- arrays of organic transistors specifically designed for electrophysiological applications,
- multi**function**al metal oxide-based nanoconstruct capable of generating highly oxidizing and cytotoxic species when specifically activated by external pressor stimuli to induce targeted cytotoxicity in target cancer cells,
- production of a nonerodable, sterilizable, biocompatible hydrogel scaffold for 3D cell culture,
- **method** for directly measuring, via a low-cost, easy-to-use, high-throughput electronic device, the activity of an enzyme involved in the aging **process** of cells,
- platform based on bacteriophages (i.e., viruses that infect bacteria), modified to selectively target and kill specific cell types of medical, veterinary, microbiological, industrial, and environmental interest,
- stimulus-responsive and **gen**etically **controll**ed artificial cell biological **system** capable of producing **protein**s and releasing synthesized pharmacological **molecul**es,
- **compound** that inhibits cancer **cell** proliferation at nanomolar concentrations by reducing **cell** growth and inducing apoptosis, in experimental **model**s of malignant peritoneal mesothelioma (DMPM);
- efficient targeted delivery **system** of **molecul**es with **therap**eutic action (e.g., cytotoxic agents) based on adipose stromal stem cells,

- multi-modular and innovative robotic **system** capable of isolating stem cells from small amounts of adipose **tissue** and automating all steps of isolation, manipulation, and cell expansion of adipose stem cells.

The cluster n.6 is more connected with the formulation of new compounds and the discovery of new molecules and therapies. Some examples are:

- innovative intranasal delivery modality through a device for the **treatment** of hypoxic-ischemic and traumatic brain injury of pediatric and neonatal age,
- multifunctional biomaterial consisting of a hydrogel (hydrogel) that is administered through an injection directly into the **tissue** to be treated,
- a microscale platform capable of generating 3D micro**tissue**s in vitro from human cells,
- a modular device for automatic drug delivery,
- oligosaccharide **molecule** capable of counteracting the motor symptomatology typical of neurodegenerative **disease**s, capable of counteracting the motor symptomatology typical of Parkinson's **disease** with biochemical and functional recovery of dopaminergic neurons, potentially resulting in improved clinical conditions of patients and in terms of quality of life,
- polymer scaffold capable of mediating cardiac regeneration,
- a **compound**, based on melanocortin agonists active on MC4 receptors, an **effecti**ve treatment in counteracting the progression of acute (e.g., stroke) and chronic neurodegenerative **disease**s,
- botulinum neurotoxin type A (BoNT/A), which is an **effecti**ve **therap**eutic **treatment** in cases of spinal cord injury,
- platform for the production of DNA vaccines that are safe and more **effecti**ve against pathogen infections, based on the fusion of a plant **protein** signal sequence with that of a viral antigen,
- mimetic peptides that, through a novel **molecular mechanism**, modulate the density and function of the L-type calcium channel and make it possible to treat those acquired or genetically-based human **disease**s associated with altered **cellular** calcium homeostasis, such as various cardiovascular, neurological and urological **disease**s,
- by means of computational virtual screening techniques, a family of pharmacologically active **molecule**s has been identified that have been shown to interfere with the **mechanism** underlying the development of rheumatoid arthritis pathology. The **compounds** identified here thus open up new possibilities in the treatment of this **disease**,
- new pathogenic **mechanism** involving the p75NTR receptor, which reduces the production of inflammatory cytokines, promising their use for the **treatment** of chronic inflammatory **disease**s of auto-inflammatory or autoimmune origin,
- innovative implantable device made for sustained release of multiple drugs.

The last cluster (n.8), named “biomedical”, includes mainly methods and techniques for disease diagnosis and monitoring:

- instrument for automatic analysis and recognition of lung sounds acquired through an electronic stethoscope,
- predictive kit for response to radio chemo**therapy** in **patient**s with locally advanced cervical cancer,
- development of an algorithm to standardize each decision-making **process** and provide the clinician with support for the **diagnosi**s of oral health conditions,
- patented **diagnost**ic kit enables rapid and early differential **diagnosis** of ulcerative recto colitis,
- computational **method**ology that allows the operator to perform a semi-quantitative analysis of MRI data,
- software that performs post-**process**ing analysis of DWI images and able to calculate the volume of the ventricles of the fetal brain,
- **method** for simulating coronary changes and/or assessing the risk of myocardial ischemia,
- device that measures the electrical characteristics of small areas of living **tissue** over a wide frequency range, with the aim of improving the accuracy of **method**s for cancer **diagnos**is; in fact, measuring **tissue** homogeneity and displaying impedance maps enables precise analysis of the region of interest by defining the margins of the neoplastic formation; this can help in many **diagnost**ic procedures currently performed by **surgica**l, endoscopic, and radiological techniques;
- **method** to support planning of linear stereotactic trajectories for implantation of intracerebral devices, such as multi-contact recording and/or stimulating electrodes, biopsy probes, laser light applicators; applicable in the field of invasive **diagnost**ics and **surgical** treatment of neurological **disease**s (Epilepsy, Parkinson's),
- process of analyzing an individual's voice to investigate his or her health status, particularly to facilitate the **diagnosi**s of **disease**s and/or disorders, both potential and overt.

We attempt to give an interpretation to these twenty most frequent words in order to make them part of an HC4.0 vocabulary.

Certainly, the goal of HC4.0 is to provide patients with better, value-added and cost-**effective** healthcare services (Al-Jaroodi et al., 2020) and improve the **effectiveness** and efficiency of the healthcare sector by trying to connect **patients**, physicians, hospitals, personal medical devices, **pharmaceutical** and medical supplier’s product and service providers in order to create a smart healthcare network along the entire healthcare value chain. There is a great **control** and monitoring (and prediction) of **patient** status thanks to the enabling **technolog**ies.

As in Industry 4.0, IoT, RFID, wearable devices, robotics, and blockchain **technolog**ies in cyber-physical **system**s create a **mechanism** for data collection, monitoring, analysis, intervention, and feedback (Li and Carayon, 2021). Linking these **technolog**ies to personalized medicine can help implement **genetic**s-based approaches to **diagnosi**s and **treatment** and improve the effectiveness of **patient** care.

In healthcare, machine learning and artificial intelligence analyze huge amounts of information (big data) to provide accurate medical **diagnosi**s before treatment is too late. Artificial intelligence **technolog**ies are increasingly being used to fight **disease** and save more lives by trying to **diagnos**e **disease**s such as **tumor** early (Haleem et al., 2022). Big Data helps determine the **effectivene**ss of medical **treatments** as well as identify effective, standardized **therap**ies for specific disorders. Big data helps improve outcomes and reduce costs through improved **disease** management strategies and the development of better **diagnosi**s and **treatment process**es (Haleem et al., 2022). Digital health tools promise to improve individual health care delivery and increase the ability to effectively detect and **treat disease**. In addition, **technolog**ies such as cell phones, Internet applications, social networks offer new **method**s for patients to monitor their health and access information.

In the field of telepathology, telemedicine and disease monitoring, even **surgery**, with the use of AI, is being increasingly digitized (**Surgery** 4.0) (Feußner and Park, 2017; Jell et al., 2019; Teber et al., 2020), with the ultimate aim of improving results and reducing costs. AI can indeed improve the speed of operations (Dahl and Kamel Boulos, 2013; Aceto et al., 2020) and support doctors by reducing the risk of tremors or other unwanted or involuntary movements during **surgery**. Tele**surgery** represents the ultimate evolution in which the surgeon and his cockpit are physically distant from the operating room (Thuemmler and Bai, 2017; Aceto et al., 2020).

Through the application of communication tools to patients and medical teams, the transfer of **treatment** from the hospital to the home is intensified, without disruption in outpatient services. Through the use of **technolog**y, classical **pharmac**ological **therapi**es are replaced with the support of digital applications provided jointly to the patient and the physician, in a perspective of better and more integrated adherence to care. Nowadays, the techno-scientific evolution of medicine is essentially taking place along at least three main axes: restorative/integrative medicine, regenerative medicine and precision medicine. The three perspectives of development have several points in common, but above all they are closely linked to the possibilities offered by the information **technolog**y revolution, linked to artificial intelligence (Cappelletti, 2018).

Integrative Medicine is originally "restorative" medicine, traditionally prosthetic (artificial devices designed to replace a missing body part) and more recently bionic, a branch of biomedical engineering that applies cybernetics to the reproduction of **function**s of living organisms.

Tissue engineering is an interdisciplinary area concerned with the development of **function**al 3D **tissue**s by mixing scaffolds, **cells**, and bioactive chemicals. It is a subset of biomedical engineering in which cell biology, materials science, chemistry, **molecular** biology, and

medicine converge. It applies to the repair, restoration, and preservation of damaged **tissue**, **articulation** or a whole organ (Haleem et al., 2022).
Regenerative Medicine is based on the principles of **genetic** engineering-isolating a **gene** from the organism that possesses it and inserting it into a host even of a different species-and is developed as **gene therapy** (reconditioning of cells in vivo), use of stem **cells** (ex vivo and in vivo cell regeneration), and **tissue** engineering (combination of artificial cells and materials) for anatomo-functional restoration of **tissue**s and organs (Jessop et al., 2016). The most promising regenerative medicine technique is that based on stem **cells** (SC) - embryonic, adult, and "induced" - with the ability to reproduce **tissue**s and organisms.
Precision Medicine tends to the **treatment** and prevention of **disease** based on individual variability of **gene**s, environment and lifestyle (personalization) and is based on deterministic understanding of **disease**, **diagnos**is of causal factors, ability to treat root causes of **disease**, using tools such as **geno**mic and post-**geno**mic biological databases, characterization methods such as "omics," cellular analysis and "mobile" **technolog**y, and bioinformatics (Cappelletti, 2018).
In HC4.0, the entire healthcare **system** and its management take on an even more key role; all stakeholders in the healthcare ecosystem are actively involved in supply chain logic. In healthcare 4.0, as automation and the use of **technolog**y increase, the participation and importance of people actually become more critical. Not only patients, physicians, and relevant support staff are included in the **system**, but also nurses, physician assistants, **pharma**cists, lab technicians etc (Li and Carayon, 2021). It is important to focus on **system** and **process**: team, team of teams, network of care, **process** spread across organizational boundaries, community involvement.

8.4 Conclusion

In conclusion, by using NLP, clustering and complex network techniques, an Italian patent platform, where the innovation of about 90 research centers and universities are collected, was analyzed in order to identify the innovative technologies in a particular research cross-domain field such as the Healthcare 4.0.
The analysis of the Italian patent heritage on the 4 healthcare-related clusters reveals a great deal of innovation focused on different areas, some of which are the most important healthcare sectors where artificial intelligence is identified to be having an impact, such as in robotic assisted surgery, personalized medicine, drug discovery, imaging&diagnostic, diseases monitoring, assisted living, automated clinical decision support, elderly care, and so on. Healthcare 4.0, in fact, is focused on patient management and care, with a key promise of the developing technology being the identification of patient-centric, data-driven tools to improve treatment regimens, hospital workflows and disease prevention.
The intrinsic multi-disciplinary nature of HC4.0, including many technical fields and deeply involving non-technical areas as well, makes it harder and harder for the operators and stakeholders in this field to keep the pace with technological progress. Therefore, precisely

in HC4.0 area, the knowledge/technology transfer, the creation of networks and ensuring easier access to IP-protected assets are especially important in order to develop applied research and make it quickly "usable" innovation according to continuous open innovation schemes. In particular, the involvement of all actors in the innovation chain, which in the health sector becomes even more important, represents the emblem of the quadruple helix model where, with public and private stakeholders and university, great importance is laid on citizens and their needs, above all in the development of health, social and other related services. This model brings greater social benefits and empowers citizens/patients who are not only passive consumers of content/services but take on the role of creators of innovation, in a patient-centered perspective.

CHAPTER 9. Discussion and conclusion

This research path revealed multiple weaknesses in the process of innovation diffusion, of the demand-supply intersection, mainly by espousing the perspective of those who produce innovation, i.e. the numerous research institutions and universities.

It is interesting, however, to try to look at the problems also from the perspective of those who adopt innovation, thus attempting to arrive at a complete multi-perspective view.

In an attempt to articulate the discussion in a coherent manner while clearly highlighting the problems and possible solutions available, this chapter is organised as follows:

- section 9.1 discusses the work by looking at what has been done from the perspective of those producing the innovation;
- section 9.2 provides a re-reading of the problems and solutions proposed during the thesis from the perspective of those who research and adopt innovations;
- section 9.3 summarises the lessons learnt.

9.1 The Producers' Perspective

This section draws conclusions regarding the problems and possible solutions proposed in the research work from the point of view of those who produce the innovation. Figure 37 describes and summarizes the logical path followed throughout the entire thesis work, as narrated in the context of the preceding 8 chapters, aimed at addressing the 5 research questions.

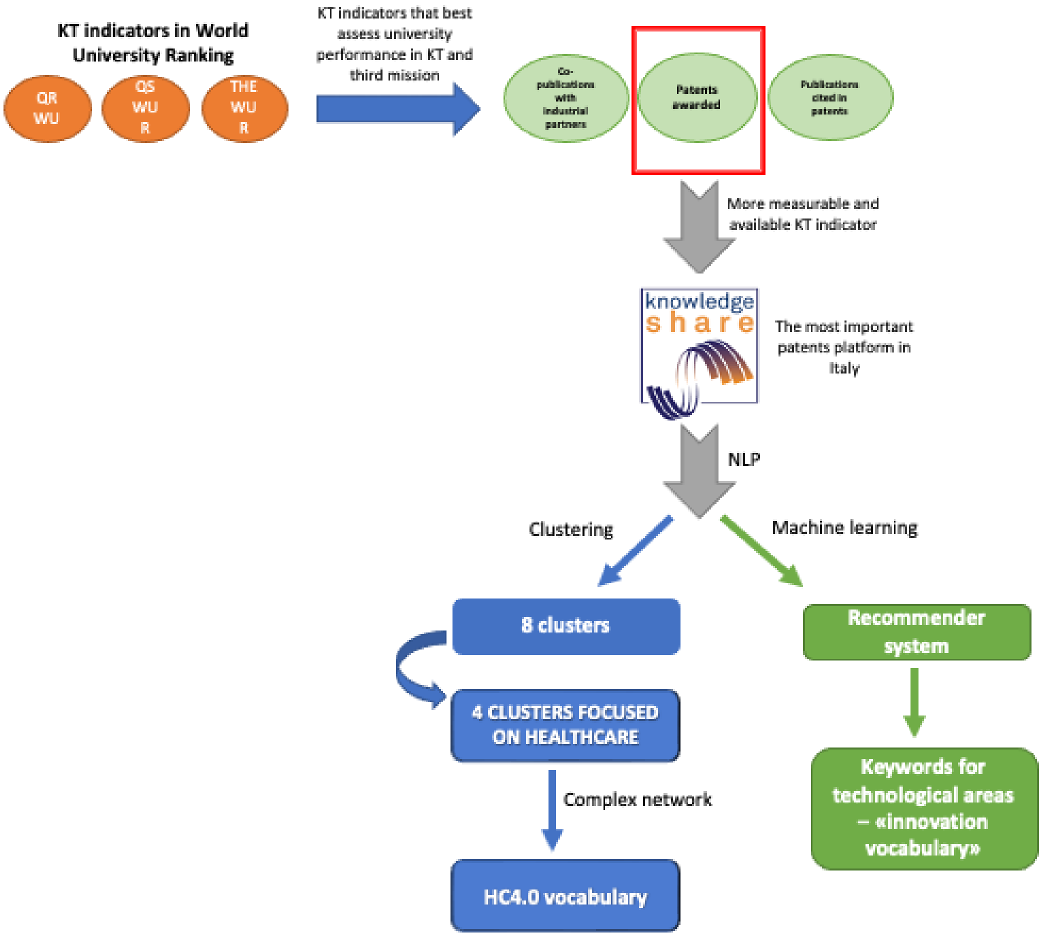


Figure 37: the logical path followed throughout the work

The research work begins with its initial research question (**RQ1**): How do the world's top universities, as per global university rankings, perform in terms of knowledge transfer? In simpler terms, can the commonly used rankings adequately assess universities' third mission performance? To address this question, we conducted an analysis of the most well-known global university rankings, including The Academic Ranking of World Universities (ARWU), The QS World University Rankings® (QSWUR), and The Times Higher Education World University Rankings (THEWUR), to identify the world's leading universities. One notable observation that arises is the absence of specific Knowledge Transfer (KT) indicators in these rankings.

The European Commission introduced U-Multirank (UMR), a ranking system that adopts a distinct approach in comparison to the established global university rankings. UMR incorporates a set of nine indicators specifically focused on Knowledge Transfer (KT). Consequently, during the research work, our initial analysis aimed to determine whether the universities identified as top performers in the global rankings also excel in terms of their third mission activities. When assessed using the dedicated Knowledge Transfer indicators

of U-Multirank 2020, it becomes evident that only three of the world's top-ranked universities manage to lead in this particular area. Out of the 30 universities that excel in Knowledge Transfer, a mere five are part of the prestigious T_{10} group of universities.
The analysis utilizing the MAD (Median Absolute Deviations) indicator has aided in identifying three key indicators that effectively assess universities' performance in terms of KT activities: Co-publications with industrial partners, Patents awarded (in absolute numbers), and Publications cited in patents. Furthermore, the research endeavors to elucidate, through a focused bibliographic analysis, why these indicators are not only interesting but also valuable for interpreting KT performance.
The evaluation of the third mission remains a complex task; the challenge often lies in our ability to measure only what can be quantified. Numerous activities, particularly those associated with KT, remain elusive to measurement and quantification, especially when they unfold through unintentional mechanisms. This issue has also sparked a scientific debate regarding the measurement of the impact and societal benefits stemming from KT activities.
Of the three indicators that best discriminate the performance of universities in terms of knowledge transfer and third mission, the work focused on the number of patents granted, which offers a more accessible dataset, with numerous patent databases being freely accessible for consultation. Given that a significant portion, ranging from 70% to 90%, of technological information remains unpublished except in patent documents, patents represent one of the most valuable sources of information.
Due to these reasons, the research primarily centered around patent analysis and patent matchmaking platforms. To achieve this, the methodologies of Intellectual Property Analytics were employed, representing a multidisciplinary approach aimed at extracting valuable insights from intellectual property data.
University patents are too often not valued and transferred, so patent matchmaking platforms have sprung up over time. This thesis focused on Italy's largest and most significant patent platform, Knowledge Share (KS), offering unrestricted access to the public and containing a total of 1694 patents, uploaded by 89 Italian Research Centers, both public and private institutions such as Universities, Research Centers, Scientific Institutes for Research, Hospitalization, and Healthcare organizations. The patents in KS are classified in 10 technological areas:

- Aerospace and aviation;
- Agrifood;
- Architecture and design;
- Chemistry, Physics, New materials and Workflows (Basic Science);
- Energy and Renewables (Green Energy);
- Environment and Constructions (Environment);
- Health and Biomedical (Biomed);
- Informatics, Electronics and Communication System (Electronics);
- Manufacturing and Packaging (Packaging);

- Transports.

Using NLP, clustering and regression techniques, the research work tried to answer the following research questions:

- are the classification taxonomies used in the patent platforms effective in classifying the whole landscape of academic patents (**RQ2**)?
- is it possible to support the user in correctly classifying a patent entered into the platforms in order to improve the matchmaking between demand and supply of innovation (**RQ3**)?
- is it possible to draw up an attempted vocabulary of technological fields from the keywords that emerged from an applied AI-based approach (**RQ4**)?

The clustering analysis reveals the presence of 8 homogeneous clusters instead of the 10 proposed by the KS platform:

1. Technologies 4.0 (mechanics and robotics)
2. Material science
3. Cancer treatment
4. Optics - Image processing
5. Sensor technology - ICT
6. New molecules - new compounds - pharmacology
7. Energy/green Technologies
8. Biomedical

This result suggests that there might be potential inconsistencies within the conventional classifications used by Knowledge Share (KS). These inconsistencies are likely attributed to the emergence of new technologies or cross-domain areas, such as Technologies 4.0. Consequently, this implies that there is considerable margin for improvement in the taxonomy used by KS to more effectively capture and represent patent content.

To assist users in the search and classification of patents, our research introduces a comprehensive automated patent multi-classification system. This system harnesses NLP techniques and Machine Learning to emphasize the keywords that define classification categories. An important feature of our methodology is its ability to assign patents to one or more labels simultaneously, acknowledging the potential for emerging technologies to transcend traditional domain boundaries. Our method, openly accessible, serves two primary purposes: firstly, it benefits patent managers by providing reliable category recommendations for patents and the associated technologies they encompass; secondly, it aids end-users of technological platforms, such as investors and companies, by offering a quantitative assessment of the similarity between the technologies they are searching for or intend to patent and the most appropriate technological domains. Consequently, our recommender system streamlines the process of matching innovation supply with demand, a critical factor in generating social and economic impact with emerging technologies.

Conversely, the methodology enables the identification of **keywords that characterize a patent** in an objective and independent manner, devoid of human subjectivity. This aspect proves highly valuable in forming an initial vocabulary composed of words extracted from patents. Ultimately, it can contribute to the redefinition of existing categories, thereby supporting portal management and enhancing the matchmaking process among users and patent owners.

Among the 8 clusters emerged, 3 ones are dedicated to the healthcare sector, a research cross-domain field. The patent analysis methods described have been employed within the particularly influential field of Healthcare 4.0 (HC4.0). This is a crucial sector characterized by ongoing evolution and by the application of cross-cutting and multidisciplinary technologies and innovations. HC4.0 is a relatively new concept that has evolved from Industry 4.0. It is used to depict the gradual integration of typical Industry 4.0 technologies, such as the Internet of Things (IoT), Industrial IoT (IIoT), cognitive computing, artificial intelligence, cloud computing, fog computing, and edge computing, into the healthcare domain. Within this emerging paradigm, Cyber-Physical Systems (CPS) play a pivotal role in shaping digital health systems that encompass products, technologies, services, and businesses.

HC4.0 is a multidisciplinary and complex field where knowledge and technology transfer, as well as the establishment of networks and collaborations, play a pivotal role in advancing applied research and swiftly converting it into practical innovations. Nonetheless, owing to its multidisciplinary and highly innovative nature, healthcare-related patents frequently find themselves within what can be termed as "monster categories". These categories encompass a broad spectrum of multidisciplinary patents that are inadequately characterized and classified. Monster categories prove ineffective in distinguishing and categorizing patents, rendering them challenging to explore and utilize efficiently.

In order to answer to **RQ5**: is it possible to define a first draft of HC4.0 vocabulary for characterizing the innovations and innovative technologies in healthcare 4.0?, by using NLP, clustering and complex network techniques, an analysis was conducted on KS patents with the aim to identify innovative technologies within the specific cross-domain field of Healthcare 4.0.

The examination of the Italian patent repository, particularly within the four healthcare-related clusters, revealed a wealth of innovation spanning various areas. Notably, these areas encompass some of the most critical sectors in healthcare where artificial intelligence is making a significant impact. These sectors include robotic-assisted surgery, personalized medicine, drug discovery, imaging and diagnostics, disease monitoring, assisted living, automated clinical decision support, elderly care, and more.

Healthcare 4.0 primarily revolves around patient management and care. A fundamental aspect of this evolving technology is the development of patient-centric, data-driven tools aimed at enhancing treatment protocols, optimizing hospital workflows, and preventing diseases.

Through the complex network approach, a first attempt of HC4.0 vocabulary was drafted. As described in Chapter 7, through the AI-based approach, keywords characterizing the innovation landscape in Italy were identified. Focusing the attention on Healthcare sector, the following compares the words that emerged from the AI-based approach and those that emerged from the complex network approach (in alphabetical order).

Keywords from complex network approach	**Keywords from AI-based approach**
Disease	Blood
Protein	Bone
Tissue	Brain
Technolog	Cancer
Cell	Cell
Method	Clinical
Effectiv	Diagnos
Patient	Diagnostic
Diagnos	Disease
Gen	Drug
Surger/Surgical	Gene
Mechani	Human
Model	Invasiv
Tumor	Molecul
Process	Phatologhy
Articul	Patient
Treat	Surgery
Function	Surgical
Pharmac	Therapeutic
Therap	Therapy
Compound	Tissue
Molecular	Treatment
Control	Tumor
System	Ultrasound

The comparison shows that the difference consists of 8 words that enrich the initial vocabulary:

- blood
- bone
- cancer (equal to tumor)
- clinical

- drug
- human
- invasive
- pathology
- ultrasound

Finally, the initial vocabulary of HC4.0 consists of 33 words that emerged from the content of Italian patents.

The rapidly evolving field of digital health is revolutionizing conventional healthcare practices. Artificial Intelligence is emerging as a pervasive technology, with applications ranging from healthcare chatbots, diagnosis, care direction, medical image enhancement, to clinical data interpretation, among others. Smartphones are being adapted for various healthcare purposes, including self-monitoring, self-care, communication with healthcare professionals, health data recording, and clinical guidance. The future landscape of healthcare is undeniably shaped by digital technologies.

9.2 The adopters' perspective

This section provides a rereading of the conclusions reached following the research process from the perspective of those who adopt an innovation.
To this end, it is interesting to look back to a reference system proposed years ago by the sociologist Rogers Everett (2003). The model allows us to approach the problem of innovation from the perspective of those who adopt it, thus giving us the opportunity to compare points of view.
Everett identifies multiple factors that influence innovation: the innovation itself, the different categories of those who adopt an innovation, the communication, the time and the territorial context. We analyze the main ones below.

With regard to **innovation**, the characteristics that are most conducive to its adoption are mainly the degree of compatibility with processes and tools already in use, the possibility of experimenting with it before finally adopting it, the availability of data and evidence on its effectiveness in relation to competing innovations, and the ease of adoption and learning.
The research carried out in this PhD unfortunately shows that, for example, the patents, on which we have focused most, are often described in an overly technical and inadequate manner, thus making it impossible to assess the key aspects identified by Everett. The Knowledge Share platform, presented in Chapter 4, with the adoption of marketing sheets, represents an attempt to overcome these critical issues, although, as the analyses conducted have shown, it suffers from further classification problems that could make the contents sometimes more difficult to find with negative repercussions on the supply-demand matching of innovation. Analyses in this thesis have shown that the multidisciplinarity and complexity

of the today innovations sometimes escape the traditional channels of research and thus technology transfer. Universities engaged in research and development in unconventional scientific domains, such as cross-domain and emerging field, are seeking alternative technology transfer processes. These should be more tailored and realistic, avoiding the one-size-fits-all approach often associated with traditional linear technology transfer models (Karanikic et al., 2019). Technology Transfer Offices should align with the evolving landscape of commercializing research outcomes in the digital economy era. They have to revise and implement their current policies for technology transfer, intellectual property protection, and management, along with commercialization strategies to effectively introduce their technologies to the market.
A more powerful supply-demand matchmaking tools and strategies are needed in order to empower technology transfer. For these reasons, the use of support systems for users, companies and technology transfer experts, such as the recommender system proposed in Chapter 7, can be crucial in reducing the gap between demand and supply of innovation, especially in highly multidisciplinary domains as HC4.0.
The transfer and commercialization of emerging and frontier technologies, such as those developed in HC4.0, may prove even more complicated. This thesis work has shown that it is also complex to monitor third mission activities, which, more broadly, involve actions and activities that are sometimes not easily quantifiable, measurable and monitorable.
Further key elements according to Everett are **communication and time**. Nowadays, most of the communication is social, and in terms of timing and content, it is incompatible with the timing of empirical evidence production and the content of academic production. Unfortunately, academic production and the research results often appear slow in the diffusion of innovation and, as described in Chapter 1, succumbs to Death Valley.
In knowledge transfer, communication is highly critical. In fact, studies conducted during the PhD revealed how IP management is not trivial and, specifically patents, are difficult to access and, therefore, poorly communicated and valorised. Not surprisingly, online patent platforms have emerged over time as convenient channels for patent transfer, with the joint effort of both academic, industrial and political partners. Unfortunately, existing initiatives/platform used for matchmaking between supply and demand for innovation are sometimes ineffective and not easily available: they often are paid services, not open, they report the patent document as such, without a usable "translation" for all that facilitates matching. The classification of the patents in a given technological area or sector is a challenging task. Often this classification and, therefore, the search and consultation method used, are based on taxonomies and keywords self-defined by users, experts or database managers and are not very effective.
These problems became clear in Chapters 6 and 7. Chapter 6 highlights how the classification adopted by KnowledgeShare based on 10 categories differs from the classification with 8 clusters obtained by applying NLP and clustering techniques. And in Chapter 7, methods are proposed to improve the classification of innovations, improving the classification taxonomies used and identifying more meaningful and discriminating keywords useful in the

search for an innovation. The most striking case of misclassification, as shown in Chapters 7 and 8, is certainly that of the health sector. It contains a significant number of poorly characterized and classified patents, and thus it represents the so called "monster category". Healthcare related innovations, due to multidisciplinary and particularly innovative nature, tends to generate classes with a large number of patents. Also for this reason, during the research period, the healthcare sector was investigated with particular attention, given its strategic importance and the poorly accurate exploration techniques. Chapter 8 present a tentative vocabulary for better characterize this critical sector.

Equally critical is the time factor. In fact, in the health sector, R&D activities are particularly complex as innovative clinical technologies are subjected to heavily regulated and supervised validation process, before they can obtain approval and placed on the market. For this reason, the development of innovation in this context can be characterised by a very long time-to-market and a very high attrition rate. These elements are often synonymous with high investments for those wishing to realise innovative solutions in the clinical field and sometimes represent a barrier to innovating autonomously as far as clinical practice.

In the past, players in the innovation ecosystems, especially in the healthcare sector, have pursued and experienced a predominantly closed innovation model, with the tendency of companies to conduct research and development internally to protect intellectual property, hoping to generate a competitive advantage. In recent years, however, it is evolving towards a digital, multi-channel paradigm, based on an increasingly patient-centred research and delivery models co-created in a multi-stakeholder perspective, part of an increasingly articulated and extended health chain.

Coming back to the analysis of the keywords carried out in Chapter 7 and 8, it is interesting to note how digital technologies are pushing innovation for both academics and practitioners, especially in the health sector. HC4.0 lies at a unique crossover of disciplines, between medicine, psychology, sociology, engineering, business management, marketing, pharmaceuticals, IT, devices, Industry 4.0. It is necessary a comprehensive and interconnected perspective, providing an integrated analysis that encompasses different domains and involving different actors: practitioners, industries, universities and public institutions. It is indispensable to foster continuous dialogue and cooperation between regions, healthcare companies, patient associations, industry, business, technology and innovation managers.

Digital technologies are improving several aspects, including diagnostic and therapeutic procedures through the use of artificial intelligence for diagnostics and surgical decision-making, as well as the development of telemedicine tools for patient counselling and monitoring. In addition, they have facilitated the redesign of more efficient organisational and administrative processes aimed at delivering cost-effective healthcare services (Biancone et al., 2021; Cobianchi et al., 2022; Miceli et al., 2021). Digital technologies such as social media platforms have been shown to enhance public health education and communication efforts. New robotic systems and surgical tools are improving the safety and efficiency of surgeries, benefiting patients, surgeons, and society as a whole. Automation

and 3D printing are introducing innovative devices for patient care, while big data is facilitating the collection of valuable information for training healthcare professionals and optimizing healthcare techniques (Dal Mas et al., 2023).
This transformation contributes to better overall health among the population. WHO Global strategy on digital health 2020-2025 states: “The vision of the global strategy is to improve health for everyone, everywhere by accelerating the development and adoption of appropriate, accessible, affordable, scalable and sustainable person-centric digital health solutions to prevent, detect and respond to epidemics and pandemics, developing infrastructure and applications that enable countries to use health data to promote health and well-being, and to achieve the health-related Sustainable Development Goals and the triple billion targets of WHO’s Thirteenth General Programme of Work, 2019–2023” (WHO Global strategy on digital health 2020-2025).
How does intellectual property contribute to the advancement of Digital Health? With the ongoing expansion of the digital health sector, IP has gained significant relevance within the industry. IP plays a fundamental role in advancing Digital Health by providing essential protection for the innovations developed by scientists and entrepreneurs. Simultaneously, it ensures that users can readily access and benefit from these innovations. Whether it involves a groundbreaking software algorithm for remote patient monitoring or a mobile application offering personalized fitness plans and health tracking features, IP safeguards the development of unique and transformative digital health solutions. These solutions can then be disseminated widely to enhance healthcare outcomes for individuals worldwide. By preserving and fostering creativity and innovation within the digital health sector, IP also aligns with the United Nations Sustainable Development Goals (SDGs), particularly SDG3 "Good Health and Well-being" and SDG9, which focuses on "Industry, innovation, and infrastructure". However, there is still much to be accomplished to make digital health solutions universally accessible, bridging the "digital divide" and ensuring equitable access for all.
A further key factor for Everett is ultimately the **context and the ecosystem** in which innovation develops. And in this regard, it is interesting to note the composition of the clusters obtained in Chapter 5. It appears to take into account the geographical dimension, i.e. the industrial, social and economic environment itself can affect the KT and third mission activities, thus suggesting that there are contextual factors that the purely quantitative analysis used by global university rankings fail to grasp or bring out. In Chapter 5 it appears evident as the Californian universities and, to a lesser extent the Japanese and some Asian ones, stand out in the KT even though they are not among the best in the traditional global university rankings.
Therefore, the particularly lively and innovation-friendly territorial context and the presence of an innovation ecosystem allow for the emergence and affirmation of universities and research centres that are particularly successful in knowledge transfer.
Today collaboration patterns acquire an extraordinary enabling value for research and technology transfer and, in particular, innovation ecosystems, i.e. networks of highly

qualified and internationally recognized public and private actors that operate in synergy with the aim of fostering the interaction, creation and promotion of innovation in a specific area of common interest, consistent with the industrial and research vocations of the reference territory (Deloitte, 2023). Definitely, IP can serve as a catalyst for collaboration and innovation.
Furthermore, we have highlighted how crucial is therefore the involvement of the entire value chain and the various actors that make it up. The Universities and Research Centers, in particular, among all the actors involved in this ecosystem, listed below, play a predominant role as a connection point and facilitator of interactions, like an enzyme in a reaction:

- Universities and public and private research centres, *in primis*, have a decisive role in the generation, nurturing and development of technologies and research;
- Enterprises, start-ups, SMEs and large companies in the pharma insurance, technology and medical devices, have the ask of integrating clinical and technological innovation into the productive context, thereby creating value for the territory and society;
- Private hospital organisations, key players in the development of research and clinical innovation, are also configured as users of new technologies, with a connecting role between scientific knowledge and its application to clinical practice;
- Incubators and accelerators act as a link between entrepreneurial realities, the state, public authorities and the private investor market and support start-ups with tools and dedicated assistance to facilitate start-up and growth phases;
- Financial stakeholders - from venture capital actors to public financial institutions - are called upon to provide capital for the growth and development of innovations;
- Public institutions, finally, have the task of defining policies to decline the process of transforming knowledge into economic benefits and direct it towards the aims of the country's industrial policy.

In this direction, in Italy, the National Recovery and Resilience Plan (PNRR), as an accelerator of innovation in multistakeholder ecosystems, has open innovation and technology transfer as key elements to enable the country's innovation strategy, enable the transmission of results, resulting from research, to the market, economy and society, and facilitate the conversion of experimental ideas into products, services and entrepreneurial initiatives, making a decisive contribution to the development and competitiveness of the scientific-industrial ecosystem.
The objectives of the PNRR are declined and implemented through six Missions, which constitute the fundamental pillars for the transition to economic development based on innovation, digitization, sustainability and social cohesion. In particular, Mission 6: Health has as its overall objective the strengthening of the health system through the creation of new infrastructures and modernization of existing ones, scientific research, technology transfer, digital and technological transition. Investments of 15.63 billion euro are planned for Mission 6.

The PNRR stands as a truly transformative project, a tool for accelerating innovation, which aims to reform the entire supply chain, from basic research to technology transfer and the actual development of pilot projects; it incentivizes new governance models, opens up participation in funding also to private investors or to public-private partnerships. In fact, five Research National Centers, fourteen Extended Partnerships and up to thirty Research and Technological Innovation Infrastructures have been financed, in which projects technology transfer plays a key role.

Focusing on the Apulian ecosystem, in this concluding section, it can be said that the Apulian research system is actively involved in all PNRR projects, with a funding of almost 300 million euro in the only province of Bari.

Apulia Region has always famously developed a growth strategy based on innovation and the valorization of young talent. Innovation is one of the pillars of regional industrial policy in the new SmartPuglia 2030 strategy, approved in 2022; Apulia ranks eighth in Italy in terms of the number of start-ups and innovative SMEs, and third and second respectively in the Mezzogiorno. In fact, it has 636 start-ups in addition to 111 innovative SMEs, both of which are registered in the relative categories of the Business Registry.

Innovation, research, incentives and innovative finance: the Puglia ecosystem offers a series of tools and interventions to companies that decide to invest in the region. Renewable energies, the aeronautical and aerospace sector, technological start-ups, biotech, and collaboration between the university and business worlds, strategies for the ERDF and ESF 2021-27 Funds: attractive drivers for those who choose entrepreneurial and scientific dynamism and quality of life 'made in Puglia'.

Today, Apulia Region is developing an innovation ecosystem that has great potential for attracting major investors, both in Italy and abroad. In fact, Cassa Depositi e Prestiti (CDP), a publicly controlled joint-stock company, which has the Italian Ministry for the Economy and Finance as its majority shareholder, opened the first accelerators in Apulia and signed agreements with Apulia's leading universities for Tech Transfer Funds. CDP Venture Capital Sgr supports research results at different stages of maturity, offering different financial instruments depending on the TRL of the research results: from Proof of Concept (PoC) to acceleration.

In recent months, the Apulia Region has launched a new instrument with the aim of fostering the growth of start-ups and innovative small and medium-sized enterprises, born from the research results, many of them from Apulia's universities and research institutions. It is called “Equity Puglia” and is the latest and most innovative financial instrument desired by the Apulia Region. The new instrument makes it possible to increase the level of capitalization and thus the equity strength of companies and does so with the collaboration of specialized investors. The initial budget is EUR 60 million and will make it possible to inject at least EUR 120 million in new capital into start-ups and innovative small and medium-sized enterprises. Co-investment funds will be provided for companies that, in line with the S3 Strategy, can be traced back to the four cross-cutting themes that determine challenges and opportunities for all sectors: environmental sustainability and the circular economy; information technologies for industry and society; life sciences and health technologies; blue growth and sea economy.

All these efforts by the regional stakeholders, together with the Apulian research system, aim to reduce the gap that is the Valley of Death and that prevents so many research results from ever reaching the market.
This effort must be even greater for the healthcare sector because without technology&knowledge transfer in this field, the innovations developed in research centers could never translate into improving people's lives and saving lives.

9.3 Lessons Learned

The lessons learned during this research work are the following:

1. It is difficult to assess, monitor and compare the knowledge transfer and third mission activities carried out by research institutions because they are varied, complex and not always quantifiable (especially when they occur through unintentional mechanisms).
2. Among the indicators that can best discriminate are patents granted, which today represent a mine of information on research results.
3. Patents are still poorly valorized and remain stuck in the Valley of Death. Less than 20 per cent of the technologies patented by research centers in Europe are exploited.
4. To cope with this under-exploitation, numerous platforms for matching innovation supply and demand have sprung up. In Italy, the most important platform created by Ministry of Enterprises and Made in Italy is Knowledge Share, where currently almost 90 universities, IRCCS and research institutions upload their patents, translating them into a simple language that everyone can understand.
5. Using the main techniques of Intellectual Property Analytics, we clustered Italian innovations and analyzed the main keywords emerging from the content of patents.
6. We developed a recommendation system for users of the KS platform to support them in ranking patents and thus improve the matchmaking between supply and demand.
7. The Healthcare sector has a strong presence in the innovations in KS; due to its multidisciplinary character, to its continuous evolution and the presence of emerging technologies, often gives rise to so-called monster classes because they are characterized by a high number of patents, which are difficult to classify.
8. The focus is on HC4.0 or healthcare in which artificial intelligence is having a significant impact. These areas include robot-assisted surgery, personalized medicine, drug discovery, imaging, disease monitoring, assisted living, automated clinical decision support, elderly care and more.

9. Using a complex network approach, the main keywords featuring Italian innovations in this field were identified, trying to draw up a first draft of a vocabulary useful to better characterize HC4.0 innovations starting from the contents of Italian patenting.
10. HC4.0 is a multidisciplinary and complex field in which knowledge and technology transfer, along with the establishment of networks and collaborations, play a crucial role in advancing applied research and rapidly translating it into practical innovations.
11. In Italy and Apulia Region, great efforts are being made to allow excellent research results to go beyond the Valley of Death and, in particular, to allow innovations in the Healthcare sector to become products, processes and services that can be used by society to improve people's lives and save human lives.

REFERENCES

1. Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, *37*, 3-13.
2. Abdi, H. (2007). Singular value decomposition (SVD) and generalized singular value decomposition. *Encyclopedia of measurement and statistics*, 907-912.
3. Abreu, M., Demirel, P., Grinevich, V., & Karataş-Özkan, M. (2016). Entrepreneurial practices in research-intensive and teaching-led universities. *Small business economics*, *47*, 695-717.
4. Academic Ranking of World Universities (ARWU) https://www.shanghairanking.com/rankings/arwu/2020 (last access 20/09/2021).
5. Aceto, G., Persico, V., & Pescapé, A. (2020). Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0. *Journal of Industrial Information Integration*, *18*, 100129.
6. Agasisti, T., Barra, C., & Zotti, R. (2019). Research, knowledge transfer, and innovation: The effect of Italian universities' efficiency on local economic development 2006− 2012. *Journal of Regional Science*, *59*(5), 819-849.
7. Aguillo, I., Bar-Ilan, J., Levene, M., & Ortega, J. (2010). Comparing university rankings. *Scientometrics*, *85*(1), 243-256.
8. Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1295.
9. Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management*, *39*(1), 45-65.
10. Al-Jaroodi, J., Mohamed, N., & Abukhousa, E. (2020). Health 4.0: on the way to realizing the healthcare of the future. *Ieee Access*, *8*, 211189-211210.
11. Alexander, A., Martin, D. P., Manolchev, C., & Miller, K. (2020). University–industry collaboration: using meta-rules to overcome barriers to knowledge transfer. *The Journal of Technology Transfer*, *45*, 371-392.
12. Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
13. Altuntas, S., Dereli, T., & Kusiak, A. (2015). Forecasting technology success based on patent data. *Technological Forecasting and Social Change*, *96*, 202-214.
14. Amoroso, N., Bellantuono, L., Monaco, A., De Nicolò, F., Somma, E., & Bellotti, R. (2021). Economic interplay forecasting business success. *Complexity*, *2021*, 1-12.
15. ANVUR - https://www.anvur.it/attivita/temi/ (last access 15/09/2023)
16. Archive: Patent statistics - Statistics Explained https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Patent_statistics&oldid=112826 (last access 09/09/2023)
17. Aristodemou, L., & Tietze, F. (2018). The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information*, *55*, 37-51.

18. Aristodemou, L., Tietze, F., Athanassopoulou, N., & Minshall, T. (2017). Exploring the future of patent analytics: A technology roadmapping approach.
19. Arora, S., Ramm, C. J., Bahekar, A. A., & Vavalle, J. P. (2017). Evaluating health of emerging economies through the eyes of heart valve disease in the transcatheter era. *Global Heart*, *12*(4), 301-304.
20. Asche, G. (2017). "80% of technical information found only in patents"–Is there proof of this?. *World Patent Information*, *48*, 16-28.
21. ASTP Survey Report (2019) https://www.astp4kt.eu//assets/documents/Report%20-%20ASTP%20Survey%20on%20KT%20Activities%20FY2019.pdf (last access 15/09/2023)
22. Atun, R. (2015). Transitioning health systems for multimorbidity. *The Lancet*, *386*(9995), 721-722.
23. AUTM "Better World" - https://autm.net/about-tech-transfer/better-world-project (last access 01/09/2023)
24. Ayer, T., Ayvaci, M. U., Karaca, Z., & Vlachy, J. (2019). The impact of health information exchanges on emergency department length of stay. *Production and operations management*, *28*(3), 740-758.
25. Azagra-Caro, J. M., Barberá-Tomás, D., Edwards-Schachter, M., & Tur, E. M. (2017). Dynamic interactions between university-industry knowledge transfer channels: A case study of the most highly cited academic patent. *Research Policy*, *46*(2), 463-474.
26. Backs, S., Günther, M., & Stummer, C. (2019). Stimulating academic patenting in a university ecosystem: An agent-based simulation approach. *The Journal of Technology Transfer*, *44*, 434-461.
27. Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., ... & Zhang, J. D. (2020). An introduction to machine learning. *Clinical pharmacology & therapeutics*, *107*(4), 871-885.
28. Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.
29. Baglieri, D., & Cesaroni, F. (2013). Capturing the real value of patent analysis for R&D strategies. *Technology Analysis & Strategic Management*, *25*(8), 971-986.
30. Bartoloni, S., Calo, E., Marinelli, L., Pascucci, F., Dezi, L., Carayannis, E., ... & Gregori, G. L. (2022). Towards designing society 5.0 solutions: The new Quintuple Helix-Design Thinking approach to technology. *Technovation*, *113*, 102413.
31. Batool, K., & Niazi, M. A. (2014). Towards a methodology for validation of centrality measures in complex networks. *PloS one*, *9*(4), e90283.
32. Bekkers, R., & Freitas, I. M. B. (2008). Analysing knowledge transfer channels between universities and industry: To what degree do sectors also matter?. *Research policy*, *37*(10), 1837-1853.

33. Bellantuono, L., Monaco, A., Amoroso, N., Aquaro, V., Bardoscia, M., Loiotile, A. D., ... & Bellotti, R. (2022). Territorial bias in university rankings: a complex network approach. *Scientific reports*, *12*(1), 4995.
34. Benneworth, P., & Hospers, G. J. (2007). The new economic geography of old industrial regions: Universities as global—local pipelines. *Environment and Planning C: Government and Policy*, *25*(6), 779-802.
35. Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937-1967.
36. Biancone, P., Secinaro, S., Marseglia, R., & Calandra, D. (2021). E-health for the future. Managerial perspectives using a multiple case study approach. *Technovation,* 102406.
37. Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC bioinformatics*, *14*, 1-16.
38. Blasi, B. (2023). Società e Università: Valutazione e Impatto Sociale.
39. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics reports*, *424*(4-5), 175-308.
40. Bock, H. H. (2007). Clustering methods: a history of k-means algorithms. *Selected contributions in data analysis and classification*, 161-172.
41. Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, *32*(1), 30-38.
42. Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future generation computer systems*, *56*, 684-700.
43. Bougnol, M. L., & Dulá, J. H. (2015). Technical pitfalls in university rankings. *Higher Education*, *69*, 859-866.
44. Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, *63*, 743-788.
45. Campbell, A., Cavalade, C., Conesa, F., Haunold, C., Karanikic, P., & Piccaluga, A. (2022). Knowledge Transfer Metrics: Phase II. Exploration of composite indicators for knowledge transfer.
46. Campbell, A., Cavalade, C., Haunold, C., Karanikic, P., Piccaluga, A., & Dinnetz, M. (2020). Knowledge transfer metrics. *Towards a European-wide set of harmonised indicators, Karlsson Dinnetz, M.(Ed.), EUR*, *30218*.
47. Cappelletti, P. (2018). Medicina 4.0. Un’introduzione. La Rivista Italiana della Medicina di Laboratorio-Italian Journal of Laboratory Medicine, 14(3), 131-135.
48. Carayannis, E. G., Barth, T. D., & Campbell, D. F. (2012). The Quintuple Helix innovation model: global warming as a challenge and driver for innovation. *Journal of innovation and entrepreneurship*, *1*, 1-12.

49. Carrión, A., García-Gutiérrez, V. R., Bas, M. C., & Carot, J. M. (2012). A new methodology for measuring third mission activities of universities. In *INTED2012 proceedings* (pp. 1218-1223). IATED.
50. Cerulli, G., Marin, G., Pierucci, E., & Potì, B. (2021). Do company-owned academic patents influence firm performance? Evidence from the Italian industry. *The Journal of Technology Transfer*, 1-28.
51. Cesaroni, F., & Piccaluga, A. (2016). The activities of university knowledge transfer offices: Towards the third mission in Italy. *The Journal of Technology Transfer*, *41*, 753-777.
52. Chang, A. C. (2020). Intelligence-based medicine. Artificial Intelligence and Human Cognition in Clinical Medicine and Healthcare, 397-412.
53. Charles, D. (2006). Universities as key knowledge infrastructures in regional innovation systems. *Innovation: the European journal of social science research*, *19*(1), 117-130.
54. Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. *Artificial Intelligence Review*, *52*(2), 803-855.
55. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.
56. Chen, H., & Deng, W. (2023). Interpretable patent recommendation with knowledge graph and deep learning. *Scientific Reports*, *13*(1), 2586.
57. Chen, J., Chen, J., Zhao, S., Zhang, Y., & Tang, J. (2020). Exploiting word embedding for heterogeneous topic model towards patent recommendation. *Scientometrics*, *125*(3), 2091-2108.
58. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
59. Choi, J., Jang, D., Jun, S., & Park, S. (2015). A predictive model of technology transfer using patent analysis. *Sustainability*, *7*(12), 16175-16195.
60. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational vision science & technology*, *9*(2), 14-14.
61. Choi, Y., Park, S., & Lee, S. (2021). Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data. *Scientometrics*, *126*, 5431-5476.
62. Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649.
63. Clark, B. R. (1998). *Creating entrepreneurial universities: organizational pathways of transformation. Issues in Higher Education*. Elsevier Science Regional Sales, 665 Avenue of the Americas, New York, NY 10010 (paperback: ISBN-0-08-0433545; hardcover: ISBN-0-08-0433421, $27)..

64. Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science advances*, *1*(1), e1400005.
65. Coates, H. A. M. I. S. H. (2016). Reporting alternatives: Future transparency mechanisms for higher education. *Global rankings and the geopolitics of higher education. Routledge, Abingdon*, 277-294.
66. Cobianchi, L., Dal Mas, F., & Ansaloni, L. (2022). New Frontiers for Artificial Intelligence in Surgical Decision Making and its Organizational Impacts. *Frontiers in Surgery*, 9, 933673.
67. Compagnucci, L., & Spigarelli, F. (2020). The Third Mission of the university: A systematic literature review on potentials and constraints. *Technological Forecasting and Social Change*, *161*, 120284
68. Costa, F. F. (2014). Big data in biomedicine. *Drug discovery today*, *19*(4), 433-440.
69. Council Recommendation (EU) 2022/2415 of 2 December 2022 on the guiding principles for knowledge valorisation (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022H2415) (last access 05/09/2023)
70. Couturier, J., Sola, D., Borioli, G., & Raiciu, C. (2012). How can the internet of things help to overcome current healthcare challenges. *Communications & Strategies*, (87), 67-81.
71. Couturier, J., Sola, D., Borioli, G., & Raiciu, C. (2012). How can the internet of things help to overcome current healthcare challenges. *Communications & Strategies*, (87), 67-81.
72. Dahl, T. S., & Kamel Boulos, M. N. (2013). Robots in health and social care: A complementary technology to home care and telehealthcare?. *Robotics*, *3*(1), 1-21.
73. Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological forecasting and social change*, *73*(8), 981-1012.
74. Dal Mas, F., Massaro, M., Rippa, P., & Secundo, G. (2023). The challenges of digital transformation in healthcare: An interdisciplinary literature review, framework, and future research agenda. *Technovation*, *123*, 102716.
75. Dasgupta, P., & David, P. A. (1987). Information disclosure and the economics of science and technology. In *Arrow and the ascent of modern economic theory* (pp. 519-542). London: Palgrave Macmillan UK.
76. De Jong, S., Barker, K., Cox, D., Sveinsdottir, T., & Van den Besselaar, P. (2014). Understanding societal impact through productive interactions: ICT research as a case. *Research Evaluation*, *23*(2), 89-102.
77. De Silva, M., Schmidt, N., Paunov, C., & Lavelle, O. (2022). How did COVID-19 shape co-creation?: Insights and policy lessons from international initiatives.
78. Decuypere, M., & Landri, P. (2021). Governing by visual shapes: University rankings, digital education platforms and cosmologies of higher education. *Critical Studies in Education*, *62*(1), 17-33.

79. Deloitte - Accelerare l'Innovazione nell'ecosistema salute - Strumenti e leve strategiche per cogliere le opportunità di investimento del PNRR (2023) - https://www2.deloitte.com/content/dam/Deloitte/it/Documents/life-sciences-health-care/accelerare-innovazione-ecosistema-salute-deloitte.pdf (last access 06/10/2023)
80. Demarinis Loiotile, A., De Nicolò, F., Agrimi, A., Bellantuono, L., La Rocca, M., Monaco, A., ... & Bellotti, R. (2022). Best Practices in Knowledge Transfer: Insights from Top Universities. *Sustainability*, *14*(22), 15427.
81. Deng, W., & Ma, J. (2021). A knowledge graph approach for recommending patents to companies. *Electronic Commerce Research*, 1-32.
82. DeVol, R., Lee, J., & Ratnatunga, M. (2017). ConCept to CommerCialization. *Milken Institute. April*.
83. Di Berardino, D., & Corsi, C. (2018). A quality evaluation approach to disclosing third mission activities and intellectual capital in Italian universities. *Journal of Intellectual Capital*, *19*(1), 178-201.
84. Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher education*, *49*, 495-533.
85. Dip, J. A. (2021). What does U-multirank tell us about knowledge transfer and research?. *Scientometrics*, *126*(4), 3011-3039.
86. Djoundourian, S., & Shahin, W. (2022). Academia–business cooperation: A strategic plan for an innovative executive education program. *Industry and Higher Education*, *36*(6), 835-845.
87. Du, W., Wang, Y., Xu, W., & Ma, J. (2021). A personalized recommendation system for high-quality patent trading by leveraging hybrid patent analysis. *Scientometrics*, *126*, 9369-9391.
88. Ebert, S., Müller, T., & Schütze, H. (2016). Lamb: A good shepherd of morphologically rich languages. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 742-752).
89. Eisenstein, J. (2019). Introduction to natural language processing. MIT press.
90. EPO (2022) - https://www.epo.org/about-us/annual-reports-statistics/statistics/2022/insight-smart-health.html (last access 20/09/2023)
91. EPO (2023) - https://www.epo.org/news-events/news/2023/20230328.html (last access 20/09/2023)
92. Ernst, H. (2003). Patent information for strategic technology management. *World patent information*, *25*(3), 233-242.
93. Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American journal of sociology*, *113*(1), 1-40.
94. Etzkowitz, H. (1998). The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. *Research policy*, *27*(8), 823-833.
95. Etzkowitz, H. (2003). Research groups as ‘quasi-firms’: the invention of the entrepreneurial university. *Research policy*, 32(1), 109-121.

96. Etzkowitz, H., & Leydesdorff, L. (1997). Universities and the global knowledge economy: A triple helix of university-industry relations. *Preprint Version of: Etzkowitz, H., & Leydesdorff, L.(1997). Universities and the Global Knowledge Economy: A Triple Helix of University-Industry-Government Relations. London: Pinter. [Archival Reprint]*.
97. Etzkowitz, H., Webster, A., Gebhardt, C., & Terra, B. R. C. (2000). The future of the university and the university of the future: evolution of ivory tower to entrepreneurial paradigm. *Research policy*, *29*(2), 313-330.
98. EU valorisation policy, https://research-and-innovation.ec.europa.eu/system/files/2020-03/ec_rtd_valorisation_factsheet.pdf (last access 25/11/2022).
99. European Commission Press release. New international university ranking: Commission welcomes launch of U-Multirank. https://ec.europa.eu/commission/presscorner/detail/en/IP_14_548 (last access 29/10/2022)
100. Everett, R. (2003). *Diffusion of Innovations, 5th Edition*. Simon and Schuster. ISBN 978-0-7432-5823-4.
101. Fall, C. J., Törcsvári, A., Benzineb, K., & Karetka, G. (2003). Automated categorization in the international patent classification. In *Acm Sigir Forum* (Vol. 37, No. 1, pp. 10-25). New York, NY, USA: ACM.
102. Feußner, H., & Park, A. (2017). Surgery 4.0: the natural culmination of the industrial revolution?. *Innovative Surgical Sciences*, *2*(3), 105-108.
103. Finne, H., Arundel, A., Balling, G., Brisson, P., & Erselius, J. (2009). Metrics for Knowledge Transfer from Public Research Organisations in Europe: Report from the European Commission's Expert Group on Knoweldge Transfer Metrics.
104. Finne, H., Day, A., Piccaluga, A., Spithoven, A., Walter, P., & Wellen, D. (2011). A Composite Indicator for Knowledge Transfer Report from the European Commission’s Expert Group on Knowledge Transfer Indicators. *European general commission for research and innovation*.
105. Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: observing Goodhart’s Law in action. *GigaScience*, *8*(6), giz053.
106. Forida, R., & Cohen, W. (1999). Engine or infrastructure? The university role in economic development. *From industrializing knowledge. University–industry linkages in Japan and the United States*, 589-610.
107. Forti, E., & Meoli, M. (2020). Indicazioni non vincolanti del gruppo di lavoro: U-MULTIRANK. In *Il Gruppo di lavoro CRUI sui ranking internazionali: attività, risultati e prospettive 2017-2020* (pp. 45-50). Fondazione CRUI.
108. Frondizi, R., Fantauzzi, C., Colasanti, N., & Fiorani, G. (2019). The evaluation of universities’ third mission and intellectual capital: Theoretical analysis and application to Italy. *Sustainability*, *11*(12), 3455.

109. Gerlach, M., Shi, H., & Amaral, L. A. N. (2019). A universal information theoretic approach to the identification of stopwords. *Nature Machine Intelligence*, *1*(12), 606-612.
110. Gherardini, A. (2015). Squarci nell'avorio: le università italiane e l'innovazione economica. *Squarci nell'avorio*, 1-177.
111. Giordano, V., Chiarello, F., Melluso, N., Fantoni, G., & Bonaccorsi, A. (2021). Text and dynamic network analysis for measuring technological convergence: A case study on defense patent data. *IEEE Transactions on Engineering Management*.
112. Giunta, A., Pericoli, F. M., & Pierucci, E. (2016). University–industry collaboration in the biopharmaceuticals: The Italian case. *The Journal of Technology Transfer*, *41*, 818-840.
113. Giuri, P., Munari, F., Scandura, A., & Toschi, L. (2019). The strategic orientation of universities in knowledge transfer activities. *Technological Forecasting and Social Change*, *138*, 261-278.
114. Goddard, J., Hazelkorn, E., & Vallance, P. (Eds.). (2016). *The civic university: The policy and leadership challenges*. Edward Elgar Publishing.
115. Godonoga, A., & Sporn, B. (2023). The conceptualisation of socially responsible universities in higher education research: a systematic literature review. *Studies in Higher Education*, *48*(3), 445-459.
116. Gomez, J. C., & Moens, M. F. (2014). A survey of automated hierarchical classification of patents. Professional Search in the Modern World: COST Action IC1002 on Multilingual and Multifaceted Interactive Information Access, 215-249.
117. Gregersen, B., Linde, L. T., & Rasmussen, J. G. (2009). Linking between Danish universities and society. *Science and public policy*, *36*(2), 151-156.
118. Gunasekara, C. (2006). The generative and developmental roles of universities in regional innovation systems. *Science and public policy*, *33*(2), 137-150.
119. Guo, X., Yin, Y., Dong, C., Yang, G., & Zhou, G. (2008). On the class imbalance problem. In *2008 Fourth international conference on natural computation* (Vol. 4, pp. 192-201). IEEE.
120. Haghighian Roudsari, A., Afshar, J., Lee, W., & Lee, S. (2022). PatentNet: multi-label classification of patent documents using deep learning based language understanding. *Scientometrics*, 1-25.
121. Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2022). Medical 4.0 technologies for healthcare: Features, capabilities, and applications. *Internet of Things and Cyber-Physical Systems*, *2*, 12-30.
122. Hamano, Y. University–Industry Collaboration – WIPO, 2018
123. Hammarfelt, B. (2021). Linking science to technology: the "patent paper citation" and the rise of patentometrics in the 1980s. *Journal of Documentation*, *77*(6), 1413-1429.

124. Hassani, K., & Lee, W. S. (2016). Visualizing natural language descriptions: A survey. *ACM Computing Surveys (CSUR)*, *49*(1), 1-34.
125. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
126. Hazelkorn, E., & Gibson, A. (2017). Global science, national research, and the question of university rankings. *Palgrave Communications*, *3*(1), 1-11.
127. Hazelkorn, E., Loukkola, T., & Zhang, T. (2014). Rankings in institutional strategies and processes: impact or illusion.
128. Heher, A. D. (2006). Return on investment in innovation: Implications for institutions and national agencies. *The Journal of Technology Transfer*, *31*, 403-414.
129. Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii international conference on system sciences* (pp. 1833-1842). IEEE.
130. Hensen, J. L., Loonen, R. C. G. M., Archontiki, M., & Kanellis, M. (2015). Using Building Simulation for Moving Innovations across the 'Valley of Death.'. *REHVA Journal*, *52*(3), 58-62.
131. Hepburn, J. (2018). Universal language model fine-tuning for patent classification. In *Proceedings of the Australasian Language Technology Association Workshop 2018* (pp. 93-96).
132. Hicks, D. (2012). Performance-based university research funding systems. *Research policy*, *41*(2), 251-261.
133. Hilbe, J. M. (2009). *Logistic regression models*. CRC press.
134. Hockaday, T. (2020). University Technology Transfer: What it is and how to Do it. JHU Press.
135. HORIZON 2020 – https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf (last access 15/09/2023)
136. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
137. Hu, J., Li, S., Hu, J., & Yang, G. (2018). A hierarchical feature extraction model for multi-label mechanical patent classification. *Sustainability*, *10*(1), 219.
138. Hu, T., & Zhang, Y. (2021). A spatial–temporal network analysis of patent transfers from US universities to firms. *Scientometrics*, *126*(1), 27-54.
139. Huang, Y., Zhu, F., Porter, A. L., Zhang, Y., Zhu, D., & Guo, Y. (2020). Exploring technology evolution pathways to facilitate technology management: From a technology life cycle perspective. *IEEE Transactions on Engineering Management*, *68*(5), 1347-1359.
140. Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z. K., & Roco, M. C. (2003). Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field. *Journal of nanoparticle research*, *5*, 333-363.

141. Hudson, J., & Khazragui, H. F. (2013). Into the valley of death: research to innovation. *Drug discovery today*, *18*(13-14), 610-613.
142. Indurkhya, N., & Damerau, F. J. (Eds.). (2010). *Handbook of natural language processing* (Vol. 2). CRC Press.
143. Intellectual property action plan implementation, 2022 https://single-market-economy.ec.europa.eu/industry/strategy/intellectual-property/intellectual-property-action-plan-implementation_en (last access 20/08/2023)
144. Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: a comprehensive survey. *IEEE access*, *3*, 678-708.
145. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
146. Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, *6*(5), 429-449.
147. Jeelani, S., Reddy, R. J., Maheswaran, T., Asokan, G. S., Dany, A., & Anand, B. (2014). Theranostics: A treasured tailor for tomorrow. *Journal of pharmacy & bioallied sciences*, *6*(Suppl 1), S6.
148. Jell, A., Vogel, T., Ostler, D., Marahrens, N., Wilhelm, D., Samm, N., ... & Kranzfelder, M. (2019). 5th-generation mobile communication: data highway for surgery 4.0. *Surgical technology international*, *35*, 36-42.
149. Jessop, Z. M., Al-Sabah, A., Francis, W. R., & Whitaker, I. S. (2016). Transforming healthcare through regenerative medicine. *BMC medicine*, *14*, 1-6.
150. Jivani, A. G. (2011). A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl*, *2*(6), 1930-1938.
151. Johnes, J. (2018). University rankings: What do they really show?. *Scientometrics*, *115*(1), 585-606.
152. Johnson Jr, A. M. (2006). The destruction of the holistic approach to admissions: The pernicious effects of rankings. *Ind. LJ*, *81*, 309.
153. Jones, S. S., Heaton, P. S., Rudin, R. S., & Schneider, E. C. (2012). Unraveling the IT productivity paradox—lessons for health care. *N Engl J Med*, *366*(24), 2243-2245.
154. Jonkers, K., & Zacharewicz, T. (2016). Research performance based funding systems: A comparative assessment. *Luxembourg: Publications Office of the European Union*.
155. Jun, S., Park, S. S., & Jang, D. S. (2014). Document clustering method using dimension reduction and support vector clustering to overcome sparseness. *Expert Systems with Applications*, *41*(7), 3204-3212.
156. Jung, G., Shin, J., & Lee, S. (2023). Impact of preprocessing and word embedding on extreme multi-label patent classification tasks. *Applied Intelligence*, *53*(4), 4047-4062.

157. Jyothi, K. C., & Khare, N. (2023). Convergence of Machine Learning and Blockchain Technology for Smart Healthcare Applications. In *AI, IoT, and Blockchain Breakthroughs in E-Governance* (pp. 72-94). IGI Global.
158. Kaiser, F., & Zeeman, N. (2017). U-Multirank: Data analytics and scientometrics. *Research Analytics*, 185-220.
159. Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139-172.
160. Kapetaniou, C., & Lee, S. H. (2017). A framework for assessing the performance of universities: The case of Cyprus. *Technological Forecasting and Social Change*, *123*, 169-180.
161. Karanikić, P., Matulja, M., & Tijan, E. (2019). The role of university technology transfer process in Digital Economy. In 2019 *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics* (MIPRO) (pp. 1419-1422). IEEE.
162. Karvonen, M., & Kässi, T. (2013). Patent citations as a tool for analysing the early stages of convergence. *Technological Forecasting and Social Change*, *80*(6), 1094-1107.
163. Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research policy*, *26*(1), 1-18.
164. Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, *117*, 228-237.
165. Knowledge Transfer Ireland - https://www.knowledgetransferireland.com/ (last access 01/09/2023)
166. Knowledge-Share - https://www.knowledge-share.eu/ (last access 01/10/2023)
167. Kocher, R., & Sahni, N. R. (2011). Rethinking health care labor. *N Engl J Med*, *365*(15), 1370-2.
168. Kohus, Z., Baracskai, Z., & Czako, K. (2020). The Relationship between University-Industry Co-Publication Outputs. *Economic and Social Development: Book of Proceedings*, 109-122.
169. Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. *World Patent Information*, *65*, 102035.
170. Kyebambe, M. N., Cheng, G., Huang, Y., He, C., & Zhang, Z. (2017). Forecasting emerging technologies: A supervised learning approach through patent analysis. *Technological Forecasting and Social Change*, *125*, 236-244.
171. Landinez, L., Kliewe, T., & Diriba, H. (2019). Entrepreneurial university indicators in global university rankings. *Developing Engaged and Entrepreneurial Universities: Theories, Concepts and Empirical Findings*, 57-85.

172. Laredo, P. (2007). Revisiting the third mission of universities: Toward a renewed categorization of university activities?. *Higher education policy*, *20*, 441-456.
173. Lee, J. J., Vance, H., Stensaker, B., & Ghosh, S. (2020). Global rankings at a local cost? The strategic pursuit of status and the third mission. *Comparative education*, *56*(2), 236-256.
174. Lee, J. S., & Hsiang, J. (2020). Patent classification by fine-tuning BERT language model. *World Patent Information*, *61*, 101965.
175. Lee, J. S., Park, J. H., & Bae, Z. T. (2017). The effects of licensing-in on innovative performance in different technological regimes. *Research Policy*, *46*(2), 485-496.
176. Lei, L., Qi, J., & Zheng, K. (2019). Patent analytics based on feature vector space model: A case of IoT. *Ieee Access*, *7*, 45705-45715.
177. Levy, R., Roux, P., & Wolff, S. (2009). An analysis of science–industry collaborative patterns in a large European University. *The Journal of technology transfer*, *34*, 1-23.
178. Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, *49*(4), 764-766.
179. Li, J., & Carayon, P. (2021). Health Care 4.0: A vision for smart and connected health care. *IISE Transactions on Healthcare Systems Engineering*, *11*(3), 171-180.
180. Li, L., Chen, Q., Jia, X., & Herrera-Viedma, E. (2021). Co-patents' commercialization: evidence from China. *Economic research-Ekonomska istraživanja*, *34*(1), 1709-1726.
181. Li, S., Hu, J., Cui, Y., & Hu, J. (2018). DeepPatent: patent classification with convolutional neural networks and word embedding. *Scientometrics*, *117*, 721-744.
182. Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature communications*, *10*(1), 5170.
183. Limoges, C., Scott, P., Schwartzman, S., Nowotny, H., & Gibbons, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. *The New Production of Knowledge*, 1-192.
184. Livan, G. (2019). Don't follow the leader: how ranking performance reduces meritocracy. *Royal Society Open Science*, *6*(11), 191255.
185. Loukkola, T. (2016). Europe: Impact and influence of rankings in higher education. In *Global Rankings and the Geopolitics of Higher Education* (pp. 127-139). Routledge.
186. Madhulatha, T. S. (2012). An overview on clustering methods. *arXiv preprint arXiv:1205.1117*.
187. Marhl, M., & Pausits, A. (2011). Third mission indicators for new ranking methodologies. *Evaluation in Higher Education*, *5*(1), 43-64.

188. Markuerkiaga, L., Caiazza, R., Igartua, J. I., & Errasti, N. (2016). Factors fostering students' spin-off firm formation: An empirical comparative study of universities from North and South Europe. *Journal of Management Development*, *35*(6), 814-846.
189. Marra, M. (2022). Connessioni virtuose. Come nasce (e cresce) un ecosistema dell'innovazione, Bologna, Il Mulino.
190. Marzocchi, C., Kitagawa, F., Rossi, F., & Uyarra, E. (2023). Reconceptualising knowledge exchange and higher education institutions: broadening our understanding of motivations, channels, and stakeholders. *Studies in Higher Education*, *48*(5), 673-682.
191. Mas Verdú, F., Roig-Tierno, N., Nieto-Alemán, P. A., & García Alvarez-Coque, J. M. (2020). Competitiveness in European regions and top-ranked universities: Do local universities matter?. *Journal of Competitiveness*, *12*(4), 91-108.
192. Masiakowski, P., & Wang, S. (2013). Integration of software tools in patent analysis. *World Patent Information*, *35*(2), 97-104.
193. McDevitt, V. L., Mendez-Hinds, J., Winwood, D., Nijhawan, V., Sherer, T., Ritter, J. F., & Sanberg, P. R. (2014). More than money: The exponential impact of academic technology transfer. *Technology & Innovation*, *16*(1), 75-84.
194. McKinsey & Company Report - Transforming healthcare with AI-The impact on the workforce and organisations (2020) - https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf (last access 04/09/2023)
195. Miceli, L., Dal Mas, F., Biancuzzi, H., Bednarova, R., Rizzardo, A., Cobianchi, L., & Holmboe, E. S. (2021). Doctor@ Home: through a telemedicine co-production and co-learning journey. *Journal of Cancer Education*, 1-3.
196. Milligan, G. W., & Cooper, M. C. (1987). Methodology review: Clustering methods. *Applied psychological measurement*, *11*(4), 329-354.
197. Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, *110*(2), 967-990.
198. Molas-Gallart, J., Salter, A., Patel, P., Scott, A., & Duran, X. (2002). Measuring third stream activities. *Final report to the Russell Group of Universities. Brighton: SPRU, University of Sussex*, *81*.
199. Monteiro, A. C. B., França, R. P., Estrela, V. V., Iano, Y., Khelassi, A., & Razmjooy, N. (2018). Health 4.0: Applications, Management, Technologies and Review. *personalized medicine*, *5*, 6.
200. Morgan, A. C., Economou, D. J., Way, S. F., & Clauset, A. (2018). Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science*, *7*(1), 40.

201. Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544-551.
202. Nakwa, K., & Zawdie, G. (2016). The 'third mission'and 'triple helix mission'of universities as evolutionary processes in the development of the network of knowledge production: Reflections on SME experiences in Thailand. *Science and Public Policy*, *43*(5), 622-629.
203. Needham, C., & Glasby, J. (2015). Personalisation–love it or hate it?. *Journal of Integrated Care*, *23*(5), 268-276.
204. O'Reilly, N. M., Robbins, P., & Scanlan, J. (2019). Dynamic capabilities and the entrepreneurial university: a perspective on the knowledge transfer capabilities of universities. *Journal of Small Business & Entrepreneurship*, *31*(3), 243-263.
205. Oancea, A. (2019). Research governance and the future (s) of research assessment. *Palgrave Communications*, *5*(1).
206. OECD Science, Technology and Industry Scoreboard 2015 : Innovation for growth and society | OECD Science, Technology and Industry Scoreboard | OECD iLibrary https://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-scoreboard-2015_sti_scoreboard-2015-en (last access 09/09/2023)
207. Olcay, G. A., & Bulu, M. (2017). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change*, *123*, 153-160.
208. Osmani, V., Balasubramaniam, S., & Botvich, D. (2008). Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of network and computer applications*, *31*(4), 628-655.
209. Oțoiu, A., & Țițan, E. (2021). To what extent ict resources influence the learning experience? an inquiry using u-multirank data. In *INTED2021 Proceedings* (pp. 40-45). IATED.
210. Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *Journal of global health*, *8*(2).
211. Panesar, A. (2019). *Machine learning and AI for healthcare* (pp. 1-73). Coventry, UK: Apress.
212. Park, H., Ree, J. J., & Kim, K. (2013). Identification of promising patents for technology transfers using TRIZ evolution trends. *Expert systems with applications*, *40*(2), 736-743.
213. Partha, D., & David, P. A. (1994). Toward a new economics of science. *Research policy*, *23*(5), 487-521.
214. PATIRIS - https://patiris.mise.gov.it/index.php/it/#summarycharts (last access 30/09/2023)
215. Peeters, H., Callaert, J., & Van Looy, B. (2020). Do firms profit from involving academics when developing technology?. *The Journal of Technology Transfer*, *45*, 494-521.

216. Peltzer, K., Williams, J. S., Kowal, P., Negin, J., Snodgrass, J. J., Yawson, A., ... & SAGE Collaboration. (2014). Universal health coverage in emerging economies: findings on health care utilization by older adults in China, Ghana, India, Mexico, the Russian Federation, and South Africa. *Global health action*, *7*(1), 25314.
217. Perkmann, M., Neely, A., & Walsh, K. (2011). How should firms evaluate success in university–industry alliances? A performance measurement system. *R&D Management*, *41*(2), 202-216.
218. Piirainen, K. A., Andersen, A. D., & Andersen, P. D. (2016). Foresight and the third mission of universities: the case for innovation system foresight. *Foresight*, *18*(1), 24-40.
219. Pinheiro, R., Langa, P. V., & Pausits, A. (2015). One and two equals three? The third mission of higher education institutions. *European journal of higher education*, *5*(3), 233-249.
220. Ponemon Report (2019) https://www.aon.com/getmedia/60fbb49a-c7a5-4027-ba98-0553b29dc89f/Ponemon-Report-V24.aspx (last access 16/12/2022)
221. Prado, A. (2021). Performances of the Brazilian Universities in the "U-MULTIRANK" in the Period 2017-2020.
222. Press release European Commission (2020). "Commission adopts Action Plan on Intellectual Property to strengthen EU's economic resilience and recovery", https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2187 (last access 16/12/2022)
223. Promoting IP valorisation through the IP platform – Knowledge Share run by the national network NETVAL - https://ec.europa.eu/research-and-innovation/en/research-area/industrial-research-and-innovation/eu-valorisation-policy/knowledge-valorisation-platform/repository/promoting-ip-valorisation-through-ip-platform-knowledge-share-run-national-network-netval (last access 20/09/2023).
224. Puccetti, G., Chiarello, F., & Fantoni, G. (2021). A simple and fast method for Named Entity context extraction from patents. *Expert Systems with Applications*, *184*, 115570.
225. Puccetti, G., Giordano, V., Spada, I., Chiarello, F., & Fantoni, G. (2023). Technology identification from patent texts: A novel named entity recognition method. *Technological Forecasting and Social Change*, *186*, 122160.
226. Pusser, B., & Marginson, S. (2013). University rankings in critical perspective. *The journal of higher education*, *84*(4), 544-568.
227. QS World University Rankings® (QSWUR) - https://www.topuniversities.com/university-rankings/world-university-rankings/2020 (last access 20/09/2021).
228. Rameshbhai, C. J., & Paulose, J. (2019). Opinion mining on newspaper headlines using SVM and NLP. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*(3), 2152-2163.
229. Rätsch, G. (2004). A brief introduction into machine learning. *Friedrich Miescher Laboratory of the Max Planck Society*, 1-6.

230. Rauhvargers, A. (2013). *Global university rankings and their impact: Report II* (pp. 21-23). Brussels: European University Association.
231. Research & innovation valorisation channels and tools - Publications Office of the EU - https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/f35fded6-bc0b-11ea-811c-01aa75ed71a1 (last access 29/10/2022)
232. Ringel, L., Espeland, W., Sauder, M., & Werron, T. (2021). Worlds of rankings. In *Worlds of rankings* (pp. 1-23). Emerald Publishing Limited.
233. Robust Statistics, 2nd Edition | Wiley https://www.wiley.com/en-us/Robust+Statistics%2C+2nd+Edition-p-9780470129906 (last access 09/09/2023).
234. Rodrigues, C. (2011). Universities, the second academic revolution and regional development: a tale (solely) made of "techvalleys"?. *European Planning Studies*, *19*(2), 179-194.
235. Rodrigues, C., da Rosa Pires, A., & de Castro, E. (2001). Innovative universities and regional institutional capacity building: The case of Aveiro, Portugal. *Industry and Higher Education*, *15*(4), 251-255.
236. Rodriguez, A., Tosyali, A., Kim, B., Choi, J., Lee, J. M., Coh, B. Y., & Jeong, M. K. (2016). Patent clustering and outlier ranking methodologies for attributed patent citation networks for technology opportunity discovery. *IEEE Transactions on Engineering Management*, *63*(4), 426-437.
237. Roessner, D., Bond, J., Okubo, S., & Planting, M. (2013). The economic impact of licensed commercialized inventions originating in university research. *Research Policy*, *42*(1), 23-34.
238. Rossi, F., & Rosli, A. (2015). Indicators of university–industry knowledge transfer performance and their implications for universities: evidence from the United Kingdom. *Studies in Higher Education*, *40*(10), 1970-1991.
239. Rothaermel, F. T., Agung, S. D., & Jiang, L. (2007). University entrepreneurship: a taxonomy of the literature. *Industrial and corporate change*, *16*(4), 691-791.
240. Roux, M. A comparative study of divisive hierarchical clustering algorithms. arXiv 2015. *arXiv preprint arXiv:1506.08977*.
241. Rtveladze, K., Marsh, T., Webber, L., Kilpi, F., Levy, D., Conde, W., ... & Brown, M. (2013). Health and economic burden of obesity in Brazil. *PloS one*, *8*(7), e68785.
242. Salomaa, M., Cinar, R., & Charles, D. (2021). Rankings and regional development: The cause or the symptom of universities' insufficient regional contributions?. *Higher Education Governance and Policy*, *2*(1), 31-44.
243. Sannino, G., De Falco, I., & De Pietro, G. (2018). A continuous noninvasive arterial pressure (CNAP) approach for health 4.0 systems. *IEEE Transactions on Industrial Informatics*, *15*(1), 498-506.
244. Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *Plos one*, *16*(8), e0254937.

245. Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American sociological review*, *74*(1), 63-82.
246. Saxenian, A. (1996). Regional advantage: Culture and competition in silicon valley and route 128, with a new preface by the author. Harvard University Press.
247. Scanlan, J. (2018). A capability maturity framework for knowledge transfer. *Industry and Higher Education*, *32*(4), 235-244.
248. Scarrà, D., & Piccaluga, A. (2022). The impact of technology transfer and knowledge spillover from Big Science: a literature review. *Technovation*, *116*, 102165.
249. Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine learning*, *13*, 135-143.
250. Schwab, K. (2017). The fourth industrial revolution. Currency.
251. Secundo, G., De Beer, C., Schutte, C. S., & Passiante, G. (2017). Mobilising intellectual capital to improve European universities' competitiveness: The technology transfer offices' role. *Journal of Intellectual Capital*, *18*(3), 607-624.
252. Seppo, M., & Lilles, A. (2012). Indicators measuring university-industry cooperation. *Discussions on Estonian Economic Policy*, *20*(1), 204.
253. Shane, S. A. (2004). Academic entrepreneurship: University spinoffs and wealth creation. Edward Elgar Publishing.
254. Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial Intelligence: Definition and Background. In *Mission AI: The New System Technology* (pp. 15-41). Cham: Springer International Publishing.
255. Shi, T., & Horvath, S. (2006). Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics*, *15*(1), 118-138.
256. Shove, E. (1998). Gaps, barriers and conceptual chasms: theories of technology transfer and energy in buildings. *Energy policy*, *26*(15), 1105-1112.
257. Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366.
258. Smith, H. L., & Bagchi-Sen, S. (2012). The research university, entrepreneurship and regional development: Research propositions and current evidence. *Entrepreneurship & Regional Development*, *24*(5-6), 383-404.
259. Smola, A. (2008). Introduction to machine learning.
260. Sobradillo, P., Pozo, F., & Agustí, Á. (2011). P4 medicine: the future around the corner. *Archivos de Bronconeumología ((English Edition))*, *47*(1), 35-40.
261. Song, K., Ran, C., & Yang, L. (2022). A digital analysis system of patents integrating natural language processing and machine learning. *Technology Analysis & Strategic Management*, 1-17.
262. Souza, C. M., Meireles, M. R., & Almeida, P. E. (2021). A comparative study of abstractive and extractive summarization techniques to label subgroups on patent dataset. *Scientometrics*, *126*(1), 135-156.

263. Stake, J. E. (2006). The interplay between law school rankings, reputations, and resource allocation: Ways rankings mislead. *Ind. LJ*, *81*, 229.
264. Sugimoto, C. R., & Larivière, V. (2018). *Measuring research: What everyone needs to know*. Oxford University Press.
265. Teber, D., Engels, C., Maier-Hein, L., Ayala, L., Onogur, S., Seitel, A., & März, K. (2020). Surgery 4.0—are we ready?. *Der Urologe*, *59*, 1035-1043.
266. Technology Transfer System Handbook https://www.polito.it/imprese/trasferimento/TTS_handbook.pdf (last access 15/12/2021)
267. Thorleuchter, D., Van den Poel, D., & Prinzie, A. (2010). A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies. *Technological Forecasting and Social Change*, *77*(7), 1037-1050.
268. Thuemmler, C., & Bai, C. (Eds.). (2017). *Health 4.0: How virtualization and big data are revolutionizing healthcare* (pp. 23-37). New York: Springer.
269. Tijssen, R. J. (2006). Universities and industrially relevant science: Towards measurement models and indicators of entrepreneurial orientation. *Research Policy*, *35*(10), 1569-1585.
270. Tijssen, R. J. (2011). Joint research publications: a performance indicator of university-industry collaboration. *Evaluation in Higher Education*, *5*(2), 19-40.
271. Tijssen, R. J., Van Leeuwen, T. N., & Van Wijk, E. (2009). Benchmarking university-industry research cooperation worldwide: performance measurements and indicators based on co-authorship data for the world's largest universities. *Research Evaluation*, *18*(1), 13-24.
272. Times Higher Education World University Rankings (THEWUR) - https://www.timeshighereducation.com/world-university-rankings/2020/world-ranking (last access 15/12/2021)
273. Tortorella, G. L., Fogliatto, F. S., Anzanello, M., Marodin, G. A., Garcia, M., & Reis Esteves, R. (2017). Making the value flow: application of value stream mapping in a Brazilian public healthcare organisation. *Total Quality Management & Business Excellence*, *28*(13-14), 1544-1558.
274. Tortorella, G. L., Fogliatto, F. S., Espôsto, K. F., Mac Cawley, A. F., Vassolo, R., Tlapa, D., & Narayanamurthy, G. (2022). Healthcare costs' reduction through the integration of Healthcare 4.0 technologies in developing economies. *Total Quality Management & Business Excellence*, *33*(3-4), 467-487.
275. Trappey, A. J., Chen, P. P., Trappey, C. V., & Ma, L. (2019). A machine learning approach for solar power technology review and patent evolution analysis. *Applied Sciences*, *9*(7), 1478.
276. Trappey, A. J., Lupu, M., & Stjepandic, J. (2020a). Embrace artificial intelligence technologies for advanced analytics and management of intellectual properties. *World Patent Information*, *61*, 101970.

277. Trappey, A. J., Trappey, C. V., & Chung, C. L. (2017). IP portfolios and evolution of biomedical additive manufacturing applications. *Scientometrics*, *111*, 139-157.
278. Trappey, A. J., Trappey, C. V., Chiang, T. A., & Huang, Y. H. (2013). Ontology-based neural network for patent knowledge management in design collaboration. *International Journal of Production Research*, *51*(7), 1992-2005.
279. Trappey, A. J., Trappey, C. V., Hsu, F. C., & Hsiao, D. W. (2009). A fuzzy ontological knowledge document clustering methodology. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *39*(3), 806-814.
280. Trappey, A. J., Trappey, C. V., Wu, J. L., & Wang, J. W. (2020b). Intelligent compilation of patent summaries using machine learning and natural language processing techniques. *Advanced Engineering Informatics*, *43*, 101027.
281. Trencher, G., Yarime, M., McCormick, K. B., Doll, C. N., & Kraines, S. B. (2014). Beyond the third mission: Exploring the emerging university function of co-creation for sustainability. *Science and Public Policy*, *41*(2), 151-179.
282. Trippe, A. (2015). Guidelines for preparing patent landscape reports. *Patent landscape reports. Geneva: WIPO*, 2015.
283. Trippe, A. J. (2003). Patinformatics: Tasks to tools. *World Patent Information*, *25*(3), 211-221.
284. Trippl, M., Sinozic, T., & Lawton Smith, H. (2015). The role of universities in regional development: Conceptual models and policy institutions in the UK, Sweden and Austria. *European Planning Studies*, *23*(9), 1722-1740.
285. Tseng, F. M., Hsieh, C. H., Peng, Y. N., & Chu, Y. W. (2011). Using patent data to analyze trends and the technological strategies of the amorphous silicon thin-film solar cell industry. *Technological forecasting and social change*, *78*(2), 332-345.
286. U-multirank - https://www.umultirank.org/about/u-multirank/the-project/ (last access 20/09/2022)
287. UN – United Nations (2019), Dichiarazione espressa da Association of Common-wealth Universities, Agence universitaire de la Francophonie e Associazione in-ternazionale delle università (IAU) all'High Level Political Forum 2019 delle Nazioni Unite, disponibile al sito: www.acu.ac.uk/news/higher-education-s-es-sential-contribution-to-the-sdgs/ (last access 08/12/2022)
288. University-industry collaboration: A Closer Look for Research Leaders https://www.elsevier.com/research-intelligence/university-industry-collaboration (last access 09/09/2023)
289. Urdari, C., Farcas, T. V., & Tiron-Tudor, A. (2017). Assessing the legitimacy of HEIs' contributions to society: the perspective of international rankings. *Sustainability Accounting, Management and Policy Journal*, *8*(2), 191-215.
290. Van de Wiel, M. A., & Di Bucchianico, A. (2001). Fast computation of the exact null distribution of Spearman's ρ and Page's L statistic for samples with and without ties. *Journal of statistical planning and inference*, *92*(1-2), 133-145.

291. Van Vught, F. (2008). Mission diversity and reputation in higher education. *Higher Education Policy*, *21*, 151-174.
292. van Vught, F., & Ziegele, F. (Eds.). (2011). *Design and testing the feasibility of a multidimensional global university ranking*. Consortium for Higher Education and Research Performance Assessment.
293. Verberne, S., D'hondt, E. K. L., Oostdijk, N. H. J., & Koster, C. H. (2010). Quantifying the challenges in parsing patent claims.
294. Visconti, R. M., Doś, A., & Gurgun, A. P. (2017). Public–private partnerships for sustainable healthcare in emerging Economies. In *The Emerald Handbook of Public–Private Partnerships in Developing and Emerging Economies* (pp. 407-437). Emerald Publishing Limited.
295. Wan, J., Tang, S., Li, D., Imran, M., Zhang, C., Liu, C., & Pang, Z. (2018). Reconfigurable smart factory for drug packing in healthcare industry 4.0. *IEEE transactions on industrial informatics*, *15*(1), 507-516.
296. Wang, Y. H., & Lin, G. Y. (2023). Exploring AI-healthcare innovation: natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes*, *52*(4), 1173-1189.
297. Wang, Y., Ning, L., & Chen, J. (2014). Product diversification through licensing: Empirical evidence from Chinese firms. *European Management Journal*, *32*(4), 577-586.
298. Way, S. F., Morgan, A. C., Larremore, D. B., & Clauset, A. (2019). Productivity, prominence, and the effects of academic environment. *Proceedings of the National Academy of Sciences*, *116*(22), 10729-10733.
299. Westerheijden, D. F., & Federkeil, G. (2018). U-Multirank: A European multidimensional transparency tool in higher education.
300. Wong, P. K., & Singh, A. (2013). Do co-publications with industry lead to higher levels of university technology commercialization activity?. *Scientometrics*, *97*, 245-265.
301. World Economic Forum - How universities can become a platform for social change https://www.weforum.org/agenda/2019/06/universities-platform-social-change-tokyo/ (last access 29/10/2022)
302. World Health Organization (2019). Countries are spending more on health, but people are still paying too much out of their own pockets. In *who. int*.
303. WHO World Health Organization - Global strategy on digital health 2020-2025 (2021) - https://iris.who.int/handle/10665/344249 (last access 09/10/2023)
304. Wu, X. Z., & Zhou, Z. H. (2017). A unified view of multi-label performance measures. In *international conference on machine learning* (pp. 3780-3788). PMLR.
305. Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, *2*(4).

306. Yamashita, Y. (2018). Exploring characteristics of patent-paper citations and development of new indicators. *Scientometrics*, *151*.
307. Yang, G., Pang, Z., Deen, M. J., Dong, M., Zhang, Y. T., Lovell, N., & Rahmani, A. M. (2020). Homecare robotic systems for healthcare 4.0: Visions and enabling technologies. *IEEE journal of biomedical and health informatics*, *24*(9), 2535-2549.
308. Yegros-Yegros, A., Azagra-Caro, J. M., López-Ferrer, M., & Tijssen, R. J. (2016). Do university–industry co-publication outputs correspond with university funding from firms?. *Research Evaluation*, *25*(2), 136-150.
309. Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management*, *54*(3), 588-599.
310. Yu, X., & Zhang, B. (2019). Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning. *Technological Forecasting and Social Change*, *145*, 273-283.
311. Yun, J., & Geum, Y. (2020). Automated classification of patents: A topic modeling approach. *Computers & Industrial Engineering*, *147*, 106636.
312. Zawdie, G. (2010). Knowledge exchange and the third mission of universities: Introduction: The triple helix and the third mission–schumpeter revisited. *Industry and Higher Education*, *24*(3), 151-155.
313. Zuva, K., & Zuva, T. (2012). Evaluation of information retrieval systems. *International journal of computer science & information technology*, *4*(3), 35.

Appendix A

In this Appendix we report the 260 most important words, as determined by Logistic Regression (LR), our top-performer classification algorithm. In Table A1 we indicate these words in descending order of their corresponding coefficients in the LR model. Moreover, we also report the category at which words have these coefficients.

Table A1: The 260 most important words for the LR model, together with the corresponding category.

Word	Category
food	Agrifood
energy	Green Energy
building	Environment
solar	Green Energy
heat	Green Energy
product	Agrifood
plant	Agrifood
patient	Biomed
cell	Biomed
battery	Green Energy
wine	Agrifood
water	Environment
electricity	Green Energy
construction	Environment
component	Packaging
oil	Agrifood
animal	Agrifood
user	Electronics
object	Packaging
material	Packaging
particle	Basic Science

extract	Agrifood
concrete	Environment
signal	Electronics
network	Electronics
information	Electronics
data	Electronics
photovoltaic	Green Energy
tissue	Biomed
power	Green Energy
thermal	Green Energy
fuel	Green Energy
packaging	Packaging
vegetable	Agrifood
gas	Green Energy
panel	Environment
milk	Agrifood
solvent	Basic Science
element	Environment
electrical	Green Energy
tumor	Biomed
seismic	Environment
communication	Electronics
circuit	Electronics
waste	Agrifood
piece	Packaging
optical	Electronics
machine	Packaging
olive	Agrifood
disease	Biomed

steel	Environment
farm	Agrifood
software	Electronics
device	Electronics
edible	Agrifood
liquid	Basic Science
chemical	Basic Science
blood	Biomed
network	Green Energy
fruit	Agrifood
tool	Packaging
structure	Environment
process	Basic Science
structural	Environment
material	Basic Science
flow	Green Energy
mechanical	Packaging
joint	Packaging
stiffness	Packaging
diagnosis	Biomed
electromagnetic	Electronics
paper	Packaging
bottle	Agrifood
metal	Basic Science
detector	Basic Science
surgical	Biomed
system	Green Energy
acid	Agrifood
image	Electronics

drug	Biomed
frame	Packaging
site	Environment
radar	Environment
human	Biomed
water	Green Energy
nanoparticles	Basic Science
polymer	Basic Science
package	Packaging
microorganism	Agrifood
environmental	Environment
air	Environment
wall	Environment
tag	Packaging
transmission	Electronics
pesticide	Agrifood
sensor	Environment
marine	Environment
content	Agrifood
generator	Green Energy
monitoring	Environment
sample	Basic Science
ion	Basic Science
exchanger	Green Energy
shaft	Packaging
reactor	Green Energy
stinger	Packaging
combustion	Green Energy
starch	Agrifood

high	Basic Science
polymeric	Packaging
lithium	Green Energy
robot	Packaging
system	Packaging
efficiency	Green Energy
mortar	Environment
clinical	Biomed
anchor	Environment
integrate	Packaging
laser	Packaging
force	Packaging
industrial	Packaging
invasive	Biomed
random	Electronics
remote	Electronics
treatment	Biomed
quantum	Electronics
rfid	Packaging
property	Basic Science
chain	Packaging
hand	Electronics
compound	Basic Science
diagnostic	Biomed
therapeutic	Biomed
part	Packaging
coli	Agrifood
limb	Packaging
hydrogen	Green Energy

defect	Packaging
microalgae	Agrifood
ph	Basic Science
code	Electronics
ceramic	Basic Science
grid	Green Energy
color	Agrifood
printing	Packaging
strain	Agrifood
maintenance	Environment
drone	Electronics
pathology	Biomed
screw	Packaging
bone	Biomed
antenna	Electronics
fiber	Packaging
mushroom	Agrifood
test	Environment
position	Electronics
gene	Biomed
composite	Basic Science
ultrasound	Biomed
cancer	Biomed
flight	Packaging
generation	Green Energy
wearable	Environment
sludge	Environment
manufacturing	Packaging
automotive	Packaging

module	Electronics
orchard	Agrifood
fat	Agrifood
therapy	Biomed
tunnel	Green Energy
cycle	Green Energy
mode	Electronics
surface	Basic Science
inspection	Environment
field	Basic Science
brain	Biomed
load	Environment
electronic	Electronics
tanning	Packaging
cell	Green Energy
beam	Basic Science
air	Green Energy
event	Electronics
produce	Green Energy
fiber	Basic Science
enzyme	Basic Science
membrane	Basic Science
system	Electronics
precursor	Basic Science
surgery	Biomed
polymer	Packaging
conversion	Green Energy
biomass	Green Energy
oxygen	Packaging

seed	Agrifood
syngas	Green Energy
gripping	Packaging
kinematic	Packaging
hot	Packaging
algorithm	Electronics
biomass	Agrifood
turbine	Green Energy
emission	Environment
material	Environment
virtual	Electronics
molecule	Biomed
production	Agrifood
cement	Environment
additive	Packaging
output	Electronics
good	Agrifood
risk	Biomed
environmental	Agrifood
cheese	Agrifood
natural	Agrifood
steam	Green Energy
biodegradable	Packaging
infection	Biomed
polyurethane	Packaging
graph	Electronics
bit	Electroncis
constraint	Packaging
connect	Green Energy

heart	Biomed
recovery	Environment
area	Electronics
ground	Environment
critical	Packaging
receive	Electronics
model	Electronics
traditional	Environment
insert	Packaging
activity	Biomed
radio	Electronics
stereolithography	Packaging
mean	Environment
automatic	Environment
treatment	Basic Science
implement	Electronics
capable	Packaging
pack	Green Energy
hmd	Packaging
characteristic	Agrifood
foam	Basic Science
conductive	Packaging
reaction	Basic Science
electric	Packaging
specific	Biomed
voltage	Green Energy
reactor	Basic Science
vitro	Biomed
noise	Electronics

soil	Agrifood
arm	Packaging
circular	Packaging
plenoptic	Basic Science
reinforcement	Environment
radiation	Basic Science
roof	Environment

RINGRAZIAMENTI

Perchè fare un dottorato a 39 anni, avendo già un lavoro a tempo indeterminato e una vita già sufficientemente incasinata? Ogni tanto continuo a chiedermelo anche io, nonostante sia giunta quasi alla fine di questo percorso, che è stato lungo e certamente non poco faticoso. La risposta? Per la voglia di continuare ad apprendere, a scoprire innovazioni, ad avventurarmi in mondi sconosciuti che potessero "allargare" la mia visione e le mie possibilità. Ecco è questa la ragione!

E la scelta di intraprendere questo percorso con persone che conoscevo già e che stimavo… Roberto e il suo gruppo di ricerca. Mi hanno supportato ed "accompagnato" in questi tre anni, guidandomi nei meandri dell'intelligenza artificiale, degli algoritmi e del "fatidico" Python, perfetti "sconosciuti" inizialmente per me.

In particolare sento davvero di ringraziare Roberto Bellotti, amico e scienziato prezioso …..nonostante il suo poco tempo a disposizione per i suoi mille impegni, ha rappresentato comunque e sempre una guida fondamentale.

E Nicola Amoroso, che negli ultimi due anni si è dedicato a me, con grande pazienza ed entusiasmo su un terreno anche per lui un pò meno conosciuto del solito, come quello della terza missione e del trasferimento tecnologico. Con Nicola abbiamo costruito da zero un rapporto basato sul rispetto, sulla stima e sulla gentilezza.

Un super ringraziamento anche al Netval grazie al quale ho potuto in questi ultimi anni avvicinarmi sempre più al mondo del trasferimento di conoscenza e costruire un network davvero importante di persone straordinarie. Non in ultimo grazie al Netval ho potuto maneggiare Knowledgeshare!

Sento ancora di ringraziare le persone che hanno creduto in me, i miei familiari in primis, con una cena sempre pronta perchè io non avevo tempo di cucinare! E le persone amiche che sul lavoro mi hanno supportato e hanno avuto pazienza per i miei tempi e le mie attività extra.

E infine… Quelle poche persone che, <u>nel silenzio</u>, mi sono state accanto dall'inizio alla fine del percorso, sopportando e supportandomi nei momenti di crisi e stanchezza, indicandomi sempre la giusta via su cui proseguire il cammino.

E dopo questo dottorato che farò? Chi può saperlo come proseguirà la vita . . .